Applied Computational Intelligence and Soft Computing

Awareness Science and Engineering

Guest Editors: Qiangfu Zhao, Cheng-Hsiung Hsieh, Keitaro Naruse, and Zhishun She

Awareness Science and Engineering

Awareness Science and Engineering

Guest Editors: Qiangfu Zhao, Cheng-Hsiung Hsieh, Keitaro Naruse, and Zhishun She

Editorial Board

Contents

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 182849, 3 pages
doi:10.1155/2012/182849

Editorial
Awareness Science and Engineering

Qiangfu Zhao,[1] Cheng-Hsiung Hsieh,[2] Keitaro Naruse,[1] and Zhishun She[3]

[1] *School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-8580, Japan*
[2] *Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung 41349, Taiwan*
[3] *Department of Electrical Engineering and Computer Science, Glyndwr University, Wrexham LL11 2AW, UK*

Correspondence should be addressed to Qiangfu Zhao, qf-zhao@u-aizu.ac.jp

Received 7 June 2012; Accepted 7 June 2012

The goal of awareness computing (AC) is to realize awareness in computing machines. Awareness is the ability to perceive, to feel, or to be conscious of events, objects, or sensory patterns. It may not lead directly to full comprehension. Awareness often implies vigilance in observing and alertness in drawing inferences from what one experiences. Compared with artificial intelligence (AI), the goal of AC is less ambitious. Nevertheless, AC could be more important for solving practical problems. We think that awareness is the bridge between perception and cognition. Without being aware, a system may never become intelligent. Moreover, awareness could make computation towards the goal of AI more efficient by shedding of irrelevant possibilities.

In the last two decades, AC has been studied mainly from an engineering perspective. To create real aware systems, however, studying different applications in an ad hoc manner is not enough. We need to study all kinds of AC-related problems in a unified framework and gain more insight about aware systems and the mechanism of awareness existing in different living beings. For this purpose, we have organized the first International Workshop on Aware Computing (IWAC2009, Aizuwakamatsu, Japan), the second International Symposium on Awareness Computing (ISAC2010, Tainan, Taiwan), and the third International Conference on Awareness Science and Technology (iCAST2011, Dalian, China). We have also established the Technical Committee on Awareness Computing under the umbrella of the IEEE Systems, Man, and Cybernetics Society.

In this special issue, we received 23 submissions on different topics of AC. From them, 11 papers are highly recommended by the referees. We can classify these papers into 3 categories. The first category is related to the fundamental theory and modeling of AC (papers by H. Takagi entitled "*Interactive evolutionary computation for analyzing human awareness mechanisms*," M. Hu and H. Liang entitled "*Variance entropy: a method for characterizing perceptual awareness of visual stimulus*," Y. Sawada entitled "*The aspects, the origin, and the merit of aware computing—suggestions from the visual hand tracking experiments*," and M. Yokota entitled "*Aware computing in spatial language understanding guided by cognitively inspired knowledge representation*"). The second one is about realization and implementation of different AC systems (papers by M. Matsuhara and S. Suzuki entitled "*Effectiveness of context-aware character input method for mobile phone based on artificial neural network*," H. Kato et al. entitled "*A real-time angle and illumination aware face recognition system based on artificial neural network*," X. Yu et al. entitled "*An application of improved Gap-BIDE algorithm for discovering access patterns*," K. Madani "*Multi-level cognitive machine-learning based concept for Artificial Awareness: application to humanoid robot's awareness using visual saliency,*" and J.-F. Wang et al. entitled "*Emotion-aware assistive system for humanistic care based on the orange computing concept*"). The third is related to AC applications (papers by S. Kikuchi and G. Chakraborty entitled "*An efficient genome fragment assembling using GA with neighborhood aware fitness function*" and X. Guo et al. entitled "*Environmental sound recognition using time-frequency intersection patterns*"). In this issue we put more weight on the scientific aspect of AC, rather than simple applications. In fact, even the two application papers are not just sensor integrations. By doing so, we hope this issue can serve as an important reference in this emerging field and provide a better understanding in establishment of a unified framework for AC.

In the paper written by H. Takagi entitled "*Interactive evolutionary computation for analyzing human awareness mechanisms*," the author attempts to analyze human awareness mechanism and to build awareness models using interactive evolutionary computation (IEC). From the related history of computational intelligence, the author paves a way to the objective. By several successful examples of IEC, the author is convinced that IEC is a possible way to awareness science. In fact, IEC can be a general tool for modeling different human-factor-related awareness.

In the paper by M. Hu and H. Liang entitled "*Variance entropy: a method for characterizing perceptual awareness of visual stimulus*," the authors propose a simple but efficient complexity measure called variance entropy, which can be important for characterizing perceptual awareness of visual stimulus. In the variance entropy both sample entropy and variance of data are considered. To show its effectiveness, the variance entropy is applied to analyze cortical local field potential data and to study neural dynamics of perceptual awareness. Results show that the variance entropy analysis is able to differentiate the perceptual visibility and is of far better discriminative performance than the sample entropy.

In the paper entitled "*The aspects, the origin, and the merit of aware computing—suggestions from the visual hand tracking experiments*," Y. Sawada studies awareness in a science perspective. He investigates several interesting aspects of awareness, including qualitative and quantitative, external and internal; awareness of thinking; awareness and experience; awareness and self-monitoring. The author also poses some profound questions for future study. For example, if a computer or robot is so aware that it can conduct self-control and self-improve based on its experiences; can we say that the computer or robot has free will? Or put it in another way, for a computer or robot to have free will, what kind of awareness should it have? After all, what is free will?

The paper written by M. Yokota entitled "*Aware computing in spatial language understanding guided by cognitively inspired knowledge representation*," is related to spatial relation awareness of objects in a given image, which is useful for multimedia retrieval. This paper, however, is not just an application paper. It introduces an omni-sensory mental image model and its description language Lmd. The language Lmd can provide multimedia expressions with intermediate semantic descriptions in predicate logic. This paper presents systematic and efficient computing guided by Lmd expression and 3D map data in crossmedia operation between linguistic and pictorial expressions as spatial language understanding.

In the paper by M. Matsuhara and S. Suzuki entitled "*Effectiveness of context-aware character input method for mobile phone based on artificial neural network*," the authors provide a character input approach for mobile phones by a context-aware mechanism. With artificial neural networks (ANNs), the proposed system becomes aware of the mapping between number segments through learning. This leads to a possibility for the system to translate the number string into the intended sentence by ANN without a dictionary. The effectiveness and feasibility of the proposed approach are verified by Twitter data on a mobile phone platform.

The paper by H. Kato et al. entitled "*A real-time angle and illumination aware face recognition system based on artificial neural network*," deals with the variations of angle and illumination in the problem of face recognition. One or two multilayer perceptrons (MLPs) are trained to map angle and illumination features to image features. By the generalization ability of MLP, the variations of angle and illumination have been taken care of. Consequently, the user awareness by face images is achieved where variations of angle and illumination are considered. The approach is justified by examples.

In the paper by X. Yu et al. entitled "*An application of improved Gap-BIDE algorithm for discovering access patterns*," the authors propose a new algorithm for discovering access patterns. The problem studied here is related to user awareness, abnormal awareness, and is an important issue in all kinds of Internet-based service systems. Compared with the previous algorithm, a process of getting a large event set is proposed in the improved Gap-BIDE algorithm. The proposed approach can find out the frequent events by discarding the infrequent events which do not occur continuously in an accessing time before generating candidate patterns.

In the paper entitled "*Multi-level cognitive machine-learning based concept for artificial awareness: application to humanoid robot's awareness using visual saliency*," K. Madani et al. study the possibility of realizing artificial awareness in a robot based on observations in human early-age skill development and early-age awareness maturation. For this purpose, a multilevel cognitive procedure is introduced. Following this procedure and the proposed algorithm, a robot can percept motion, be aware of the environment visually, and pay attention to an interesting object.

In the paper by J.-F. Wang et al. "*Emotion-aware assistive system for humanistic care based on the orange computing concept*," studies emotion awareness and its application to mental care. In fact, mental care is one of the main objectives of the so-called orange computing. Its main purpose is to help people, not only physically but also mentally, to reduce different mental diseases which are often difficult to heal in the modern societies. In this paper, a case study on a human-machine interactive and assistive system for emotion care is conducted. The system can detect emotional states of users by analyzing their facial expressions, emotional speeches, and laughter in a ubiquitous environment. In addition, the system can provide corresponding feedback to users according to the results.

In the paper by S. Kikuchi and G. Chakraborty entitled "*An efficient genome fragment assembling using GA with neighborhood aware fitness function*," the authors propose an interesting algorithm for solving the genome fragment assembling problem. The neighbor aware fitness function, although simple, is very efficient and effective for solving this NP-hard problem. Thus, different kinds of awareness are important not only for producing more intelligent systems, but also for solving different problems more intelligently. The point is that how to be aware of the necessary awareness types for solving a given problem.

In the paper by X. Guo et al. entitled "*Environmental sound recognition using time-frequency intersection patterns*,"

the authors provide a method that can be useful for a patrol robot to be situationally aware even in a dark environment. In this paper, a two-stage method for environmental sound recognition using neural networks (NNs) is proposed. It includes a classification stage and a recognition stage. At the classification stage, the environmental sounds are classified into three categories based on their long-term power-variance patterns, and the recognition stage recognizes the sound type based on both the short-term power-variance pattern and the instantaneous spectrum at the power peak.

Qiangfu Zhao
Cheng-Hsiung Hsieh
Keitaro Naruse
Zhishun She

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 896948, 6 pages
doi:10.1155/2012/896948

Research Article

Effectiveness of Context-Aware Character Input Method for Mobile Phone Based on Artificial Neural Network

Masafumi Matsuhara[1] and Satoshi Suzuki[2]

[1] *Department of Software and Information Science, Iwate Prefectural University, 152-52, Takizawa, Iwate 020-0193, Japan*
[2] *Supernet Department, System Consultant Co., Ltd., 2-14-6, Kinshi, Sumida, Tokyo 130-0013, Japan*

Correspondence should be addressed to Masafumi Matsuhara, masafumi@iwate-pu.ac.jp

Received 10 February 2012; Revised 19 April 2012; Accepted 26 April 2012

Academic Editor: Cheng-Hsiung Hsieh

Opportunities and needs are increasing to input Japanese sentences on mobile phones since performance of mobile phones is improving. Applications like E-mail, Web search, and so on are widely used on mobile phones now. We need to input Japanese sentences using only 12 keys on mobile phones. We have proposed a method to input Japanese sentences on mobile phones quickly and easily. We call this method number-*Kanji* translation method. The number string inputted by a user is translated into *Kanji-Kana* mixed sentence in our proposed method. Number string to *Kana* string is a one-to-many mapping. Therefore, it is difficult to translate a number string into the correct sentence intended by the user. The proposed context-aware mapping method is able to disambiguate a number string by artificial neural network (ANN). The system is able to translate number segments into the intended words because the system becomes aware of the correspondence of number segments with Japanese words through learning by ANN. The system does not need a dictionary. We also show the effectiveness of our proposed method for practical use by the result of the evaluation experiment in Twitter data.

1. Introduction

Ordinary Japanese sentences are expressed by two kinds of characters, that is, *Kana* and *Kanji*. *Kana* is Japanese phonogramic characters and has about fifty kinds. *Kanji* is ideographic Chinese characters and has about several thousand kinds. Therefore, we need to use some *Kanji* input methods in order to input Japanese sentences into computers. A typical method is the *Kana-Kanji* translation method of nonsegmented Japanese sentences. This method translates nonsegmented *Kana* sentences into *Kanji-Kana* mixed sentences. Since one *Kana* character is generally inputted by combination of a few alphabets, this method needs twenty six keys for the alphabets.

Recently, performance of mobile computing devices is greatly improving. We consider that the devices are grouped into two by their quality. One gives importance to easy operation, the other gives importance to good mobility. Mobile phones are usable as mobile computers and belong to the latter group. Their mobility is very good because typical size of them is small. However, a general mobile phone has only 12 keys, which are $0, 1, \ldots, 9, *$, and #, because of the limited size. A growing number of Smartphones, for example, iPhones, Blackberries, and so on, have full QWERTY keyboards. It is not easy to press the intended key because the key size is small. Moreover, a user needs to press a few keys per *Kana* character since one *Kana* character generally consists of a few alphabets. Therefore, we focus on 12 keys layout on the mobile phones.

The letter cycling input method is most commonly used for the input of sentences on mobile phones. In this input method, a chosen key represents a consonant, and the number of pressing it represents a vowel in Japanese. For example, the chosen key "7" represents "m", and three presses of the key represent "u". Then, the number of key presses is three for the input character "む (*mu*)". Since this input method needs several key presses per *Kana* character, it is troublesome for a user. Opportunities and needs are rapidly increasing to input Japanese sentences into a small device such as a mobile phone since performance of mobile phones

1: あ い う え お a i u e o	2: か き く け こ ka ki ku ke ko	3: さ し す せ そ sa si su se so
4: た ち つ て と ta ti tu te to	5: な に ぬ ね の na ni nu ne no	6: は ひ ふ へ ほ ha hi hu he ho
7: ま み む め も ma mi mu me mo	8: や ゆ よ ya yu yo	9: ら り る れ ろ ra ri ru re ro
*: ゛ ゜ Voiced consonant, P-sound	0: わ を ん wa wo n	#: 、 。 Punctuation marks

Figure 1: Correspondance of number to KANA and its pronunciation.

		k	s	t	n	h	m	y	r	w
a	あ	か	さ	た	な	は	ま	や	ら	わ
i	い	き	し	ち	に	ひ	み		り	
u	う	く	す	つ	ぬ	ふ	む	ゆ	る	
e	え	け	せ	て	ね	へ	め		れ	
o	お	こ	そ	と	の	ほ	も	よ	ろ	を
n										ん

Figure 2: 50-sound table of KANA.

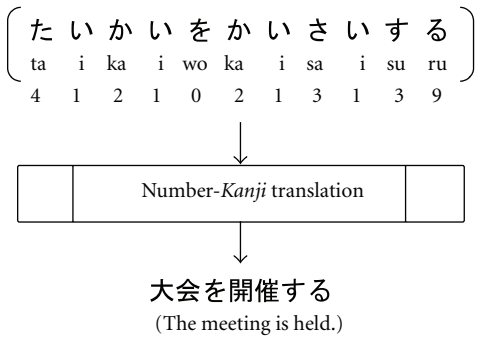

Figure 3: Example of translation.

is improving. Applications like E-mail, Web search, and so on are widely used on mobile phones now. Therefore, methods are demanded which enable us to promptly and easily input Japanese sentences on mobile phones.

Some input methods for mobile phones have been proposed [1, 2], and the systems have been developed: for example, T9 (Nuance Communications, Inc. has developed T9. http://www.t9.com/). T9 enables us to input one alphabet per key press on the keypad of 9 keys. Since three or four letters are assigned to each key of 9 keys, the specific letter intended by one key press is ambiguous. This system disambiguates the pressed keys on word level. However, the system is for English mainly. Some input methods have been proposed for Japanese [3–5]. The methods enable us to input one *Kana* character per key press. Since about five *Kana* characters are assigned to each key on a mobile phone, the specific character intended by one key press is ambiguous. The methods disambiguate by dictionaries. Therefore, they are not able to translate the number strings into words not included into the dictionary. Moreover, the methods spend a lot of memory as the inputted data increases because the words are acquired and registered into the dictionary in some methods. Some predictive input methods have been proposed [6–8]. The methods output word candidates by prediction or completion. The number of key presses increases to select the intended word because there are many word candidates. Therefore, we focus on a number-*Kanji* translation method without prediction.

We have proposed a number-*Kanji* translation method based on artificial neural network (ANN) [9]. The system becomes aware of the correspondence of number segments with Japanese words through learning by ANN. Then, the system translates an inputted number string by ANN. The system does not use dictionaries for translation. Therefore, the system may translate the number-segments into unknown words without dictionaries. Moreover, the system requires the only fixed memory determined by the size of ANN. Because of reduced memory requirement, we consider that our proposed method is especially suitable for a mobile phone.

This paper shows the outline of the number-*Kanji* translation, the processes of our proposed method, the evaluation experiment, its result, and the effectiveness of our proposed method for practical use.

2. Outline of Number-*Kanji* Translation

Figure 3 shows an example of the number-*Kanji* translation. A user inputs the number-string "41210213139" for the *Kanji-Kana* mixed sentence "大会を開催する (The meeting is held.)". A user is able to input rapidly and easily because one key stroke corresponds to one *Kana* character. The number-string is translated into the intended Japanese sentence by a number-*Kanji* translation method.

A user inputs a string of numbers corresponding to the pronunciation of an intended Japanese sentence based on Figure 1. The *Kana-Kanji* translation method translates a *Kana* sentence, whereas the number-*Kanji* translation method translates a string of numbers. A key pressed on the keypad of 12 keys represents a line of the 50-sound table of *Kana*, which is the Japanese syllabary. Figure 2 shows the 50-sound table. It is set in a five-by-ten matrix. The matrix has five vowels and ten consonants. Almost all *Kana* characters are composed of a consonant plus a vowel. A user is able to input one *Kana* character per key press.

Figure 1 shows the correspondence of the number with *Kana* characters: for example, the key "4" represents "た (*ta*)" or "ち (*ti*)" or "つ (*tu*)" or "て (*te*)" or "と (*to*)" of *Kana* characters. The characters in parentheses represent the pronunciation of *Kana*. Then, a number character of 12 keys generally corresponds to a consonant. Since the vowel information degenerates, the string of numbers has ambiguity: for example, the number-string "4121" corresponds to not

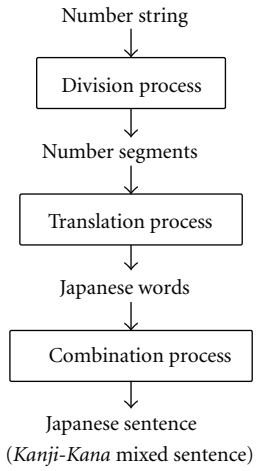

Figure 4: Procedure.

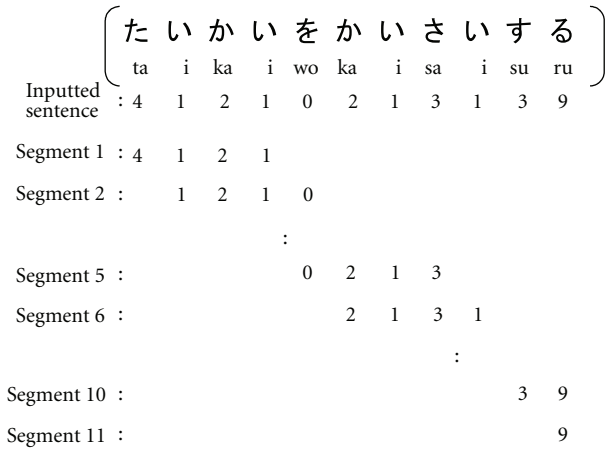

Figure 5: Example of division process.

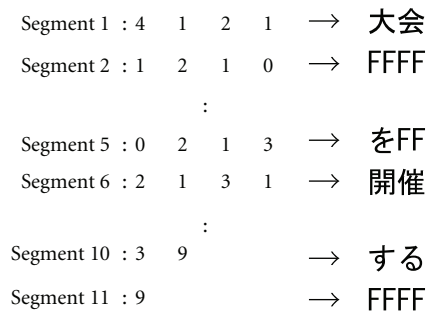

Figure 6: Example of translation process.

only the *Kana* characters "たいかい (*taikai*)" but also "ていこう (*teikou*)", "とうこう (*toukou*)", and so on. Moreover, a string of *Kana* character means some Japanese words: for example, the *Kana* characters "たいかい (*taikai*)" mean not only the Japanese word "大会 (the meeting)" but also "退会 (withdrawal)", "大海 (ocean)", and so on. Our proposed method uses ANN for the disambiguation.

The user presses the key "∗" for a voiced consonant and a p-sound in our proposed method. For example, the user inputs the number-string "4 ∗ 12" for the Japanese word "大工 (a carpenter)" of which the pronunciation is "だいく (*ta∗iku*)" ("*ta∗iku*" is generally expressed as "*daiku*" in Japanese. However, "*da*" is translated into "4∗", and the "4∗" also corresponds to "*ta*∗" in the system. Therefore, "*daiku*" is expressed as "*ta∗iku*" in this paper).

3. Processes

Our proposed method has the learning stage and the translation stage. Figure 4 shows the procedure in the translation stage. The procedure consists of the division process, the translation process, and the combination process in this order.

3.1. Division Process. Our proposed method uses ANN, and the size of ANN needs to be fixed basically. A user inputs a string of numbers corresponding to the pronunciation of an intended Japanese sentence. It is difficult to design ANN because the length of a natural language sentence is indefinite and a Japanese sentence is not segmented. Therefore, the system based on our proposed method divides the inputted number-string into the number-segments with a fixed length.

Figure 5 shows an example of the division process. The inputted number-string is divided into 11 segments, that is, from segment 1 to segment 11. The fixed length of each segment is 4 in Figure 5.

It is easy to design ANN because the length of the segments is fixed. However, the segmentations are not always correct. The segments may include incorrect words. Therefore, the system needs to select the correct words and to combine them for making up the Japanese sentence intended by the user in the combination process.

3.2. Translation Process. The system becomes aware of the correspondence of number-segments with Japanese words through learning by ANN in the learning process. The system translates each divided segment by the ANN. The system needs to translate the correct segments into the correct Japanese words and to decide the incorrect segments.

Figure 6 shows an example of the translation process. Each segment divided in the division process is translated by ANN. The segment 1 needs to be translated into the correct word "大会 (the meeting)" because its segmentation is correct. The segment 2 needs to be decided as the incorrect segment because its segmentation is incorrect. Then, the segment 2 is translated into "FFFF" as a noncharacter code in Figure 6.

3.3. Combination Process. The system based on our proposed method makes up the Japanese sentence to combine the translation result because the translation result is divided into segments.

Figure 7 shows an example of the combination process. The segment 2, the segment 11, and soon are decided as the incorrect words. Then, the system makes up the Japanese

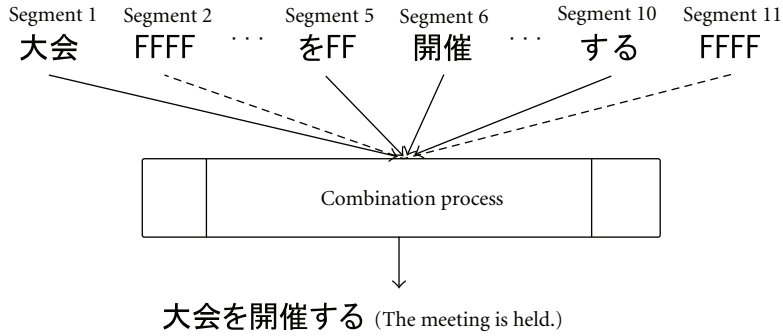

Figure 7: Example of combination process.

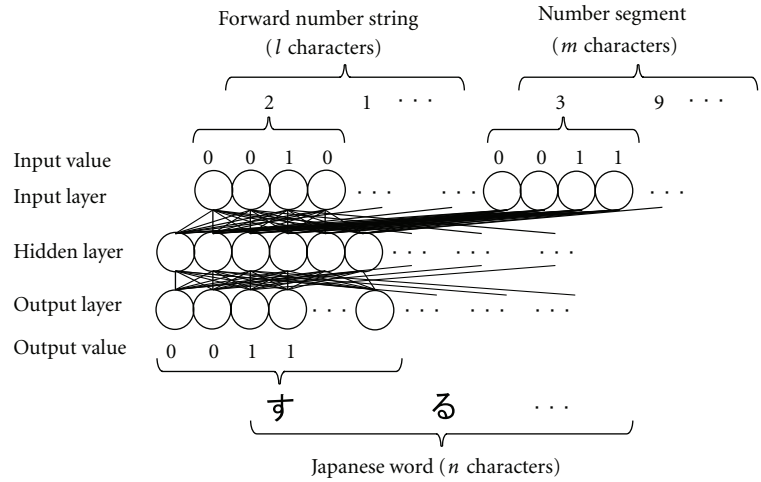

Figure 8: Structure of ANN.

sentence "大会を開催する" to combine the segment 1, the segment 5, the segment 6, and the segment 10 in Figure 7.

3.4. Learning Stage. The learning stage is performed independent of the translation stage. The system becomes aware of the correspondence of number-segments with Japanese words through learning by ANN.

We use multilayer feed-forward neural network trained by error backpropagation. The excitations propagate in a single direction, from the input layer to the output layer, through multiple intermediate layers, often called hidden layers. The connection weights, which mimic the synapses, are initialized with random values and gradually trained for the task in hand using a gradient descent training algorithm. The most common one is known as error backpropagation [10]. Thus, the functionality of the network is stored among the connection weights of different neuron nodes in a distributed manner.

The structure of ANN is shown in Figure 8. A number-string is inputted to the input layer as the input value. The number-string has 12 kinds of characters, that is, $0, 1, \ldots, 9, *$, and #. Since each input value is a binary digit, the input layer needs 4 nodes per character. The number-string consists of the forward number-string and the number-segment. A forward number-string has l characters. A number-segment has m characters. Therefore, the input layer has $4 \times (l + m)$ nodes. A Japanese word is outputted to the output layer as the output value. The output value is a binary digit also. Since a Japanese character needs 2 Bytes = 16 nodes, the output layer has $16 \times n$ nodes for n Japanese characters. The network is adjusted by evaluating the difference of a predicted character and a given character as nodes (=binary digits) in the output layer.

For example, the correspondence of the number-segment "4121" with the Japanese word "大会" is learned by ANN. Then, the system is able to translate the number-segment "4121" into the Japanese word "大会" without a dictionary. Not only a segment but also its forward number-string is learned by ANN. For example, the forward number-string "2131" of the segment "39" is learned. Then, the backward segment "39" of the number-string "2131" is able to translate into the correct word "する". Thus, our proposed method uses a context.

4. Evaluation Experiment

The system based on our proposed method has been developed for an experiment. The system is not able to make up the correct Japanese sentence in the combination process

Table 1: Experiment data.

No. of characters	55,951
No. of different words	4,199
No. of character code segments	20,674
No. of noncharacter code segments	28,565

Table 2: Parameter of ANN.

No. of input nodes	40
No. of hidden nodes	144
No. of output nodes	144
Learning rate	0.01

Table 3: Accuracy of translation per node.

Japanese character code	93.4 [%]
Noncharacter code	98.8 [%]
Total	96.5 [%]

Table 4: Mean number of erroneous node per segment.

Japanese character code	10.64
Noncharacter code	1.97
Total	5.62

if the number-segments are not translated into the correct Japanese words in the translation process. Therefore, we evaluated the translation accuracy in the translation process.

4.1. Experiment Data and Procedure. The data for the experiment is text a user inputted on Twitter (an online social networking service http://twitter.com/). The detail is shown in Table 1. The character code segments correspond to the correct words. They have to be translated into the Japanese words. The noncharacter code segments correspond to the incorrect words. They have to be translated into "FFFF" in the translation process.

The parameter of ANN is shown in Table 2. The input nodes are for the divided number-segments and the forward number-string. The max length of the segments is 6 (=m in Figure 8), and the length of the forward string is 4 (=l in Figure 8). The value is decided by the preliminary experiment. The number of input nodes is 40 because a number character needs 4 nodes in the network. The output nodes are for the character codes of the Japanese words. The max length of the words is 9 (=n in Figure 8), and a Japanese character needs 16 nodes (2 Bytes) in the network. Then, the number of output nodes is 144. The number of hidden nodes is equal to the number of output nodes. The learning rate is 0.01.

The data is divided into 5 sets for K-fold cross-validation. Each of the 4 sets is used to train the network, and the rest 1 set is used to test.

4.2. Results and Considerations. First of all, we evaluated the root mean square errors (RMSEs) in the learning stage for confirmation of the learning times. Figure 9 shows RMSE for each set of 5 sets for K-fold cross-validation in the learning stage. In Figure 9, the errors are decreasing as the learning times are increasing. The value of RMSE is below 0.005, and the changes are convergent finally. Therefore, it is shown that the system is able to learn the data normally. 10, 000 epochs are sufficient for the training of the data.

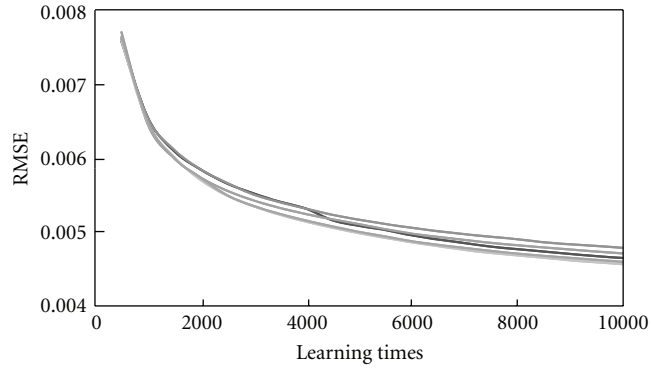

Figure 9: Changes in RMSE.

Table 3 shows the mean rate for the correct translation per node in the network of the Japanese character code, the noncharacter code, and total in the translation process. In Table 3, the accuracy of translation for noncharacter code is higher than that for Japanese character code. This is because the segments of noncharacter code are larger than ones of Japanese character code. Ordinarily, the translation accuracy tends to be higher when the data is large for the learning.

The translation accuracy of Japanese *Kana-Kanji* translation method is about 95 [%] per character in general. Therefore, we consider that 6 [%] translation error for Japanese character code is not always large. The *Kana-Kanji* translation method translates a *Kana* sentence, whereas our proposed method translates a string of numbers. It is difficult to translate a number-string because a number-string is more ambiguous than a *Kana* sentence. The accuracy of the number-*Kanji* translation method is about 85 [%] per character in our previous work [3]. Therefore, the accuracy of our proposed method is never low even though the accuracy is per node. We consider that the accuracy achieves a practical level.

Table 4 shows the mean number of the erroneous nodes per segment for the Japanese character code, the non-character code, and total. The non-character code means the segmentation is wrong, and the number-segment does not correspond to a Japanese word. The system needs to distinguish the segments with Japanese character code from ones with non-character code. The distinction is never easy because the non-character code segment may correspond to another Japanese word.

In Table 3, the accuracy of translation for non-character code is 98.8 [%]. In Table 4, the mean number of erroneous nodes is 1.97. Then, the translation accuracy of the segment for non-character code is high. The accuracy for Japanese character code is 93.4 [%]. Although the rate is high, the translation result has errors. The mean number of erroneous nodes is 10.64 in Table 4. The value is low relatively because

the size of the output nodes is 144. Therefore, we consider that it is possible to translate the erroneous nodes into the correct words by increasing learning data or adding the correction process and so on.

We are able to calculate the total number of links in the network. The number of links is defined as

$$\begin{aligned}\text{no. of links} = {} & (\text{no. of input nodes} + 1) \\ & \times \text{no. of hidden nodes} \\ & + (\text{no. of hidden nodes} + 1) \\ & \times \text{no. of output nodes},\end{aligned} \tag{1}$$

where "+1" means an additional node for a bias of ANN. The total number of links in the system of the evaluation experiment is calculated as

$$(40 + 1) \times 144 + (144 + 1) \times 144 = 26,784. \tag{2}$$

If the size for a weight is 4 Bytes per link in the network, the size of memory is about 107 KB. The size is small and fixed. The memory size does not change when the learning data increases. Therefore, it is easy to implement our proposed method on a mobile phone.

5. Conclusion

In this paper, we proposed a context-aware number-*Kanji* translation method using ANN and have shown the effectiveness of the method by the actual experiment for practical use.

The algorithm enables to input one *Kana* character per key stroke. Then, a user is able to input a Japanese text rapidly and easily. However, a string of numbers inputted by the user is ambiguous. Our proposed method disambiguates the number-string and translates it into the Japanese sentence intended by the user using ANN. The system becomes aware of the correspondence of number-segments with Japanese words through learning. Therefore, the system is able to translate the number-string into the intended sentence by ANN without a dictionary. The system requires the fixed memory determined by the size of ANN. Because of reduced memory requirement, our proposed method is especially suitable for a mobile phone.

In the experiment, we use Twitter data to confirm the effectiveness of our proposed method for practical use. The accuracy of the translation per node is high. The mean number of the erroneous nodes is about 11 per segment for Japanese character code. The value is low in comparison with the size of the output nodes in the network. Therefore, we consider that it is possible to translate the erroneous segments into the correct words. By the actual experiment, it is shown that our proposed method is effective for practical use.

One of future works is to add the correction process for recovering the erroneous nodes. Then, we need to evaluate the translation accuracy in the combination process and compare with current popular methods.

References

[1] C. Kushler, "AAC: using a reduced keyboard," in *Proceedings of the Technology & Persons with Disabilities Conference (CSUN '98)*, Los Angeles, Calif, USA, March 1998.

[2] S. Hasan and K. Harbusch, "N-Best hidden Markov model super tagging to improve typing on an ambiguous keyboard," in *Proceedings of Seventh International Workshop on Tree Adjoining Grammar and Related Formalisms*, pp. 24–31, Vancouver, BC, Canada, May 2004.

[3] M. MaMatsuhara, K. Araki, Y. Momouchi, and K. Tochinai, "Evaluation of number-Kanji translation method of non-segmented Japanese sentences using inductive learning with degenerated input," in *Proceedings of the 12th Australian Joint Conference on Artificial Intelligence: Advanced Topics in Artificial Intelligence*, vol. 1747 of *Lecture Note in Artificial Intelligence*, pp. 474–475, Springer, December 1999.

[4] M. Matsuhara, K. Araki, and K. Tochinai, "Evaluation of number-Kanji translation method using inductive learning on E-mail," in *Proceedings of 3rd IASTED International Conference on Artificial Intelligence and Soft Computing (ASC '00)*, pp. 487–493, Alberta, Canada, July 2000.

[5] K. Tanaka-Ishii, Y. Inutsuka, and M. Takeichi, "Personalization of text entry systems for mobile phones," in *Proceedings of 6th Natural Processing Pacific Rim Symposium*, pp. 177–184, Tokyo, Japan, November 2001.

[6] K. Tanaka-Ishii, "Word-based predictive text entry using adaptive language models," *Natural Language Engineering*, vol. 13, no. 1, pp. 51–74, 2007.

[7] A. Van Den Bosch and T. Bogers, "Efficient context-sensitive word completion for mobile devices," in *Proceedings of the10th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '08)*, pp. 465–470, September 2008.

[8] M. D. Dunlop and M. Montgomery Masters, "Investigating five key predictive text entry with combined distance and key stroke modelling," *Personal and Ubiquitous Computing*, vol. 12, no. 8, pp. 589–598, 2008.

[9] M. Matsuhara and S. Suzuki, "An efficient context-aware character input algorithm for mobile phone based on artificial neural network," in *Proceedings of the 3rd International Conference on Awareness Science and Technology (iCAST '11)*, pp. 314–318, Dalian, China, September 2011.

[10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in Microstructures of Cognition*, vol. 1, pp. 318–362, MIT Press, Cambridge, UK, 1986.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 274617, 9 pages
doi:10.1155/2012/274617

Research Article

A Real-Time Angle- and Illumination-Aware Face Recognition System Based on Artificial Neural Network

Hisateru Kato, Goutam Chakraborty, and Basabi Chakraborty

Faculty of Software and Information Science, Iwate Prefectural University, Iwate, Takizawamura 020-0193, Japan

Correspondence should be addressed to Hisateru Kato, nd4y2518@docomo.ne.jp

Received 10 March 2012; Revised 19 May 2012; Accepted 23 May 2012

Academic Editor: Cheng-Hsiung Hsieh

Automatic authentication systems, using biometric technology, are becoming increasingly important with the increased need for person verification in our daily life. A few years back, fingerprint verification was done only in criminal investigations. Now fingerprints and face images are widely used in bank tellers, airports, and building entrances. Face images are easy to obtain, but successful recognition depends on proper orientation and illumination of the image, compared to the one taken at registration time. Facial features heavily change with illumination and orientation angle, leading to increased false rejection as well as false acceptance. Registering face images for all possible angles and illumination is impossible. In this work, we proposed a memory efficient way to register (store) multiple angle and changing illumination face image data, and a computationally efficient authentication technique, using multilayer perceptron (MLP). Though MLP is trained using a few registered images with different orientation, due to generalization property of MLP, interpolation of features for intermediate orientation angles was possible. The algorithm is further extended to include illumination robust authentication system. Results of extensive experiments verify the effectiveness of the proposed algorithm.

1. Introduction

The need for personal identification has grown enormously in the last two decades. Previously, biometric identification using fingerprints or face images was restricted to criminal prosecution only. A few experts could serve the demand. With increased terrorist activities, stricter security requirements for entering buildings, and other related applications, need for automatic biometric machine-authentication systems is getting more and more important.

Recognizing people from face (face image) is the most natural and widely used method we human do always and effortlessly. Due to ease of collection without disturbing the subject, it is one of the most popular ways of automatic machine authentication. An excellent survey of face-recognition algorithms is available in [1].

In automatic face recognition, the first step is to identify the boundary of the face and separate it from the photographed image. Next, recognition algorithms extract feature vectors from the input (probe) image. These features are then compared with the set of such features stored in the database. The database (gallery image) contains same set of features already extracted and stored during registration phase for all persons required to be authenticated.

There are two classes of algorithms to extract features from the image—model based and appearance based. Model-based algorithms use explicit 2D or 3D models of the face. In model-based algorithms, geometrical features like relative positions of important facial components, for example, eyes, nose, mouth, and so forth, and their shapes are used as features. These features are robust to lighting conditions but weak for change in the orientation of the face. We use a subset of such features as "Angle-feature" in our previous work [2]. In appearance-based methods pattern of the light and shade distribution in the facial image is used to derive features. Being computationally simpler, appearance-based paradigm is more popular. One of the significant works is the eigenface approach [3] by Turk and Pentland. We also used appearance-based algorithms to extract facial features.

Though automated face recognition by computers for frontal face images taken under controlled lighting conditions is more or less successful, recognition in uncontrolled environment is an extremely complex and difficult task. Lots of researchers are trying to develop unconstrained face recognition system [4], specially for pose and illumination invariant face recognition [5], for a wide variety of real-time applications.

For most of the biometric applications, we need to authenticate a particular person in *real time* from his/her *quickly taken* face image. The face image has to be already registered in the system. For proper verification, the input image (probe image) should exactly match the registered image (gallery image) of that particular person (to avoid false rejection of the genuine person) and not with anyone else's face image (to avoid false acceptance). The algorithm has to be efficient to work in real time. The task becomes difficult because the quickly taken probe image may differ in illumination and pose (and therefore features) from the image of the individual registered in the data base.

Even though the person is same, the automatic authentication system may fail due to angle orientation, ambient lighting, age, make-up, glasses, expression of the face, and so forth which are different from the stored gallery image of the individual. It is said that about 75% of the authentication failure is due to the fact that angle of orientation of the probe face image is different from the stored image. It is impossible and very inefficient to store the images (i.e., image features) of an individual taken at all possible angles and at different illuminations in the gallery. But we need that information for correct recognition. In this work, we focus on angle-aware face recognition, and then the proposed algorithm is extended to include ambient light-aware face recognition. In the proposed angle- and illumination-aware face recognition, we store the available (training) information in a trained Artificial Neural Network. Retrieval of the features for any intermediate angle and illumination from the trained ANN is very efficient. The algorithm can be used in real time. We experimented with a benchmark database. Our system could achieve excellent results both for false-acceptance rate (FAR) as well as for false-rejection rate (FRR).

In the next section we briefly discuss related works on orientation and illumination robust face recognition. In Section 3 we represent our proposed idea for angle-aware face recognition and its extension to illumination-aware recognition which is followed by Section 4 containing simulation experiments and results. Section 5 contains conclusion and discussion.

2. Related Works on Angle and Illumination Invariant Face Recognition

According to FERET and FRVT [6] test reports, performance of face recognition systems drops significantly when large pose variations are present in the input images. Though the registration image is a frontal face image, the probe image is more often than not a perfect frontal image. Angle-aware face recognition is a major research issue. Approaches to address the pose variation problem are mainly classified into three categories.

(1) Single-view approach in which invariant features or 3D model based methods are used to produce a canonical frontal view from various poses. In [7] a Gabor wavelet-based feature extraction method is proposed which is robust to small angle variations. This approach did not receive much attention due to high computational cost.

(2) Multiview face recognition is an extension of appearance-based frontal image recognition. Here, gallery images of every subject at many different poses are needed. Earlier works on pose invariant appearance based on multiview algorithms are reported in [8–10]. Most algorithms in this category require several images of each subject in the data base and consequently require much more computation for searching and memory for storage.

(3) Class-based hybrid methods in which multiview training images are available during training but only one gallery image per person is available for recognition. The popular eigenface approach [3] has been extended in [11] in order to achieve pose invariance. In [12] a robust face recognition scheme based on graph matching has been proposed.

More recent methods to address pose and illumination are proposed in [2, 13–21].

The simplest approach is to look for a feature which is invariant to variation of pose. But, till now such a feature is not found. Reference [7] works only for very small range of angle variation, and the algorithm is too heavy to be used real time. Geometrical features are very weak to angle variation. Variation of image-pattern-based features due to angle variation exceeds variation of features across individuals, jeopardizing the recognition process and would lead to high FAR and FRR. Prince and Elder [22] presented a heuristic algorithm to construct a single feature which does not vary with pose. Murase and Nayer [23] have used principle components of many views to visualize the change due to pose variation. Graham and Allison [24] sampled input sequences of varying pose to form eigensignature when projected into an eigenspace. A good review of these approaches can be found in [5, 25].

3. Angle- and Illumination-Aware Face Recognition

Our approach is to store multiple pose image features in a single trained MLP, so that both storage and searching for intermediate angles are efficient. We do not overload the database by adding features for the same face at different angles. We train an artificial neural network to store them all as a function of the orientation angle. Due to good generalization property of MLP, it can give feature values at intermediate angles and very efficiently too. Through experiments, we realized that geometrical features are fragile to angle variation. We used a subset of geometrical features

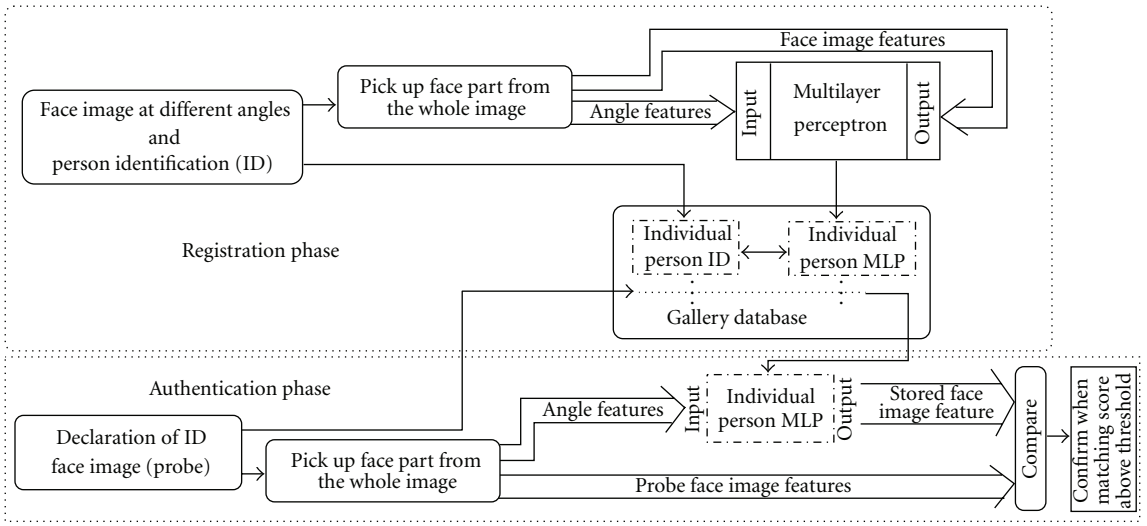

Figure 1: Block diagram of registration and authentication phase.

to express the pose angle. The following important aspects were investigated while selecting the efficient angle features:

(1) low computational complexity to extract the angle feature, so that the algorithm can run real time

(2) the pose-angle feature contains enough information about the angle

(3) the feature values vary smoothly with angle variation, so that MLP can be trained easily and with little error.

The two main contributions regarding pose invariant face authentication, over our previous work [17], are to automatize the angle-feature extraction from the face image and enrich the angle feature vector with more relevant features. We also verified that artificial neural network could achieve good generalization for intermediate angles of orientation, for which data were not available during training phase. A brief description of the whole algorithm, with an emphasis on the new contribution, is presented in this section.

Figure 1 shows the block diagram of the proposed angle invariant face recognition system. The system consists of two phases, registration phase and recognition phase. In registration phase, a set of face images are taken from equal distance but at different angles.

If the number of cameras is n, we get n training samples to train the individual person's MLP at the time of registration. From all the n photographs, taken by n cameras, first the training data is created to train that individual's MLP. The input vector of the training data is the angle feature vector, and the output vector is the image feature. Procedures to extract angle features and image features are explained in Sections 3.1 and 3.2, respectively.

A person's identification (ID) and the corresponding trained MLP (using her/his face image angle feature and image feature) are stored as a pair. Such ID-MLP pair forms gallery image "DATA BASE." In the recognition phase, the individual's face image (probe image) is presented with her/his ID. From gallery "DATA BASE" of MLPs, the particular trained MLP for the claimed ID is retrieved. Angle features from the image are extracted and used as input to that person's MLP retrieved from the data-base. Image-feature from the probe image and that obtained as output from the MLP are compared. If the distance between two feature vectors are below some predefined threshold, the decision is accept, otherwise reject. The implicit assumption here in that the MLP would be able to deliver correct image feature for any intermediate face orientation due to its good interpolation (generalization) property.

In the following section, we will discuss how angle features are extracted from the face image. We will also show what angle features are finally selected for our system and why.

3.1. Angle Feature Extraction. Angle feature should contain the information of the orientation angle of the face image. Geometrical features of a face image, which uses distances between important parts of the face and angle between connecting lines, are capable of expressing the orientation of the face image. We used cues from those approaches of feature extraction. In our previous work [17], we used three points, the left and right eye locations and the middle of mouth. The distances between them and the slope of the lines connecting them are used as elements of the feature vector. The distance between the two eyes decreases as the orientation angle increases. Similarly, the slope of the line connecting eyes and the mouth changes as the face turns towards right or left. The results obtained using these six elements of angle-feature vector gave reasonably good results. But, in our previous work, the three points from the face image were manually identified, and feature vectors were manually evaluated from all face images under investigation. In total we used 10 facial images, each for 21 different angles. Therefore, 210 angle feature vectors were hand calculated.

In the present work, we wrote algorithm to automatically identify the important points on the face. This facilitated

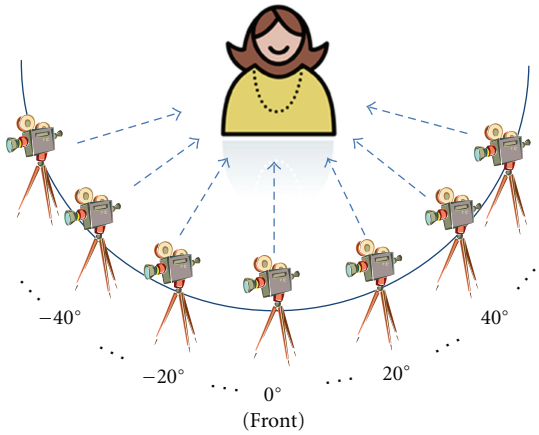

Figure 2: Collection of several face images taken from equal distance but at different angles.

Figure 3: Images after different filtering steps.

working with a larger data set. Moreover, after filtering, we could always identify the eyebrows, eyes, nostrils, and mouth. All possible identifying points can be listed as two end points of left eyebrow, two end points of right eyebrow, two end points of left eye, two end points of right eye, nostril (sometimes two), and two end points of mouth. This is clear from the picture after binary conversion (the best result obtained with a threshold of 0.75), as shown in Figure 3. We used the database with oriental faces only. A lot of angle vector elements can be identified whose values change as the angle changes. We tried different combinations taking care that the procedure is simple and efficient.

The important parts from the face image are separated as follows. At first minimum value filter is used. The minimum value filter emphasizes the part where the image is dark, because important parts on face are darker than that of surrounding skin. Through experiments, we ensured that this technique is effective to identify locations of eyebrows, eyes, nostrils, and mouth. After using the minimum value filter, binarization is performed to clearly identify important parts of the face. In addition to our targeted important parts of the face, hair also is filtered out. First the hair part is detected. Though it is an important element too to profile the face image, we do not use it. We delete the hair part and the background. We then identify eyebrows, eyes, nose, and mouth, with heuristic algorithm using knowledge of their relative positions. As we do not use eyebrows to create the angle-feature vector, eyebrows are also deleted after identification. Once both eyes, nose and mouth are located on the face image, we generate the angle features.

Table 1: Distance components of angle feature.

Description	Symbol
Distance between LE and RE	D_1
Distance between LE and N	D_2
Distance between RE and N	D_3
Distance between LE and M	D_4
Distance between RE and M	D_5

Table 2: Gradient components of angle feature.

Description	Symbol
Gradient of line joining LE and N	m_1
Gradient of line joining LE and M	m_2
Gradient of line joining RE and N	m_3
Gradient of line joining RE and M	m_4

First we will give the details of the elements of angle-feature vector and then explain the rationality of choosing them. The angle feature vector is

$$\mathrm{AF} = (W_1, W_2, W_3, D_1, D_2, D_3, D_4, D_5, m_1, m_2, m_3, m_4). \quad (1)$$

It consists of 12 elements. W_1, W_2, and W_3 are the widths of the left eye, the right eye, and the mouth. Following that, we find the center for left eye, right eye, nostrils, and mouth. Let us denote the coordinates of these four points of left eye (LE) as (x_1,y_1), right eye (RE) as (x_2,y_2), mouth (M) as (x_3,y_3), and nose (N) as (x_4,y_4). We have six distances taking any two points from the above four points. Except the distance between N and M, all other distances change with face angle orientation. We use the five distances shown in Table 1 as components of angle feature vector. The remaining four features are the gradient of lines described in Table 2.

All these features are easy to calculate and change more or less smoothly with angle variation. We did not include the distance between mouth and nose, the gradient of the line joining mouth and nose, and the line joining the two eyes. This is because these parameters do not change with angle change.

In order to ensure how our angle feature vector changes with change in the orientation of the face image, we plotted the Euclidean distance between angle vectors against the angle of orientation. It is shown in Figure 4(a). We have not discussed about the face-image feature yet. But in Figure 4(b), we have shown the Euclidean distance between face image feature vectors as the orientation angle changes.

The plots were for all training samples. It shows the smooth changes, though nonlinear but monotonic. From this plot, we can ensure that our angle feature is suitably chosen, and an MLP could be trained in a small number of epochs. Of course, during registration period, this training will be done off-line, and a longer training time is permissible. At the time of authentication, the MLP will give out the face image feature, from the input angle feature, instantly. That will ensure real-time application.

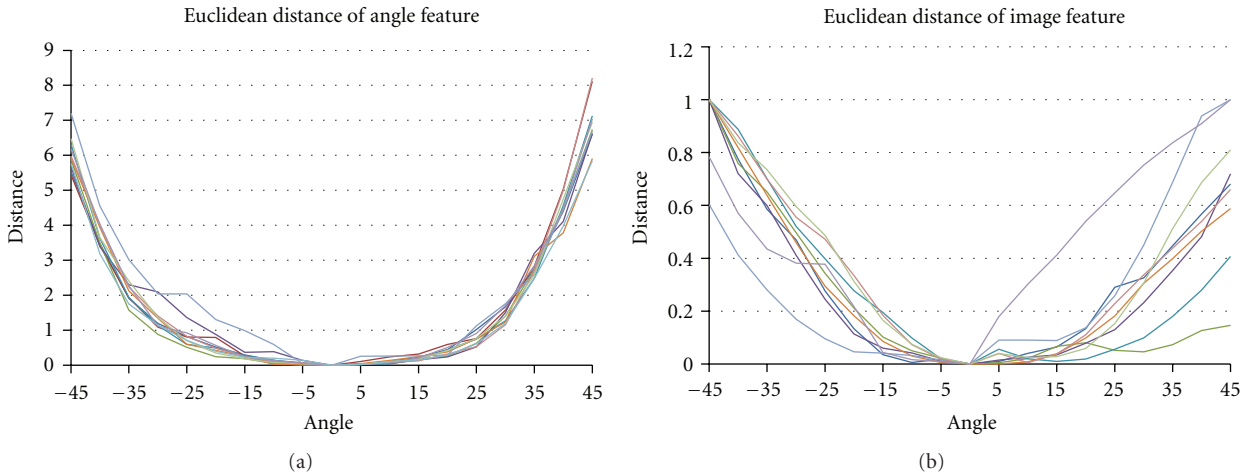

Figure 4: Euclidean distance of (a) angle feature and (b) image feature.

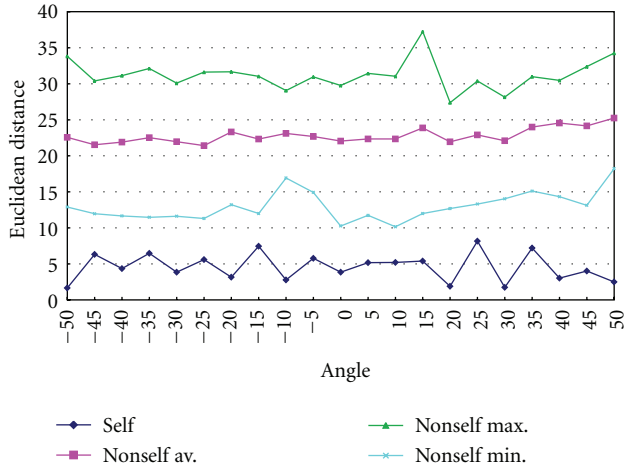

Figure 5: The results of distances between self and non-self-face image features at different angles of orientation.

In summary, compared to our previous work, we have improved our angle feature extraction technique not only by automating it but also by adding six more elements in the angle-feature vector to capture the angle of orientation information more faithfully. This also enables us to work with larger data set of face images.

3.2. Image Feature Extraction. The image feature captures the characteristic of the entire image, the spatial distribution of the pixel values. The most widely used method is eigenface, first proposed byTurk and pentland in [3]. It is based on principal component analysis. First few principal components are used as features, and every face image is expressed as a vector with values of the few principal components. We used the same technique to create image feature vector.

In our experiments 8 principal components, which carry 99% of the image information, were used. We further extended our experiments using independent components on image feature. As independent component feature of the image gave better results, in this paper we will only present those results.

3.3. Neural Network for Mapping Angle Feature to Image Feature. Multilayer neural network, trained with error back-propagation, is used as a mapping function—to map an individual's face orientation angle to his/her face image features for that particular angle. As angle feature vector consists of 12 elements, the MLP has 12 input nodes plus one bias node. We use a single hidden layer with 15 hidden nodes. Experiments were tried with different number of hidden nodes. The training is fast and quickly converges to very low MSE. Even with hidden nodes 10, it is possible to get low error after training, but we need more numbers of training epochs. The number of output nodes is eight, equal to the number of image features by using independent component analysis.

As already mentioned, we have separate MLP for every individual. For every registered individual, we have face images taken with orientation angle from −50 degrees to +50 degrees, at an interval of 5 degrees. In total, we have 21 image data for any individual. Out of the available 21 data, we use those taken at orientation −50, −40, −30, −20, −10, 0, +10, +20, +30, +40, and +50, that is, in total 11, for training the MLP. The rest 10 images, taken at angles −45, −35, −25, and so forth, were used for testing the trained MLP. Figure 5 shows the result after averaging over all images against a single self-image. A very good generalization is obtained. We can notice that at testing points the error is a little more than the points where it is trained. Yet, the distance between self and non-self-images is quite large, ensuring low values for both FAR as well as FRR, when threshold is properly chosen.

3.4. Robust Systems to Illumination Variation. In this work we also proposed an extension of our system to include correction for illumination variation. Two alternative systems are proposed shown in Figures 6 and 7. In System I, only one MLP is used as in the case of angle invariant system. The only difference is that one input to the MLP

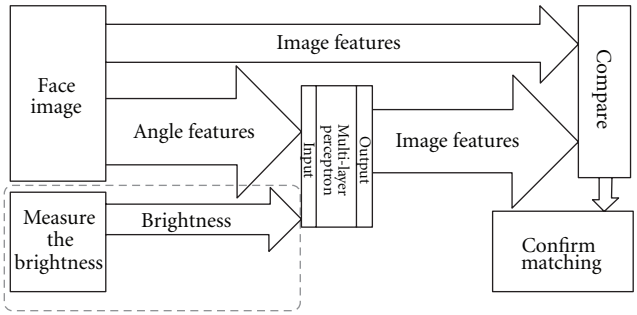

Figure 6: Block diagram of System I.

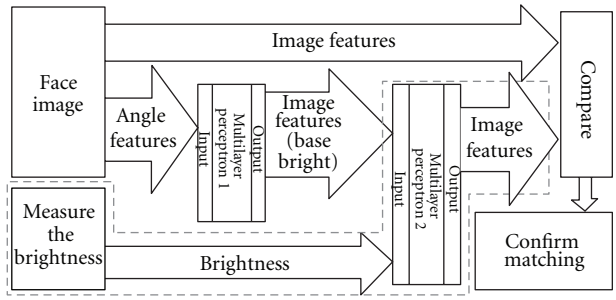

Figure 7: Block diagram of System II.

is added to include image brightness information. The rest of the algorithm remains the same. In System II, two MLPs were used. Both of them were trained separately. The first MLP (MLP1) output the image feature using angle feature as input. While training this, we train with image using base brightness, that is, 0% darkness. When darker images are input to this MLP, the output image features will be incorrect. The second MLP (MLP2) takes the output of MLP1 and brightness information. It is trained to give correct image feature for the darker image. Finally, the output of MLP2 is compared to image feature of the probe face image to take the authentication decision.

4. Simulation Experiments and Results

As already mentioned, the system consists of two stages—learning of MLP, that is, the registration phase, and using the learned MLP in the authentication stage.

4.1. Registration Phase. When person "A" is to be registered, face photograph of person "A" is taken using multiple cameras set at different angles, as shown in Figure 2. We use the database [26] from Softopia, Japan. The database has images taken at an interval of 5 degrees. For registration, we use face image data at intervals of 10 degrees, from −50 degrees to +50 degrees. The registration system is shown in Figure 8. First, the image is converted to grey-scale image, face part is cut out, and the angle and independent component features of the face image are extracted. The angle feature is used as input to the MLP and the image features as teacher signal. From the database, 11 of such data are used for training. The training is converged within 5000 epochs, with very low mean square error.

4.2. Authentication Phase. In authentication phase, the person announces his/her identification and let the image be taken. The angle is arbitrary, depending on how the person poses in front of the camera. We assume this angle to be within −50 to +50 degrees. The mapping task of MLP is to interpolate. The layout of the authentication system is shown in Figure 9. From the camera image, the face part is cut out. The angle features are extracted and input to the MLP trained for the person, as retrieved from the database according to identification declaration. The image feature taken from the image and that obtained as output of the MLP are compared. The Euclidean distance is calculated. If the distance is below a threshold value, the person is accepted, and otherwise rejected. The Euclidean distance is calculated by (2) as follows:

$$R = \sqrt{\sum_{i=1}^{m} (NN_i - IN_i)^2}. \quad (2)$$

Here, $IN = \langle IN_1, IN_2, IN_3, IN_4, IN_5, IN_6, IN_7, IN_8 \rangle$ is image feature vector from input image. $NN = \langle NN_1, NN_2, NN_3, NN_4, NN_5, NN_6, NN_7, NN_8 \rangle$ is the image vector from MLP output. Judgment of the proper threshold value is important. If the threshold is too low, false accept rate (FAR) will increase. On the other hand, if the threshold is set too high, false reject rate (FRR) will be high. Depending on the application, the threshold is fixed. For a heavily secured place, where false acceptance is not tolerable at the cost of a few misjudgment in face rejection, the threshold is kept high. In general, the threshold is kept at a value where FAR is equal to FRR.

4.3. Experimental Setup and Results. Compared to our previous work, in the present work the angle feature vector has changed, from 6 elements to 12 elements. The image feature vector is also changed from PCA to ICA, the number of elements remaining the same 8. As the number of input nodes is increased, we increased the hidden nodes to 16 for faster training. We used face image data, taken in same illumination condition, with orientation angle from −50 degrees to +50 degrees, taken at intervals of 5 degrees. Image data at intervals 10 degrees was used for training, and the intermediate is for testing. In total, face image of 15 individuals was used. Experiments were performed by varying the threshold in steps.

Experimental results, for the angle variation from −50 degrees to +50 degrees, are summarized in Figure 10. Average FAR and FRR for all the images were calculated and plotted in this figure. The value of FAR and FRR at proper threshold is improved from our previous work about 20% to 10%, that is, an overall improvement of 10% in recognition rate over the whole range of angle variation. We attribute this to our improved angle feature vector. It is also important to note that the optimum threshold value is now increased from 9 to 12, and the slope around that threshold is lower. In the previous work, as shifting of threshold value greatly changed FAR and FRR, it is difficult to select proper threshold, as it would be different for different individuals. The new result

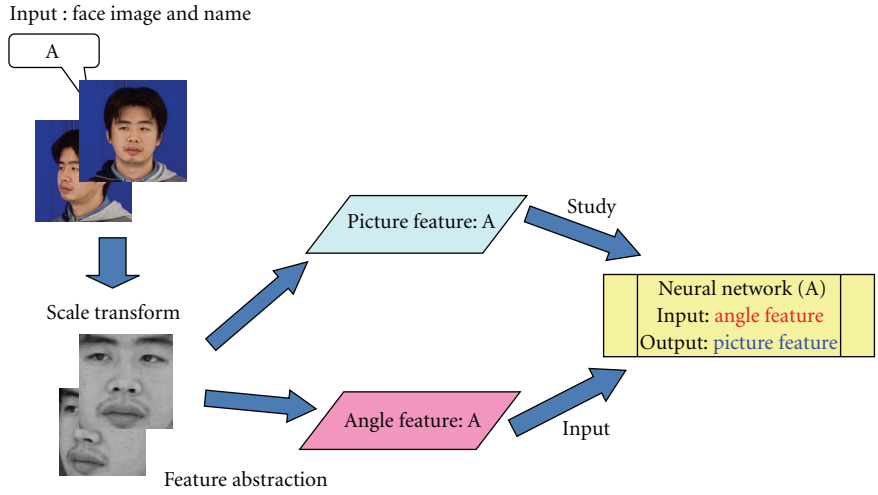

Figure 8: Registration phase which consists of taking the face image at different angles and use them to train an MLP.

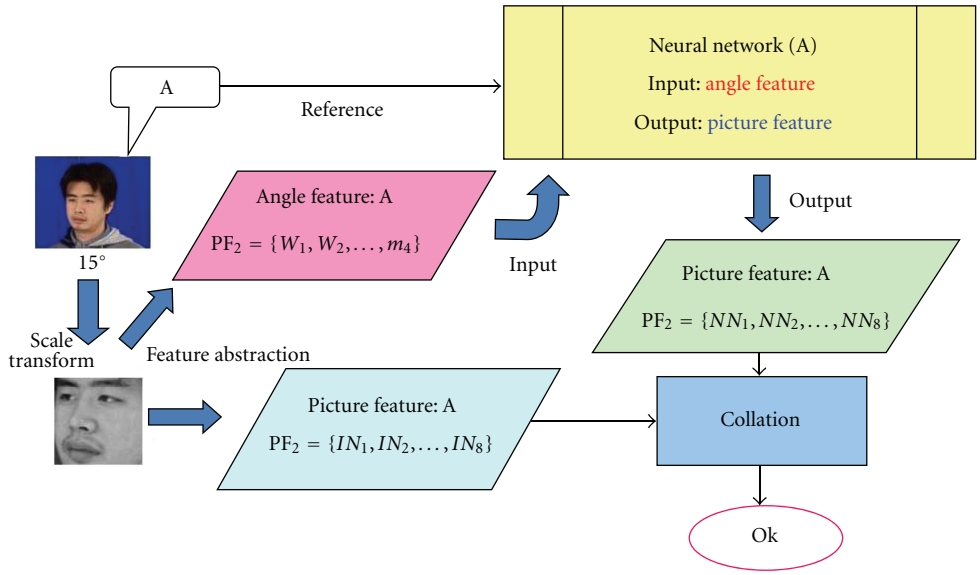

Figure 9: Description of the authentication phase.

shows that FAR and FRR do not change much when the threshold is changed.

4.4. Experiments with Changing Illumination. The image feature changes also with illumination condition. We did a preliminary experiment to investigate the image feature change with brightness and on the basis of our investigation proposed the robust systems for illumination variation presented in the earlier section.

To investigate the pattern of change, we varied the brightness of face image by steps of 4% (of the original brightness) to a level up to −80% of the original value. Here, maximum value of the brightness is considered to be 0%. The image features at different illumination levels are compared, in terms of Euclidean distance, with respect to the brightest image, that is, 0%. The results are summarized in Figure 11. Though the variation of image feature is different for different images, the nature is same.

As shown, the Euclidean distances are larger with the decrease of brightness values. The nature of variation is easy to be learned by ANN. From this, we conclude that, we can extend the proposed system to be able to perform well in case of illumination variation too.

4.5. Experiments with Extended System and Results. We compared our results for System I and System II. We use brightness of different image features at intervals of 4%, from −80% to 0%. The image features are the same. ICA features are used in Section 4.

All the experimental results are summarized in Figure 12 and Table 3. Figure 12 shows the average value of misidentification with variation of both orientation angle and brightness. Least misidentification remains almost unchanged

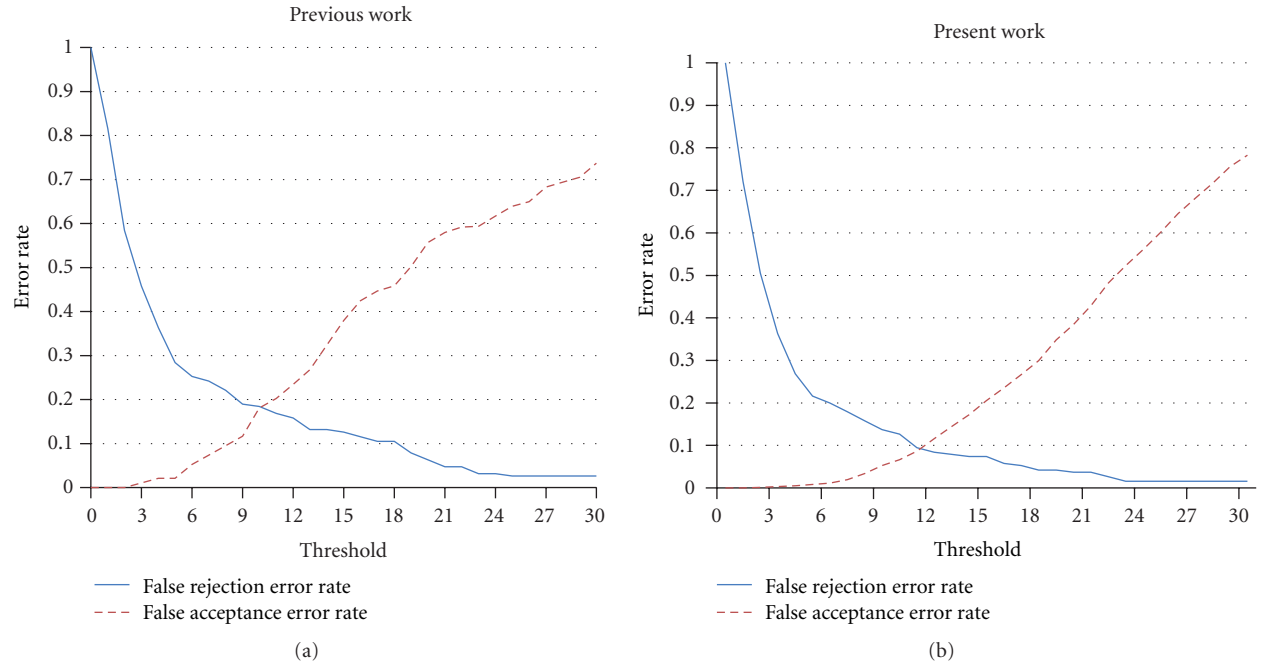

Figure 10: False rejection rate and false acceptance rate.

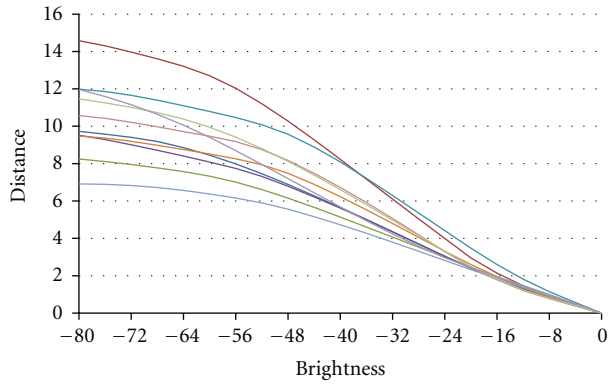

Figure 11: Euclidean distance of image feature as brightness changes.

Table 3: Misidentification at different brightness levels.

	Only angle	System I	System II
0%	0.090	0.182	0.127
–20%	0.114	0.205	0.114
–40%	0.287	0.199	0.172

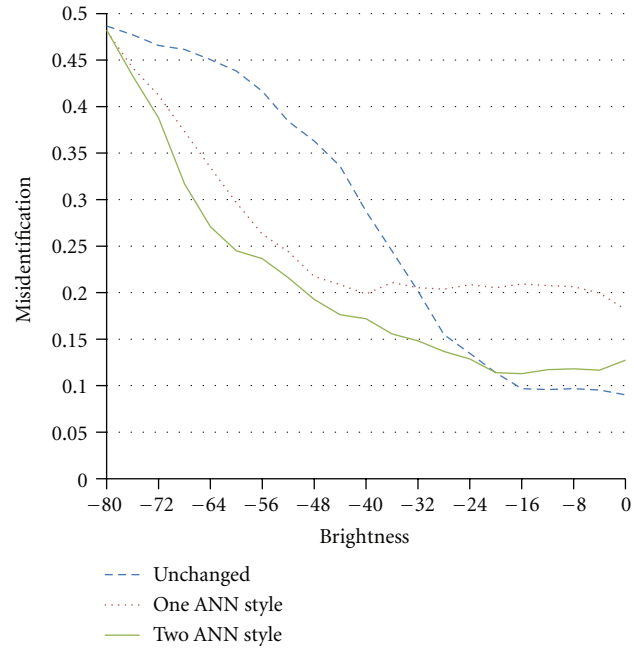

Figure 12: Error rate due to changes in brightness.

when brightness is reduced from 0% to 20%. It shows that when illumination is strong, there is no need to correct the original system. Misidentification using System I and System II is much less compared to the original system when the image is dark. To train System II, it takes more time and memory. But System II gives much better result. It is also found that System II's performance is consistent, and correct authentication rate steadily improves as image brightness decreases more and more.

5. Conclusion

In this work we have proposed an efficient technique for angle-aware face recognition and extended the same technique to take care of the effect of illumination variation. Though there are lots of works on angle invariant and illumination invariant face recognition proposed in the literature so far, there is a very few work in which same

framework is used for taking care of both the problems simultaneously. Our proposed system can take care of angle variation from −50 degrees to + 50 degrees and at the same time 40 sets and different image feature set. The results are now reliable to work with larger data set. We used only one data set and currently are engaged in using other data sets for simulation experiments.

In this work, we considered the change in angle orientation in the horizontal plane, but orientation in the vertical plane may also vary and affect face recognition. We would like to extend our work to take care of the change in orientation in the vertical plane. Further experiments to work with more bench mark data sets are also our future target.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[2] B. Chakraborty, "A novel ANN based approach for angle invariant face verification," in *Proceedings of the IEEE Symposium on Computational Intelligence in Image and Signal Processing (CIISP '07)*, pp. 72–76, April 2007.

[3] M. A. Turk and A. P. Pentland, "Face recognition using eigen faces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '91)*, pp. 586–591, June 1991.

[4] S. K. Zhou, R. Chellappa, and W. Zhao, *Unconstrained Face Recognition*, Springer, 2006.

[5] R. Gross, S. Baker, I. Matthews, and T. Kanade, "Face recognition scross pose and illumination," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds., Springer, 2004.

[6] P. J. Phillips et al., "Face recognition vendor test 2002: evaluation report," NISTIR 6965, 2003, http://www.frvt.org/.

[7] T. Maurer and C. von der Malsburg, "Single-view based recognition of faces rotated in depth," in *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pp. 248–253, 1995.

[8] D. Beymer, "Face recognition under varying pose," Technical Report 1461, MIT AI Laboratory, Cambridge, Mass, USA, 1995.

[9] S. Ullman and R. Basri, "Recognition by linear combinations of models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 10, pp. 992–1006, 1991.

[10] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[11] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pp. 84–91, June 1994.

[12] L. Wiskott, J. M. Fellous, N. Krüger, and C. D. Von Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[13] R. Gross, I. Matthews, and S. Baker, "Eigen Light-Fields and Face Recognition across Pose," in *Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition*, 2002.

[14] T. Kanade and A. Yamada, "Multi-Subregion based probabilistic approach toward pose invariant face recognition," in *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA '3)*, pp. 954–959, 2003.

[15] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 4, pp. 449–465, 2004.

[16] S. K. Zhou and R. Chellappa, "Image-based face recognition under illumination and pose variations," *Journal of the Optical Society of America A*, vol. 22, no. 2, pp. 217–229, 2005.

[17] G. Chakraborty, B. Chakraborty, J. C. Patra, and C. Pornavalai, "An MLP-based face authentication technique robust to orientation," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '09)*, pp. 481–488, Atlanta, Ga, USA, June 2009.

[18] J. Shermina and V. Vasudevan, "An efficient face recognition system based on the hybridization of invariant pose and illumination process," *European Journal of Scientific Research*, vol. 64, no. 2, pp. 225–243, 2011.

[19] H. F. Liau and D. Isa, "New illumination compensation method for face recognition," *International Journal of Computer and Network Security*, vol. 2, no. 3, pp. 5–12, 2010.

[20] J. Shermina, "Impact of locally linear regression and fisher linear discriminant analysis in pose invariant face recognition," *International Journal of Computer Science and Network Security*, vol. 10, no. 10, 2010.

[21] J. Shermina, "Illumination invariant face recognition using discrete cosine transform and principal component analysis," in *Proceedings of the International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT '11)*, pp. 826–830, March 2011.

[22] S. J. D. Prince and J. H. Elder, "Creating invariance to "nuisance parameters" in face recognition," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 446–453, June 2005.

[23] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-d objects from appearance," *International Journal of Computer Vision*, vol. 14, no. 1, pp. 5–24, 1995.

[24] D. Graham and N. Allison, "Face recognition from unfamiliar views: subspace methods and pose dependency," in *Proceedings of the International Conference of Automatic Face and Gesture Recognition*, pp. 348–353, 1998.

[25] W. Zhao and R. Chellappa, "A guided tour of face processing," in *Face Processing*, Zhaoo and Chellappa, Eds., Academic Press, 2006.

[26] http://www.softopia.or.jp/rd/facedb/top.html.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 593147, 7 pages
doi:10.1155/2012/593147

Research Article

An Application of Improved Gap-BIDE Algorithm for Discovering Access Patterns

Xiuming Yu,[1] Meijing Li,[1] Taewook Kim,[1] Seon-phil Jeong,[2] and Keun Ho Ryu[1,2,3]

[1] *Database and Bioinformatics Laboratory, Chungbuk National University, Cheongju 361-763, Republic of Korea*
[2] *Division of Science and Technology, BNU-HKBU United International College, Zhuhai 519-085, China*
[3] *Multimedia Systems Laboratory, School of Computer Science and Engineering, The University of Aizu, Aizu-Wakamatsu, Fukushima 965-8580, Japan*

Correspondence should be addressed to Xiuming Yu, yuxiuming@dblab.chungbuk.ac.kr

Received 9 March 2012; Revised 14 May 2012; Accepted 20 May 2012

Academic Editor: Qiangfu Zhao

Discovering access patterns from web log data is a typical sequential pattern mining application, and a lot of access pattern mining algorithms have been proposed. In this paper, we propose an improved approach of Gap-BIDE algorithm to extract user access patterns from web log data. Compared with the previous Gap-BIDE algorithm, a process of getting a large event set is proposed in the provided algorithm; the proposed approach can find out the frequent events by discarding the infrequent events which do not occur continuously in an accessing time before generating candidate patterns. In the experiment, we compare the previous access pattern mining algorithm with the proposed one, which shows that our approach is very efficient in discovering access patterns in large database.

1. Introduction

The web has become an important channel for conducting business transactions and e-commerce. Also, it provides a convenient means for us to communicate with each other worldwide. With the rapid development of web technology, the web has become an important and preferred platform for distributing and acquiring information. The data collected automatically by the web and application web servers represent the navigational behavior of web users, and such data is called web log data.

Web mining is a technology to discover and extract useful information from web log data. Because of the tremendous growth of information sources, increasing interest of various research communities, and the recent interest in e-commerce, the area of web mining has become vast and more interesting. It deals with data related to the web, such as data hidden in web contents, data presented on web pages, and data stored on web servers. Based on the kinds of data, there are three categories of web mining: web content mining, web structure mining, and web usage mining [1]. The Web usage data includes the data from web server access logs, proxy server logs, and browser logs. It is also known as web access patterns. Web usage mining tries to discover the access patterns from web log files. Web access tracking can be defined as web page history [2]; the mining task is a process of extracting interesting patterns in web access logs. There are so many techniques of mining web usage data including statistical analysis [3], association rules [4], sequential patterns [5–7], classification [8–10], and clustering [11–13]. Access pattern mining is a popular approach of sequential pattern mining, which extracts frequent subsequences from a sequence database [14]. Further, discovering access patterns is an important challenge in the field of web mining. And the popular applications of access patterns mining are obtaining useful information of web users' behavior.

A lot of studies have been proposed on access pattern mining for finding valuable knowledge from web log data, such as AprioriAll algorithm [15, 16] and GSP (generalized sequential pattern) algorithm [17]. All of above algorithms mine sequential patterns using a paradigm of candidate generate-and-test maintain a candidate set of already mined patterns in the mining process. When the data set is huge, it will generates a lot of candidate patterns. In other words, GSP algorithm needs much memory while the data set is large. The BIDE algorithms [18] mine frequent patterns without

ALGORITHM: gap-Bide (SDB, t_session, min_sup_les, min_sup, M, N)
INPUT: (1) SDB: An input sequence database with time, (2) t_session: the time user session, (3) min_sup_les: the minimum support threshold of getting large event set, (4) min_sup: the minimum support threshold of getting closed sequential pattern, (5) M and N: the parame-ters of a gap constraint.
OUTPUT: the set of gap-constrained closed sequential patterns.
(1) call getLargeEventSet (SDB, t_session, min_sup_les);
(2) select sequence from input database only contained in LES
(3) find the set of length-1 frequent sequential patterns, $L1$;
(4) for each item i in $L1$
(5) call patternGrowth(i);
(6) return

Algorithm 1: Improved Gap-BIDE algorithm.

keeping the candidate pattern sets, therefore it needs less space during the mining task. And above algorithms focus on finding out the patterns which are adjacent and that may miss some hidden relationships among noncontinuous patterns. So the constraint of gap should be considered. In the paper [19], the author proposed an improved BIDE algorithm (Gap-BIDE) for mining closed sequential patterns with gap constraint and considers the patterns that are not only adjacent but also noncontiguous; Gap-BIDE algorithm had been applied to web mining in [20]. And in the previous work [21], we have improved the Gap-BIDE algorithm by discarding infrequent events before generating frequent candidate events and applying the improved algorithm to access pattern mining and discussed the efficient of parameter of the values of gap. In this paper, we perform the improved algorithm and compare the efficiency with previous access pattern mining algorithms, such as GSP algorithm.

The rest of this paper is organized as follows. Section 2 presents the precedent of our algorithm compared with the original algorithm. Section 3 focuses on discovering access patterns, namely, preprocessing, pattern discovery, and result analysis, and it focuses on the efficiency of the proposed approach in terms of access pattern mining. In Section 4, we present an extensive performance study. Finally, we conclude this study in Section 5.

2. Algorithm of Improved Gap-BIDE

2.1. Gap-BIDE Algorithm. Gap-BIDE algorithm is presented in paper [19], and it inherits the same design philosophy as BIDE algorithm. It shares the same merit, that is, it does not need to maintain a candidate pattern set, which saves space consumption, and it can find some hidden relationships among the patterns that contend for the gap constraint.

The algorithm first finds the set of all frequent patterns, and it then mines the gap-constrained closed sequential patterns with pattern P as the prefix. In this process, it first scans the backward spaces of prefix pattern P, uses the gap-constrained backscan pruning method to prune search space, scans the forward spaces of prefix P, and uses the gap-constrained pattern closure checking scheme to check whether or not pattern P is closed; finally, it scans each forward space of all appearances of pattern P and finds the set of all locally frequent items, L, uses each item in L to extend P, and mines the gap-constrained closed sequential patterns for the new prefix by calling subroutine again.

In the algorithm, forward space is defined as that given an appearance of pattern $P[M, N]$ with triple (sid, beginPos, and endPos). The forward space of appearance is part of the sequence of range [endPos + M, endPos + N] $\cap$ [endPos, l), where l is the length of sequence sid. Here, the definition of forward space (FS) is induced for getting frequent subsequence patterns. We can get the sequence support of every subsequence by scanning the forward spaces of the appearances of a prefix pattern. The sequences whose supports are greater than or equal to the minimal support threshold Minsup will be the frequent subsequences patterns of a prefix pattern.

The definition of backward space (BS) is important, and it is defined as that given an appearance of pattern $P[M, N]$ with triple (sid, beginPos, and end-Pos). The backward space of appearance is part of the sequence sid that is of the range [beginPos $-N$, beginPos $-M$] $\cap$[0, beginPos).

Performance of proposed approach shows that Gap-BIDE is both runtime and space efficient in mining frequent, closed sequences with gap constraints.

2.2. Improved Gap-BIDE Algorithm. Although Gap-BIDE algorithm is advanced in the algorithms of sequential pattern mining, there are still a lot of fool's errands are done during the mining task, such as generating some candidate patterns for infrequent events in the original data set. To avoid the unnecessary memory use, an improved algorithm is proposed. Our algorithm is designed based on the Gap-BIDE algorithm; the main idea is to discard infrequent events before generating frequent candidate events; we call this process as getting a large event set.

Algorithm 1 is the main algorithm. The Algorithm 2 is a subroutine of Algorithm 1; it proposes the process of

ALGORITHM: getLargeEventSet (SDB, t_session, min_sup_les)
INPUT: (1) SDB: An input sequence database with time, (2) t_session: the time user session, (3) min_sup_les: the minimum support threshold of getting large event set.
OUTPUT: LES: large event set.
(7) scan sequence database; find all candidate events [$\langle E1\rangle$, $\langle E2\rangle$,. . .,$\langle Ej\rangle$]
(8) group sequences by IP address and t-session; find all sessions [$S1$, $S2$,. . .,Sm]
(9) for each candidate event Ej in session Sm
(10) calculate support for Ej
(11) if (support of $Ej \geq$ min_sup_les)
(12) output event Ej to LES
(13) return

Algorithm 2: Get large event set.

ALGORITHM: patternGrowth (P)
INPUT: (1) P: prefix sequence pattern.
OUTPUT: the set of gap-constrained closed sequential patterns with prefix P.
(14) backward_check (P needPruning, hasBackwardExtension)
(15) if (needPruning)
(16) return;
(17) forward_check(P, hasForwardExtension);
(18) if ! (hasBackwardExtension || hasForwardExtension)
(19) output pattern P;
(20) search each forward space of all appearances of P, and find the set of all local frequent items, L;
(21) for each item i in L
(22) build new pattern $P_{\text{new}} = P + i$;
(23) call patternGrowth (P_{new});
(24) return.

Algorithm 3: Generate closed sequential patterns.

getting a large event set. A large event set (LES) is an event set that contains the events that satisfy a user specified minimum support threshold. The events in LES represent the transactions or objects with large proportion in the entire data set. In this paper, a web log file denotes the data set, and one web page is defined as an event; thus, LES denotes the set of web pages that are accessed by web users with enough frequency in a period of time. In this mining process, the generation sequence through LES can reduce the number of test data to improve the efficiency and accuracy of the mining task. After obtaining large event set, sequence data with only large events are generated. Then the algorithm scans the generated database, finds the set of all frequent items with length (length-1), and calls Algorithm 3 iteratively. Algorithm 3 patternGrowth (P) is the other subroutine of Algorithm 1; it proposes the process to mine the gap-constrained closed sequential patterns with pattern P as the prefix.

An important definition for generating LES is the user session. The user session is an activity that a user with a unique IP address spends on a web page during a specified period of time. It can be used to identify a continuous access to user statistics visits by this measure. The specified period of time is determined via a cookie, also known as web cookie and HTTP cookie, which can be set by the server with or without an expiration date, modified by web designer and is set to a default value of 600 seconds. Within the expiration date, the access of web user is effective.

3. Discovery of Access Patterns

In this section, the process of mining task is discussed.

3.1. Data Preprocessing. Web log files reside on the web servers that record the activities of clients who access the web server via a web browser. Traditionally, there have been many types of web log files including error logs, access logs, and referrer logs. In this paper, data in the web access log is defined as the raw data. The web access log records all requests that are processed by the web server. Data in the

log file contains some missing value data and irrelevant attributes; it cannot be directly used for the mining task. In this section, we describe the process of data cleaning and attribute selection to remove unwanted data.

(1) *Data cleaning*: removing irrelevant data.

(a) *Remove the records with URLs of jpg, png, gif, js, css, and so on, which are automatically generated when a web page is requested.*

(b) *Remove the data with wrong statue numbers that start with the numbers 4 or 5*. These wrong records are caused by the error of requests or server. For example, the HTTP client error: 400 Bad Request and 404 Not Found and HTTP server error: 500 Internal Server Error and 505 HTTP Version Not Supported.

(c) Discard missing value data that are caused by breaking a web page while loading.

(2) *Attribute selection*: removing the irrelevant attributes. There are many attributes in one record of web log file. In this paper, we need the attributes of IP Address, Time, and URL; thus, the rest of attributes of method, status, size, and so on, need to be discarded.

(3) *Transformed URLs into code numbers.*

It is difficult to distinguish the requested URLs of web log data in thousands of records. There are typically dozens of kinds of web pages in thousands of records. So, the URLs can be transformed into code numbers for simplicity. For example, a web log data that comes from the server of website http://www.vtsns.edu.rs/, and there are 31 different kinds of web pages that have been accessed. We transform their URLs into code numbers, such as galerija.php → 1, nenastavno_osoblje.php → 15, and rezultati_ispita.php → 21.

We choose a set of data from a web log file as an example data. After data preprocessing, we get the clean data shown in Table 1.

3.2. Process of Discovering Access Patterns. In this section, we present the process of discovering access patterns with an example.

After data preprocessing, we apply the algorithm to web log data. Then, LES is generated with sorting the data in Table 1 by the attributes of IP Address and Time; here, the time of user session is defined as one hour for simplicity. Then, these data are grouped by one hour for each web user; finally, the sorted data is shown in Table 2.

Then, we calculate the support of each event. For example, for the event ⟨2⟩, it occurs three times, which are in "82.117.202.158" at time 2, in "82.208.207.41" at time 2, and in "82.208.255.125" at time 2. After calculating of events support, the candidate event set is obtained as shown in Table 3.

Finally, a user specified minimum support threshold (MinSup) must be defined. MinSup denotes a kind of abstract level that is a degree of generalization. Choosing MinSup is very important; if it is low, then we can get a detailed event. If it is high, then we can get general events. In this example, MinSup is defined as 75%. In other words, if a web page is accessed by greater than or equal to 75% web users, then this web page can be denoted as a large event. After the process of getting large event set, the LES is obtained as shown in Table 4.

Table 1: Example data.

No.	IP address	Time	URL
1	82.117.202.158	01:12:18	4
2	82.117.202.158	01:12:22	1
3	82.208.207.41	01:12:43	4
4	83.136.179.11	01:22:43	4
5	83.136.179.11	01:23:43	3
6	82.208.207.41	02:12:23	4
7	82.208.207.41	02:12:25	7
8	82.208.207.41	02:13:43	2
9	82.117.202.158	02:17:26	6
10	82.117.202.158	02:17:39	2
11	83.136.179.11	02:17:41	6
12	82.208.255.125	02:17:44	6
13	82.208.255.125	02:17:53	2
14	82.117.202.158	03:12:42	7
15	83.136.179.11	03:27:23	4
16	83.136.179.11	03:37:32	5
17	82.208.255.125	03:37:44	7
18	83.136.179.11	04:13:43	7
19	82.117.202.158	04:17:26	4
20	82.208.255.125	05:17:39	6
21	82.208.255.125	05:17:41	7
22	82.208.207.41	05:18:40	7
23	82.117.202.158	05:37:53	6
24	82.117.202.158	05:39:42	7
25	83.136.179.11	06:27:23	6
26	83.136.179.11	06:37:32	7
27	82.117.202.158	01:12:18	4

Table 2: Sorted data.

IP address	Time	Event
82.117.202.158	1	4, 1
82.117.202.158	2	6, 2
82.117.202.158	3	7
82.117.202.158	4	4
82.117.202.158	5	6, 7
83.136.179.11	1	4, 3
83.136.179.11	2	6
83.136.179.11	3	4, 5
83.136.179.11	4	7
83.136.179.11	6	6, 7
82.208.207.41	1	4
82.208.207.41	2	4, 7, 2
82.208.207.41	3	7
82.208.255.125	2	6, 2
82.208.255.125	3	7
82.208.255.125	5	6, 7

Table 3: Candidate event set.

Event	Support
1	1
2	3
3	1
4	3
5	1
6	3
7	4

Table 4: Large event set.

Event	Support
2	3
4	3
6	3
7	4

After obtaining LES, the infrequent events ⟨1⟩, ⟨3⟩, and ⟨5⟩ are removed from Table 2, and the events are then transformed into a set of tuples (sequence identifier, sequence). We define the IP Address as the sequence identifier and define the event as a sequence. The sequence set is shown in Table 5.

Then, we call the original Gap-BIDE algorithm to find the frequent sequential pattern and prune the patterns. Here, gap is defined as $g(M, N)$, where M is the value of minimum gap, and N is the value of the maximum gap. Assume a pattern P with $g(M, N)$, which can be expressed as $P[M, N]$. This approach is presented like the description of timing constrains with the mingap and maxgap. If the value of M-N is D, then the events in a sequence must occur within D of the events occurring in the previous event.

After calling our improved algorithm, we get the closed patterns as shown in Table 6.

Useful information can be found from the experimental result. The relationships of web pages are known easily, and user behavior information is shown directly. Each number in the output sequential patterns represents a website or a web user request. For example, the numbers 6 and 7 represent web pages ispit_raspored_god.php and upis_prva.php, respectively. For the closed sequential pattern [6, 7] shown in Table 6, it means 75% (3 out of 4 user sessions) of the web users who access web page upis_prva.php tend to always visit web page ispit_raspored_god.php first. According to the relationship between these two web pages, the design of web pages can be improved. For example, the web designer can add a hyperlink into web page ispit_raspored_god.php that points to web page upis_prva.php. This approach can be applied in many areas. For instance, in the electronic shopping cart, when customers complete their shopping, there can be some hyperlinks in the finished web page that point to some related web pages according to the mining result of purchase history. When web users watch a movie, some hyperlinks that point to some web pages of related movies on the site must be present.

Table 5: Sequence set.

Sequence identifier	Sequence
82.117.202.158	4
82.117.202.158	6, 2
82.117.202.158	7
82.117.202.158	4
82.117.202.158	6, 7
83.136.179.11	4
83.136.179.11	6
83.136.179.11	4
83.136.179.11	7
83.136.179.11	6, 7
82.208.207.41	4
82.208.207.41	4, 7, 2
82.208.207.41	7
82.208.255.125	6, 2
82.208.255.125	7
82.208.255.125	6, 7

Table 6: Closed patterns.

No.	Pattern	Support
1	[4, 7]	3
2	[6, 7]	3
3	[6, 7, 7]	3
4	[7, 7]	4

4. Experimental Result and Analysis

4.1. Effect of Parameter in the Process of Getting Large Event Set. The process of getting a large event set aims at extracting the events that satisfy a user defined minimum support of large event set. It can discard the infrequent events to reduce the size of experimental database for reducing the search space and time and maintaining the accuracy of the whole process of mining task. To evaluate the parameter effect, we compare the numbers of large events by changing the values of the minimum support of large event set (MSLE). In this experiment, the experimental data records the access information of website (http://www.vtsns.edu.rs/), which is an institution's official website. The number of original records in the web log file is 5999, and after data preprocessing, there are 269 user sessions in the records. The experimental result is shown in Figure 1. We can see that the smaller the minimum support are, the more generalized the obtained LES becomes. There always exists a value of minimum support, and from the value, the number of large events will not change, or will change very little. This value is always selected to be used as the value of minimum support in the experiment.

4.2. Comparing with Original Gap-BIDE Algorithm. In this section, we compare our algorithm with the original Gap-BIDE algorithm [19]. The experimental data come from internet information server (IIS) logs for msnbc.com and

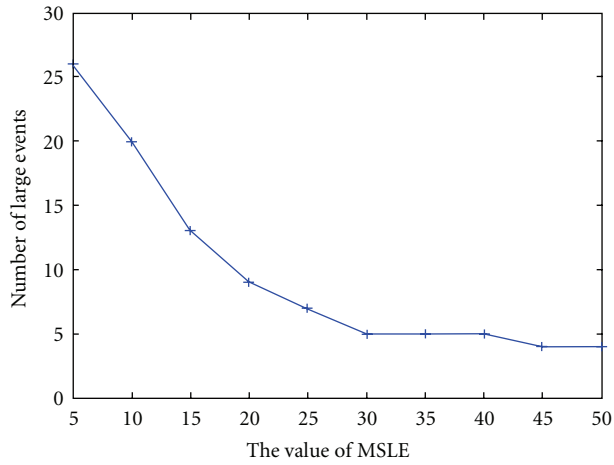

Figure 1: Effect of parameter in the process of getting large event set.

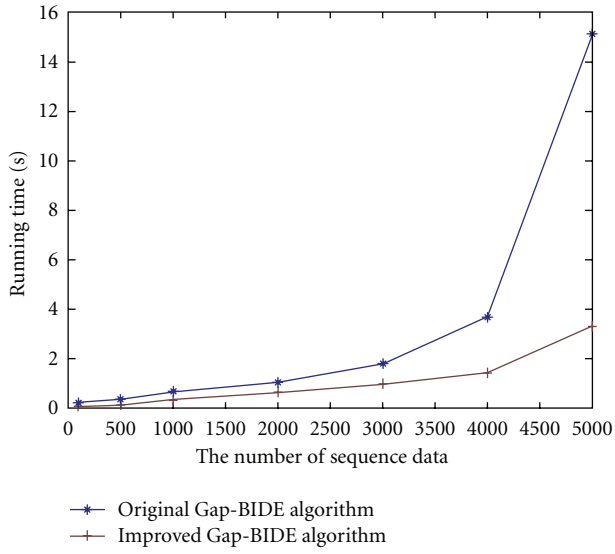

Figure 2: Comparing with original Gap-BIDE.

news-related portions of msn.com for the entire day of September 28, 1999. Each sequence in the dataset corresponds to page views of a user during that twenty-four hour period. Each event in the sequence corresponds to a user's request for a page. There are 989818 anonymous user sessions; we choose the test data by the approach of simple random sampling without replacement from these data. In the experiment, we define minimum support threshold of large event set as 20, minimum support of closed sequential pattern as 10, and the value of gap as $[0, 2]$. We implemented the experiment on a 2.40-GHz Pentium PC machine with 4.00 GB main memory and ran the algorithm in Python 2.7 with JDK 1.6.0. Then, the experimental result is shown in Figure 2. It shows that when applying our proposed algorithm, the cost of time is less than that of the original Gap-BIDE algorithm.

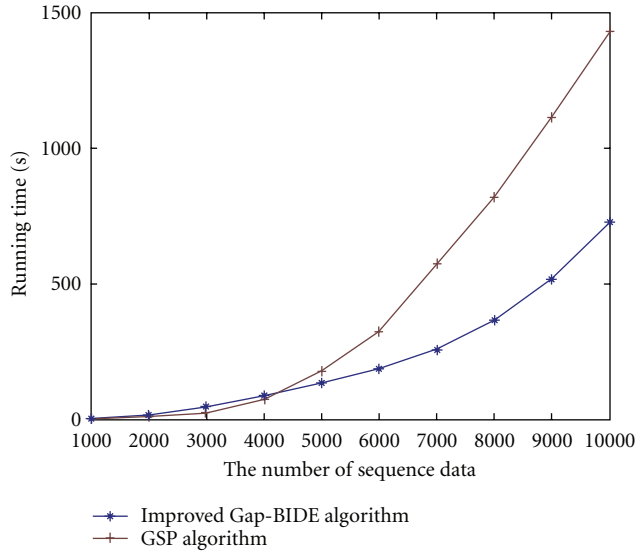

Figure 3: Comparing with GSP algorithm.

4.3. Comparing with GSP Algorithm. Previous studies have shown that our proposed algorithm is more effective than original Gap-BIDE algorithm when we apply the algorithms on discovering access patterns. In this section, we want to prove that our proposed algorithm is more effective than previous access pattern mining algorithm. To validate it, we compare our algorithm and GSP algorithm proposed in [17] with an experiment. The experimental data come from Internet information server (IIS) logs for msnbc.com and news-related portions of msn.com for the entire day of September 28, 1999, and we choose the test data by the approach of simple random sampling without replacement from these data. In the experiment, we define minimum support of closed sequential pattern as 10 and the experimental result is shown in Figure 3. It shows that when applying our proposed algorithm to large database, the cost of time is less than that GSP algorithm.

5. Conclusion

In this paper, we presented the application of improved Gap-BIDE algorithm for discovering closed sequential patterns in web log data. We improve the algorithm by discarding all infrequent events before generating the frequent candidate events. In the process of data preprocessing, we removed the irrelevant attributes and transformed URLs into code numbers for simplicity, and we removed the missing value data to improve the quality of data. For getting experimental data for the mining task, we transformed the web log data into sequences based on the time constraint. The value of time is determined by an expiration date of the cookies. As a result, we obtained new web access patterns that expressed the order in which websites were access based on the Gap-BIDE algorithm. Compared with the previous web mining approaches, the proposed approach achieves the best performance in terms of getting a large event set of sequence. It reduces the sequences to get more effective and accurate

results. We performed some experiments to compare our algorithm with previous algorithms. The experiments show that our algorithm uses less time than the original Gap-BIDE algorithm and cost less time than GSP algorithm in discovering access patterns in large database. In future work, we will try to find a more efficient algorithm for mining the closed gap constraint sequential patterns and will try to achieve a more efficient way for transforming web log files into sequence patterns.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (no. 2012-0000478).

References

[1] L. K. J. Grace, V. Maheswari, and D. Nagamalai, "Analysis of web logs and web user in web mining," *International Journal of Network Security & Its Applications*, vol. 3, no. 1, 2011.

[2] K. Saxena and R. Shukla, "Significant interval and frequent pattern discovery in web log data," *International Journal of Computer Science Issue*, vol. 7, no. 1, 2010.

[3] K. Suresh and S. Paul, "Distributed linear programming for weblog data using mining techniques in distributed environment," *International Journal of Computer Applications (0975–8887)*, vol. 11, no. 7, 2010.

[4] Y. Wang, J. Le, and D. Huang, "A method for privacy preserving mining of association rules based on web usage mining," in *International Conference on Web Information Systems and Mining (WISM '10)*, vol. 1, pp. 33–37, IEEE Computer Society Washington, Washington, DC, USA, 2010.

[5] C. Wei, W. Sen, Z. Yuan, and L. C. Chang, "Algorithm of mining sequential patterns for web personalization services," *ACM SIGMIS Database*, vol. 40, no. 2, pp. 57–66, 2009.

[6] J. Zhu, H. Wu, and G. Gao, "An efficient method of web sequential pattern mining based on session filter and transaction identification," *Journal of Networks*, vol. 5, no. 9, pp. 1017–1024, 2010.

[7] X. Yu, M. Li, and H. Kim, "Mining access patterns using temporal interval relational rules from web logs," in *Proceedings of the 4th International Conference (FITAT/DBMI '11)*, pp. 80–83, 2011.

[8] M. Santini, "Cross-testing a genre classification model for the web," *Genres on the Web*, vol. 42, Part 3, pp. 87–128, 2011.

[9] J. J. Rho, B. J. Moon, Y. J. Kim, and D. H. Yang, "Internet customer segmentation using web log data," *Journal of Business & Economics Research*, vol. 2, no. 11, 2004.

[10] N. Kejžar, S. K. Èerne, and V. Batagelj, "Network analysis of works on clustering and classification from web of science," in *Proceedings of the 11th Conference of the International Federation of Classification Societies (IFCS '10)*, Part 3, pp. 525–536, 2010.

[11] G. Xu, Y. Zong, and P. Dolog, "Co-clustering analysis of weblogs using bipartite spectral projection approach," in *Proceedings of the 14th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES '10)*, vol. 6278, pp. 398–407, 2010.

[12] A. A. O. Makanju, A. N. Zincir-Heywood, and E. E. Milios, "Clustering event logs using iterative partitioning," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 1255–1263, July 2009.

[13] J. Wang, Y. Mo, B. Huang, and J. Wen, "Web search results clustering based on a novel suffix tree structure," in *Proceedings of the 5th International Conference on Autonomic and Trusted Computing (ATC '08)*, vol. 5060, pp. 540–554, 2008.

[14] J. Chen and T. Cook, "Mining contiguous sequential patterns from web logs," in *Proceedings of the 16th International World Wide Web Conference (WWW '07)*, pp. 1177–1178, May 2007.

[15] M. Saravanan and B. Valaramathi, "Generalization of web log datas using WUM technique," in *Proceedings of the 12th International Conference on Networking, VLSI and signal processing (ICNVS '10)*, pp. 157–165, 2010.

[16] N. R. Mabroukeh and C. I. Ezeife, "A taxonomy of sequential pattern mining algorithms," *ACM Computing Surveys*, vol. 43, no. 1, article 3, 2010.

[17] S. Ramakrishnan and A. Rakesh, "Mining sequential patterns: generalizations and performance improvements," *Lecture Notes in Computer Science*, vol. 1057, pp. 3–17, 1996.

[18] J. Wang, J. Han, and C. Li, "Frequent closed sequence mining without candidate maintenance," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 8, pp. 1042–1056, 2007.

[19] C. Li and J. Wang, "Efficiently mining closed subsequences with gap constraints," in *Proceedings of International Conference on Data Mining (SIAM '08)*, April 2008.

[20] X. Yu, M. Li, D. G. Lee, K. D. Kim, and K. H. Ryu, "Application of closed gap-constrained sequential pattern mining in web log data," in *Proceedings of the 2nd International Conference of Electrical and Electronics Engineering (ICEEE '11)*, pp. 649–657, 2011.

[21] X. Yu, M. Li, H. Kim, D. G. Lee, and K. H. Ryu, "A novel approach to mining access patterns," in *Proceedings of the 3rd International Conference on Awareness Science and Technology*, pp. 346–352, 2011.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 183610, 8 pages
doi:10.1155/2012/183610

Research Article

Emotion-Aware Assistive System for Humanistic Care Based on the Orange Computing Concept

Jhing-Fa Wang, Bo-Wei Chen, Wei-Kang Fan, and Chih-Hung Li

Department of Electrical Engineering, National Cheng Kung University, Tainan 70101, Taiwan

Correspondence should be addressed to Jhing-Fa Wang, wangjf@mail.ncku.edu.tw

Received 10 February 2012; Accepted 12 April 2012

Academic Editor: Qiangfu Zhao

Mental care has become crucial with the rapid growth of economy and technology. However, recent movements, such as green technologies, place more emphasis on environmental issues than on mental care. Therefore, this study presents an emerging technology called orange computing for mental care applications. Orange computing refers to health, happiness, and physiopsychological care computing, which focuses on designing algorithms and systems for enhancing body and mind balance. The representative color of orange computing originates from a harmonic fusion of passion, love, happiness, and warmth. A case study on a human-machine interactive and assistive system for emotion care was conducted in this study to demonstrate the concept of orange computing. The system can detect emotional states of users by analyzing their facial expressions, emotional speech, and laughter in a ubiquitous environment. In addition, the system can provide corresponding feedback to users according to the results. Experimental results show that the system can achieve an accurate audiovisual recognition rate of 81.8% on average, thereby demonstrating the feasibility of the system. Compared with traditional questionnaire-based approaches, the proposed system can offer real-time analysis of emotional status more efficiently.

1. Introduction

During the past 200 years, the industrial revolution has caused a considerable effect on human lifestyles [1, 2]. A number of changes occurred [3] with the rapid growth of the economy and technology, including the information revolution [3], the second industrial revolution [4], and biotechnology development. Although such evolution was considerably beneficial to humans, it has caused a number of problems, such as capitalism, utilitarianism, poverty gap, global warming, and an aging population [1, 2]. Because of recent changes, a number of people recognized these crises and appealed for effective solutions [5], for example, the green movement [6], which successfully creates awareness of environmental protection and leads to the development of green technology or green computing. However, the green movement does not concentrate on body and mind balance. Therefore, a solution that is feasible for shortening the discrepancy between technology and humanity is of utmost concern.

In 1972, the King of Bhutan proposed a new concept that used gross national happiness (GNH) [7] to describe the standard of living of a country, instead of using gross domestic product (GDP). The GNH has attracted considerable attention because it measured the mental health of people. Similar ideas were also proposed in other works. For example, Andrew Oswald advocated Happiness Economics [8] by combining economics with other research fields, such as psychology and sociology. Moreover, a book entitled "Well-Being" [9], which was written by Daniel Kahneman (a Nobel Prize winner in Economic Sciences in 2002) explained the fundamentals of happy psychology. The common objective of those theories is to upgrade the living quality of humans and to bring more happiness into our daily lives. Recently, the IEEE launched the humanitarian technology challenge (HTC) project (http://www.ieeehtc.org/) [10] by sponsoring resource-constrained areas to build reliable electricity and medical facilities. Such an action also highlights the importance of humanistic care. Similar to the HTC project, Intel has supported a center for aging services technologies

(CAST) (http://www.agingtech.org/), and its objective is to accelerate development of innovative healthcare technologies. Several academic institutes responded to the trend and subsequently initiated medical care research, such as the "CodeBlue" project at Harvard University [11] and "Computers in the Human Interaction Loop" (CHIL) at Carnegie Mellon University [12]. Inspired by those related concepts [1, 2, 6, 8–12], this study devised a research project for studying the new interdisciplinary "Orange Technology" to promote health, happiness, and humanistic care.

Instead of emphasizing the relations between environments and humans, as proposed by green technology, the objective of the orange computing project is to bring more care or happiness to humans and to promote mental wellness for the well-being of society.

Orange computing is an interdisciplinary field that includes computer science, electrical engineering, biomedical engineering, psychology, physiology, cognitive science, and social science.The research scope of orange computing contains the following.

(1) Health and security care for the elderly, children, and infants.

(2) Care and disaster relief for people in disaster-stricken areas.

(3) Care for low-income families.

(4) Body-mind care for people with physiological and psychological problems.

(5) Happiness indicator measurement and happiness enhancement.

To demonstrate the concept of orange computing, a case study on a human-machine interactive and assistive system for emotion care was investigated in this study. The proposed system is capable of recognizing human emotions by analyzing facial expressions and speech. When the detected emotion status exceeds a threshold, an alarm will be send to a doctor or a nurse for further diagnosis and treatment.

The remainder of this paper is organized as follows: Section 2 introduces the orange computing models; Section 3 presents a discussion of a case study on the emotion recognition system for care services; Section 4 summarizes the performance of the proposed method and the analysis results; lastly, Section 5 offers conclusions.

2. Related Work and Orange Computing Concept

Orange computing originates from health informatics, and it contains two research topics: one is physiological care and the other psychological care. Both of the two topics focus on enhancing humans' physical and mental health, enriching positive emotions and finally bring more happiness to others [13, 14]. The physiological and psychological care models of orange computing are similar to the health model in medical expert systems [15, 16], which have been well developed and commonly used in health informatics over several decades.

In a medical expert system, when a user inputs a query through the interface, the system can automatically search predefined knowledge databases and consult with relevant experts or doctors. After querying databases or merging opinions of experts, the system subsequently replies to the user with an appropriate response. In traditional medical expert systems, database querying and feedback usually involve semantic understanding techniques and delicate interface design [17–19], so that users do not feel inconvenient during the process. However, in some telemedical care systems, such as [20], knowledge databases and feedback mechanisms are replaced with caregivers for better interactivity. Recently, expert systems have gradually integrated knowledge-based information management systems with pervasive computing [21]. Although such systems have been prototyped and modeled in several studies [22, 23], they have not been deployed. However, the abovementioned ideas have spurred the development of orange computing.

Happiness informatics, or the happiness model, is the key characteristic of orange computing. Similar to the health model, the happiness model also requires a user input and a predefined database. The input is commonly measured from the biosignals or behavior of a user, for example, facial expressions, emotional speech, laughter, body gestures, gaits, blood pressure, heartbeat rates, electroencephalograms (EEGs), electrocardiograms (ECGs), and electromyograms (EMGs) [24, 25]. With such information, the happiness model can help users evaluate their emotional status in various applications. Nevertheless, it is quite challenging to determine the manner in which to combine those data and determine emotional status [26–28].

3. Case Study

This section demonstrates a technological application for daily humanistic care in home environments. The system uses contactless multimodal recognition techniques to measure positive emotion degree of users. The recognition results can be logged into the database and sent to analysts for further processing. As shown in Figure 1, the ambient devices of the proposed system include multiple audiovisual sensors, a service robot, and a smart TV. The robot is a self-propelled machine with four wheels and serves as a remote agent between users and the server. To interact with users, it is equipped with audiovisual sensors, loudspeakers, and a touch screen. Similar to the robot, the TV is also used for interacting with users.

After the ambient sensors receive signals from users, data are subsequently sent to a processing server through a cloud network. The workflow of the data processing procedures comprises three stages, as follows: the first and second stages are the audiovisual recognition, and the last stage is the feedback stage. The detail of each stage is described as follows.

3.1. Visual Recognition. At the image processing stage, as shown in Figure 2, after video streams are captured by the camera, Haar-like features [29] are extracted and sent to

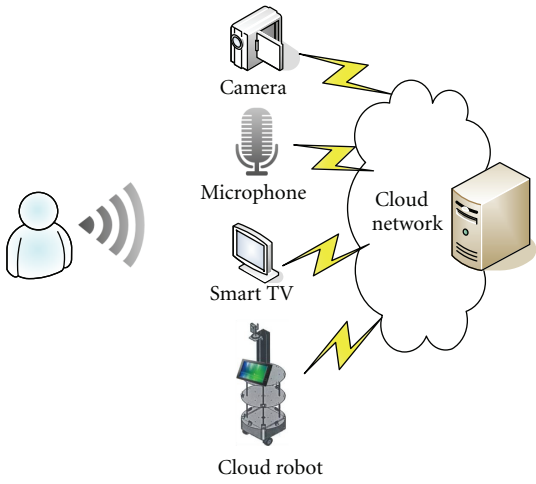

Figure 1: Framework of the system.

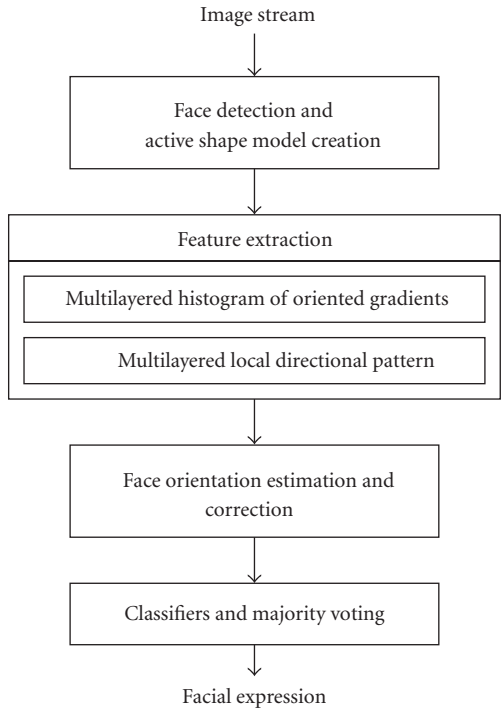

Figure 2: Workflow of the image processing stage.

AdaBoost classifiers [29] to detect user faces. Subsequently, the system uses the Active Shape Model, which was proposed by Cootes et al. [30], to model facial regions. Thus, facial regions can be represented by a set of points using the point distribution model.

A novel feature called "Multilayered Histogram of Oriented Gradients" (MLHOGs) is proposed in this study to generate reliable characteristics for estimating facial expressions. The MLHOGs are derived from Histograms of Oriented Gradients (HOGs) [31] and Pyramid Histograms of Oriented Gradients (PHOGs) [32]. Let $f(x, y)$ represent the pixel of coordinate x and y, G denote gradients, W refer

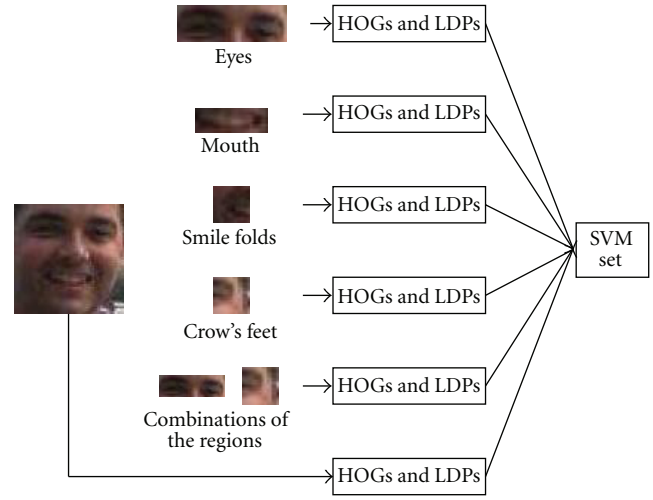

Figure 3: Concept of multilayered histogram of oriented gradients and multilayered local directional patterns (the facial image is extracted from the MPLab GENKI database).

to the weight of a coordinate, and θ be edge directions. The histogram of oriented gradients can be expressed as follows:

$$\begin{aligned} G_{x,y}^{\text{Horizontal}} &= 2f(x+1,y) - 2f(x-1,y) + f(x+1,y+1) \\ &\quad - f(x-1,y+1) + f(x+1,y-1) \\ &\quad - f(x-1,y-1), \\ G_{x,y}^{\text{Vertical}} &= 2f(x,y+1) - 2f(x,y-1) + f(x+1,y+1) \\ &\quad - f(x+1,y-1) + f(x-1,y+1) \\ &\quad - f(x-1,y-1), \\ W_{x,y} &= \left(\left(G_{x,y}^{\text{Horizontal}}\right)^2 + \left(G_{x,y}^{\text{Vertical}}\right)^2\right)^{1/2}, \\ \theta_{x,y} &= \tan^{-1}\left(\frac{G_{x,y}^{\text{Horizontal}}}{G_{x,y}^{\text{Vertical}}}\right). \end{aligned} \tag{1}$$

After gradients are computed, a histogram of edge directions is subsequently created to collect the number of pixels that belongs to a direction.

Unlike pyramid histogram of oriented gradients, which concentrates on fixed rectangular shapes inside an image, the proposed MLHOGs are modeled by object-based regions of interest (ROIs), such as eyes, mouths, noses, and combinations of ROIs. Furthermore, each objected-based ROI has a dedicated classifier for recognizing the same type of ROIs. A concept example of multilayered histogram of oriented gradients and multilayered local directional patterns is illustrated in Figure 3.

Similar to the proposed MLHOGs, our study also develops a new texture descriptor called "Multilayered Local Directional Pattern" for enhancing recognition rates. Such multilayered directional patterns are computed according to

"edge responses" of pixels, which are based on the same concept of Jabid's feature, "Local Directional Patterns (LDPs)" [33]. The difference is that the proposed method focuses on patterns at various ROI levels. Computation of multilayered local directional patterns is listed as follows:

$$\mathbf{R}_{\psi} = \mathbf{F} * \mathbf{M}_{\psi}, \tag{2}$$

$$\varepsilon_{\psi} = \sum_{\forall \text{block}_{3\times 3}} \text{LDP}_{\text{Binary Code}}\left(\mathbf{R}_{\psi}\right), \tag{3}$$

where **F** is the input image, **M** means eight-directional Kirsch edge masks like Sobel operators, **R** stands for edge responses of **F**, ψ represents eight directions, and ε is the number of edge responses in a designated direction. Before the system accumulates the edge responses of **R** using (3), an LDP binary operation [33] is imposed on **R** to generate an invariant code. A one-by-eight histogram is adopted to collect the edge responses in the eight directions. In the proposed multilayered local directional patterns, only edge responses in objects of interest are collected, so that the histogram differs from ROIs to ROIs.

In addition to upright and full frontal faces, this work also supports roll/yaw angle estimation and correction. The active shape model can label facial regions. Relative positions, proportions of facial regions, and orientations of nonfrontal faces can be measured properly with the use of spatial geometry. Once the direction is determined, corresponding transformation matrices are applied to the nonfrontal faces for pose correction.

At the end of the image processing stage, multiple Support Vector Machines (SVMs) are used to classify facial expressions. Each of the SVMs is trained to recognize a specific facial region. The classification result is generated by majority voting.

3.2. Audio Recognition. Audio signals and visual data have a considerable effect on deciphering human emotions. Therefore, the audio processing stage focuses on detecting emotional speech and laughter to extract emotional cues from acoustic signals. The workflow at this stage is illustrated in Figure 4.

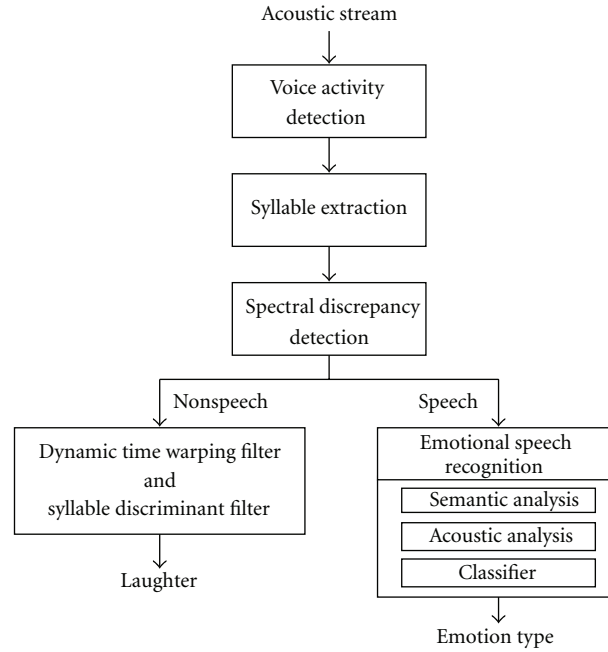

Figure 4: Workflow of the audio processing stage.

First, silence segments in audio streams are removed by using voice activity detection (VAD) algorithm. Subsequently, an autocorrelation method called "Average Magnitude Difference Function" (AMDF) [34] is used to extract phoneme information from acoustic data. The AMDF can effectively estimate periodical signals, which are the main characteristics of speech, laughter, and other vowel-based nonspeech sounds. The AMDF is derived as follows:

$$\begin{aligned} \tau^{*} &= \arg\min_{\tau} \text{AMDF}(\tau), \\ \text{AMDF}(\tau) &= \sum_{t=0}^{T-t-1} |S(t) - S(t+\tau)|, \end{aligned} \tag{4}$$

where S represents one of the segments in the acoustic signal, T is the length of S, t denotes the time index, and τ is the shifting length. After AMDF(τ) reaches the minimum, a phoneme $P = [S(\tau^{*}), S(\tau^{*}+1), \ldots, S(2\tau^{*})]$ can be acquired by extracting indices from $S(\tau^{*})$ to $S(2\tau^{*})$.

Algorithm 1 expresses the process of syllable extraction when phonemes of a signal are determined.

In the next step, to classify signals into their respective categories, energy and frequency changes are used as the first criteria to separate speech from vowel-based nonspeech because spectral discrepancy of speech is relatively smaller in most cases.

Compared with other vowel-based nonspeech, the temporal pattern of laughter usually exhibits repetitiveness. To detect such patterns, this study uses cascade filters, which consist of a Dynamic Time Warping (DTW) filter [35] and a syllable discriminant filter, to compute similarities of the input data. With the use of Mel-frequency cepstral coefficients (MFCCs), the Dynamic Time Warping filter can find out desired signals by matching them with the samples in the database. The signals that successfully pass through the first filter are subsequently input to the second filter. The syllable discriminant filter compares each input sequence with predefined patterns by using the inner product operation. When the score of an input is higher than a threshold, the input is labeled as laughter.

For emotional speech recognition, this study follows previous works [36–38] and extracts prosodic and timbre features from speech to recognize emotional information in voices. Tables 1 and 2 show the acoustic features used in this system.

In addition to the acoustic features, this study also uses the keyword spotting technique to detect predefined keywords in speech because textual data offer more emotion clues than acoustic data. After detecting predefined keywords in utterances, the system iteratively computes the association

Initialization
designating the beginning phonemes P_{start};
For each phoneme P_m
begin
If Similarity $(P_{\text{start}}, P_m) < \delta_{\text{Similarity}}$
If Distance $(P_{m-1}, P_m) > \delta_{\text{Distance}}$
m is the end of asyllable;
end

Algorithm 1: Algorithm for syllable extraction.

Table 1: Timbre features.

Type	Parameter
1st–3rd formants	Frequency
	Mean
	Standard deviation
	Medium
	Bandwidth
Spectrum related	Centroid
	Spread
	Flatness

Table 2: Prosodic features.

Type	Parameter
Pitch and energy related	Maximum value
	Minimum value
	Mean
	Medium
	Standard deviation
	Range
	Coefficients of the linear regression
Duration related	Speech rate
	Ratio between voiced and unvoiced regions
	Duration of the longest voiced speech

degree between the detected keyword and each emotion category.

Let i represent the index of the emotion categories; j denote the index of the detected keyword in the sentence corpus; ϖ_j refer to the detected keyword; $\Gamma(\varpi_j, c_i)$ represent the occurrence of ϖ_j in category c_i; $\Gamma(\varpi_j)$ denote the number of sentences containing (ϖ_j).

The association degree can be defined as

$$e_i\left(\varpi_j\right) = \frac{\Gamma\left(\varpi_j, c_i\right)}{\Gamma\left(\varpi_j\right)} \times \frac{\sum_i \Gamma\left(\varpi_j, c_i\right)^2}{\Gamma\left(\varpi_j\right)^2}, \tag{5}$$

where the first part of the equation is the weighting score, and the second part is the confidence score of ϖ_j (see [39] for detailed information). The textual feature vector is subsequently combined with the acoustic feature vector and sent into a classifier (AdaBoost) for training and recognition.

3.3. Feedback Mechanism. After completion of the audiovisual recognition stage, the system generates three results along with their classification scores. One of the three results is the detected facial expression, another is the detected vocal emotion type, and the other is laughter. The classification scores are linearly combined with the recognition rates of the corresponding classifiers and finally output to users. Additionally, the recognition result is logged in the database 24 hours a day. A user can browse the curve of emotion changes by viewing the display. The system is also equipped with a telehealthcare module. Personal emotion status can be sent to family psychologists or psychiatrists for mental care. The service robot can serve as an agent between the cloud system and users, providing a remote interactive interface.

4. Experimental Results

This study conducted an experiment to test audiovisual emotion recognition to assess the performance of our system. Only positive emotions, including smiling faces, laughter, and joyful voices, were tested in the experiment.

At the evaluation of the facial expression stage, 500 facial images containing smiles and nonsmiles were manually selected from the MPLab GENKI database (http://mplab.ucsd.edu/). The kernel function of the SVM was the radial basis function, and the penalty constant was empirically set to one. Furthermore, 50% of the dataset was used for training, and 50% was used for testing. During the evaluation of laughter recognition, a database consisting of 84 sound clips was created by recording the utterances of six people. Eighteen samples from these 84 clips were the sound of people laughing. After removing silence parts from all of the clips, the entire dataset was subsequently sent into the system for recognition. For emotional speech recognition, this research used the same database as that in our previous work [39]. The speech containing joyful and nonjoyful emotions was manually chosen and parsed to obtain their literal information and acoustic features. Finally, these features were inputted into an AdaBoost classifier for training and testing.

Figure 5 shows a summary of the experimental results of our system, in which the vertical axis denotes accuracy rates,

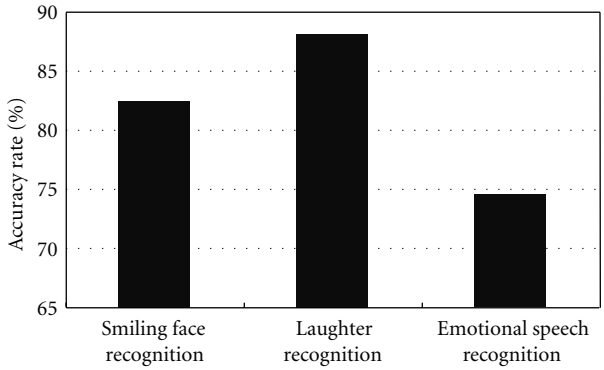

Figure 5: Accuracy rates of the audiovisual recognition.

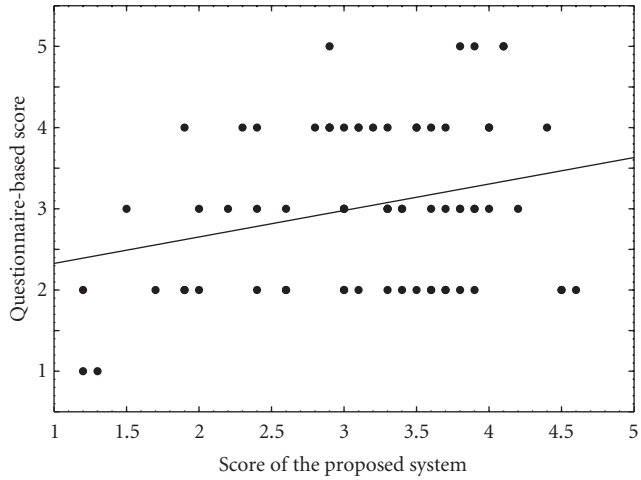

Figure 6: Correlation test of the scores between the questionnaire approach and the proposed system. The horizontal axis represents the evaluation result of the proposed system, whereas the vertical axis means the result of the questionnaire. The slope of the regression line is 0.33, and Pearson's correlation coeffcient is 0.27.

and the horizontal axis represents recognition modules. As shown in the figure, the accuracy rate of smile detection can reach 82.5%. The performance of laughter recognition can also achieve an accuracy rate as high as 88.2%. Compared with smile and laughter recognition, although the result of emotional speech recognition reached 74.6%, such performance is comparable to those of related emotional speech recognition systems. When combined with the test result of emotional speech recognition, the overall accuracy rate can reach an average of 81.8%.

The following experiment tests whether the proposed system can help testees remind and evaluate their emotional health status as caregivers do. During the experiment, total ten persons were selected from the sanatorium and the hospital to test the system for a week. The age of the participants ranges from 40 to 70 years old. The audiovisual sensors were installed in their living space, so that the emotional data can be acquired and analyzed in real time. For privacy, the sensors captured behavior only during 10:00 to 16:00. To avoid generating biased data, each testee was not aware of the locations of the sensors and the testing details of the experiment. Furthermore, after the system analyzed the data, the medical doctors and nurses helped testees complete questionnaires. The questionnaire contained total ten questions, nine of which were irrelevant to this experiment. The remaining question was the key criterion that allowed the testees to give a score (one(unhappy)–five(happy)) to their daily moods.

The questionnaire scores are subsequently compared with the estimated emotional status of the proposed system. To obtain the estimated emotional score, the proposed method firstly calculates the duration of smiling face expressions, joyful speech, and laughter of the testees. Next, a ratio can be computed by converting the duration into a one-to-five rating scale based on the test period.

The correlation test in Figure 6 shows performance of the questionnaire approach and the proposed system. The vertical axis represents the questionnaire result, whereas the score of the proposed system is listed on the horizontal axis. All the samples are collected from the testees. Closely examining the scatterness in this figure reveals that Pearson's correlation coefficient reaches as high as 0.27. This implies that our method is analogous with the questionnaire-based approach. Moreover, two groups of the scores in the linear regression analysis reflect a linear rate of 0.33. Above findings indicate that the proposed method can allow computers to monitor users' emotional health, subsequently assisting caregivers in reminding users' psychological status and saving more human resources.

5. Conclusion

This paper presents a new concept called orange computing for health, happiness, and humanistic care. To demonstrate the concept, a case study on the audiovisual emotion recognition system for care services is also conducted. The system uses multimodal recognition techniques, including facial expression, laughter, and emotional speech recognition to capture human behavior.

At the facial expression recognition stage, multilayered histograms of oriented gradients and multilayered local directional patterns are proposed to model facial features. To detect patterns of laughing sound, two cascade filters consisting of a Dynamic Time Warping filter and a syllable discriminant filter are used in the acoustic processing phase. Furthermore, when classifying emotional speech, the system combines textual, timbre, and prosodic features to calculate association degree to predefined emotion classes.

Three analyses are conducted for evaluating recognition performance of the proposed methods. Experimental results show that our system can reach an average accuracy rate of 81.8%. Concerning the feedback mechanism, data from the real-life test indicate that our method is comparable to the questionnaire-based approach. Additionally, correlation degree between two methods is as high as 0.27. The above results demonstrate that the proposed system is capable of recognizing users' emotional health and thereby providing an in-time reminder for them.

In summary, orange computing hopes to arouse awareness of the importance of mental wellness (health, happiness, and warming care), subsequently leading more people to join the movement, to share happiness with others, and finally to enhance the well-being of society.

Acknowledgments

This work was supported in part by the National Science Council of the Republic of China under Grant no. 100-2218-E-006-017. The authors would like to thank Yan-You Chen, Yi-Cheng Chen, Wei-Kang Fan, Chih-Hung Li, and Da-Yu Kwan for contributing the experimental data and supporting this research.

References

[1] M. C. Jensen, "The modern industrial revolution, exit, and the failure of internal control systems," *Journal of Applied Corporate Finance*, vol. 22, no. 1, pp. 43–58, 1993.

[2] F. Dunachie, "The success of the industrial revolution and the failure of political revolutions: how Britain got lucky," *Historical Notes*, vol. 26, pp. 1–7, 1996.

[3] Y. Veneris, "Modelling the transition from the industrial to the informational revolution," *Environment & Planning A*, vol. 21, no. 3, pp. 399–416, 1990.

[4] J. Hull, "The second industrial revolution: the history of a concept," *Storia Della Storiografia*, vol. 36, pp. 81–90, 1999.

[5] P. Ashworth, "High technology and humanity for intensive care," *Intensive Care Nursing*, vol. 6, no. 3, pp. 150–160, 1990.

[6] P. Gilk, *Green Politics is Eutopian*, Lutterworth Press, Cambridge, UK, 2009.

[7] S. B. F. Hargens, "Integral development—taking the middle path towards gross national happiness," *Journal of Bhutan Studies*, vol. 6, pp. 24–87, 2002.

[8] A. J. Oswald, "Happiness and economic performance," *Economic Journal*, vol. 107, no. 445, pp. 1815–1831, 1997.

[9] D. Kahneman, E. Diener, and N. Schwarz, *Well-Being : The Foundations of Hedonic Psychology*, Russell Sage Foundation Publications, New York, NY, USA, 1998.

[10] K. Passino, "World-wide education for the humanitarian technology challenge," *IEEE Technology and Society Magazine*, vol. 29, no. 2, p. 4, 2010.

[11] K. Lorincz, D. J. Malan, T. R. F. Fulford-Jones et al., "Sensor networks for emergency response: Challenges and opportunities," *IEEE Pervasive Computing*, vol. 3, no. 4, pp. 16–23, 2004.

[12] A. Waibel, "Speech processing in support of human-human communication," in *Proceedings of the 2nd International Symposium on Universal Communication (ISUC '08)*, p. 11, Osaka, Japan, December 2008.

[13] J.-F. Wang, B.-W. Chen, Y.-Y. Chen, and Y.-C. Chen, "Orange computing: challenges and opportunities for affective signal processing," in *Proceedings of the International Conference on Signal Processing, Communications and Computing*, pp. 1–4, Xian, China, September 2011.

[14] J.-F. Wang and B.-W. Chen, "Orange computing: challenges and opportunities for awareness science and technology," in *Proceedings of the 3rd International Conference on Awareness Science and Technology*, pp. 538–540, Dalian, China, September 2011.

[15] Y. Hata, S. Kobashi, and H. Nakajima, "Human health care system of systems," *IEEE Systems Journal*, vol. 3, no. 2, pp. 231–238, 2009.

[16] K. Siau, "Health care informatics," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 1, pp. 1–7, 2003.

[17] K. Kawamura, W. Dodd, and P. Ratanaswasd, "Robotic body-mind integration: Next grand challenge in robotics," in *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN '04)*, pp. 23–28, Kurashiki, Okayama, Japan, September 2004.

[18] P. Belimpasakis and S. Moloney, "A platform for proving family oriented RESTful services hosted at home," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 690–698, 2009.

[19] L. S. A. Low, N. C. Maddage, M. Lech, L. B. Sheeber, and N. B. Allen, "Detection of clinical depression in adolescents' speech during family interactions," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 574–586, 2011.

[20] C. Yu, J.-J. Yang, J.-C. Chen et al., "The development and evaluation of the citizen telehealth care service system: case study in Taipei," in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '09)*, pp. 6095–6098, Minneapolis, Minn, USA, September 2009.

[21] J. B. Jørgensen and C. Bossen, "Executable use cases: requirements for a pervasive health care system," *IEEE Software*, vol. 21, no. 2, pp. 34–41, 2004.

[22] A. Mihailidis, B. Carmichael, and J. Boger, "The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 3, pp. 238–247, 2004.

[23] Y. Hata, S. Kobashi, and H. Nakajima, "Human health care system of systems," *IEEE Systems Journal*, vol. 3, no. 2, pp. 231–238, 2009.

[24] Y. Gizatdinova and V. Surakka, "Feature-based detection of facial landmarks from neutral and expressive facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 135–139, 2006.

[25] R. A. Calvo and S. D'Mello, "Affect detection: an interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010.

[26] C. Busso and S. S. Narayanan, "Interrelation between speech and facial gestures in emotional utterances: a single subject study," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2331–2347, 2007.

[27] Y. Wang and L. Guan, "Recognizing human emotional state from audiovisual signals," *IEEE Transactions on Multimedia*, vol. 10, no. 4, pp. 659–668, 2008.

[28] Z. Zeng, J. Tu, B. M. Pianfetti, and T. S. Huang, "Audio-visual affective expression recognition through multistream fused HMM," *IEEE Transactions on Multimedia*, vol. 10, no. 4, pp. 570–577, 2008.

[29] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. I511–I518, Kauai, Hawaii, USA, December 2001.

[30] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[31] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 886–893, San Diego, Calif, USA, June 2005.

[32] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proceedings of the 6th ACM International Conference on Image and Video Retrieval (CIVR '07)*, pp. 401–408, Amsterdam, Netherlands, July 2007.

[33] T. Jabid, M. H. Kabir, and O. Chae, "Local Directional Pattern (LDP)—a robust image descriptor for object recognition," in *Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '10)*, pp. 482–487, Boston, Mass, USA, September 2010.

[34] C. K. Un and S.-C. Yang, "A pitch extraction algorithm based on LPC inverse filtering and AMDF," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 6, pp. 565–572, 1977.

[35] J.-F. Wang, J.-C. Wang, M.-H. Mo, C.-I. Tu, and S.-C. Lin, "The design of a speech interactivity embedded module and its applications for mobile consumer devices," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 870–876, 2008.

[36] S. Casale, A. Russo, G. Scebba, and S. Serrano, "Speech emotion classification using Machine Learning algorithms," in *Proceedings of the 2nd Annual IEEE International Conference on Semantic Computing (ICSC '08)*, pp. 158–165, Santa Clara, Calif, USA, August 2008.

[37] C. Busso, S. Lee, and S. Narayanan, "Analysis of emotionally salient aspects of fundamental frequency for emotion detection," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 4, pp. 582–596, 2009.

[38] N. D. Cook, T. X. Fujisawa, and K. Takami, "Evaluation of the affective valence of speech using pitch substructure," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 142–151, 2006.

[39] Y.-Y. Chen, B.-W. Chen, J.-F. Wang, and Y.-C. Chen, "Emotion aware system based on acoustic and textual features from speech," in *Proceedings of the 2nd International Symposium on Aware Computing (ISAC '10)*, pp. 92–96, Tainan, Taiwan, November 2010.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 184103, 10 pages
doi:10.1155/2012/184103

Research Article

Aware Computing in Spatial Language Understanding Guided by Cognitively Inspired Knowledge Representation

Masao Yokota

Department of System Management, Fukuoka Institute of Technology, Fukuoka 811-0295, Japan

Correspondence should be addressed to Masao Yokota, yokota@fit.ac.jp

Received 23 January 2012; Accepted 29 March 2012

Academic Editor: Keitaro Naruse

Mental image directed semantic theory (MIDST) has proposed an omnisensory mental image model and its description language L_{md}. This language is designed to represent and compute human intuitive knowledge of space and can provide multimedia expressions with intermediate semantic descriptions in predicate logic. It is hypothesized that such knowledge and semantic descriptions are controlled by human attention toward the world and therefore subjective to each human individual. This paper describes L_{md} expression of human subjective knowledge of space and its application to aware computing in cross-media operation between linguistic and pictorial expressions as spatial language understanding.

1. Introduction

The serious need for more human-friendly intelligent systems has been brought by rapid increase of aged societies, floods of multimedia information over the WWW, development of robots for practical use, and so on. For example, it is very difficult for people to exploit necessary information from the immense multimedia contents over the WWW. It is still more difficult to search for desirable contents by queries in different media, for example, text queries for pictorial contents. In this case, intelligent systems facilitating cross-media references are helpful and worth developing. In this research area so far, it has been most conventional that conceptual contents conveyed by information media such as languages and pictures are represented in computable forms independent of each other and translated via so-called "transfer" processes which are often ad hoc and very specific to task domains [1–3].

In order to systematize cross-media operation, however, it is needed to develop such a computable knowledge representation language for multimedia contents that should have at least a good capability of representing spatiotemporal events perceived by people in the real world. For this purpose, mental image directed semantic theory (MIDST) has proposed a model of human mental image and its description language L_{md} (Language for mental-image description) [4]. This language is capable of formalizing human omnisensory mental images (equal to multimedia contents, here) in predicate logic, while other knowledge description schema [5, 6] are too coarse or linguistic (or English-like) to formalize them in an integrative way as intended here. L_{md} is employed for many-sorted predicate logic and has been implemented on several versions of the intelligent system IMAGES [4, 7] and there is a feedback loop between them for their mutual refinement unlike other similar theories [8, 9].

As detailed in the following sections, MIDST was rigidly formalized as a deductive system [10] in the formal language L_{md}, which is remarkably distinguished from other work (e.g., [5, 8]). However, its application to computerized systems is another thing because computational cost of logical formulas is very high in general. In fact, however, the deductive system contains a considerable number of theses or postulates much easier to realize in imperative programming (e.g., in C) than in declarative programming (e.g., in Prolog) because L_{md} expressions normalized by atomic locus formulas are very suitable to structure and operate in table so-called Hitree [11]. Conventionally, it is as well convinced that hybrid computation based on both the programming paradigms is more flexible and efficient than that based on only one of them. This is also the case

for each version of IMAGES so far and therefore the author has been promoting to replace declarative programs with imperative ones considering the benefit of L_{md} expression. This paper focuses as well on the hybrid computation guided by L_{md} expression and 3D map data, here so-called partially symbolized direct knowledge of space (PSDKS), in cross-media operation between linguistic and pictorial expressions as spatial language understanding. That is, static spatial relations among objects as 3D map data for imperative programming are utilized as well as those in L_{md} for declarative programming.

The remainder of this paper is organized as follows. Section 2 presents the omnisensory mental image model and its relation to the formal language L_{md}. Section 3 describes representation of subjective spatial knowledge in L_{md}. In Sections 4 and 5 are sketched several cognitive hypotheses on mental images for their systematic computation. Section 6 describes the systematic cross-media operation based on L_{md} expression. Section 7 gives the details of direct knowledge of space. In Section 8, is described an example of cross-media operation by IMAGES. Some discussion and conclusion are given in the final section.

2. Mental Image Model and L_{md}

An attribute space corresponds with a sensory system and can be compared to a certain measuring instrument just like a barometer, thermometer or so, and the loci represent the movements of its indicator. A general locus is to be articulated by "Atomic Locus" over a certain absolute time interval $[t_i, t_f]$ as depicted in Figure 1 and formulated as (1) in L_{md}, where the interval is suppressed because people are not aware of absolute time (nor always consult a chronograph).

$$L(x, y, p, q, a, g, k). \tag{1}$$

This is a formula in many-sorted predicate logic, where "L" is a predicate constant with five types of terms: "Matter" (at "x" and "y"), "Value (of Attribute)" (at "p" and "q"), "Attribute" (at "a"), "Pattern (of Event)" (at "g"), and "Standard" (at "k"). Conventionally, Matter variables are headed by "x", "y," and "z."

This formula is called "Atomic Locus Formula" whose first two arguments are sometimes referred to as "Event Causer (EC)" and "Attribute Carrier (AC)," respectively, while ECs are often optional in natural concepts such as intransitive verbs. By the way, hereafter, the terms at AC and Standard are often replaced by "_" when they are of little significance to discern one another. The parameters "g" and "k" cannot be denoted explicitly in Figure 1 because their roles vary drastically depending on its interpretation.

The intuitive interpretation of (1) is given as follows.

> *"Matter " x" causes Attribute " a" of Matter " y" to keep ($p = q$) or change ($p \neq q$) its values temporally ($g = G_t$) or spatially ($g = G_s$) over an absolute time-interval, where the values " p" and " q" are relative to the standard " k". "*

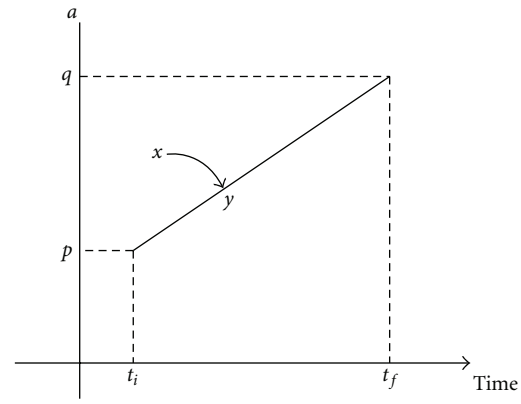

Figure 1: Graphical interpretation of Atomic Locus—the curved arrow indicates the abstract effect from "x" to "y."

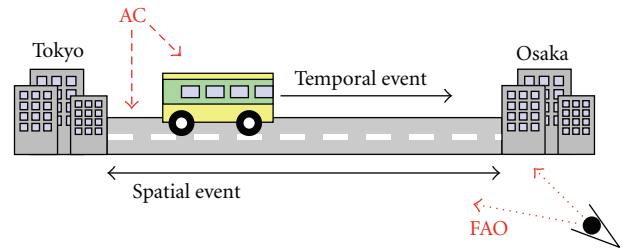

Figure 2: FAO movements and Event types.

When $g = G_t$ and $g = G_s$, the locus indicates monotonic change or constancy of the attribute in time domain and that in space domain, respectively. The former is called "temporal change event" and the latter, "spatial change event," which are assumed to correspond with temporal and spatial gestalt in psychology, respectively. For example, the motion of the "bus" represented by (S1) is a temporal change event and the ranging or extension of the "road" by (S2) is a spatial change event whose meanings or concepts are formulated as (2) and (3), respectively, where "A_{12}" denotes the attribute "Physical Location". These two formulas are different only at the term "Pattern."

(S1) The bus runs from Tokyo to Osaka.

$$(\exists x, y, k)L(x, y, \mathrm{Tokyo}, \mathrm{Osaka}, A_{12}, G_t, k) \wedge \mathrm{bus}(y). \tag{2}$$

(S2) The road runs from Tokyo to Osaka.

$$(\exists x, y, k)L(x, y, \mathrm{Tokyo}, \mathrm{Osaka}, A_{12}, G_s, k) \wedge \mathrm{road}(y). \tag{3}$$

The difference between temporal and spatial change event concepts can be attributed to the relationship between the Attribute Carrier (AC) and the Focus of the Attention of the Observer (FAO). To be brief, FAO is fixed on the whole AC in a temporal change event but runs about on the AC in a spatial change event. Consequently, as shown in Figure 2, the bus and the FAO move together in the case of (S1) while FAO

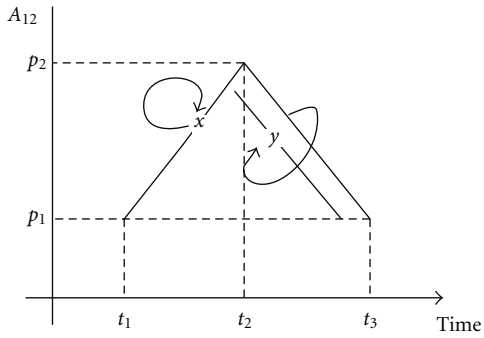

Figure 3: Conceptual image of "fetch."

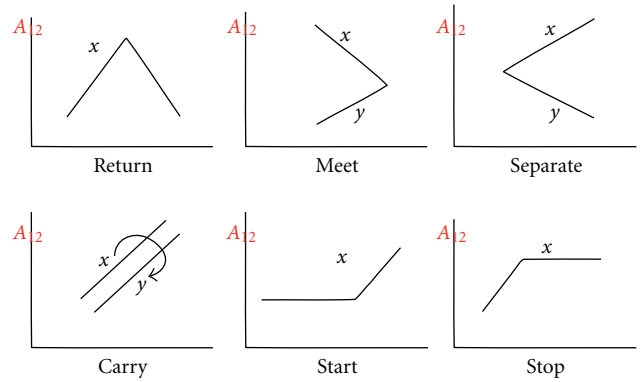

Figure 4: Event patterns of physical location (A_{12}).

solely moves along the road in the case of (S2). That is, *all loci in attribute spaces correspond one to one with movements or, more generally, temporal change events of FAO.*

Articulated loci are combined with tempological conjunctions, where "SAND ($\wedge_0$)" and "CAND ($\wedge_1$)" are most frequently utilized, standing for "Simultaneous AND" and "Consecutive AND", conventionally symbolized as "$\sqcap$" and "$\cdot$," respectively. The formula (4) refers to a temporal change event depicted as Figure 3, implying that "x" goes to some location and then comes back with "y" and corresponding to such a verbal expression as "x fetches y from some location":

$$(\exists x, y, p_1, p_2, k)L(x, x, p_1, p_2, A_{12}, G_t, k) \cdot (L(x, x, p_2, p_1, A_{12}, G_t, k) \sqcap L(x, y, p_2, p_1, A_{12}, G_t, k)) \wedge x \neq y \wedge p_1 \neq p_2. \quad (4)$$

As easily imagined, an event expressed in L_{md} is compared to a movie film taken through a floating camera where both temporal and spatial extensions of the event are recorded as a time sequence of snapshots because it is necessarily grounded in FAO's movement over the event. This is one of the most remarkable features of L_{md}, clearly distinguished from other knowledge representation languages (KRLs).

The attribute spaces for humans correspond to the sensory receptive fields in their brains. At present, about 50 attributes and 6 categories of standards concerning the physical world have been extracted from thesauri. Event patterns are the most important for our approach and have been already reported concerning several kinds of attributes [4, 7]. Figure 4 shows several examples of event patterns in the attribute space of "physical location (A_{12})."

3. Representation of Subjective Spatial Knowledge

MIDST can provide human knowledge pieces with flat L_{md} expressions as human mental images, not concerning whether they are concepts meant by certain symbols (i.e., semantic) or not. Therefore, such a distinction is not denoted explicitly hereafter. There are assumed two major hypotheses

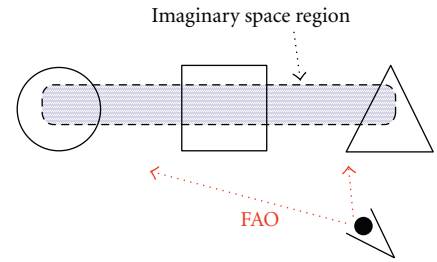

Figure 5: Row as spatial change event.

on mental image. One is that mental image is in one-to-one correspondence with FAO movement as mentioned above. And, the other is that it is not one-to-one reflection of the real world. It is well known that people perceive more than reality, for example, so-called "Gestalt" in psychology. A psychological matter here is not a real matter but a product of human mental functions, including Gestalt and abstract matters such as "society" and "information" in a broad sense. For example, Figure 5 concerns the perception of the formation of multiple objects, where FAO runs along an imaginary object so called "*Imaginary Space Region (ISR).*" This spatial change event can be verbalized as (S3) using the preposition "between" and formulated as (5) or (6), corresponding also to such concepts as "row," and "line-up," where A_{13} denotes the attribute "Direction".

Employing ISRs and the 9-intersection model [12], all the topological relations between two objects can be formulated in such expressions as (7) or (8) for (S4), and (9) for (S5), where "In," "Cont," and "Dis" are the values "inside", "contains" and "disjoint" of the attribute "Topology (A_{44})" with the standard "9-intersection model ($K_{9\mathrm{IM}}$)," respectively. Practically, these topological values are given as 3×3 matrices with each element equal to 0 or 1 and therefore, for example, "In" and "Cont" are transposes each other. That is, Cont = In^T.

(S3) The square is between the triangle and the circle.

(S4) Tom is in the room.

(S5) Tom exits the room.

$$
\begin{aligned}
&(\exists x_1, x_2, x_3, y, p, q)\\
&\quad (L(\_, y, x_1, x_2, A_{12}, G_s, \_) \sqcap L(\_, y, p, p, A_{13}, G_s, \_))\\
&\quad \cdot (L(\_, y, x_2, x_3, A_{12}, G_s, \_) \sqcap L(\_, y, q, q, A_{13}, G_s, \_))\\
&\quad \wedge\ \mathrm{ISR}(y) \wedge p = q \wedge \mathrm{triangle}(x_1)\\
&\quad \wedge \mathrm{square}(x_2) \wedge \mathrm{circle}(x_3),
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
&(\exists x_1, x_2, x_3, y, p)\\
&\quad (L(\_, y, x_1, x_2, A_{12}, G_s, \_) \cdot L(\_, y, x_2, x_3, A_{12}, G_s, \_))\\
&\quad \sqcap L(\_, y, p, p, A_{13}, G_s, \_) \wedge \mathrm{ISR}(y)\\
&\quad \wedge \mathrm{triangle}(x_1) \wedge \mathrm{square}(x_2) \wedge \mathrm{circle}(x_3),
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
&(\exists x, y) L(\mathrm{Tom}, x, y, \mathrm{Tom}, A_{12}, G_s, \_)\\
&\quad \sqcap L(\mathrm{Tom}, x, \mathrm{In}, \mathrm{In}, A_{44}, G_t, K_{9\mathrm{IM}}) \wedge \mathrm{ISR}(x) \wedge \mathrm{room}(y),
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
&(\exists x, y) L(\mathrm{Tom}, x, \mathrm{Tom}, y, A_{12}, G_s, \_)\\
&\quad \sqcap L(\mathrm{Tom}, x, \mathrm{Cont}, \mathrm{Cont}, A_{44}, G_t, K_{9\mathrm{IM}})\\
&\quad \wedge \mathrm{ISR}(x) \wedge \mathrm{room}(y),
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
&(\exists x, y, p, q) L(\mathrm{Tom}, \mathrm{Tom}, p, q, A_{12}, G_t, \_)\\
&\quad \sqcap L(\mathrm{Tom}, x, y, \mathrm{Tom}, A_{12}, G_s, \_)\\
&\quad \sqcap L(\mathrm{Tom}, x, \mathrm{In}, \mathrm{Dis}, A_{44}, G_t, K_{9\mathrm{IM}}) \wedge \mathrm{ISR}(x)\\
&\quad \wedge \mathrm{room}(y) \wedge p \neq q.
\end{aligned}
\tag{9}
$$

With a special attention, the author has analyzed a considerable number of spatial terms over various kinds of English words such as prepositions, verbs, adverbs, and so forth, categorized as "Dimensions," "Form," and "Motion" in the class "SPACE" of the Roget's thesaurus [13], and found that almost all the concepts of spatial change events can be defined in exclusive use of five kinds of attributes for FAOs, namely, "Physical location (A_{12})," "Direction (A_{13})," "Trajectory (A_{15})," "Mileage (A_{17})," and "Topology (A_{44})."

4. Hypothetical Operations upon Mental Images

People can transform their mental images in several ways such as mental rotation [14]. Here are introduced and defined 3 kinds of mental operations, namely, "reversing," "duplicating," and "converting."

4.1. Image Reversing. It is easy for people to imagine the reversal of an event just like "rise" versus "sink." This mental operation is here denoted as "R" and recursively defined as O_R, where χ_i stands for a image. The reversed values p^R and q^R depend on the properties of the attribute values p and q. For example, $p^R = p$, $q^R = q$ for A_{12}; $p^R = -p$, $q^R = -q$ for A_{13}; $p^R = p^T$, $q^R = q^T$ for A_{44}.

O_R:

$$
\begin{aligned}
\left(\chi_1 \cdot \chi_2\right)^R &\Longleftrightarrow \chi_2^R \cdot \chi_1^R,\\
\left(\chi_1 \sqcap \chi_2\right)^R &\Longleftrightarrow \chi_1^R \sqcap \chi_2^R,\\
L^R(x, y, p, q, a, g, k) &\Longleftrightarrow L\left(x, y, q^R, p^R, a, g, k\right).
\end{aligned}
\tag{10}
$$

4.2. Image Duplicating. Humans can easily imagine the repetition of an event just like "visit twice" versus "visit once." This operation is also recursively defined as O_n, where "n" is an integer representing the frequency of an image χ.

O_n:

$$
\begin{aligned}
\chi^n &\Longleftrightarrow \chi \quad (n = 1),\\
\chi^n &\Longleftrightarrow \chi \cdot \chi^{n-1} \quad (n > 1).
\end{aligned}
\tag{11}
$$

4.3. Image Converting. We can convert temporal and spatial change event images each other and this is the reason why it is easy for us to understand instantly such an expression as (S2). This mental operation is here denoted as "C" and recursively defined as O_C, which will help a robot to cope with such a somewhat queer expression as "The road jumps up at the point. Be careful!".

O_C:

$$
\begin{aligned}
\left(\chi_1 \cdot \chi_2\right)^C &\Longleftrightarrow \chi_1^C \cdot \chi_2^C,\\
\left(\chi_1 \sqcap \chi_2\right)^C &\Longleftrightarrow \chi_1^C \sqcap \chi_2^C,\\
L^C(x, y, p, q, a, g, k) &\Longleftrightarrow L\left(x, y, p, q, a, g^C, k\right),
\end{aligned}
\tag{12}
$$

where $g^C = G_s$ for $g = G_t$ and $g^C = G_t$ for $g = G_s$.

5. Hypothetical Properties of Mental Images

Properties or laws of mental images as spatial knowledge pieces are formalized in L_{md} and introduced as postulates and their derivatives in a deductive system [10] to be employed in theorem proving there. Here are described two examples of such postulates, namely, "Postulate of Reversibility of Spatial Change Event" and "Postulate of Partiality of Matter."

5.1. Postulate of Reversibility of Spatial Change Event. As already mentioned in Section 2, all loci in attribute spaces are assumed to correspond one to one with movements or, more generally, temporal change events of the FAO. Therefore, the L_{md} expression of an event is compared to a movie film recorded through a floating camera over the event. And this is why (S6) and (S7) can refer to the same scene in spite of their appearances, where what "sinks" or "rises" is the FAO as illustrated in Figure 6 and whose conceptual descriptions are given as (13) and (14), respectively, where "A_{13}," "↑," and "↓" refer to the attribute "Direction" and its values "upward" and "downward" (practically as 3D unit vectors), respectively.

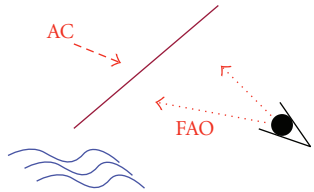

Figure 6: Slope as spatial change event.

(S6) The path *sinks* to the brook.

(S7) The path *rises* from the brook.

$$(\exists y,z,p)L(\_,y,p,z,A_{12},G_s,\_)\sqcap L(\_,y,\downarrow,\downarrow,A_{13},G_s,\_)$$
$$\wedge\ \text{path}(y)\wedge\text{brook}(z)\wedge z\neq p, \tag{13}$$

$$(\exists y,z,p)L(\_,y,z,p,A_{12},G_s,\_)\sqcap L(\_,y,\uparrow,\uparrow,A_{13},G_s,\_)$$
$$\wedge\ \text{path}(y)\wedge\text{brook}(z)\wedge z\neq p. \tag{14}$$

Such a fact is generalized as P_{RS} (postulate of reversibility of spatial change event), where χ_s and χ_s^R are an image and its "reversal" for a certain spatial change event, respectively, and they are substitutable with each other because of the property of "$\equiv_0$." This postulate can be one of the principal inference rules belonging to people's common-sense knowledge about geography.

P_{RS}:

$$\chi_s^R\cdot\equiv_0\chi_s. \tag{15}$$

This postulation is also valid for such a pair of (S8) and (S9) as interpreted approximately into (16) and (17), respectively. These pairs of conceptual descriptions are called equivalent in the P_{RS}, and the paired sentences are treated as paraphrases each other.

(S8) Route A and Route B separate at the city.

(S9) Route A and Route B meet at the city.

$$(\exists p,y,q)L(\_,\text{Route\_}A,p,y,A_{12},G_s,\_)$$
$$\sqcap L(\_,\text{Route\_}B,q,y,A_{12},G_s,\_)\wedge\text{city}(y)\wedge p\neq q, \tag{16}$$

$$(\exists p,y,q)L(\_,\text{Route\_}A,y,p,A_{12},G_s,\_)$$
$$\sqcap L(\_,\text{Route\_}B,y,q,A_{12},G_s,\_)\wedge\text{city}(y)\wedge p\neq q. \tag{17}$$

Of course, P_{RS} is as well applicable to such an inference that "if x is to the right of y, then y is to the left of x," which is conventionally based on a considerably large set of such *linguistic* axioms as (18) regardless of *time*. Furthermore, it is notable that there are an infinite number of directions without good correspondence with single words such as "right."

$$(\forall x,y)\text{right}(x,y)\supset\text{left}(y,x),$$
$$(\forall x,y)\text{under}(x,y)\supset\text{above}(y,x). \tag{18}$$

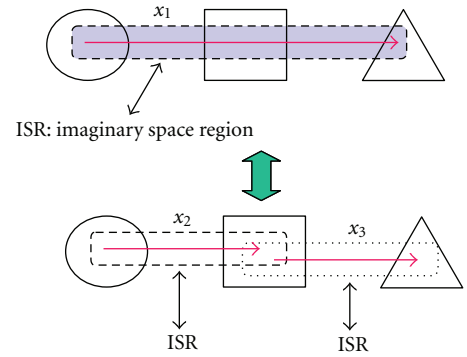

Figure 7: Partiality of ISR—the arrows represent the directions of FAO.

5.2. Postulate of Partiality of Matter. Any matter is assumed to consist of its parts in a structure (i.e., spatial change event) and generalized as P_{PM} (postulate of partiality of matter) here. For example, Figure 7 shows that an ISR x_1 can be deemed as a complex of ISRs x_2 and x_3.

P_{PM}:

$$(\forall y,x_1,p,q,a,k)L(y,x_1,p,q,a,G_s,k)$$
$$\cdot L(y,x_1,q,r,a,G_s,k).\supset_0(\exists x_2,x_3)L(y,x_2,p,q,a,G_s,k)$$
$$\sqcap L(y,x_3,q,r,a,G_s,k),$$

$$(\forall y,x_2,x_3,p,q,a,k)$$
$$L(y,x_2,p,q,a,G_s,k)\sqcap L(y,x_3,q,r,a,G_s,k)$$
$$.\supset_0(\exists x_1)L(y,x_1,p,q,a,G_s,k)\cdot L(y,x_1,q,r,a,G_s,k). \tag{19}$$

We often refer to parts of an image especially for deductive inference upon it. For example, we can easily deduce from Figure 7 (Top) the two facts "the square is to the left of the triangle" and "the circle is to the left of the square." As its reversal, we can merge these two partial images into one meaningful image such as Figure 7 (Bottom). That is, P_{PM} is very useful to compute static spatial relations that are expressed by English spatial terms and conventionally formalized by a large set of such *linguistic* axioms as (20) regardless of *time* just like the case of P_{RS}. Furthermore, it is notable that the reversals of these axioms (i.e., $(\forall x,y,z)$ between $(y,z,x)\supset w(y,x)\wedge w(z,y)$) do not always exist in good correspondence with words (e.g., "left" for the predicate w).

$$(\forall x,y,z)\text{left}(y,x)\wedge\text{left}(z,y)\supset\text{between}(y,z,x),$$
$$(\forall x,y,z)\text{under}(y,x)\wedge\text{under}(z,y)\supset\text{between}(y,z,x). \tag{20}$$

Besides its orthodox usage above, P_{PM}, in cooperation with P_{RS}, can be utilized for translating such a paradoxical sentence as "The Andes Mountains run north and south." into such a plausible interpretation as "Some part of the Andes Mountains run north (from somewhere) and the other part run south."

6. Cross-Media Translation

As easily understood by its definition, an atomic formula corresponds with a pair of snapshots at the beginning and the ending of a monotonic change in an attribute. Viewed from pictorial representation, temporal and spatial change events correspond to animated and still pictures, respectively. Furthermore, the L_{md} expression of a spatial change event as the locus of FAO can be related to the sequence of pen-down and pen-up in line drawing. This section describes cross-media translation in general, focusing on that between text and map, one kind of still picture, as the core of spatial language understanding.

6.1. Functional Requirements. Systematic cross-media translation here is defined by the functions (F1)–(F4) as follows.

(F1) To translate source representations into target ones as for contents describable by both source and target media. For example, positional relations between/among physical objects such as "in", "around." are describable by both linguistic and pictorial media.

(F2) To filter out such contents that are describable by source medium but not by target one. For example, linguistic representations of "taste" and "smell" such as "sweet candy" and "pungent gas" are not describable by usual pictorial media although they would be seemingly describable by cartoons, and so forth.

(F3) To supplement default contents, that is, such contents that need to be described in target representations but not explicitly described in source representations. For example, the shape of a physical object is necessarily described in pictorial representations but not in linguistic ones.

(F4) To replace default contents by definite ones given in the following contexts. For example, in such a context as "There is a box to the left of the pot. The box is red. ...," the color of the box in a pictorial representation must be changed from default one to red.

For example, the text consisting of such two sentences as "There is a hard cubic object" and "The object is large and gray" can be translated into a still picture in such a way as shown in Figure 8.

6.2. Formalization. According to the MIDST, any content conveyed by an information medium is assumed to be associated with the loci in certain attribute spaces and in turn the world describable by each medium can be characterized by the maximal set of such attributes. This relation is conceptually formalized by (21), where Wm, Am_i, and F mean "the world describable by the information medium m," "an attribute of the world," and "a certain function for determining the maximal set of attributes of Wm," respectively,

$$F(\mathrm{Wm}) = \{\mathrm{Am}_1, \mathrm{Am}_2, \ldots, \mathrm{Am}_n\}. \tag{21}$$

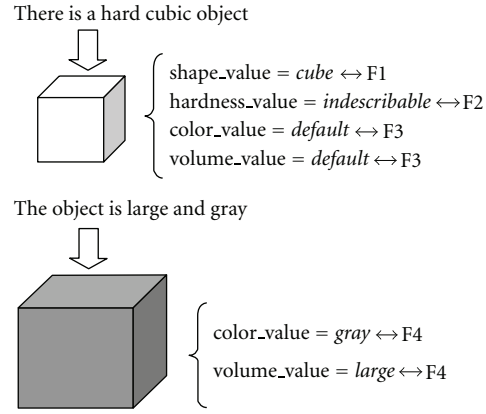

Figure 8: Systematic cross-media translation.

Considering this relation, cross-media translation is one kind of mapping from the world describable by the source medium (ms) to that by the target medium (mt) and can be defined by the following equation:

$$Y(S_{\mathrm{mt}}) = \psi(X(S_{\mathrm{ms}})), \tag{22}$$

where S_{ms}: maximal set of attributes of the world describable by the source medium ms, S_{mt}: maximal set of attributes of the world describable by the target medium mt, $X(S_{\mathrm{ms}})$: L_{md} expression about the attributes belonging to S_{ms}, $Y(S_{\mathrm{mt}})$: L_{md} expression about the attributes belonging to S_{mt}, and ψ: function for transforming X into Y, so called, "L_{md} expression paraphrasing function."

The function ψ is designed to clear all the requirements (F1)–(F4) by inference processing at the level of L_{md} expression.

6.3. L_{md} Expression Paraphrasing Function ψ. In order to realize the function (F1), a certain set of "*Attribute paraphrasing rules* (*APRs*)," so called, are defined *at every pair of source and target media*. The function (F2) is realized by detecting L_{md} expressions about *the attributes without any corresponding APRs* from the content of each input representation and replacing them by *empty events* [10].

For (F3), *default reasoning* is employed. That is, such an inference rule as defined by (23) is introduced, which states if *X is deducible and it is consistent to assume Y then conclude Z*. This rule is applied typically to such instantiations of X, Y, and Z as specified by (24) which means that the indefinite attribute value " p" with the indefinite standard " k" of the indefinite matter " y" is substitutable by the constant attribute value " P" with the constant standard "K"of the definite matter "$O\#$" of the same kind "M":

$$X \circ Y \longrightarrow Z, \tag{23}$$

$$\begin{aligned}&\{X/(L(x, y, p, p, A, G, k) \wedge M(y))\\ &\quad \wedge (L(z, O\#, P, P, A, G, K) \wedge M(O\#)),\\ &\quad Y/p = P \wedge k = K, Z/L(x, y, P, P, A, G, K) \wedge M(y)\}.\end{aligned} \tag{24}$$

Table 1: APRs for text-picture translation (A_{12}: physical location, A_{13}: direction, A_{17}: mileage, A_{10}: volume, A_{11}: shape, A_{32}: color, A_{44}: topology).

APRs	Correspondences of attributes (text : picture)	Value conversion schema (text ↔ picture)
APR-01	$A_{12} : A_{12}$	$p \leftrightarrow p'$
APR-02	$\{A_{12}, A_{13}, A_{17}\} : A_{12}$	$\{p, d, l\} \leftrightarrow p' + l'd'$
APR-03	$\{A_{11}, A_{10}\} : A_{11}$	$\{s, v\} \leftrightarrow v's'$
APR-04	$A_{32} : A_{32}$	$c \leftrightarrow c'$
APR-05	$\{A_{12}, A_{44}\} : A_{12}$	$\{p_a, m\} \leftrightarrow \{p'_a, p'_b\}$

The function (F4) is realized quite easily by memorizing the history of applications of default reasoning.

6.4. Attribute Paraphrasing Rules for Text and Picture. Five kinds of APRs for this case are shown in Table 1 where $p, s, c, \ldots$ and $p', s', c', \ldots$ are linguistic expressions and their corresponding pictorial expressions of attribute values, respectively. Further details are as follows.

(i) APR-02 is used especially for a sentence such as "The box is 3 meters to the left of the chair." The symbols p, d and l correspond to "the location of the chair," "left," and "3 meters," respectively, yielding the pictorial expression of "the location of the box," namely, "$p' + l'd'$."

(ii) APR-03 is used especially for a sentence such as "The pot is big." The symbols $s$and v correspond to "the shape of the pot (default value)" and "the volume of the pot ("big")," respectively. In pictorial expression, the shape and the volume of an object is inseparable and therefore they are represented only by the value of the attribute "shape", namely, $v's'$.

(iii) APR-05 is used especially for a sentence such as "The cat is in the box." The symbols p_a, p_b and m correspond to "the location of the desk," "the location of the cat," and "in," respectively, yielding a pair of pictorial expressions of the locations of the two objects.

7. Direct Knowledge of Space

Partially symbolized direct knowledge of space (PSDKS in short) introduced here is one of the data structures for imperative programming in IMAGES as well as Hitree [11]. PSDKS is a map for directional and metric relations among objects while Hitree is intended to be a complete substitute of L_{md} expression. That is, the relation between L_{md} expression and PSDSK is what is formalized by APR-02 in Table 1. For example, consider the scene of a room shown in Figure 9, where the FAO is posed on the formation of the flower-pot, box lamp, chair, and cat. PSDKS here does not mean any kind of live image perceived by a human (or snapshot by a system) at a time point but somewhat abstract 3D map resulted from its recognition as depicted in Figure 10.

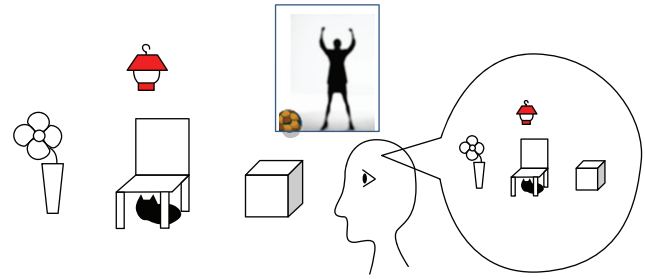

Figure 9: Scene of a room and its live image in human.

That is, PSDKS is defined as a set of points representing the 3D locations (i.e., A_{12}) of the involved objects linked to the corresponding L_{md} expression and therefore directly reusable for computation without recognizing them unlike the memory of their live image or snapshot.

In turn, consider verbalization of the PSDKS. In this case, any system must be forced to articulate it in accordance with existing word concepts and may utter such a set of sentences (S10)–(S13). These are to be generated from such L_{md} expressions as (25)–(28), respectively, where I_n, Fp, Ch, Bx, Lp and Ct stand for ISR, flower-pot, chair, box lamp, and cat, respectively.

(S10) The chair is 3 meters to the right of the flower-pot.

(S11) The flower-pot is 6 meters to the left of the box.

(S12) The lamp hangs above the chair.

(S13) The cat lies under the chair.

$$L(\_, I_1, \mathrm{Fp}, \mathrm{Ch}, A_{12}, G_s, \_) \sqcap L(\_, I_1, \rightarrow, \rightarrow, A_{13}, G_s, \_) \sqcap L(\_, I_1, 3m, 3m, A_{17}, G_s, \_), \tag{25}$$

$$L(\_, I_2, \mathrm{Bx}, \mathrm{Fp}, A_{12}, G_s, \_) \sqcap L(\_, I_2, \leftarrow, \leftarrow, A_{13}, G_s, \_) \sqcap L(\_, I_2, 6m, 6m, A_{17}, G_s, \_), \tag{26}$$

$$L(\_, I_3, \mathrm{Ch}, \mathrm{Lp}, A_{12}, G_s, \_) \sqcap L(\_, I_3, \uparrow, \uparrow, A_{13}, G_s, \_), \tag{27}$$

$$L(\_, I_4, \mathrm{Ch}, \mathrm{Ct}, A_{12}, G_s, \_) \sqcap L(\_, I_4, \downarrow, \downarrow, A_{13}, G_s, \_). \tag{28}$$

Even only for directional and metric relationships between two objects out of the five objects in Figure 10, there can be at least 20 ($=_5P_2$) expressions in English including (S10)–(S13) that correspond with such formulas in conventional logic as (29)–(32), respectively.

$$\mathrm{right}(\mathrm{Ch}, \mathrm{Fp}, 3\_\mathrm{meters}), \tag{29}$$

$$\mathrm{left}(\mathrm{Fp}, \mathrm{Bx}, 6\_\mathrm{meters}), \tag{30}$$

$$\mathrm{above}(\mathrm{Lp}, \mathrm{Ch}), \tag{31}$$

$$\mathrm{under}(\mathrm{Ct}, \mathrm{Ch}). \tag{32}$$

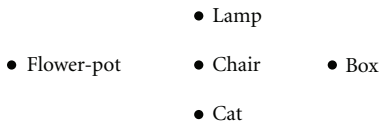

Figure 10: PSDKS resulted from the live image in Figure 9.

This fact implies that conventional declarative programs must employ numerous theses including the axioms (18) and (20) even for solving rather simple problems associated with this scene such as "What is between the box and the flower-pot?". The meaning of this question is conventionally notated as (33). However, it must be noted that the axioms like (18) and (20) cannot be applied to the assertions (29)–(32) for the answer to this question (i.e., $?x$).

On the contrary, it is much easier to search in the PSDKS for the event pattern specified by the L_{md} expression (34) for the question. This formula, a locus of FAO, can be procedurally interpreted as the command "Find "$?x$" by scanning *straight* from the *box* to the *flower-pot*." In case of understanding (S10)–(S13), the system is to apply APR-02 to (25)–(28) and synthesize the partial scenes into one whole scene similar to (not always the same as) the PSDKS shown in Figure 10, that is to say, *reconstructed* direct knowledge of space:

$$\mathrm{between}(?x, \mathrm{Bx}, \mathrm{Fp}), \tag{33}$$

$$\begin{aligned}&(L(\_, y, \mathrm{Bx}, ?x, A_{12}, G_s, \_) \cdot L(\_, y, ?x, \mathrm{Fp}, A_{12}, G_s, \_))\\ &\quad \sqcap L(\_, y, p, p, A_{13}, G_s, \_).\end{aligned} \tag{34}$$

At summarization of this section, PSDKS is very much compact in memory size compared with conventional declaration about space and L_{md} expression can systematically indicate how to search PSDKS for an event pattern.

8. Implementation

IMAGES-M, the last version of intelligent system IMAGES, has recently adopted the multiparadigm language Python in place of PROLOG to facilitate both declarative and imperative programming. IMAGES-M is one kind of expert system with five kinds of user interfaces besides the inference engine (IE) and the knowledge base (KB) as follows.

(i) Text Processing Unit (TPU).

(ii) Speech Processing Unit (SPU).

(iii) Picture Processing Unit (PPU).

(iv) Action Data Processing Unit (ADPU).

(v) Sensory Data Processing Unit (SDPU).

These user interfaces can mutually convert information media and L_{md} expressions in the collaboration with IE and KB, and miscellaneous combinations among them bring forth various types of cross-media operations. The further details about mutual conversion between language and picture can be found in other papers (e.g., [15, 16]).

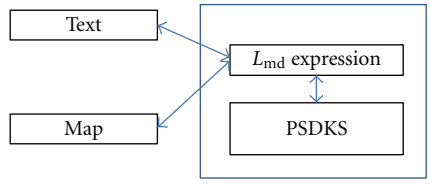

Figure 11: Text-map operation via L_{md} expression and PSDKS.

Figure 12: Transactions between human user and IMAGES-M while text understanding, map composition and question-answering on the map (At headers: "u…" = human user, "s…" = IMAGES-M).

The methodology mentioned above has been implemented on IMAGES-M for spatial language understanding. Here, distinguished from others, spatial language understanding is defined as cross-media operation between spatial language and map such as mutual translation and question-answering between them. The author has confirmed that the hybrid program in Python employing L_{md} expression mainly and PSDKS auxiliarly as shown in Figure 11 is much more flexible and efficient than the previous one [4] in PROLOG for solving problems expressed in spatial language.

Here is presented an example of cross-operation between text and picture performed by IMAGES-M.

IMAGES-M understood the human user's assertions or questions and answered them in picture or word. Figure 12 shows the transactions exchanged between the human user and the system, where the headers "u…." and "s…." stand for the human user's inputs and the system's responses,

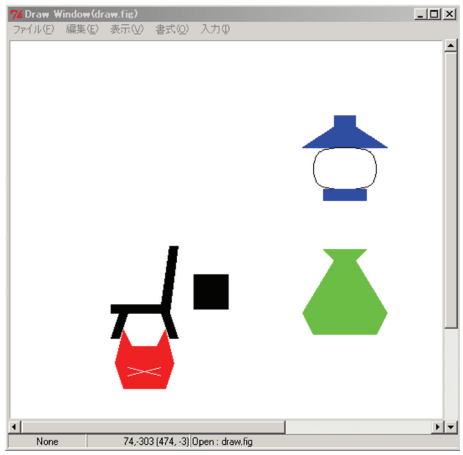

Figure 13: Map finally composed by IMAGES-M for u0001–u0006.

respectively. IMAGES-M can accept 3 kinds of natural language besides English, namely, Japanese (e.g., u0002, u0008 and s0029), Chinese (e.g., u0007 and s0026 in Pinyin) and Albanian (e.g., u0003, u0010 and s0035) as shown in Figure 12, where

u0002 = "The cat is 1 m under the chair,"

u0003 = "The cat is red,"

u0008 = "What is between the chair and the pot?,"

s0029 = "Box,"

u0007 = "Is the cat red?,"

s0026= "yes,"

u0010 = "Is the box between the cat and the lamp?,"

s0035 = "yes."

The map shown in Figure 13 was the final version of those which IMAGES-M composed at each of the user's assertions. IMAGES-M interpreted the assertions u0001–u0006 into L_{md}, and in turn into map and PSDKS (exactly, reconstructed PSDKS), where the system updated them assertion by assertion, responding so by s0002–s0022. In the process of text to map, default reasoning about color, and so forth. was performed in such a way as shown in Figure 8, where only the default locations of the objects within the map are significant for PSDKS.

On the other hand, during the question-answering (i.e., u0007-s0035), IMAGES-M translated each of the user's questions (i.e., u0007–u0010) into L_{md} and consulted the reconstructed PSDKS about Location (A_{12}) within the map or the corresponding L_{md} expression about the other attributes such as Color (A_{32}). In this process, the postulates P_{RS} and P_{PM} were utilized as procedures in Python, which could reduce remarkably the number of axioms such as (18) and (20) that are necessarily employed in conventional systems.

9. Discussion and Conclusion

MIDST is still under development and intended to provide a formal system, represented in L_{md}, for natural semantics of space and time. This formal system is one kind of applied predicate logic consisting of axioms and postulates subject to human perceptive processes of space and time, while the other similar systems in Artificial Intelligence [17–19] are objective, namely, independent of human perception and do not necessarily keep tight correspondences with natural language. This paper showed that L_{md} expressions can contribute to aware computing of spatial relations leading to representational and computational cost reduction in aid of Partially Symbolized Direct Knowledge of Space (PSDKS) while some further quantitative elaboration is needed on this point.

The author has already reported that cross-media operation between texts in several languages (Japanese, Chinese, Albanian, and English) and pictorial patterns like maps were successfully implemented on IMAGES-M [4]. As detailed in this paper, IMAGES-M has recently adopted the multiparadigm language Python in place of PROLOG to facilitate both declarative and imperative programming, and the author has confirmed that the hybrid program in Python employing L_{md} expression mainly and PSDKS auxiliarly is much more flexible and efficient than the previous one in PROLOG for solving problems expressed in spatial language. To our best knowledge, there is no other system (e.g., [20, 21]) that can perform cross-media operations in such a seamless way as described here. This leads to the conclusion that L_{md} has made the logical expressions of event concepts remarkably computable and has proved to be very adequate to systematize cross-media operations. This adequacy is due to its medium-freeness and its good correspondence with the performances of human sensory systems in both spatial and temporal extents while almost all other knowledge representation schemes are ontology-dependent, computing- unconscious or spatial-change-event unconscious (e.g., [8, 9]).

The author deems that aware science or technology is still on the way to maturation and therefore that now it should foster various kinds of approaches. The model of human cognition employed in MIDST is formalized based on declarative knowledge representation in symbolic logic which has almost been discarded in this research area so far and instead certain approaches based on procedural knowledge representation has been prevalent. The author's very intention here is to present some prospective possibility of his original theory MIDST in aware science. The example presented in Section 8 is rather simple but one of the most complicated spatial relations displayable in this version of the intelligent system IMAGES-M because it was programmed exclusively to check the efficacy of PSDKS. Another extended version of the system is now under construction and some examples of further complicated human-system interaction in natural language have already been presented in another paper [15].

Our future work will include establishment of learning facilities for automatic acquisition of word concepts from

sensory data [7] and human-robot communication by natural language under real environments [22].

Acknowledgment

This work was partially funded by the Grants from Computer Science Laboratory, Fukuoka Institute of Technology and Ministry of Education, Culture, Sports, Science and Technology, Japanese Government, nos. 14580436, 17500132, and 23500195.

References

[1] A. Yamada, A. Yamada, H. Ikrda et al., "Reconstructing spatial image from natural language texts," in *Proceedings of the 15th International Conference on Computational Linguistics (COLING '90)*, Nantes, France, 1992.

[2] P. Olivier and J. Tsujii, "A computational view of the cognitive semantics of spatial expressions," in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics (ACL '94)*, Las Cruces, New Mexico, 1994.

[3] G. Adorni, M. Di Manzo, and F. Giunchiglia, "Natural language driven image generation," in *Proceedings of the 10th International Conference on Computational Linguistics (COLING '84)*, pp. 495–500, 1984.

[4] M. Yokota and G. Capi, "Cross-media operations between text and picture based on mental image directed semantic theory," *WSEAS Transactions on Information Science and Applications*, vol. 2, no. 10, pp. 1541–1550, 2005.

[5] J. F. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole, Pacific Grove, Calif, USA, 2000..

[6] G. P. Zarri, "NKRL, a knowledge representation tool for encoding the "Meaning" of complex narrative texts," *Natural Language Engineering—Special Issue on Knowledge Representation for Natural Language Processing in Implemented Systems*, vol. 3, pp. 231–253, 1997.

[7] S. Oda, M. Oda, and M. Yokota, "Conceptual analysis and description of words for color and lightness for grounding them on sensory data," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 16, no. 5, pp. 436–444, 2001.

[8] R. W. Langacker, *Concept, Image and Symbol, Mouton de Gruyter*, Berlin, Germany, 1991.

[9] G. A. Miller and P. N. Johnson-Laird, *Language and Perception*, Harvard University Press, 1976.

[10] M. Yokota, "Systematic formulation and computation of subjective spatiotemporal knowledge based on mental image directed semantic theory: toward a formal system for natural intelligence," in *Proceedings of the 6th International Workshop on Natural Language Processing and Cognitive Science (NLPCS '09)*, pp. 133–143, Milan, Italy, May 2009.

[11] M. Yokota, "Towards awareness computing under control by world knowledge grounded in sensory data," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '10)*, pp. 769–775, October 2010.

[12] B. M. Shariff, M. J. Egenhofer, and D. M. Mark, "Natural-language spatial relations between linear and areal objects: the topology and metric of English-language terms," *International Journal of Geographical Information Science*, vol. 12, no. 3, pp. 215–245, 1998.

[13] P. Roget, *Thesaurus of English Words and Phrases*, J.M. Dent & Sons Ltd, London, UK, 1975.

[14] R. Shepard and J. Metzler, "Mental rotation of three-dimensional objects," *Science*, vol. 171, no. 3972, pp. 701–703, 1971.

[15] M. Yokota, "Systematic analysis and synthesis of human subjective knowledge of space and time for intuitive human-robot interaction," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '11)*, pp. 208–215, 2011.

[16] M. Yokota, "Towards artificial communication partners with a multiagent mind model based on mental image directed semantic theory," in *Humanoid Robots*, B. Choi, Ed., pp. 333–364, I-Tech Press, 2009.

[17] J. F. Allen, "Towards a general theory of action and time," *Artificial Intelligence*, vol. 23, no. 2, pp. 123–154, 1984.

[18] D. V. McDermott, "A temporal logic for reasoning about processes and plans," *Cognitive Science*, vol. 6, no. 2, pp. 101–155, 1982.

[19] Y. Shoham, "Time for actions: on the relationship between time, knowledge, and action," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 954–959, Detroit, Mich, USA, 1989.

[20] J. P. Eakins and M. E. Graham, "Content-based Image Retrieval: A report to the JISC Technology Applications Programme," Institute for Image Data Research, University of Northumbria at Newcastle, 1999.

[21] M. L. Kherfi, D. Ziou, and A. Bernardi, "Image retrieval from the World Wide Web: issues, techniques, and systems," *ACM Computing Surveys*, vol. 36, no. 1, pp. 35–67, 2004.

[22] M. Yokota, M. Shiraishi, and G. Capi, "Human-robot communication through a mind model based on the mental image directed semantic theory," in *Proceedings of the 10th International Symposium on Artificial Life and Robotics (AROB '05)*, pp. 695–698, Oita, Japan, 2005.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 354785, 11 pages
doi:10.1155/2012/354785

Research Article

Multilevel Cognitive Machine-Learning-Based Concept for Artificial Awareness: Application to Humanoid Robot Awareness Using Visual Saliency

Kurosh Madani, Dominik M. Ramik, and Cristophe Sabourin

Images, Signals and Intelligence Systems Laboratory (LISSI/EA 3956) and Senart-FB Institute of Technology, University Paris-EST Créteil (UPEC), Bât.A, avenue Pierre Point, 77127 Lieusaint, France

Correspondence should be addressed to Kurosh Madani, madani@u-pec.fr

Received 11 March 2012; Revised 12 May 2012; Accepted 20 May 2012

Academic Editor: Qiangfu Zhao

As part of "intelligence," the "awareness" is the state or ability to perceive, feel, or be mindful of events, objects, or sensory patterns: in other words, to be conscious of the surrounding environment and its interactions. Inspired by early-ages human skills developments and especially by early-ages awareness maturation, the present paper accosts the robots intelligence from a different slant directing the attention to combining both "cognitive" and "perceptual" abilities. Within such a slant, the machine (robot) shrewdness is constructed on the basis of a multilevel cognitive concept attempting to handle complex artificial behaviors. The intended complex behavior is the autonomous discovering of objects by robot exploring an unknown environment: in other words, proffering the robot autonomy and awareness in and about unknown backdrop.

1. Introduction and Problem Stating

The term "cognition" refers to the ability for the processing of information applying knowledge. If the word "cognition" has been and continues to be used within quite a large number of different contexts, in the field of computer science, it often intends artificial intellectual activities and processes relating "machine learning" and accomplishment of knowledge-based "intelligent" artificial functions. However, the cognitive process of "knowledge construction" (and in more general way "intelligence") requires "awareness" about the surrounding environment and, thus, the ability to perceive information from it in order to interact with the surrounding milieu. So, if "cognition" and "perception" remain inseparable ingredients toward machines intelligence and thus toward machines (robots', etc.) autonomy, the "awareness" skill is a key spot in reaching the above-mentioned autonomy.

Concerning most of the works relating modern robotics, and especially humanoid robots, it is pertinent to note that they either have concerned the design of controllers controlling different devices of such machines [1, 2] or have focused the navigation aspects of such robots [3–5]. In the same way, the major part of the work dealing with human-like, or in more general terms intelligent, behavior, has connected abstract tasks, as those relating reasoning inference, interactive deduction mechanisms, and so forth. [6–10]. Inspired by early-ages human skills developments [11–15] and especially human early-ages walking [16–19], the present work accosts the robots intelligence from a different slant directing the attention to emergence of "machine awareness" from both "cognitive" and "perceptual" traits. It is important to note that neither the presented work nor its related issues (concepts, architectures, techniques, or algorithms) pretend being "artificial versions" of the complex natural (e.g., biological, psychological, etc.) mechanisms discovered, pointed out, or described by the above-referenced authors or by numerous other scientists working within the aforementioned areas whose works are not referenced in this paper. In [20] Andersen wrote concerning artificial neural networks: "*It is not absolutely necessary to believe that neural network models have anything to do with the nervous*

system, but it helps. Because, if they do, we are able to use a large body of ideas, experiments, and facts from cognitive science and neuroscience to design, construct, and test networks. Otherwise, we would have to suggest functions and mechanism for intelligent behavior without any examples of successful operation." In the same way, those natural mechanisms help us to look for plausible analogies between our down-to-earth models and those complex cognitive mechanisms.

Combining cognitive and perceptual abilities, the machine (robot) shrewdness is constructed on the basis of two kinds of functions: "unconscious cognitive functions" (UCFs) and "conscious cognitive functions" (CCFs). We identify UCFs as activities belonging to the "instinctive" cognition level handling reflexive abilities. Beside this, we distinguish CCFs as functions belonging to the "intentional" cognition level handling thought-out abilities. The two above-mentioned kinds of functions have been used as basis of a multilevel cognitive concept attempting to handle complex artificial behaviors [21]. The intended complex behavior is the autonomous discovering of objects by robot exploring an unknown environment. The present paper will not itemize the motion-related aspect that has been widely presented, analyzed, discussed and validated (on different examples) in [21]. It will focus on perceptual skill and awareness emergence. Regarding perceptual skill, it is developed on the basis of artificial vision and "salient" object detection. The paper will center this foremost skill and show how the intentional cognitive level of the above-mentioned concept could be used to proffer a kind of artificial awareness skill. The concept has been applied in design of "motion-perception-" based control architecture of a humanoid robot. The designed control architecture takes advantage of visual intention allowing the robot some kind of artificial awareness regarding its surrounding environment.

The paper is organized in five sections. The next section briefly introduces the multilevel cognitive concept. Section 3 describes the general structure of a cognitive function and depicts the suggested motion-perception control strategy. Section 4 presents the visual intention bloc. Validation results, obtained from implementation on a real humanoid-like robot, are reported in this section. Finally, the last section concludes the paper.

2. Brief Overview of Multilevel Cognitive Concept

Within the frame of this concept, we consider a process (mainly a complex process) as a multimodel structure where involved components (models), constructed as a result of machine learning (ML), handle two categories of operational levels: reflexive and intentional [21]. This means that ML and related techniques play a central role in this concept and the issued architectures. According to what has been mentioned in the introductory section, two kinds of functions, so-called UCFs and CCFs, build up functional elements ruling the complex task or the complex behavior. Figure 1 illustrates the bloc diagram of the proposed cognitive conception. As it is noticeable from this figure, within the proposed concept,

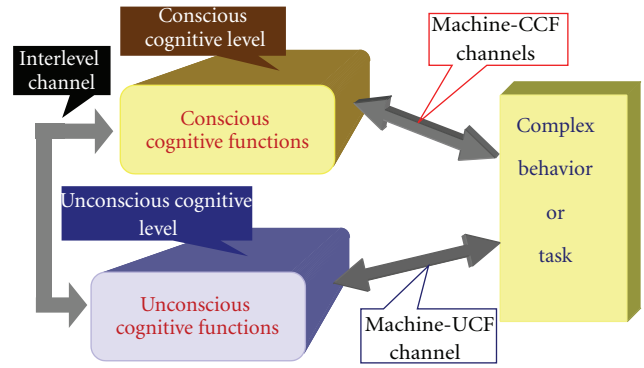

Figure 1: Robot coordinates described by a triplet as $P(x, y, \theta)$.

the overall architecture is obtained by building up cognitive layers (levels) corresponding to different skills fashioning the complex task. It is pertinent to remind that, as well as UCFs, CCFs enclose a number of "elementary functions" (EFs). Within such a scheme, a cognitive layer may fulfil a skill either independently of other layers (typically, the case of unconscious cognitive levels) or using one or several talents developed by other layers (characteristically, the case of conscious cognitive levels) [21].

The first key advantage of conceptualizing the problem within such an incline is to detach the modelling of robots complex artificial behaviours from the type of robot. In other words, models built within such conceptualizing could be used for modelling the same kind of complex behaviours for different kinds of robots. An example of analogy (similarity) with natural cognitive mechanisms could be found in early-ages human walking development. In fact, in its global achievement, the early-ages human abilities development does not depend on the kind of "baby." The second chief benefit of such a concept is that the issued artificial structures are based on "machine learning" paradigms (artificial neural networks, fuzzy logic, reinforcement learning, etc.), taking advantage of "learning" capacity and "generalization" propensity of such approaches. This offers a precious potential to deal with high dimensionality, nonlinearity, and empirical (non-analytical) proprioceptive or exteroceptive information.

3. From Cognitive Function to Motion-Perception Architecture

As it has been mentioned above, a cognitive function (either UCF or CCF) is constructed by a number of EFs. EF is defined as a function (learning-based or conventional) realizing an operational aptitude composing (necessary for) the skill accomplished by the concerned cognitive function. An EF is composed of "elementary components" (ECs). An EC is the lowest level component (module, transfer function, etc.) realizing some elementary aptitude contributing in EF operational aptitude. Two kinds of ECs could be defined (identified): the first corresponding to elementary action that we call "action elementary component" (AEC) and

the second corresponding to elementary decision that we call "decision elementary component" (DEC). An EF may include one or both kinds of the above-defined EC. In the same way, a cognitive function may include one or several ECs. Figure 2 gives the general structure of a cognitive function. However, it is pertinent to notice that there is any restriction to the fact that when it may be necessary, an EC could play the role of an EF. In the same way, when necessary, a cognitive function could include only one EF.

Supposing that a given cognitive function (either conscious or unconscious) includes K ($K \in N$, where N represents the "natural numbers ensemble) elementary functions, considering the kth EF (with $k \in N$ and $k \leq K$) composing this cognitive function, we define the following notations.

Ψ_k is the input of kth EF: $\Psi_k = [\psi_1, \ldots, \psi_j, \ldots, \psi_M]^T$, where ψ_j represents the input component of the jth EC of this EF, $j \leq M$, and M the total number of elementary components composing this EF.

O_k is the output of kth EF.

o_j is the output of the jth EC of the kth EF, with $j \leq M$, and M the total number of elementary components composing the kth EF.

$F_k(\cdot)$ is the skill performed by the kth EF.

$f_j^A(\cdot)$ is the function (transformation, etc.) performed by jth AEC.

$f^D(\cdot)$ is the decision (matching, rule, etc.) performed by DEC.

Within the above-defined notation, the output of kth EF is formalized as shown in (1) with o_j given by (2). In a general case, the output of an EC may also depend on some internal (specific) parameters particular to that EC [21]:

$$O_k = F_k(\Psi_k) = f^D\left(\Psi_k, o_1, \ldots, o_j, \ldots, o_M\right), \qquad o_j = f_j^A\left(\psi_j\right). \tag{1}$$

Based on the aforementioned cognitive concept, the control scheme of a robot could be considered within the frame of "motion-perception-" (MP-) based architecture. Consequently, as well as the robot motions its perception of the environment is obtained combining UCF and CCF. Robot sway is achieved combining unconscious and conscious cognitive motion functions (UCMFs and CCMFs, resp.). In the same way, essentially based on vision, robot perceptual ability is constructed combining unconscious and conscious cognitive visual functions (UCVFs and CCVFs, resp.). Figure 3 shows such an MP-based robot cognitive control scheme. It is pertinent to notice that the proposed control scheme takes advantage of some universality, conceptualizing the build-up of both robot motion and perception abilities independently of the type of robot. It is also relevant to emphasize that the proposed cognitive

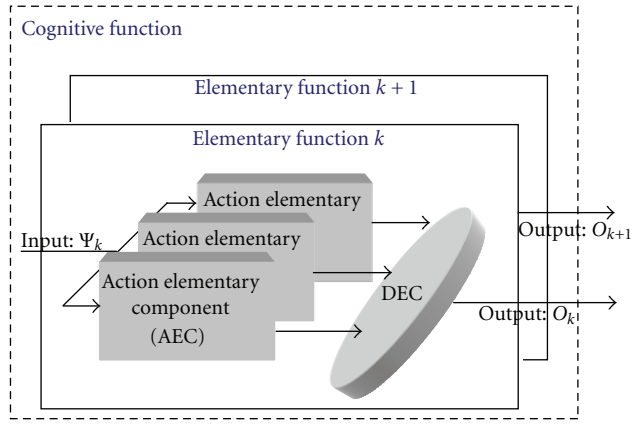

Figure 2: General bloc diagram of a cognitive function.

scheme links the behavior control construction to perception constructing the robot action from and with perceptual data and interaction with the context. This slant of view lays the robot way of doing (e.g., robot knowledge construction) to the human way of learning and knowledge construction: humans or animals learn and construct the knowledge by interacting with the environment. In other words, these natural intelligent beings operate using "awareness" about the surrounding environment in which they live.

If the question of how humans learn, represent, and recognize objects under a wide variety of viewing conditions is still a great challenge to both neurophysiology and cognitive researchers [22], a number of works relating the human early-ages cognitive walking ability construction process highlighting a number of key mechanisms. As shows clinical experiments (as those shown by [23]), one them is the strong linkage between visual and motor mechanisms. This corroborates the pertinence of the suggested cognitive MP-based scheme. Beside this, [24, 25] show that apart of shaping (e.g., recognizing objects and associating shapes with them), we (human) see the world by bringing our attention to visually important objects first. This means that the visual attention mechanism plays also one of the key roles in human infants learning of the encountered objects. Thus, it appears appropriate to draw inspiration from studies on human infants visual learning in constructing robots awareness on the basis of learning by visual revelation.

Making an intelligent system perceive the environment in which it evolves and construct the knowledge by learning unknown objects present in that environment makes a clear need appear relating the ability to select from the overwhelming flow of sensory information only the pertinent ones. This foremost ability is known as "visual saliency," sometimes called in the literature "visual attention," unpredictability, or surprise. It is described as a perceptual quality that makes a part of an image stand out relative to the rest of the image and to capture attention of observer [26]. It may be generalized that it is the saliency (in terms of motion, colors, etc.) that lets the pertinent information "stand out" from the context [27]. We argue that in this context visual saliency may be helpful to enable unsupervised extraction and subsequent learning of

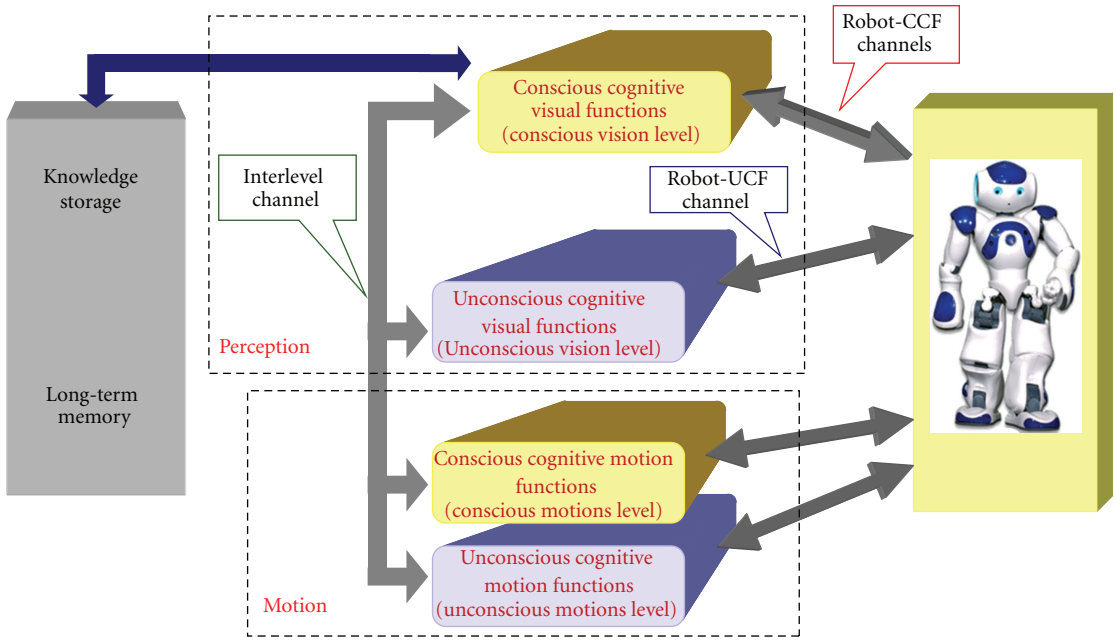

Figure 3: Bloc diagram of motion-perception-based robot cognitive control scheme.

a previously unknown object by a machine, in other words, proffering to the machine (robot) the awareness about its environment.

Referring to the perception bloc of Figure 3, the visual perception is composed of an unconscious visual level including UCVF and one conscious visual level containing CCVF. Unconscious visual level handles reflexive visual tasks, namely, the preprocessing of acquired images, the salient objects detection, and the detected salient objects storage. If the preprocessing could appear as an independent UCVF, it also may be an EF of one of UCVFs composing the unconscious visual level. In this second way of organizing the unconscious visual level, the UCVF including the preprocessing task will deliver the preprocessing results (as those relating image segmentation, different extracted features, etc.) as well to other UCVFs composing the unconscious level as to those CCVFs of conscious level which need the aforementioned results, using the interlevel channel. Conscious visual level conducts intentional visual tasks, namely, the objects learning (including learning detected salient objects), the knowledge construction by carrying out an intentional storage (in unconscious visual level) of new detected salient objects, the detected salient objects recognition in robot surrounding environment (those already known and the visual target (recognized salient object) tracking) allowing the robot self-orientation and motion toward a desired recognized salient object. Consequently, the conscious level communicates (e.g., delivers the outputs of concerned CCVF) with unconscious level (e.g., to the concerned UCVF) as well as with unconscious motion and conscious motion levels (e.g., with the bloc in MP-based robot cognitive control scheme in charge of robot motions).

4. From Salient Objects Detection to Visual Awareness

This section is devoted to description of two principle cognitive visual functions. The first subsection will detail the main UCVF, called "salient vision," which allows robot to self-discover (automatically detect) pertinent objects within the surrounding environment. While, the second subsection will spell out one of the core CCVFs, called "visual intention," which proffers the robot artificial visual intention ability and allows it to construct the knowledge about the surrounding environment proffering the robot the awareness regarding its surrounding environment.

Before describing the above-mentioned functions, it is pertinent to note that a recurrent operation in extracting visually salient objects (from images) is image segmentation. Generally speaking, one can use any available image segmentation technique. However, the quality of segmentation may be weighty for an accurate extraction of salient objects in images. In fact, most of the usual segmentation techniques (used beside standard image salient object extraction techniques) using manual or automatic thresholding remain limited because they do not respect the original image features. That is why we made use of the algorithm proposed recently by [28]. It is based on K-means clustering of color space with an adaptive selection of K and a spatial filter removing meaningless segments. The used algorithm is very fast (tens of milliseconds for a 320×240 pixels image on a standard PC) and it claims to have results close to human perception. The execution speed is a major condition in effective implementation in robotics applications reinforcing our choice for this already available algorithm, which keeps upright between execution speed and achieved segmentation quality.

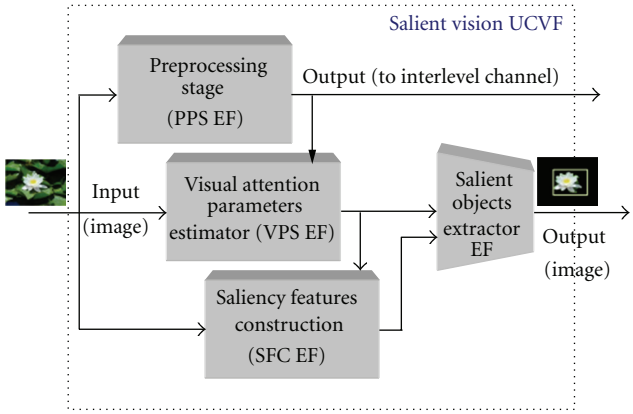

Figure 4: Bloc-diagram of Salient Vision UCVF, handling the automated detection of salient objects.

4.1. Salient Vision UCVF and Salient Objects Detection. The bloc diagram detailing the structure of "salient vision" UCVF is given in Figure 4. As it is visible from this figure, the "salient vision" UCVF includes also the preprocessing stage (defined as one of its constituting EFs), meaning that this UCVC handles the image segmentation and common image features' extraction tasks, delivering the issued results to other UCVFs as well as to conscious visual level. Beside this EF, it includes three other EF: the "visual Attention parameters estimator" (VAPE) EF, the "salient features construction" (SFC) EF, and the "salient objects extraction" (SOE) EF. This last EF plays the role of a decision-like elementary component, implemented as an independent EF.

4.1.1. Visual Attention Parameters Estimation Elementary Function. The visual attention parameter estimation (VAPE) elementary function determines what could be assimilated to some kind of "visual attention degree." Computed on the basis of preprocessing bloc issued results and controlling local salient features, visual attention parameter p constructs a top-down control of the attention and of the sensitivity of the feature in scale space. High value of p (resulting in a large sliding window size) with respect to the image size will make the local saliency feature more sensitive to large objects. In the same way, low values of p allow focusing the visual attention on smaller objects and details. The value of visual attention parameter p can be hard-set to a fixed value based on a heuristic according to [29]. However, as different images usually present salient objects in different scales, this way of doing will limit the performance of the system. Thus, a new automated cognitive estimation of the parameter p has been designed. The estimation is based, on the one hand, on calculation (inspired from the work presented in [30]) of the histogram of segment sizes from the input image, and on the other hand, on using of an artificial neural network (ANN). The ANN receives (as input) the feature vector issued from the above-mentioned histogram and provides the sliding window value. The weights of the neural network are adapted in training stage using a genetic algorithm.

To obtain the aforementioned histogram, the input image is segmented into n segments $(S_1, S_2, \ldots, S_n)$. For each one of the found segments S_i (where $S_i \in \{S_1, S_2, \ldots, S_n\}$), its size $|S_i|$ (measured in number of pixels) is divided by the overall image size $|I|$. An absolute histogram H_{SA} of segment sizes is constructed according to (2), avoiding leading to a too sparse histogram. This ensures that the first histogram bin contains the number of segments with area larger than 1/10 of the image size, the second contains segments from 1/10 to 1/100 of the image size, and so forth. For practical reasons we use a 4-bin histogram. Then, this absolute histogram leads to a relative histogram H_{SR} computed according to relation (3):

$$H_{SA}(i) = \sum_{j=1}^{n} \begin{cases} 1 & \text{if } 10^{i-1} \leq \dfrac{|S_i|}{|I|} \leq 10^{i}, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

$$H_{SR}(i) = \frac{H_{SA}(i)}{\sum_j H_{SA}(j)}. \tag{3}$$

The core of the proposed visual attention parameter estimator is a fully connected three-layer feed-forward MLP-like ANN, with a sigmoidal activation function, including 4 input nodes, 3 hidden neurons, and 1 output neuron. The four input nodes are connected each to its respective bin from the H_{SR} histogram. The value of the output node, belonging to the continuous interval $[0, 1]$, could be interpreted as the ratio of the estimated sliding window size p and the long side size of the image. The ANN is trained making use of a genetic algorithm described in [31]. Each organism in the population consists of a genome representing an array of floating point numbers whose length corresponds with the number of weights in MLP. To calculate the fitness of each organism, the MLP weights are set according to its current genome. Once visual attention parameter p is available (according to the MLP output) saliency is computed over the image and salient objects are extracted. The result is compared with ground truth and the precision, the recall and the F-ratio (representing the overall quality of the extraction) are calculated (according to [32] and using the measures proposed in the same work to evaluate quantitatively the salient object extraction). The F-ratio is then used as the measure of fitness. In each generation, the elitism rule is used to explicitly preserve the best solution found so far. Organisms are mutated with 5% of probability. As learning data set, we use 10% of the MSRA-B data set (described in [32]). The remaining 90% of the above-indicated data set has been used for validation.

4.1.2. Salient Features Construction Elementary Function. The salient features construction (SFC) elementary function performs two kinds of features (both used for salient objects detection). The first kind is global saliency features and the second local saliency features. Global saliency features capture global properties of image in terms of distribution of colors. The global saliency is obtained combining "intensity saliency" and the "chromatic saliency." Intensity saliency $M_l(x)$, given by relation (4), is defined as Euclidean distance of intensity I to the mean of the entire image. Index l stands

Figure 5: Examples of global and local saliency features: original image (a), global saliency map (b), local saliency map (c), and final saliency map.

for intensity channel of the image, and $I_{\mu l}$ is the average intensity of the channel. In the same way, chromatic saliency, given by relation (6), is defined as Euclidean distance of azimuth and zenith components intensities (e.g., azimuth ϕ and zenith θ, resp.) to their means ($I_{\mu\phi}$ and $I_{\mu\theta}$ resp.) in the entire image. Term (x) denotes coordinates of a given pixel on the image:

$$M_l(x) = \left\| I_{\mu l} - I_l(x) \right\|,$$
$$M_{\phi\theta}(x) = \sqrt{\left(I_{\mu\phi} - I_\phi(x)\right)^2 + \left(I_{\mu\theta} - I_\theta(x)\right)^2}. \tag{4}$$

The global saliency map $M(x)$, given by relation (5) is a hybrid result of combination of maps resulted from (1) according to logistic sigmoid blending function. Blending of the two saliency maps together is driven by a function of color saturation C of each pixel. It is calculated from RGB color model for each pixel as pseudonorm, given by $C = \mathrm{Max}[\mathrm{R}, \mathrm{G}, \mathrm{B}] - \mathrm{Min}[\mathrm{R}, \mathrm{G}, \mathrm{B}]$. When C is low, importance is given to intensity saliency. When C is high, chromatic saliency is emphasized:

$$M(x) = \frac{1}{1 - e^{-C}} M_{\phi\theta}(x) + \left(1 - \frac{1}{1 + e^{-C}}\right) M_l(x). \tag{5}$$

The global saliency (and related features) captures the visual saliency with respect to the colors. However, in real cases, the object visual saliency may also consist in its particular shape or texture, distinct to its surroundings, either beside or rather than simply in its color. To capture this aspect of visual saliency, a local feature over the image is determined. Inspired from a similar kind of feature introduced in [32], the local saliency has been defined as a centre-surround difference of histograms. The idea relating the local saliency is to go through the entire image and to compare the content of a sliding window with its surroundings to determine how similar the two are. If similarity is low, it may be a sign of a salient region within the sliding window.

To formalize this idea leading local saliency features, let us have a sliding window P of size p, centered over pixel (x). Define a (centre) histogram H_C of pixel intensities inside it. Then, let us define a (surround) histogram H_S as histogram of intensities in a window Q surrounding P in a manner that the area of $(Q - P) = p^2$. The centre-surround feature $d(x)$ is then given as (6) over all histogram bins (i):

$$d(x) = \sum_i \frac{|H_C(i) - H_S(i)|}{p^2}. \tag{6}$$

Resulting from computation of the $d(x)$ throughout all the l, ϕ, and θ channels, the centre-surround saliency $D(x)$ on a given position (x) is defined according to (7). Similarly to (5), a logistic sigmoid blending function has been used to combine chromaticity and intensity in order to improve the performance of this feature on images with mixed achromatic and chromatic content. However, here the color saturation C refers to average saturation of the content of the sliding window P:

$$D(x) = \frac{1}{1 - e^{-C}} d_l(x) + \left(1 - \frac{1}{1 + e^{-C}}\right) \mathrm{Max}\left(d_\phi(x), d_\theta(x)\right). \tag{7}$$

4.1.3. Salient Objects Extraction Elementary Function. Salient objects extraction (SOE) elementary function acts as the last step of saliency map calculation and salient objects detection. The extracted global and local salient features (e.g., $M(x)$ and $D(x)$, resp.) are combined using (8), resulting in final saliency map $M_{\mathrm{final}}(x)$, which is then smoothed by Gaussian filter. The upper part of the condition in (8) describes a particular case, where a part of image consists of a color that is not considered salient (i.e., pixels with low $M(x)$ measure) but which is distinct from the surroundings by virtue of its shape.

(a) (b) (c)

Figure 6: Examples of salient object detection: input image (a), detected salient objects (b), and ground truth salient objects (c).

(a) (b) (c)

Figure 7: Effect of the visual attention parameter p: input image (a), detected salient objects with high values of p (b) and small values of p (c).

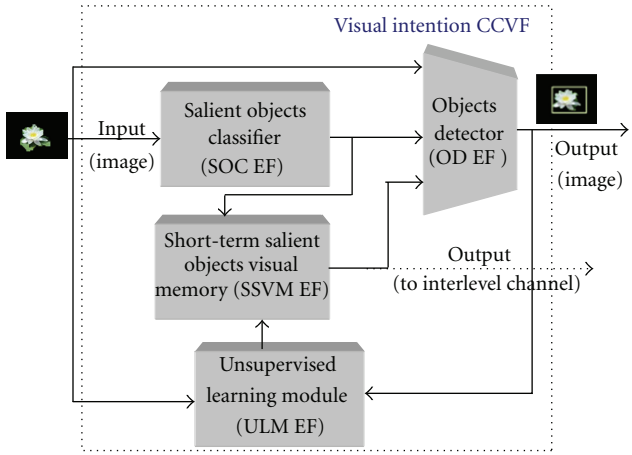

Figure 8: Bloc diagram of visional intention CCVF.

The final saliency map samples are shown on the column d of Figure 5:

$$M_{\text{final}}(x) = \begin{cases} D(x) & \text{if } M(x) < D(x), \\ \sqrt{M(x)D(x)} & \text{otherwise.} \end{cases} \quad (8)$$

Accordingly to segmentation and detection algorithms described in [30, 33], the segmentation splits an image into a set of chromatically coherent regions. Objects present on the scene are composed of one or multiple such segments. For visually salient objects, the segments forming them should cover areas of saliency map with high overall saliency, while visually unimportant objects and background should have this measure comparatively low. Conformably to [33], input image is thus segmented into connected subsets of pixels or segments $(S_1, S_2, \ldots, S_n)$. For each one of the found segments S_i (where $S_i \in \{S_1, S_2, \ldots, S_n\}$), its average saliency $\overline{S_i}$ and variance (of saliency values) $\text{Var}(S_i)$ are computed over the final saliency map $M_{\text{final}}(x)$. All the pixel values $p(x, y) \in S_i p(z, y)$ of the segment are then set following (9), where $\tau_{\overline{S_i}}$ and τ_{Var} are thresholds for average saliency and its variance, respectively. The result is a binary map containing a set of connected components $C = \{C_1, C_2, \ldots, C_n\}$ formed by adjacent segments S_i evaluated by (9) as binary value "1". To remove noise, a membership condition is imposed that any $C_i \in C$ has its area larger than a given threshold. Finally, the binary map is projected on the original image leading to a result that is part (areas) of the original image containing its salient objects. References [33, 34] give different values for the aforementioned parameters and thresholds:

$$p(x, y) = \begin{cases} 1 & \text{if } \overline{S_i} > \tau_{\overline{S_i}},\ \text{Var}(S_i) > \tau_{\text{Var}}, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Figure 5 shows examples of global and local saliency features extracted from two images. In the first image the global salient feature (upper image of column b) is enough to track salient objects, while for the second, where the salient object (leopard) is partially available, chromatic saliency is not enough to extract the object. Figures 6 and 7 show examples

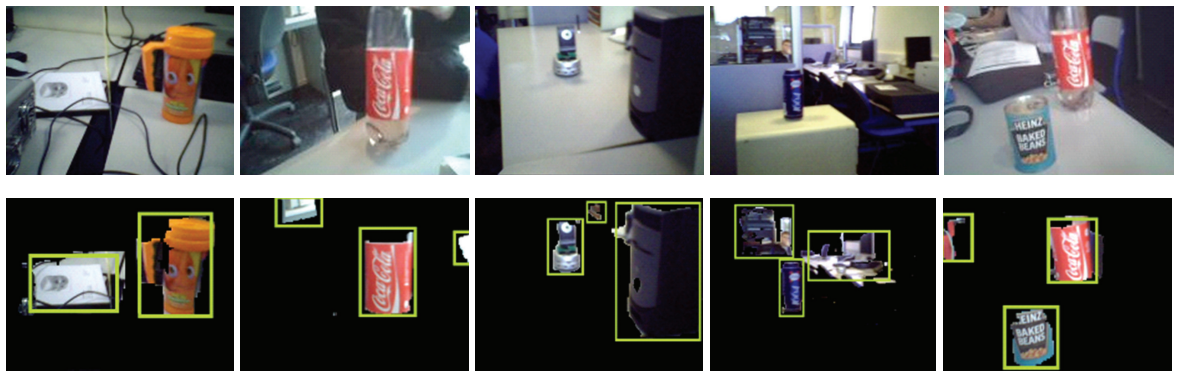

Figure 9: NAO robot camera issued images (upper images) and corresponding salient objects found and segmented by NAO (lower images).

Figure 10: Results relative to a set of objects detection by robot.

of salient object detection as well as effect of the visual attention parameter p on extracted salient regions, respectively.

4.2. Visual Intention CCVF. As it has previously been stated, composed of conscious cognitive visual functions (CCVFs), the conscious visual level conducts intentional visual tasks. One of the core functions of this level is "visual intention" CCVF, proffering the robot some kind of "artificial visual intention ability" and allowing the machine to construct its first knowledge about the surrounding environment. Figure 8 gives the bloc diagram of visional intention CCVF. As it could be seen from this figure, this CCVF is composed of four elementary functions: "short-term salient objects visual memory" (SSVM) EF, "unsupervised learning module" (ULM) EF, "salient objects classifier" (SOC) EF, and "object detector" (OD) EF.

The main task of short-term salient objects visual memory (SSVM) EF is to provide already known objects and store currently recognized or detected salient objects. It could also be seen as the first knowledge construction of surrounding environment because it contains the clusters of salient objects resulting from unsupervised learning. Its content (e.g., stored salient objects or groups of salient objects) could supply the main knowledge base (a long-term memory). That is why its output is also connected to interlevel channel.

The role of unsupervised learning (performed by ULM EF) is to cluster the detected (new) salient objects. The learning process is carried out on line. When an agent (e.g., robot) takes images while it encounters a new object, if the objects are recognized to be salient (e.g., extracted) they are grouped incrementally while new images are acquired. The action flow of the learning process is given below. In the first time, the algorithm classifies each found fragment, and, in a second time, the learning process is updated (online learning)

```
acquire image
extract fragments by salient object detector
for each fragment F
  if(F is classified into one group)
    populate the group by F
  if(F is classified into multiple groups)
    populate by F the closest group by Euclidian distance of features
    if(F is not classified to any group)
        create a new group and place F inside
select the most populated group G
use fragments from G as learning samples for object detection algorithm
```

The salient objects classifier is a combination of four weak classifiers $\{w_1, w_2, w_3, w_4\}$, each classifying a fragment as belonging or not belonging to a certain class. F denotes the currently processed fragment, and G denotes an instance of the group in question. The first classifier w_1, defined by (10), separates fragments with too different areas. In experiments $t_{\text{area}} = 10$. The w_2, defined by (11), separates fragments whose aspects are too different to belong to the same object.

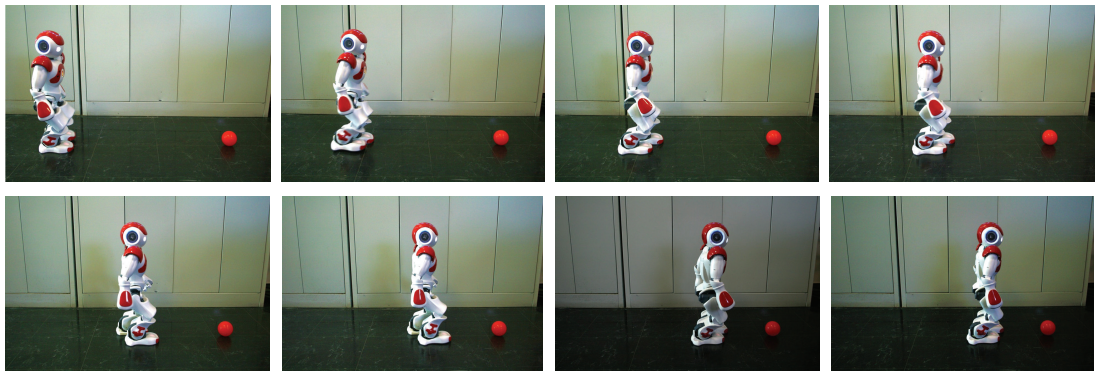

Figure 11: Results relative to an intentional object tracking: the robot tracks a red ball moving toward it.

In experiments, t_{aspect} has been set to 0.3. The classifier w_3, defined by (12), separates fragments with clearly different chromaticity. It works over 2D normalized histograms of ϕ and θ component denoted by $G_{\phi\theta}$ and $F_{\phi\theta}$, respectively, with L bins, calculating their intersection. We use $L = 32$ to avoid too sparse histogram and $t_{\phi\theta}$ equal to 0.35. Finally, w_4 (defined by (13)) separates fragments whose texture is too different. We use the measure of texture uniformity calculated over the l channel of fragment. $p(z_i)$, where $i \in \{0, 1, 2, \ldots, L-1\}$, is a normalized histogram of l channel of the fragment, and L is the number of histogram bins. In experiments, 32 histogram bins have been used to avoid too sparse histogram and value $t_{\text{uniformity}}$ of 0.02. A fragment belongs to a class if $\prod_{i=1}^{n} w_i = 1$:

$$w_1 = \begin{cases} 1 & \text{if } c_{w1} < t_{\text{area}}, \\ 0 & \text{otherwise,} \end{cases} \quad c_{w1} = \frac{\max(G_{\text{area}}, F_{\text{area}})}{\min(G_{\text{area}}, F_{\text{area}})}, \tag{10}$$

$$w_2 = \begin{cases} 1 & \text{if } c_{w2} < t_{\text{aspect}}, \\ 0 & \text{otherwise,} \end{cases}$$
$$c_{w2} = \left\| \log\left(\frac{G_{\text{width}}}{G_{\text{hight}}}\right) - \log\left(\frac{F_{\text{width}}}{F_{\text{hight}}}\right) \right\|, \tag{11}$$

$$w_3 = \begin{cases} 1 & \text{if } c_{w3} < t_{\phi\theta}, \\ 0 & \text{otherwise,} \end{cases} \text{ with}$$
$$c_{w3} = \frac{\sum_{j=1}^{L-1} \sum_{k=1}^{L-1} \min\left(G_{\phi\theta}(j,k) - F_{\phi\theta}(j,k)\right)}{L^2}, \tag{12}$$

$$w_4 = \begin{cases} 1 & \text{if } c_{w4} < t_{\text{uniformity}}, \\ 0 & \text{otherwise,} \end{cases}$$
$$c_{w4} = \left\| \sum_{j=0}^{L-1} p_G^2(z_j) - \sum_{k=0}^{L-1} p_F^2(z_k) \right\|. \tag{13}$$

4.3. Implementation on Real Robot and Experimental Validation. The above-described concept has been implemented on NAO robot, which includes vision devices and a number of onboard preimplemented motion skills. It also includes a number of basic standard functions that have not been used. For experimental verification, the robot has been introduced in a real environment with different common objects (representing different surface, shapes, and properties). Several objects were exposed in robots field of view, presented in a number of contexts different from those used in the learning phase. The number of images acquired for each object varied between 100 and 600 for learning sequences and between 50 and 300 for testing sequences, with multiple objects occurring on the same scene. During the learning process, the success rate of 96% has been achieved concerning pertinent learned objects (e.g., those identified by the robot as salient and then learned), that is, only 4% of image fragments were associated with wrong groups. During the testing process, objects were correctly extracted reaching 82% success rate.

To demonstrate real-time abilities of the system, the NAO robot was required to find some learned objects in its environment and then to track them. It is pertinent to emphasize that those objects have been learned in different environment. A sample of results of those experiments is shown in Figures 9 to 12. Figures 9 and 10 show results relating robot ability to detect and extract salient objects from its surrounding environment. It is pertinent to notice the multiple salient objects detection ability of the implemented strategy representing different shapes and various natures. Figure 11 shows the expected robot ability to detect and to follow a simple object in real environment, validating the correct operation of unconscious and intentional cognitive levels transitions in accomplishing the required task. Finally, Figure 12 shows the robot ability to detect, isolate, and follow a previously detected and learned salient object in a complex surrounding environment. The video of this experiment could be seen using the link indicated in the legend of this figure. It is pertinent to emphasize the fact that the object (a "book" in the experiment shown by the Figure 12) has been detected and learned in different conditions (as one could see this from the above-indicated video). Thus, this experiment shows the emergence of a kind of robot "artificial awareness" about the surrounding environment validating the presented cognitive multilevel concept and issued "perception-motion" architecture.

Figure 12: Tracking a previously learned moving object (upper images: video http://www.youtube.com/watch?v=xxz3wm3L1pE). The upper right corner of each image shows robot camera picture.

5. Conclusion

By supplanting the modeling of robots complex behavior from the "control theory" backdrop to the "cognitive machine learning" backcloth, the proposed machine-learning-based multilevel cognitive motion-perception concept attempts to offer a unified model of robot autonomous evolution, slotting in two kinds of cognitive levels: "unconscious" and "conscious" cognitive levels, answerable of its reflexive and intentional visual and motor skills, respectively.

The first key advantage of conceptualizing the problem within such incline is to detach the build-up of robot perception and motion from the type of machine (robot). The second chief benefit of the concept is that the issued structure is "machine-learning-" based foundation taking advantage from "learning" capacity and "generalization" propensity of such models.

The "visual intention" built-in CCVF proffers the robot artificial visual intention ability and allows it to construct the knowledge about the surrounding environment. This intentional cognitive function holds out the robot awareness regarding its surrounding environment. The extracted knowledge is first stored in (and recalled from) short-term memory. It could then be stored in a long-term memory proffering the robot some kind of learning issued knowledge about previously (already) explored environments or already known objects in a new environment. Beside this appealing ability, the unconscious visual level realizing the salient objects detection plays a key role in the so-called "artificial awareness" emergence. In fact, the ability of automatic detection of pertinent items in surrounding environment proffers the robot some kind of "unconscious awareness" about potentially significant objects in the surrounding environment. The importance of this key skill appears not only in emergence of "intentional awareness" but also in construction of new knowledge (versus the already learned items) upgrading the robot (or machine's) awareness regarding its surrounding environment.

References

[1] E. R. Westervelt, G. Buche, and J. W. Grizzle, "Experimental validation of a framework for the design of controllers that induce stable walking in planar bipeds," *International Journal of Robotics Research*, vol. 23, no. 6, pp. 559–582, 2004.

[2] J. H. Park and O. Kwon, "Reflex control of biped robot locomotion on a slippery surface," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '01)*, pp. 4134–4139, May 2001.

[3] J. Chestnutt and J. J. Kuffner, "A tiered planning strategy for biped navigation," in *Proceedings of the 4th IEEE-RAS International Conference on Humanoid Robots (Humanoids '04)*, vol. 1, pp. 422–436, November 2004.

[4] Q. Huang, K. Yokoi, S. Kajita et al., "Planning walking patterns for a biped robot," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 280–289, 2001.

[5] K. Sabe, M. Fukuchi, J. S. Gutmann, T. Ohashi, K. Kawamoto, and T. Yoshigahara, "Obstacle avoidance and path planning for humanoid robots using stereo vision," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '4)*, pp. 592–597, May 2004.

[6] R. Holmes, *Acts of War: The Behavior of Men in Battle*, The Free Press, New York, NY, USA, 1st American edition, 1985.

[7] M. Tambe, W. L. Johnson, R. M. Jones et al., "Intelligent agents for interactive simulation environments," *AI Magazine*, vol. 16, no. 1, pp. 15–40, 1995.

[8] P. Langley, "An abstract computational model of learning selective sensing skills," in *Proceedings of the 18th Conference of the Cognitive Science Society*, pp. 385–390, 1996.

[9] C. Bauckhage, C. Thurau, and G. Sagerer, "Learning human-like opponent behavior for interactive computer games," *Lecture Notes in Computer Science*, vol. 2781, pp. 148–155, 2003.

[10] V. Potkonjak, D. Kostic, S. Tzafestas, M. Popovic, M. Lazarevic, and G. Djordjevic, "Human-like behavior of robot arms: general considerations and the handwriting task—part II: the robot arm in handwriting," *Robotics and Computer-Integrated Manufacturing*, vol. 17, no. 4, pp. 317–327, 2001.

[11] J. Edlund, J. Gustafson, M. Heldner, and A. Hjalmarsson, "Towards human-like spoken dialogue systems," *Speech Communication*, vol. 50, no. 8-9, pp. 630–645, 2008.

[12] A. Lubin, N. Poirel, S. Rossi, A. Pineau, and O. Houdé, "Math in actions: actor mode reveals the true arithmetic abilities of french-speaking 2-year-olds in a magic task," *Journal of Experimental Child Psychology*, vol. 103, no. 3, pp. 376–385, 2009.

[13] F. A. Campbell, E. P. Pungello, S. Miller-Johnson, M. Burchinal, and C. T. Ramey, "The development of cognitive and academic abilities: growth curves from an early childhood educational experiment," *Developmental Psychology*, vol. 37, no. 2, pp. 231–242, 2001.

[14] G. Leroux, M. Joliot, S. Dubal, B. Mazoyer, N. Tzourio-Mazoyer, and O. Houdé, "Cognitive inhibition of number/length interference in a Piaget-like task in young adults: evidence from ERPs and fMRI," *Human Brain Mapping*, vol. 27, no. 6, pp. 498–509, 2006.

[15] A. Lubin, N. Poirel, S. Rossi, C. Lanoë, A. Pineau, and O. Houdé, "Pedagogical effect of action on arithmetic performances in Wynn-like tasks solved by 2-year-olds," *Experimental Psychology*, vol. 57, no. 6, pp. 405–411, 2010.

[16] O. C. S. Cassell, M. Hubble, M. A. P. Milling, and W. A. Dickson, "Baby walkers—still a major cause of infant burns," *Burns*, vol. 23, no. 5, pp. 451–453, 1997.

[17] M. Crouchman, "The effects of babywalkers on early locomotor development," *Developmental Medicine and Child Neurology*, vol. 28, no. 6, pp. 757–761, 1986.

[18] A. Siegel and R. Burton, "Effects of babywalkers on early locomotor development in human infants," *Developmental & Behavioral Pediatrics*, vol. 20, pp. 355–361, 1999.

[19] I. B. Kauffman and M. Ridenour, "Influence of an infant walker on onset and quality of walking pattern of locomotion: an electromyographic investigation," *Perceptual and Motor Skills*, vol. 45, no. 3, pp. 1323–1329, 1977.

[20] J. A. Andersen, *An Introduction to Neural Network*, MIT Press, Cambridge, Mass, USA, 1995.

[21] K. Madani and C. Sabourin, "Multi-level cognitive machine-learning based concept for human-like "artificial" walking: application to autonomous stroll of humanoid robots," *Neurocomputing*, vol. 74, no. 8, pp. 1213–1228, 2011.

[22] H. Bülthoff, C. Wallraven, and M. Giese, "Perceptual robotic," in *Handbook of Robotics*, B. Siciliano and O. Khatib, Eds., Springer, 2007.

[23] http://www.universcience-vod.fr/media/577/la-marche-des-bebes.html.

[24] P. Zukow-Goldring and M. A. Arbib, "Affordances, effectivities, and assisted imitation: caregivers and the directing of attention," *Neurocomputing*, vol. 70, no. 13–15, pp. 2181–2193, 2007.

[25] R. J. Brand, D. A. Baldwin, and L. A. Ashburn, "Evidence for "motionese": modifications in mothers' infant-directed action," *Developmental Science*, vol. 5, no. 1, pp. 72–83, 2002.

[26] R. Achanta, S. Hemami, E. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '09)*, 2009.

[27] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?" *Nature Reviews Neuroscience*, vol. 5, no. 6, pp. 495–501, 2004.

[28] T. W. Chen, Y. L. Chen, and S. Y. Chien, "Fast image segmentation based on K-means clustering with histograms in HSV color space," in *Proceedings of the IEEE 10th Workshop on Multimedia Signal Processing (MMSP '08)*, pp. 322–325, October 2008.

[29] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, vol. 2, pp. 1–8, June 2007.

[30] R. Moreno, M. Graña, D. M. Ramik, and K. Madani, "Image segmentation by spherical coordinates," in *Proceedings of the 11th International Conference on Pattern Recognition and Information Processing (PRIP '11)*, pp. 112–115, 2011.

[31] J. H. Holland, *Adaptation in Natural anti Artificial Systems: An introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, MIT Press, 1992.

[32] T. Liu, Z. Yuan, J. Sun et al., "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.

[33] D. M. Ramík, C. Sabourin, and K. Madani, "Hybrid salient object extraction approach with automatic estimation of visual attention scale," in *Proceedings of the 7th International Conference on Signal Image Technology & Internet-Based Systems (IEEE—SITIS '11)*, pp. 438–445, 2011.

[34] D. M. Ramík, C. Sabourin, and K. Madani, "A cognitive approach for robots' vision using unsupervised learning and visual saliency," in *Advances in Computational Intelligence*, vol. 6691 of *LNCS*, pp. 65–72, Springer, 2011.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 945401, 11 pages
doi:10.1155/2012/945401

Research Article

An Efficient Genome Fragment Assembling Using GA with Neighborhood Aware Fitness Function

Satoko Kikuchi and Goutam Chakraborty

Faculty of Software and Information Science, Iwate Prefectural University, Takizawa-mura, Iwate 020-0193, Japan

Correspondence should be addressed to Goutam Chakraborty, goutam@iwate-pu.ac.jp

Received 12 March 2012; Accepted 11 May 2012

Academic Editor: Qiangfu Zhao

To decode a long genome sequence, shotgun sequencing is the state-of-the-art technique. It needs to properly sequence a very large number, sometimes as large as millions, of short partially readable strings (fragments). Arranging those fragments in correct sequence is known as fragment assembling, which is an NP-problem. Presently used methods require enormous computational cost. In this work, we have shown how our modified genetic algorithm (GA) could solve this problem efficiently. In the proposed GA, the length of the chromosome, which represents the volume of the search space, is reduced with advancing generations, and thereby improves search efficiency. We also introduced a greedy mutation, by swapping nearby fragments using some heuristics, to improve the fitness of chromosomes. We compared results with Parsons' algorithm which is based on GA too. We used fragments with partial reads on both sides, mimicking fragments in real genome assembling process. In Parsons' work base-pair array of the whole fragment is known. Even then, we could obtain much better results, and we succeeded in restructuring contigs covering 100% of the genome sequences.

1. Introduction

1.1. What Is Genome? The study of bioinformatics is one of the most vibrant area of research, whose important applications are growing exponentially. A good starting point is the introductory book by Neil and Pavel [1]. Varieties of interesting applications are reported, where computational algorithms play an important role, as [2, 3]. Of all bioinformatics researches, genome sequencing received high importance from the beginning of this century, starting with human genome [4], to sequencing of crops [5].

Genome is the complete genetic sequence made from an alphabet of four elements, Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). The letters A, T, C, G represent molecules called nucleotides or bases. In a living cell, it appears in a double helix structure [6]. Every base A in one strand is paired with a T on the other strand, and every base C is similarly paired with G. These pairs are called base-pairs, or simply bp [7, 8].

The genome (DNA) sequences are enormously long from a few thousand nucleotides for small viruses to more than 3 giga nucleotides for human. Genomes like that of wheat (1.7×10^{10} bp) and lily (1.2×10^{11} bp) are longer than human genome (from NCBI database (the National Center for Biotechnology Information—http://www.ncbi.nlm.nih.gov/)). Obviously, deciphering them, though important, is very complex.

1.2. Why Is It Important to Decipher DNA Sequence? DNA sequence, responsible for producing different proteins, is at the root of functioning of a living organism. Decoding genome sequence is thus the first step to understand the function as well as malfunction of living things, for medical, agricultural, and many other research areas. Investigation of human genome could lead to the cause of inherited diseases and the development of medical treatments for various illnesses. The genome analysis is useful for the breeding of improved crops. Moreover, it is the key information for investigations in evolutionary biology. It is hoped that the platypus genome, which is very recently decoded in 2008, would provide a valuable resource for in-depth comparative analysis of mammals [9].

1.3. Existing Fragment Assembly Methods. Sanger sequencing [10] method is well known and most commonly used for reading genome sequences. It uses *Gel electrophoresis* that facilitates reading the bases due to reaction of the fluorescent dye staining different bases differently. But it is possible to read only a few hundred base-pairs. To overcome this problem, the target genome is cloned into copies and cut into small *fragments*. Base sequences of those fragments are read and the original genome is reassembled using information of the overlapping portions of the fragments. This procedure is called *shotgun sequencing* and is the most commonly used method for genome sequencing [11–13]. In fact, Lander et al. [4], using shotgun sequencing, presented an initial sequencing of human genome of ≈3.5 Gbps length by 2001.

Most of the existing fragment assembly systems read the fragment base-sequence by Sanger technique and reconstruct the original genome sequence with their proprietary assembling algorithms. Many assembling algorithms were proposed, the important ones being TIGR assembler [14], RAMEN [15], Celera Assembler [16], CAP3 [17], and Phrap [18]. To reconstruct the target genome, lots of fragments are needed and assembling those into correct sequence is an NP-hard problem.

1.4. GA-Based Genome Fragment Assembling Works. The present work is based on genetic algorithm (GA), which is modified to be efficient for such array assembling problem. During last ten years a few works were reported to use genetic algorithm or similar techniques like *clustering algorithm* and *pattern matching algorithm*, to solve fragment assembling problem. A survey with comparison of their respective performances is reported by Li and Khuri [19]. Most of the GA based works are simple modifications of Parsons et al.'s works [20]. Recent works [21, 22] used distributed GA. Fang et al.'s work [23] was also based on standard GA, but they used toy problems of very small genome length of 100 base-pairs. The main difference between our work and other published works based on GA is in the definition of fragments. The base-pairs of the fragments used in Parsons' and others' works are fully read. They used GenFrag [24] to generate such fragments. In reality, by Sagner technique, we can read only small portions on both sides of the fragment. In our experiments (including our previous work [25]) a small portion of base-pairs, only at the two ends of the fragment, are known. Thus we handle a problem which is more realistic and difficult compared to the case where the base-pair sequence of the whole fragment is known. In addition, due to the use of such fragments, we could realize scaffolding in our proposed method.

Several deterministic algorithms, based on graph-theory, and greedy heuristic algorithms are proposed. But they are extremely computationally involved and need large scale parallel processing computational environment which is very costly. Worldwide only a few such installations are available, and they are owned by large research facilities. Yet, the need for genome sequencing is felt more and more strongly at every small medical research centers, drug development centers, agricultural research centers, and so forth. To help progress of their researches we need an efficient fragment assembling algorithm, which could run on an inexpensive computational platform. Moreover, on many occasions what one needs is only a partial sequencing, or to know whether a particular sequence is present in the genome or not, not the whole genome sequence.

The main motivation of this work is to find an efficient fragment assembling algorithm that could run on desktops, yet be able to find nearly correct draft sequences.

In the proposed method, fragment matching, contig formation, and scaffolding all are embedded in one process. Moreover if the researcher needs to know/confirm only certain gene sequence, information of which is available in the draft sequence (in the contig pool defined in Section 3.4), she/he may stop the moment it appears in the contig pool instead of continuing more generations of genetic search.

The rest of the paper is organized as follows. In Section 2, shotgun sequencing and problems of the existing techniques are briefly explained. Section 3 is devoted to explain the proposed algorithm. In Section 4, we state the experimental setup and results of the experiments using two actual genome sequences and discuss them. Conclusion is in Section 5.

2. Shotgun Sequencing Method

2.1. Shotgun Sequencing. In this section, we explain fragment assembling based on Sanger sequencing. To decode a long DNA sequence one needs to clone it to a few copies, split it up into fragments, read the individual fragments, and then assemble them in correct sequence to reconstruct the target DNA. This process is called shotgun sequencing and is the basis of all sequencing strategies. In 2000 Myers et al. successfully sequenced the fruit fly *drosophila* genome of length ≈125 Mbps using whole genome shotgun sequencing (WGSS) [16], and consequently WGSS was established as the generally accepted technique.

WGSS was used in the determination of draft human genome in 2001 by Celera Genomics [26]. In recent years, WGSS also decoded several biotic genomes, the chimpanzee by Mita in 2005 [27], the honeybee *Apis mellifera* by Weinstock in 2006 [28], and the Sea Urchin *Strongylocentrotus purpuratus* by Sodergren et al. in 2006 [29].

Our target is similar, to create the sequence based on WGSS, doing the assembling part using GA.

2.2. Outline of WGSS. The whole process of WGSS is divided into two steps—one is the biological part of cloning, fragmenting, and reading. The other one is the computational part of assembling the fragments.

2.2.1. Biological Part. The basic shotgun procedure starts with a number of copies of DNA whose sequence is cut into a large number of random fragments of different lengths. Fragments that are too large or too small are discarded. Of the remaining fragments, that is, those used for assembling, the length of short ones is about 2 kbp and of the long ones is about 10 kbp [11]. The base-pair at both ends of all the fragments is read with DNA sequencer, shown as dark parts

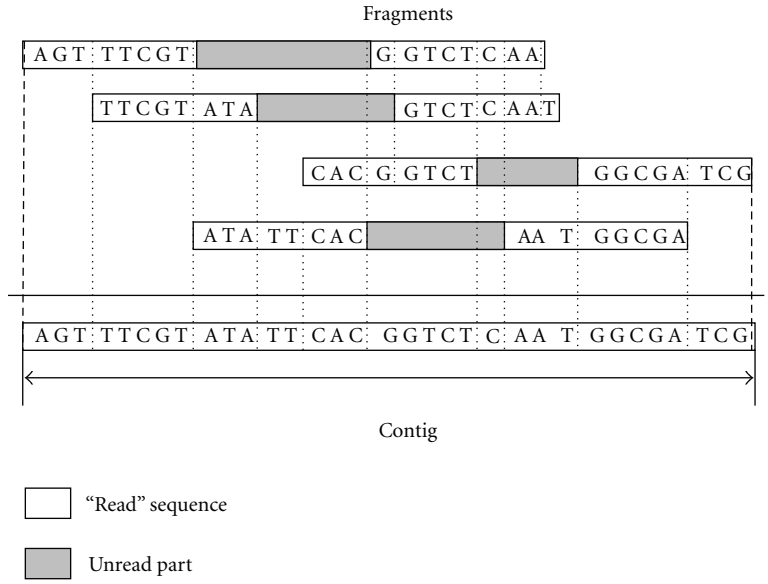

Figure 1: Formation of contigs.

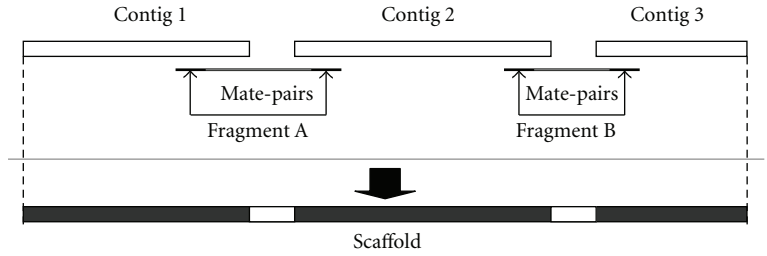

Figure 2: Scaffolding.

of fragments in Figure 1. Only about 500 to 1000 bp can be read using present sequencer technology. The base sequence at both ends of a fragment read by the sequencer is called *read*, and a pair of reads from two ends of a fragment is called *mate-pairs*.

Starting with a fair number of clones, the total base-pair reads of fragments are several times the number of bases in the original genome. Commonly, a term *coverage* is used to measure the redundancy of the fragment read data. It is defined as the total number of base reads from fragments as a ratio of the length of the source DNA [30]:

$$\text{Coverage} = \frac{\sum_{i=1}^{N} \text{reads\_of\_fragment}_i}{\text{target\_genome\_length}}, \quad (1)$$

where N is the total number of fragments. The genome is fragmented randomly. Their *read* parts together may not even include the whole genome if the *coverage* is low. To be able to reconstruct the original genome, the *coverage* needs to be set at around 8 to 10 (described as 8X∼10X). If coverage is high, the probability of covering original genome is higher and the accuracy of the assembled parts is improved. However, the number of fragments and therefore the computational complexity is also increased. Even though the *coverage* is 10X, some part of the original genome may not be available in the fragment *reads*.

2.2.2. Computational Part. To sequence the original DNA, we first identify overlapping sections by comparing the already read base sequences at both ends of the fragments, as shown in Figure 1. Long lengths of base sequences without gaps, obtained by assembling the *reads*, are called contigs. Figure 1 shows how two contigs are formed. Here, it is presumed that two overlapping *reads*, one a prefix of a fragment and the other the suffix, originate from the same region of the genome. This is however true only when the overlapped *base-pair* length is sufficiently long, as small sequences appear repetitively in the genome.

Two *reads* on two sides of a single fragment are called *mate-pair*. The position and distance between contigs are determined from the mate pair of fragments (Figure 2). Thus, subset of contigs with known order is grouped together and this process is called scaffolding. This is done by constructing a graph in which the nodes correspond to contigs, and a directed edge links two nodes when mate-pairs bridge the gap between them. Most of the recent assemblers include a separate scaffolding step. A rough frame of original genome sequence is made by this scaffolding process. After

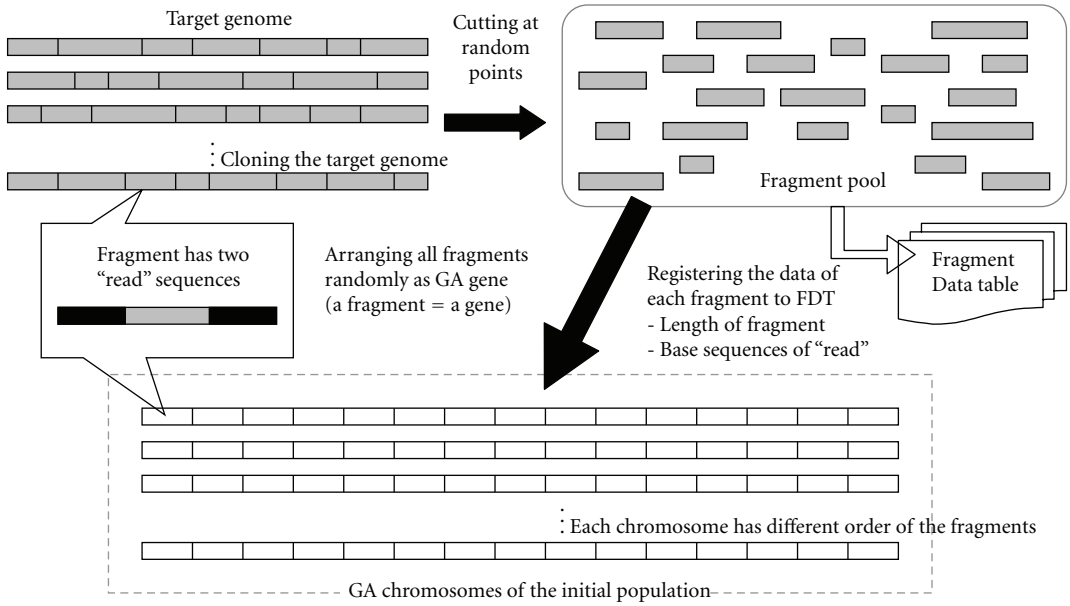

Figure 3: Chromosomes of GA for fragments assembly.

all contigs are oriented and ordered correctly, we can close gaps between two contigs. This process is called gap closer or finishing.

Celera assembler [26] employs scaffolding algorithm based on graph theory using mate pairs. TIGR assembler [14] employs a greedy heuristic algorithm.

3. Proposed Algorithm

Our proposed genetic algorithm technique is specialized for fragments assembling and similar problems. The main contribution here is Chromosome Reduction Step (CRed), which reduces the length of GA chromosome with progressive generation. As the chromosome length reduces, so does the search space, and the searching is more and more efficient. The other contribution is Chromosome Refinement Step (CRef), which is a greedy mutation to improve the correctness of the solution by local rearrangement of genes. We were able to combine the phase of overlap (contig formation) and scaffolding by the way we defined the structure of the GA chromosome and CRed. The details are explained in the following sections.

3.1. The Structure of GA Chromosome. Genes of our GA chromosome are genome fragments, where one fragment is one gene. In a GA chromosome gene, there is the information of two *read* sequences which are mate-pair, and the gap of unknown length in between. This fragment structure is different from previous works, where the *base-pair* array of the whole fragment is known. This makes our problem more realistic and complex.

The fragments generated by shotgun sequencing method are labeled in serial numbers, 1 to N, where, N is the total number of fragments. The read information corresponding to different fragments is stored in Fragment Data Table (FDT). The chromosome is composed of all these N fragments sequenced in random. Thus a chromosome is actually a permutation of numbers 1 to N, which are labels of different fragments. A number of such chromosomes, equal to the population size, are created. The formation of the initial population of GA chromosome is illustrated in Figure 3. A fragment is not cut on the way of genetic operations like crossover, because crossover and mutation are done at the boundary of fragments (genes of the chromosome). Thus, the information of mate-pair is retained without break. The flow of genetic search algorithm is shown in Figure 4. Different blocks of the algorithm are explained below in Section 3.2 to Section 3.5.

3.2. Evaluation Function. The goal of the search is to bring closer the fragments generated from the same region of the original chromosome. Fitness of a GA chromosome increases as adjacent genes match in their *base-pair* arrays. Similarities of all adjoining pairs of GA chromosome genes (which are actually the genome fragments) are calculated to find the fitness as follows:

$$\text{Fitness}(c) = \sum_{i=0}^{N-2} \text{similarity}\ (i, i+1). \tag{2}$$

The genes in a chromosome are numbered 0 to $N-1$, from left to right. In (2), i and $i+1$ are adjoining fragments and N is the total number of genes in the chromosome. To calculate the similarity (overlap), we use Smith-Waterman algorithm [31] that detects alignment of a pair of genes by dynamic programming. Here we set a threshold value mp of overlap to judge whether there is a real match between two fragments. If two fragments i and $i+1$ have the same sequence of bp

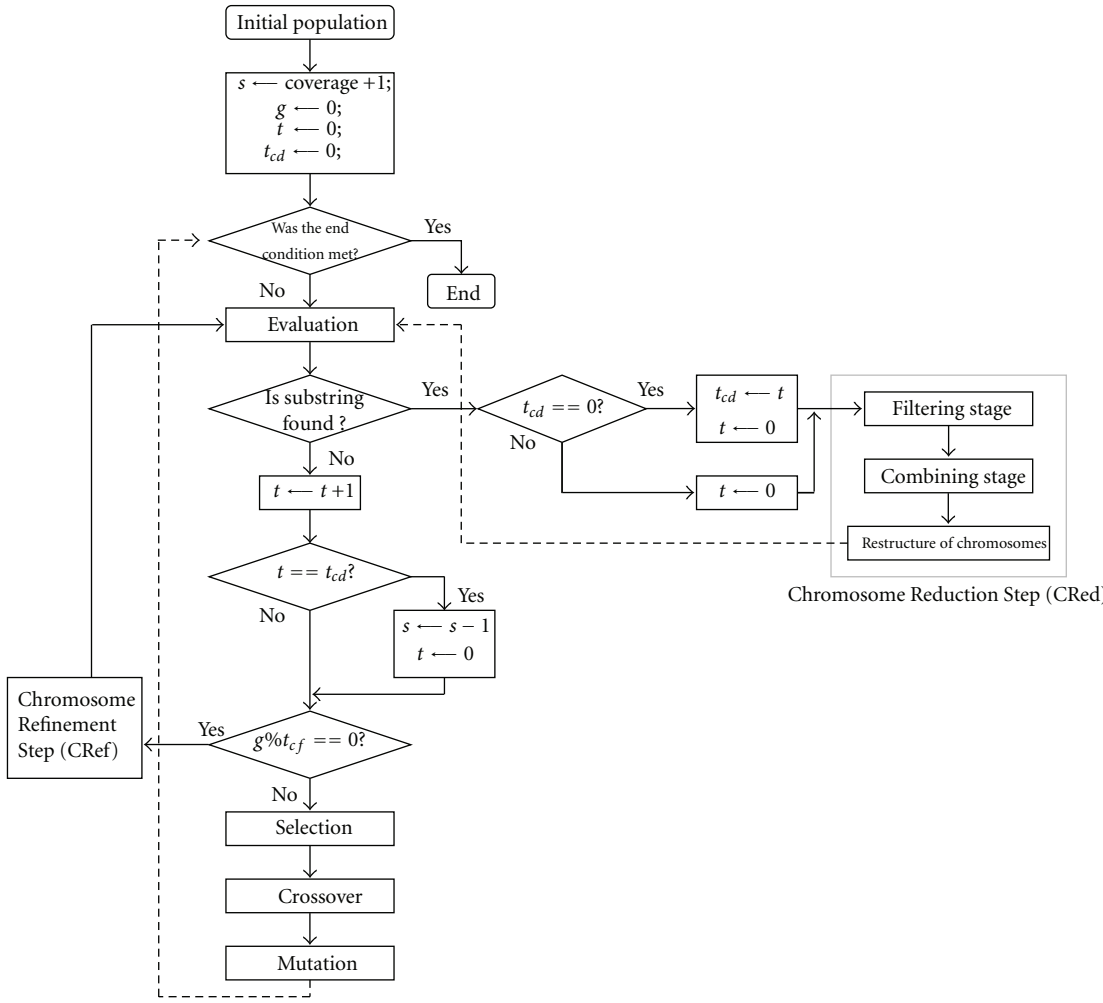

Figure 4: Algorithm including CRed and CRef.

of length more than or equal to mp, similarity $(i, i+1)$ score is 1, and otherwise it is 0. Similarity $(i, i+1)$ could have discrete values 0, 1, or 2. Because each fragment has two *reads*, similarity is 2 if two *reads* on both sides of the fragments matched. If the *reads* on only one side matches, the score is 1. If we set mp low, the misassembling (error) probability increases. On the other hand, if we set mp high, the probability of successful match would be low slowing the progress of genetic search.

3.3. Selection, Crossover, and Mutation

3.3.1. Selection. We use roulette-wheel selection and elitist preservation. Roulette-wheel selection tends to converge on local maximum when a few chromosomes have much better fitness. On the other hand, the selection is more fair as it properly takes care of individuals' fitness. Other selection methods, like ranking and tournament selection, have less selection pressure and therefore less probability to early convergence. The variance of fitness among chromosomes is low in our chromosome design, and due to CRed operation. (At the end of CRed operation every chromosomes fitness again resets to a low narrow range). In fact, in a preparatory experiment we have verified that roulette-selection is more efficient for the proposed algorithm.

3.3.2. Crossover and Mutation. We do not allow multiple copies of the same fragment in our GA chromosome. To ensure that, we used order-based crossover (OX) and swap, often used in solving TSP [32]. In OX, *offspring* 1 directly copies genes from *parent* 1, from the beginning up to the crossover point. The rest of the genes are copied from *parent* 2 preserving the sequential order of *parent* 2 and skipping the genes already copied from *parent* 1. *Offspring* 2 is constructed similarly. Here, two point crossover is also possible, but we used one point crossover.

Mutation is done by simple swapping. Two genes in a chromosome are selected at random and swapped over. In swap mutation, it is also possible to swap a subset where the selected gene is included in the subset. By doing so we can avoid breaking the subset already formed. But we did simple one pair gene swapping.

3.4. Chromosome Reduction (CRed) Step. Through generations, chromosomes bring individual fragments with long matched *base-pairs* to adjacent positions by evaluation function and selection. Once overlapping fragments are brought closer, we use CRed operation to separate out formed contigs and reorganize array of genes in the chromosome. This is done in two stages, filtering stage and combining stage. These two stages together is called Chromosome Reduction Step (CRed).

In filtering stage we search for contigs already formed in GA chromosome. The search is performed on the elite chromosome. If contig over a certain threshold length is formed, all fragments contained within that contig are extracted from all chromosomes. This shortens the length of chromosomes.

Further detail is as follows. Here, s is the threshold length of the target substring, expressed as number of fragments. g is the generation number. t is a counter for counting generations. The threshold length s is reduced from its initial value, as formation of longer substring become more and more difficult as GA generation progresses. t_{cd} is the interval in numbers of generations, which determines when s should be decreased.

First, s is initialized with the value (coverage + 1). g and t are also initialized at 0 and increments with each generation. We search for the subset/s consisting of s fragments in the best chromosome, after fitness evaluation. If such subset is not found, CRed is not started and $g \leftarrow g + 1$ and $t \leftarrow t + 1$. First time such subset/s is/are found, mark the fragments which are contained within the subset/s. Those fragments are deleted from all the chromosomes, and Fragment Data Table (FDT) is updated. At the same time, $t_{cd} \leftarrow t$ and $t \leftarrow 0$.

Marked fragments are combined based on their overlaps. The contig/s is/are stored in a separate database that we call "contig pool" which is indexed in Contig Data Table (CDT). This stage is called combining stage. If the other contig/s is/are already in the contig pool, newly formed contig is compared with those contig/s and is combined with those to get longer contigs whenever possible. Accordingly "contig pool" and CDT are updated.

As mentioned, the length of subset to be extracted, in terms of number of fragments, is initialized to $s =$ (coverage + 1). At every generation all subsets of length (coverage + 1) fragments are extracted from the best chromosome. If no such subset is assembled for consecutive t_{cd} generations, since when a contig consisting of s fragments was found, s is reduced by 1. The flow of the CRed algorithm is shown as "yes" part of the decision diamond "*is substring formed?*".

The first time, when s is initially set to (coverage + 1) and is the longest, number of generations taken for contig formation is also longest. We set t_{cd} as number of generations we allow for new substring of length s to be formed. As long as we find contigs of length s within t_{cd} generations, s is not reduced. If, even after t_{cd} generations, subset of length s is not formed, we reduce our expectations to subset of smaller lengths. We decrement the parameter s by one, as well as $t \leftarrow 0$. Once a substring is found, the filtering is done by which fragments corresponding to the newly assembled substring are deleted from all chromosomes.

After filtering stage, combining stage is executed. When a new contig is added to the contig pool, we try to combine it with the existing contigs, if possible, to make longer contigs. Once a longer contig is formed, further genes (genome fragments) could be shed off from the chromosomes the way it is done in the filtering stage.

In filtering stage of CRed, the fragments in the substring extracted from the chromosome, may join to one end of an existing contig, or it may join two contigs on two sides to form a very long contig. Information of contigs after filtering operation is updated to Contig Data Table (CDT). Information about their relationship, if any, obtained from *mate-pairs* are added to the Scaffold Data Table (SDT). Thus, SDT holds the information about the label of contigs and their relative positions. Every time combining stage starts, new contigs are compared with existing contigs and combined when possible. CDT and SDT are renewed after that. Using simple user interface, the formation of contigs and scaffolds can be visualized and it is possible to manipulate them manually by the user or an expert, when available.

As the contigs become longer and chromosomes shorter, GA runs more efficiently. After every combining stage, the user could check whether the available results are good enough (long enough) for her/his purpose. If not, the genetic search continues.

3.5. Chromosome Refinement (CRef) Step. Instead of depending on genetic search alone, we add a step to facilitate proper sequencing more efficiently by manual greedy swapping. This is a simple and fast heuristic that we named CRef.

CRef improves the quality of solution by rearranging the sequence of fragments in a GA chromosome to correspond to the base sequence in the target genome. When two fragments A and B are sequentially positioned in a chromosome due to high-degree of overlap, the following overlap patterns, as shown in Figure 5, are possible.

If two fragments have overlap of pattern 4, it is obvious that their sequential order is wrong in the chromosome. We swap the positions of these two fragments. With this, the positions of fragments in GA chromosome are arranged to correspond to their positions in the original genome as shown in Figure 6.

This concept could be extended by expanding the scope of fragment comparison, beyond that of adjacent fragments only. We set a numeric parameter f_{cf} which represents the scope of comparison of fragments. For example, when $f_{cf} = 3$, all three neighboring fragments are compared. Though, while evaluating the fitness of a GA chromosome, we compare two adjoining fragments only, it is possible to find higher similarity with a fragment which is one or two fragments away. This could be explored by setting f_{cf} to a value more than 2 and running CRef.

A detail explanation of the CRef operation is as follows. Here, g is the number of generations, and t_{cf} is the parameter specifying the interval, in number of generations,

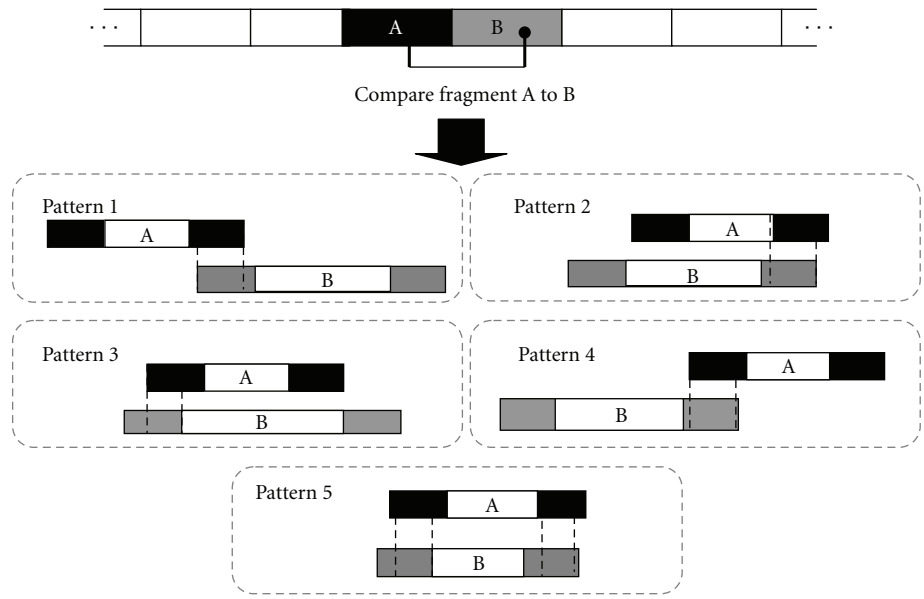

Figure 5: Matching pattern of two fragments. Pattern 1: tail-part of fragment A overlaps with the beginning of fragment B. Pattern 2: tail-parts of the two fragments overlap. Pattern 3: beginning of the two fragments overlap. Pattern 4: beginning of fragment A overlaps with end of fragment B. Pattern 5: both beginning and end of the two fragments overlap.

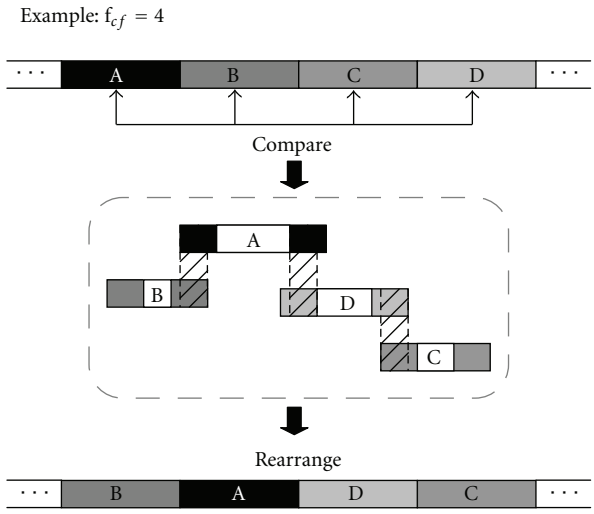

Figure 6: Chromosome Refinement (CRef) Step.

of the periodicity of execution of CRef. Thus, if t_{cf} is 100, CRef will be executed every 100 generations. N is the total number of gene in the chromosome. As the CRef operation is computationally heavy, it is run over the best few chromosomes in the population. We set a parameter rc which specifies the number of chromosomes on which CRef operation will be undertaken.

First, we set the values of f_{cf}, t_{cf}, and rc. Typical values are $f_{cf} = 3$, $t_{cf} = 100$, and $rc = 10$. At intervals of t_{cf} generations, we compare matching portions from fragment 1 to N, taking consecutive f_{cf} fragments at a time, from best rc chromosomes. If pattern 4 type matching (Figure 5) or fragments of high similarity (but not adjacent) are found, we rearrange them properly as shown in Figure 6.

The computation cost is low much because CRef is neither executed on all chromosomes nor is executed at every generation. It is limited to a few high-fitness chromosomes (set by the parameter rc) and the periodicity of execution is set by t_{cf}. These two parameters are set depending on the available computational power and time.

With these two steps of CRed and CRef, both the efficiency and quality of result of our genetic search greatly improved.

4. Experiments and Results

In this section, we describe the details of our experimental setup, discuss the results, and compare them with a the most frequently referred GA-based assembling proposed by Parsons et al. [20].

4.1. Experimental Genome Data. We used two real genome sequence data, also frequently used by other researchers, to test the effectiveness of our algorithm. They are available in the NCBI database [33]. The important features are shown in Table 1. POBF is the human apolopoprotein, which is 10089 bp long. AMCG is the initial 40% of the bases from LAMCG which is the complete genome of bacteriophage lambda and its length is 20100 bp. We scaled down number of fragments, fragment length, and *read* size compared to actual shotgun sequencing experiments to reduce the computation time. In actual experiment of shotgun sequencing, the genome fragments are of length 2 kbp~10 kbp and *read* lengths are 800 bp~1000 bp on both sides of the fragment. We cloned each genome sequence and splitted them into

Table 1: Experimental genome data.

POBF	AMCG
10089 bp	20100 bp
Accession no.: M15421	Accession no.: J02459
Human apolipoprotein B-100 mRNA, complete cds.	Bacteriophage lambda, complete genome (initial 40%)
Number of fragments: about 500	Number of fragments: about 1000

Table 2: The differences between proposed GA and Parsons' GA.

	Our GA	Parsons' GA
Gene of GA chromosome	Fragment with 2 *read*s	*read*s only
Fitness function	Equation (2)	
Crossover	Order-based crossover	
Mutation	swap mutation + greedy mutation	Swap mutation
Heuristic part	CRed and CRef	None
Scaffolding	possible	Not possible

fragments of length 200 bp to 500 bp imitating the shotgun method, but scaling down the length of a fragment by a factor of 10. Each *read* is set to 50 bp, scaled down by a factor of 12 to 20. Thus both the *read* and the fragment length are scaled down by similar factor compared to actual shotgun fragment assembly [11]. We also reduced the coverage from the standard value of 8X to 10X, so that the number of fragments is less. This reduces the computation load, but at the same time reducing the success rate of assembly the whole genome. There is further explanation in Section 4.2.

4.2. Experimental Setup. We implemented Parsons' GA-based algorithm and compared results with proposed algorithm under same experimental conditions. The basic differences between our GA and Parsons' GA are shown in Table 2.

In the experiment described in Parsons' paper, they used the some POBF and AMCG data. But the fragment lengths were different, and the whole fragment was readable. In our case, the *read* is only of small length (at the two ends) of the fragment—which is more akin to actual whole genome shotgun sequencing method and obviously more difficult. We ran both Parsons' and our algorithm on the same genome data. To reduce computation load, we set the *coverage* to 4X and 5X for POBF and AMCG, respectively, a much lower value compared to 8X~10X used in actual genome sequencing.

4.3. Setup of GA Parameters. Population size is set at 100 chromosomes which are generated by technique explained in Section 3.1. The crossover rate and the mutation rate are set at 0.8 and 0.05, respectively. In the experiment with POBF sequence, we ran genetic operation for 40 hours (about 2,000,000~2,500,000 generations), and 100 hours (about 5,000,000~6,000,000 generations) for AMCG experiment, using a desktop PC (CPU is Intel Xeon 3.40 GHz and 3.00 GB RAM). The threshold parameter mp (Section 3.2) is set at 25 bp. mp length of 50% of *read* is a strict setting. But a shorter mp could lead to error in the final assembled array from unwarranted matching. In fact, in other experiments, a much lower value of 5% (of the read which is 40~50) is generally used [26]. But if we set mp around 40 to 50, we could hardly get any match, with a *read* of length 50.

We defined s in Section 3.4. This parameter, which triggers the starting of CRed operation, is initially set at a value equal to (coverage + 1). The value of s is decreased in steps of 1 till $s = 2$.

In another preparatory experiment we examined how to set the proper value of the parameter f_{cf}, used in CRef. We experimented setting the value of f_{cf} between 2 to 5. Repeated experiments showed that $f_{cf} = 3$ gives the best result.

4.4. Results. We experimented 20 trials with different sets of fragments, with POBF and AMCG genome data. In our proposed technique, the number and length of contigs were checked in the "contig pool" (Section 3.4). The reconstruction ratio for the proposed algorithms as well as Parsons' algorithm is the percentage of base sequences which could be reconstructed in a certain period of time. We counted the number of bases of all obtained contigs. It is good if the reconstruction ratio is high. However even if the reconstruction ratio is high, a result with many gaps is not good. The number of contigs is that measure. Less number of contigs are better. If gaps are frequent, the results may be useless. Error is the number of the mismatches. Mismatch means fragments are combined incorrectly, that is, not the way they are in the original genome. This may happen when mp is chosen to be too short. The contigs generated using our algorithm and Parsons' GA, for POBF and AMCG data, are compared in Table 3. The value in parenthesis is the number of times 100% reconstruction is achieved.

With our proposed GA, we could reconstruct the complete original genome of POBF twice. Though Parsons' GA could obtain the complete genome in their paper [20], the experimental genome data used by them did not have any gap and the whole fragment was *readable*, making the problem less complex. With an unread gap in the fragment between two *reads*, the fragment data used in our experiment is realistic though reconstruction of the original genome sequence was more difficult.

Because fragmentation is done randomly, though the overall coverage was 4X, some part of the genome may be covered only once by the fragment *reads*. In Parsons' algorithm, if those fragments could not be matched to a combinable contig in the early stage, they are left alone by the other subsets already formed. This is because the fragment matching is tried for the adjoining fragments only. Those fragments are left alone, and only mutation can combine them. In contrast, our proposed technique has higher chance of combining those fragments. One reason is that, because we have two *reads* separated by an *unread* portion, it is unlikely that base-pairs in both *reads* are sparse in number. Moreover,

Table 3: Results about contigs.

	POBF				AMCG			
	Proposed GA		Parsons' GA		Proposed GA		Parsons' GA	
	Average	Best	Average	Best	Average	Best	Average	Best
Number of contig	27.6	1	19.8	9	38.3	8	35.1	11
Length of contig	317.3	10089	342.6	1008	349.4	2795	316.4	1341
Reconstruction ratio	86.8	100 (2)	67.2	74.3	66.5	72.1	55.2	61.1
Error	0	0	0	0	0	0	0	0

Table 4: Results about scaffolds.

	POBF		AMCG	
	Average	Best	Average	Best
Number of scaffold	4.1	1	12.4	6
Length of scaffold	2135.9	10089	1079.1	3888

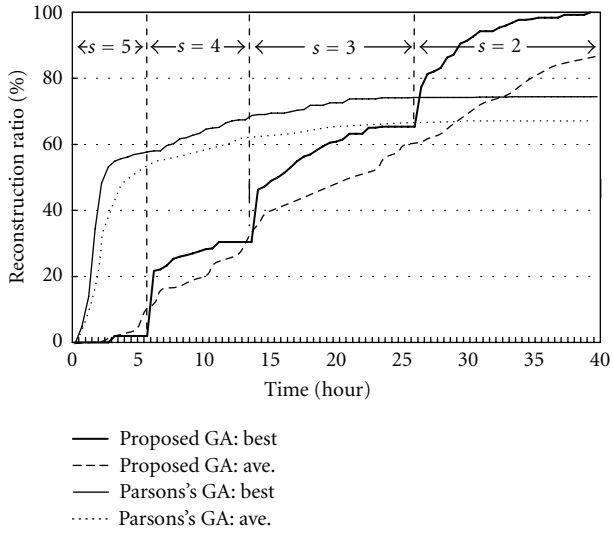

Figure 7: The improvement of the reconstruction ratio: POBF dataset.

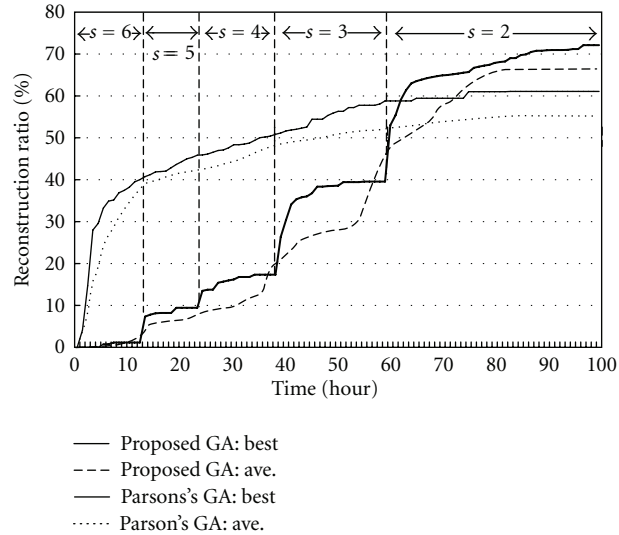

Figure 8: The improvement of the reconstruction ratio: AMCG dataset.

due to CRed operation, there are more realignments of fragments as well as shortening of the whole chromosome. They together improve the chance of combining the lone *reads*.

The average length of contigs using POBF data in our proposed GA is slightly lower than that of Parsons' GA. However, the scaffolds generated by our proposed technique, as shown in Table 4, are longer. The length of scaffold shown is the length of known contig parts only. Scaffolding is achieved as part of the proposed algorithm, which is a necessary step in shotgun sequencing.

The improvement of the reconstruction ratio versus execution time is shown in Figures 7 and 8 for POBF and AMCG, respectively. We checked the reconstruction ratio every 30 minutes for POBF and every hour for AMCG.

We calculated the reconstruction ratio from the total length of all contigs in the contig pool. It was 0% in the beginning until CRed started its operation. Because we set a threshold value for substring length to be taken out from chromosome and transferred to contig pool, improvement of reconstruction ratio was stagnant periodically. When the stagnation continued over a certain period of time, CRed parameter s is lowered to break the stagnation. At every such period, s is set high with the hope of getting long contigs. When it appears that such long contigs could not be constructed, s is reduced by 1. Then new contigs were found in the GA chromosome and the reconstruction ratio increased again. We can clearly observe such surges in the curve of reconstruction ratio. For POBF, there are four steps. Initially s was set to 5 (coverage + 1), and subsequently reduced to 4, 3, and 2. For AMCG, there are five steps, starting with $s = 6$ and finally reduced to $s = 2$.

Even at the end of the predefined length of execution time (40 hours and 100 hours), the results were improving. By increasing f_{cd} and the total length of execution time, better results could be obtained. Our algorithm would be able to get much better results with actual fragment assembly data of high coverage, longer, *reads* and lower value of mp (in percentage of *read*). We could not perform such experiments due to lack of computation resource.

5. Conclusion

We proposed a genetic-algorithm-based approach to assemble DNA fragments to construct the genome sequence. Our GA chromosomes were different from previous approaches. We also added two modifications, Chromosome Reduction step (CRed) and Chromosome Refinement Step (CRef), to improve the efficiency of GA optimization for fragment assembly. Experimenting with actual genome data, we could obtain 100% of the POBF genome sequences. We compared our proposed algorithm with Parsons's algorithm and have shown that the proposed algorithm delivered better results.

We used a coverage of only 4X instead of more practical ≈10X. More fragments covering the same part of the sequence are required for better reconstruction ratio and accuracy. Also, introducing biological knowledge is helpful for effective fragment assembling and improving its accuracy, which, in future, we will link with our genetic search, in CRed and CRef operations. In amino acid and protein formation, there are some distinct rules in the alignment of bases. Using that pattern knowledge we could decrease the computational cost during similarity measurements in addition to dynamic programming. We plan to include domain knowledge to design more efficient dynamic programming specific to this problem. In fact, if we know the threshold length *mp*, it is much efficient to compare blocks of nucleotides of length *mp*. There is another drawback in our approach. In chromosome refinement stage, we evaluate the degree of matching using the way the two fragments are oriented in the GA chromosome. In fact, the orientation may be reversed, from the way they appear in the actual genome. Comparing both orientations for neighboring fragments would be computationally more complex, but we can achieve the whole genome structure in less number of GA generations, as well as with less number of fragments. These two modifications are our future work.

Though CRed step is proposed for fragment assembling problem, it is applicable to similar problems like clustering, path search and other combinatorial optimization.

References

[1] J. C. Neil and P. A. Pavel, *An Introduction to Bioinformatics Algorithms*, A Bradford Book, 2004.

[2] Y. Li and J. C. Patra, "Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network," *Bioinformatics*, vol. 26, no. 9, Article ID btq108, pp. 1219–1224, 2010.

[3] Y. Li and J. C. Patra, "Integration of multiple data sources to prioritize candidate genes using discounted rating system," *BMC Bioinformatics*, vol. 11, no. 1, article S20, 2010.

[4] E. S. Lander, L. M. Linton, B. Birren et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[5] T. Sasaki, "The map-based sequence of the rice genome," *Nature*, vol. 436, no. 7052, pp. 793–800, 2005.

[6] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[7] P. A. Benjamin, *Genetics—A Conceptual Approach*, W. H. Freeman and Company, 2005.

[8] P. P. Vaidyanathan, "Genomics and proteomics: a signal processor's tour," *IEEE Circuits and Systems Magazine*, vol. 4, no. 4, pp. 6–29, 2004.

[9] W. C. Warren, L. W. Hillier, J. A. Marshall Graves et al., "Genome analysis of the platypus reveals unique signatures of evolution," *Nature*, vol. 453, no. 7192, pp. 175–183, 2008.

[10] F. Sanger, A. R. Coulson, G. F. Hong, D. Hill, and G. Petersen,, "Nucleotide sequence of bacteriophage λ DNA," *Journal of Molecular Biology*, vol. 162, no. 4, pp. 729–773, 1982.

[11] M. Pop, "Shotgun sequence assembly," *Advances in Computers*, vol. 60, pp. 193–248, 2004.

[12] M. T. Tammi, *The Principles of Shotgun Sequencing and Automates Fragment Assembly*, Center for Genomics and Bioinformatics, Karolinska Institute, Stockholm, Sweden, 2003.

[13] S. Kim, "A survey of computational techniques for genome sequencing," Project Report, Korea Institute of Science and Technology Information, 2002.

[14] G. G. Sutton, O. White, M. D. Adams, and A. R. Kerlavage, "TIGR assembler: a new tool for assembling large shotgun sequencing projects," *Genome Science and Technology*, vol. 1, pp. 9–19, 1995.

[15] K. Mita, M. Kasahara, S. Sasaki et al., "The genome sequence of silkworm, Bombyx mori," *DNA Research*, vol. 11, no. 1, pp. 27–35, 2004.

[16] E. W. Myers, G. G. Sutton, A. L. Delcher et al., "A whole-genome assembly of *Drosophila*," *Science*, vol. 287, no. 5461, pp. 2196–2204, 2000.

[17] X. Huang and A. Madan, "CAP3: A DNA sequence assembly program," *Genome Research*, vol. 9, no. 9, pp. 868–877, 1999.

[18] P. Green, "Phrap documentation : algorithms," Phred/Phrap/Consed System Home Page, April 2006, http://www.phrap.org/.

[19] L. Li and S. Khuri, "A comparison of DNA fragment assembly algorithms," in *Proceedings of the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS '04)*, pp. 329–335, June 2004.

[20] R. J. Parsons, S. Forrest, and C. Burks, "Genetic algorithms, operators, and DNA fragment assembly," *Machine Learning*, vol. 21, no. 1-2, pp. 11–33, 1995.

[21] K. Kim and C. K. Mohan, "Parallel hierarchical adaptive genetic algorithm for fragment assembly," *IEEE Congress on Evolutionary Computation*, vol. 1, pp. 600–607, 2003.

[22] E. Alba, G. Luque, and S. Khuri, "Assembling DNA fragments with parallel algorithms," in *Proceedings of IEEE Congress on Evolutionary Computation (CEC '05)*, vol. 1, pp. 57–64, September 2005.

[23] S. C. Fang, Y. Wang, and J. Zhong, "A genetic algorithm approach to solving DNA fragment assembly problem," *Journal of Computational and Theoretical Nanoscience*, vol. 2, no. 4, pp. 499–505, 2005.

[24] M. L. Engle and C. Burks, "Artificially generated data sets for testing DNA sequence assembly algorithms," *Genomics*, vol. 16, no. 1, pp. 286–288, 1993.

[25] S. Kikuchi and G. Chakraborty, "Efficient assembling of genome fragments using genetic algorithm enhanced by heuristic search," in *Proceedings of IEEE Congress on Evolutionary Computation (CEC '07)*, vol. 1, pp. 305–312, September 2007.

[26] J. Craig Venter, M. D. Adams, E. W. Myers et al., "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, 2001.

[27] K. Mita, "Initial sequence of the chimpanzee genome and comparison with the human genome," *Nature*, vol. 437, no. 7055, pp. 69–87, 2005.

[28] G. M. Weinstock, G. E. Robinson, R. A. Gibbs et al., "Insights into social insects from the genome of the honeybee *Apis mellifera*," *Nature*, vol. 443, no. 7114, pp. 931–949, 2006.

[29] E. Sodergren, G. M. Weinstock, E. H. Davidson et al., "The genome of the sea urchin Strongylocentrotus purpuratus," *Science*, vol. 314, no. 5801, pp. 941–952, 2006.

[30] J. Setubal and J. Meidanis, *Introduction to Computational Molecular Biology*, PWS Publishing Company, 1997.

[31] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *Journal of Molecular Biology*, vol. 147, no. 1, pp. 195–197, 1981.

[32] Z. Michalewicz, *Algorithms + Data Structures = Evolution Programs*, Springer, 1999.

[33] The National Center for Biotechnology Information, Home page at http://www.phrap.org/.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 760908, 5 pages
doi:10.1155/2012/760908

Research Article

The Aspects, the Origin, and the Merit of Aware Computing

Yasuji Sawada

Tohoku Institute of Technology, 35-1 Yagiyama-Kasumi, Taihaku, Sendai 982-8577, Japan

Correspondence should be addressed to Yasuji Sawada, sawada@tohtech.ac.jp

Received 24 February 2012; Accepted 2 May 2012

Academic Editor: Qiangfu Zhao

In this paper we tried to understand scientifically the awareness, a daily word. Some aspects of awareness, such as qualitative or quantitative, the targets of awareness, either the external world or the internal world, were discussed. Suggestion on the human awareness was described from the experimental results of visual hand tracking. The origin and the merit of awareness in the process of evolution of animals were discussed. Finally some characters of possible aware computers and aware robots were studied.

1. Introduction

For the scientists and the engineers to understand and to make use of the human ability, it is needed to translate into scientific terms, the words of human science expressing the human ability, which has been created in a long history. *awareness* is a word in human science. Now we are asking what it is explained in scientific terms.

Awareness in human science term naturally implies existence of a subject. There is no awareness in a system which has no subject. Subject implies existence of a central information processing system and an exterior self-expression device. There is no subject which has no mind and body. In the history of biological evolution, brain was originally created to produce information for body motion effective for survival against external change. A sensorial system was also created simultaneously to check the effectiveness of the motion.

In this paper, we survey how this word has been used in our daily life. We analyze them and try to extract essential factors and try to discuss them by scientific terms (Section 2). We also refer to the results of our recent experiments of visual tracking [1–4]. These experimental results suggested that two kinds of awareness exist in the human sensorial-motor system, which functions in the mutually exclusive manner (Section 3). Combination of the various aspects of the awareness, which we used in daily life and some experimental implication, has led us to believe that "usefulness" may be a keyword to understand the awareness and to apply the concept of awareness for the aware computers and aware robots. Thus, the paper is constructed in the following order:

(1) some aspects of awareness,

(2) experimental suggestions on the concept of awareness,

(3) usefulness of awareness,

(4) awareness of the internal world and mind,

(5) aware computer,

(6) an aware robot which behaves as if it had a free will.

2. Some Aspects of Awareness

In this section, we examine some sentences including awareness used in daily life and try to find some aspects among them.

2.1. Qualitative and Quantitative.

> *"I am aware of following the preceding car within the distance of ten meters."* (S1)

This sentence means existence of a subject, an action, and an evaluation. The evaluation is quantitative. We may call this awareness "quantitative awareness." On the other hand, human civilization has sorted the complex phenomena of the external world into a countable number of the concepts. Awareness is also used to identify the kind of the body action with one of the known concepts. For example,

> *"I am aware of walking in spite of my original intention of running."* (S2)

We might call this kind of awareness as "qualitative awareness." These examples show that for the awareness of any kind, body motion seems indispensable as an exterior self-expression device of self. We examine if this is true in the sections below.

2.2. External and Internal. Being aware of external phenomena is daily, as

> *"I am aware of snow falling outside the window."* (S3)

When we are aware of some phenomena which occurs with some distance from the observer's body, the target of awareness is external. However, it is important to realize the difference between (S3) and a sentence such as

> *"A camera is ready to take the record of falling snow outside."* (S4)

The difference is essential for understanding an aspect of awareness, which we discuss in Sections 2.4 and 4. Furthermore, it is also important to understand that the target of awareness is external even when the target of awareness is within our body.

> *"I am aware of my stomach aching."* (S5)

> *"I am aware of my skin hurting as the result of my falling down on the floor."* (S6)

One may think that the stomach is internal, whereas the skin is external. But there is no principal difference between the pain in the skin and the pain in the stomach, by assuming that all the body parts are external from the central information processor (CIP). As we mention later, even the memory system is external to the subject. In that sense, stomach is not internal. The stomach is painful because it was attacked by the external virus. Information of pain is sent to the CIP from a sensor system in a part of the body. This information was created to inform the damage of the parts of the body for survival. The information on the condition of the parts of the system is needed for the evaluation of the action. For survival purpose repair or replacement of the part may be important.

The targets of sensing a part of body (S5) and (S6) are considered also external.

2.3. Awareness of Thinking.

> *"I am aware of imagining the falling snow outside of the house."* (S7)

Thinking or imagining is definitely internal. But the time when he imagines the falling snow and the time when he is aware of imagining it is not simultaneous. Most of the time, he may be imagining the falling snow, and this fact is constantly memorized in the memory system. From time to time, the CIP is switched from the imagination to the memory system, and he became aware he had been imagining up to this moment. The switching is done so smoothly and so quickly that he does not realize if the information is from past memory or real time. Although awareness is used in various ways, it can be unified by considering that the CIP is watching various parts of the external world including other body parts of the subject and even itself through the memory system.

2.4. Awareness and Experience. The awareness is identification of either observed phenomena or the image in the brain as one of the patterns experienced in the past. The awareness sometimes identifies information of the state of the local part of body, such as stomach, muscle, as one of the experienced patterns such as pain or itch. When we observe the dynamics of some object, we are aware what is going on. In contrast, a video-camera observes the same thing, but it is not aware what is going on. The difference clearly tells us that awareness is related to our knowledge.

When we say we are aware of something, the object of the awareness is either the phenomena in the external world, or the image of our internal world projected from the real world. In other words, we are never aware of the image which we never experienced. When the input signals into our brain are cutoff and when we are computing anticipatively what is going on in the external real world, the image is the imitation of the real world. Then, the brain commands the motor system to respond to the external world for adaptation or optimization.

2.5. Difference between the Awareness and Self-Monitoring. An important conceptual question arises: is the awareness different from self-monitoring? If a part failed and it was repaired automatically by a new part quickly enough to be in time for real-time processing, it is certainly a self-monitoring computer, but is it an aware computer?

Awareness is a kind of self-monitoring. But awareness is special in the sense that it monitors its own function by a CIP not by a distributed system. Any central monitor system observes the function of the total system by reducing information, either by identifying the pattern of function with one of the stored pattern, or by segmenting the total system into a relatively small number of groups and by identifying the type of mal-function with the stored pattern. The aware computer may say in the former case "I am aware that I am writing a paper." It may say in the latter case "I have a pain in my stomach." No CIP can be aware that the 31586th cell in my stomach has trouble in calcium channel.

3. Experimental Suggestions on the Concept of Awareness

Hand tracking is the experiments [1–4] in which a subject is asked to follow by a cursor as accurately as possible a target moving on a screen programmed by a computer. When the target is shown it is a visual tracking experiment, while it is an intermittently blind target tracking when the target is intermittently hidden. Among various facts found in the tracking experiment, two facts are relevant here. In the visible tracking experiments, the cursor is always in

an error-corrective mode, and therefore is retarded with respect to the target. On the other hand, the cursor moves in an anticipatory mode and precedes the target in the blind tracking [2–4].

A question we ask here is whether visible tracking mode and blind tracking mode are both the aware computing? This question was a start which has lead me to a general question this present paper is studying. It should be most natural to say that a visual tracking is a typical aware computing, because it is an experiment in which a subject's brain computes a proper hand motion which should minimize the error he is aware of with respect to an external target. On the other hand, it should be discussed whether a blind tracking experiment is aware computing or not. From the discussion in Section 2 we can say that this mode is also aware computing. The brain computes the hand motion to follow the hidden target using the memory of target motion observed in the visible region.

Another experiment proved that the same acceleration occurs even in a fully visible target tracking, if the target motion is fast [1]. This result was explained that the information processing speed of the vision system is not fast enough to take in the positional information incessantly, and that the percept-motor system is controlled to partly use the predictive mode instead of error-corrective mode. In addition, it is shown that two modes do not operate simultaneously. They are mutually exclusive. Why it has to be mutually exclusive? Is it possible to make a system which is aware of the both? This question is probably related to the question why internal world exists first of all. It might be that the internal world is created to compensate information which is not always available by some reason, for example, when the external motion is too fast that the visual system cannot process the information of the instantaneous information which evolves too quickly, as shown experimentally [1]. If a computing system is fast enough by using a fast processor, there may be no need to have an internal world.

When the speed of the external world is not very fast, one may think it may be possible to use both external information and the internal information at the same time. But it is not the case, because it must creates its own consistent body motion. For this purpose, CPS must have one and only one awareness at a time. They are thus mutually exclusive.

4. Usefulness of Awareness

What is the merit of aware computing? The awareness has been developed in biological system for survival against both external enemies and internal troubles. A number of perceptive systems and predictive capabilities were developed to protect biological systems from the external enemies, and a variety of feelings and internal neural systems were developed for warning of the possible internal system troubles.

Are they necessary for the artificial computer systems, too? We shall discuss this problem in a later section.

The sentences from (S1) to (S7) except (S4) used in daily life shown in the previous section include awareness, but does not always seem to be related to evaluation of his action. Present human society is deviating from the severe biological evolutional society. Each action of individual human is not necessarily related to the survival evaluation and therefore, awareness under these circumstances is not accompanied by evaluation. It is not important for survival whether he runs or walks in an ordinary life. But it is certainly important if he is trying to escape from a tsunami. An aware system has advantage for surviving, because it can sometimes predict the external world and can give a chance to move to avoid it.

Awareness is the real-time knowledge on the evaluation of the action with respect to the effectiveness for survival.

5. Awareness of the Internal World and Mind

The awareness of the internal world discussed so far is closely related to the mind, which is also a deep human language, not a scientific one. How close we are now to the concept of mind? We discussed awareness of thinking and awareness of the status of the body parts, pain of the stomach, pleasantness of the whole body, and so forth.

Perhaps only remaining part of mind is the problem of free will.

Proactive hand motion was observed in tracking experiments [1–3]. When they found that their hand moved proactively with respect to an object which they were supposed to follow as accurately as possible, many subjects reported that they felt as if they were intending to lead the object. This results pushed me to imagine a possible physical interpretation of the "inversion of causality," which is against principles of physics, but necessary for understanding the "free will" of human being.

The mechanism of proactive motion was clarified by a recent intermittently visible tracking experiments [2]. When the target is visible and moves slow enough, hand motion is error corrective and retarded with respect to the target motion (error-corrective mode). On the other hand, when the target is moving invisibly on an already known orbit, the hand motion was found to precede the target. It was measured that the hand motion is accelerated as soon as the target is hidden. It was understood that the hand is then controlled by a predictive mechanism (predictive mode).

As discussed in a previous section, the awareness of the internal world may be created to compensate a slow processing of the human perceptive information. The preceding research [2] showed that the internal clock moved faster than the evolution of the real physical world, and the proactiveness of the body motion caused by the faster internal clock helps to optimize the dynamic error [1]. Internal world is the predicted projection of the external world using the information obtained in the past. This evolution of the internal world referring to the external world is a part of the activity in the brain called mind. The other part of the mind is the awareness of the local part of the body such as pain, discussed in Section 2.2.

6. Aware Computer

Already, most of the present computer systems are equipped with aware functions to some extent. One of the external

enemies is virus, and antivirus vaccination software was developed greatly. Some alarm softwares informing of possible local trouble have also been developed. Nevertheless, the security of the computers equipped with these functions are monitored and taken care by human being not by the computer itself. Computers may be considered to be exposed to a severe survival society, unlike humans whose awareness is not directly related to survival. But we realize that a computer itself is not exposed to the severe market competition in real time. It is the manager of the computer company who is aware of the competition. In this sense, the computer is not a self-closed machine. It is not automotile either, like a future aware robot. When a computer is installed in a future robot which may move independently of the human control, the security must be controlled by a CPS of the robot. For this purpose, the concept of the self-closed real time aware computing will be indispensable.

From the following examples that we notice, it requires evaluation of the function and choice for some unknown factors for a computer or for a robot to be aware. The computer watches its own performance, and if the CPS of the computer itself can modify the system, either hardwares or softwares to improve the function, it is an aware computer. In other words, there can be no aware computer if there is no evaluation and action by the computer itself.

6.1. Self-Improving Computer. Let us imagine an "aware computer" which is designed to compute various optimization problems, and it is equipped with an evaluation counter whose number changes by the performance of the computer for constantly changing request. It is designed to change the system somewhat randomly, when the index of the evaluation counter goes down, and when the number of the counter becomes below a critical number, the power line is cutoff. Among an ensemble of computers which has a software and a hardware different from each other, a small number of computers with good performance will survive. In this case, if the computer can search and change the software by itself, constantly monitoring the number of the counter, it will be one of the aware computers.

6.2. Self-Sequencing Computer. Let us imagine a computer which performs a sequence of many programs by choosing out of many other sequences. If it has a memory which records the process and speed of computing, and if the computer from time to time stops computing and checks the previous process of computing, using the memory system, and if it can change the sequence of the job to achieve a higher global performance, then this computer might be called one of the aware computers.

7. An Aware Robot Which Behaves as If It Had a Free Will

Awareness implies existence of a CPS which can identify macroscopically the type of the present function with one stored. When an aware computer fails to identify the state of the machine, the computing system is not aware of what is going on.

It seems reasonable to define the aware computing as such having a CPS like ourselves. Some examples of sentiment we feel such as pleasantness, painfulness, happiness, and sadness are the awareness and identification of the present function of the system with one of the patterns we experienced in the past. When we say that the human being is an aware computer, how do we explain free will?

Even if we construct a very good aware robot which will monitor perfectly the macroscopic functions of the computer itself, it will not have free will. To implement free will, I propose here, based on the results of our experiments discussed in the Section 3, an aware robot with double time; one is the physical time and the other is a brain time. The robot is assumed to operate by the physical clock when the computer is functioning with external signal as a reference signal, and to operate with a brain clock which is a little faster than the physical clock when the computer is simulating the external world without the signals from the external world.

Such aware robot will find that his internal world is leading the real world when he compares both from time to time. He would feel that he is not following the change of the external world, but the world is following him. There are two conditions for this mechanism to work. One is that the dynamics of the world is simple enough that the computer can simulate it by learning. The second condition is that the brain clock moves faster than the physical clock when the external input is off. The experimental evidence [2–4] showed that this is really the case, when human is asked to track an object moving on a simple orbit.

8. Epilogue

Terms and concepts which the author discussed in the paper are

awareness,

external world,

internal world,

awareness of external world,

awareness of internal world,

mind,

free will,

self-closed system.

As the author mentioned in the Introduction, those words are terminologies in the human science domain, not in the natural science domain. The author tried in this paper to discuss them by the words of natural science, not to translate nor define them into natural science with the scientific precision, nor to create new terms which have strict definitions [5]. The terms in the human science have wide and various aspects because of the length of history it was used, compared to the terms used in the modern natural science. I believe it is more useful at present to discuss the

aspects of the concept, not trying to define it in modern science terminology.

In the future, one will be able to make an aware robot which behaves as if it is implemented all the functions that a human has, and which stops functioning by cutting off the power when some number reaches to a critical value which is a measure of competition between the other robots. However, there remains a fundamental difference between the aware human and an aware robot. The former has a deep desire to keep living, while not for the latter.

Thus, we came finally to face a most fundamental question why we desire to live, and a question if we can implement the desire into an aware robot. At the present moment, we do not know why we desire to live. The only thing we know is these species which survived through a severe natural selection survived by obtaining the instinctive wish to live. It would be wonderful if we can understand this question either from genetic information or from brain structure. But, I am afraid that all these effort will fail.

Creators, by definition, whether it is a creator of a computer or the creator of animals design their products to survive as much as possible against their enemies. How have the creators of animals implemented the desire to live? I believe we can find some suggestions by translating this question into the scientific terms. To do so, we must look for the causal relation between the key word such as awareness, desire to live, natural selection, mind, and survival.

A following scenario seems most natural to the author.

(1) Animals have obtained awareness through the severe survival race through the evolution. Humans are most aware, hopefully.

(2) As a result, they obtained their internal world and mind.

(3) Then, they came to "think" that they have desire to live, because the mind they obtained was the kind which desire to live through the selection, although we do not know the physical mechanism of the desire yet.

(4) Important thing is to notice that it is not that we are aware because we have desire to live, but we desire to live because we are aware.

References

[1] F. Ishida and Y.E. Sawada, "Human hand moves proactively to the external stimulus; an evolutional strategy for minimizing transient error," *Physical Review Letters*, vol. 93, no. 16, Article ID 168105, 2004.

[2] Y. Hayashi, Y. Tamura, K. Sase, K. Sugawara, and Y. Sawada, "Intermittently-visual tracking experiments reveal the roles of error-correction and predictive mechanisms in the human visual-motor control system," *Transactions of The Society of Instrument and Control Engineers*, vol. 46, no. 7, pp. 391–400, 2010.

[3] Y. Hayashi, Y. Tamura, K. Sugawara, and Y. Sawada, "Why the hand motion proceeds the target in tracking experiment?" in *Proceedings of the 3rd International Symposium on Mobiligence*, vol. 34, 2009.

[4] Y. Hayashi and Y. Sawada, "A transition from an alternative error-correction mode to a synchronization mode in the mutual hand tracking and the mutual finger tapping," *Physical Review E*. Submitted.

[5] H. R. Maturana and F. J. Varela, *Autopoiesis and Cognition*, D. Reidel, Dordrecht, The Netherlands, 1980.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 694836, 8 pages
doi:10.1155/2012/694836

Research Article

Interactive Evolutionary Computation for Analyzing Human Awareness Mechanisms

Hideyuki Takagi

Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

Correspondence should be addressed to Hideyuki Takagi, takagi@design.kyushu-u.ac.jp

Received 6 March 2012; Accepted 3 May 2012

Academic Editor: Cheng-Hsiung Hsieh

We discuss the importance of establishing awareness science and show the idea of using interactive evolutionary computation (IEC) as a tool for analyzing awareness mechanism and making awareness models. First, we describe the importance of human factors in computational intelligence and that IEC is one of approaches for the so-called humanized computational intelligence. Second, we show examples that IEC is used as an analysis tool for human science. As analyzing human awareness mechanism is in this kind of analyzing human characteristics and capabilities, IEC may be able to be used for this purpose. Based on this expectation, we express one idea for analyzing the awareness mechanism. This idea is to make an equivalent model of an IEC user using a learning model and find latent variables that connect inputs and outputs of the user model and that help to understand or explain the inputs-outputs relationship. Although there must be several definitions of awareness, this idea is based on one definition that awareness is to find out unknown variables that helps our understanding. If we establish a method for finding the latent variables automatically, we can realize an awareness model in computer.

1. Introduction

The number of papers using the keywords context awareness, power awareness, location awareness, and situation awareness in the SciVerse Scopus database of Elsevier is, respectively, 6,383, 1,749, 1,688, and 257 as of February 2012. Engineering interest in these areas has increased.

These engineering approaches call obtaining unknown knowledge or facts *awareness*. However, how do these engineering approaches differ from data mining or knowledge acquisition? Although these kind of applications are useful and important, we need other scientific approaches not only to support the engineering applications of awareness but also to extend awareness science and engineering.

One such scientific approach would be analyzing the awareness mechanisms of human beings and/or animals and constructing awareness models based on these mechanisms. Once we establish their core technologies, we may be able to make a computer with using such a model *be aware of* something. As the result, we can expect not only to develop data mining-like applications as has been done until now but also to progress human-machine communications, the monitoring of social networks, and new areas.

Analyzing awareness mechanisms and modeling them are the first important step to research performed in this direction. It is important to integrate ideas from and cooperate with those in ethology, psychology, mathematical modeling, engineering analytical methods, and other interdisciplinary areas. Cooperation with human sciences is especially, important though cooperation between awareness computing and human science has not been so active.

This paper has two objectives; one is to show that IEC can be an analytical tool for human sciences and the other is to discuss how we should use IEC for the analysis of awareness mechanisms and awareness modeling. For the first objective, we introduce some research using IEC as a tool for analyzing humans in Section 3 and show IEC's potential as a tool for awareness science. For the second objective, we show an application of idea in a tentative trial for further discussions in Section 4, though concrete approaches have not been proposed yet. As IEC is a core technology of this paper, we also draw a big picture to explain why IEC is necessary

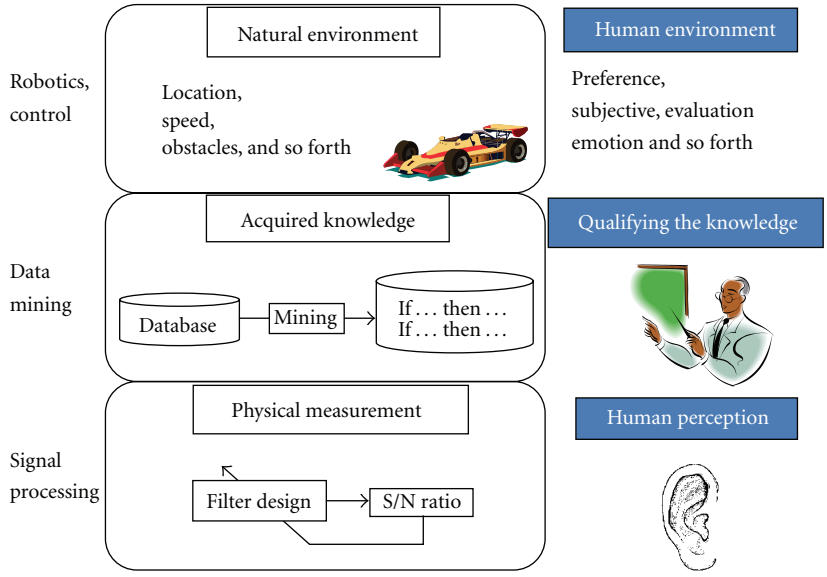

Figure 1: Human factors have become important in conventional engineering.

and introduce IEC briefly in Section 2 before these main sections.

2. Humanized Computational Intelligence and IEC

In this section, we would like to emphasize how exploiting human factors has become increasingly important in computational intelligence. We will refer to the integration of human factors into computational intelligence as *Humanized Computational Intelligence*. An engineering solution is designed to optimize some task, and conventionally the engineers would create the design based on their knowledge, experience, and even preferences. This design approach shares some similarity with the approaches taken by artists or craftspeople. We may call this kind of design, based on the engineer's capabilities, a first-generation engineering design. In the second generation of engineering designs, the human engineer is replaced with optimization techniques or computational intelligence algorithms using autodesign methods. To clarify, computer-aided design (CAD) is not an autodesign method but rather a tool used by engineers to input their designs into a computer; designs using CAD belong to the first-generation. In the third generation of engineering designs, there exists cooperation between human engineers and autodesign methods. If perfect autodesign methods existed, the second generation would be the optimal. Unfortunately, no such methods exist. Since autodesign methods and human experts have different strong points, the best choice is to combine them and blend their strengths. Figure 1 illustrates some examples of this third generation of design. In conventional robotics and control applications targeting the natural environment, designs get information from sensors and make decisions regarding actions. Recently, consumer robotics have achieved considerable success and emotional reactions such as *cute*, *friendly*, or *safety* have become the most important factors in determining the sales for this kind of product. Conventional approaches cannot design for such emotional reactions without a human's subjective evaluations during the design process. Data mining acquires knowledge from a database, and practical techniques in this area have been established. However, a computer cannot evaluate how the obtained knowledge is important and therefore cannot obtain qualified knowledge without the evaluations of domain experts. Filters for sound or image can be implemented by mathematical signal processing process algorithms on the computer. Although specifications, such as signal-to-noise ratio or error between processed signals and target signals, can be used to design filters for a given target and optimization techniques can be used during the design, really, the best processed signals are frequently decided based on human vision or auditory inspection rather than the numerical specifications alone.

These examples show how human factors have become important in engineering design and how computational intelligence would benefit from the embedding of these human factors. That is, we need *Humanized Computational Intelligence*. Historically speaking, engineering approaches in artificial intelligence have handled only human logic or knowledge, while human beings have two aspects: the logic or knowledge aspect handling reasoning, knowledge expression, knowledge acquisition, associative learning, associative memory, and so on and subjective aspect handling intuition, preference, subjective evaluation, perception, cognition, and so on. The *Humanized Computational Intelligence* should deal with the latter.

Interactive Evolutionary Computation (IEC) is a tool well suited for realizing *Humanized Computational Intelligence*. IEC is an optimization based on an IEC user's subjective evaluations instead of a fitness function or measured fitness (see Figure 2). There are many tasks that are difficult

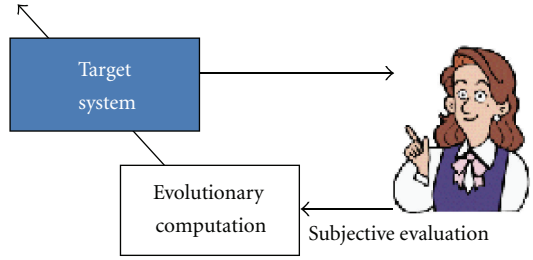

Figure 2: Human factors became important in conventional engineering.

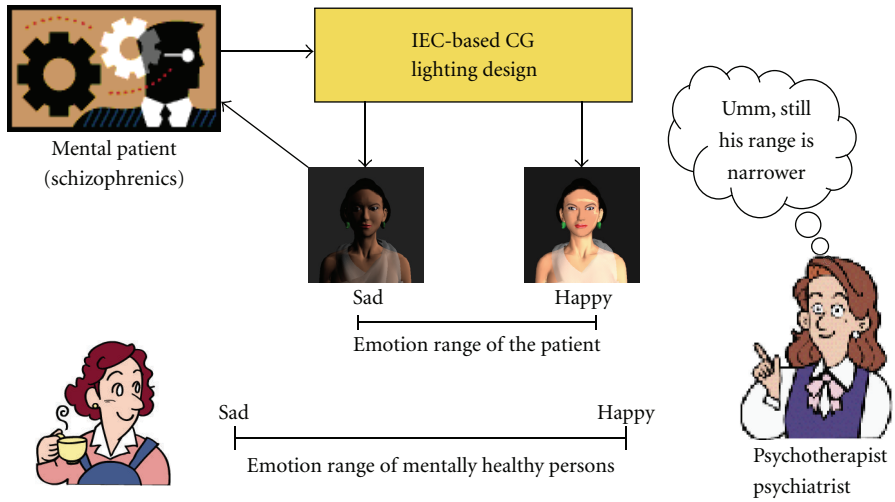

Figure 3: Mental-state measurement with IEC-based design support system. A subject designs impressions, such as *happy* and *sad*, using an IEC-based design support system, and mental experts compare his/her impression range of *happy-sad* with that of mentally healthy subjects.

or impossible to evaluate quantitatively; for example, there is no way to measure how sounds from a hearing aid are good to its user, but we must optimize its fitting even if we have to fit it based on a trial-and-error strategy. IEC can solve these kinds of optimization tasks.

Since the first IEC application by Dawkins in 1986 [1], many IEC papers have been presented. IEC research is roughly categorized into two: finding new IEC applications and reducing IEC user fatigue. IEC applications can be roughly categorized into three areas: artistic applications, engineering applications, and others. These applications include artistic applications such as graphics, music, industrial design, and facial design, engineering applications such as acoustics, image signal processing, data mining, robotics, control, and other applications such as geology, education and games. For more details, see the IEC tutorial and big IEC survey in [2].

Recently, new types of IEC applications have been proposed though most IEC applications are still optimizations of target systems. A new approach in IEC research is to use IEC as a tool for analyzing humans. As mentioned, evolutionary computation optimizes a target system based on human evaluations in an IEC system. We may understand the human's evaluation metrics or mechanisms by analyzing the target system optimized by the human evaluation. This approach has a similarity to reverse engineering which estimates inputs from outputs. We describe this approach in Section 3, and the approach is the background to the main idea presented in this paper described in Section 4.

3. IEC for Human Science

3.1. Measuring Emotional Dynamic Range of Human Mind. Some therapists have proposed, through their experiences, that the range of emotional expressions in the faces of schizophrenics is smaller than those of mentally healthy persons. There has been, however, no way to confirm this empirically. This was a motivation for measuring them indirectly using IEC. Although it is difficult to measure the dynamic emotion ranges from facial images, it is possible to measure patients' decision makings for emotions. Here, we can use IEC.

Schizophrenics patients and mentally healthy university students used IEC to create *happy* and *sad* lighting impressions based on their subjective degrees of the *happy* and *sad* impressions. We conducted a human subjective test and 33 subjects compared created lighting impressions by schizophrenics users and those by mentally healthy users, and we applied a statistical test to the human subjective test results (Figure 3). If there is no significant difference between two user groups, we may say that their dynamic ranges

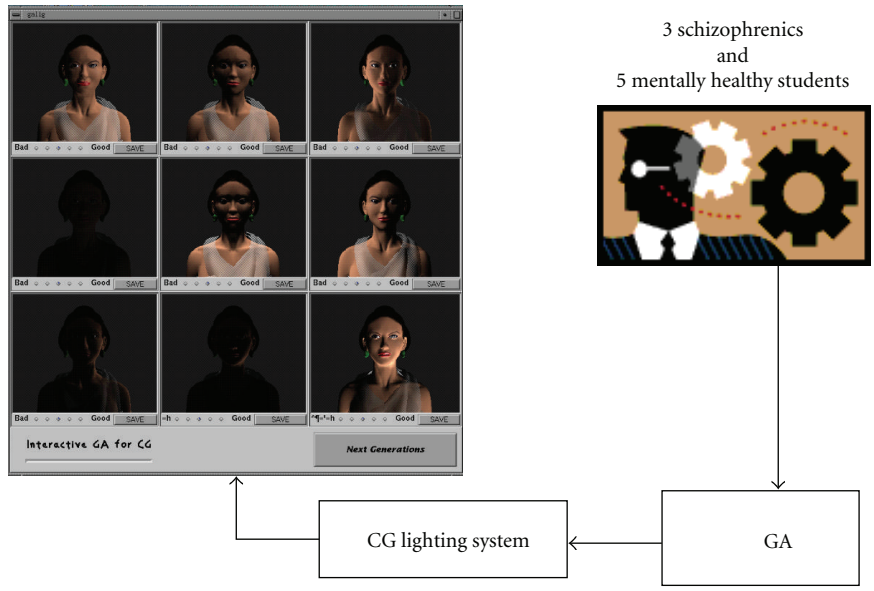

Figure 4: The graphical user interface of the IEC-based CG lighting design support system. Three schizophrenics and five mentally healthy subjects compared nine lighting designs and evaluated them according to a scale with five grades. A genetic algorithm optimized a CG lighting system based on their evaluations.

of *happy-sad* are almost the same. We measure emotional dynamic ranges from this relative comparison.

An IEC-based 3-dimensional computer graphics (3D CG) lighting design support system [3, 4] was used for this experiment. This system is used to design CG impressions matching to the given design concept by optimizing the coordinates of the lights in a 3D space, on/off of the lights, lighting strength, and types of light sources. Lighting colors are optimized, too, for color lighting design.

A genetic algorithm (GA) was used as an EC and generated nine 3D CG images. The IEC user gave fitness values in five levels for each image, and GA optimized the mentioned parameters based on the fitness. IEC users consist of three schizophrenics and five mentally healthy subjects. The experimental system is shown in Figure 4.

After obtaining 3 + 5 CG images per design concept, 33 subjects used a five-level rating scale to compare 28 ($= {}_8C_2$) pairs of each of *happy* and *sad* impressions using Scheffé's method of paired comparison that is a statistical method, analysis of variance (ANOVA), from paired comparison data.

The obtained psychological scales are shown in Figure 5. Psychological yardsticks for a *happy* constructed scale are 0.19 and 0.16 for ($P < 0.01$) and ($P < 0.05$), respectively, and those for a *sad* constructed scale are 0.24 and 0.21 for ($P < 0.01$) and ($P < 0.05$), respectively. These experimental results show statistical significances as PT, NH, PT < PM < NY < NK, and NN < NS on a *happy* constructed scale and NY, PK, NH < NK, NS < NN, PM, and PT on a *sad* constructed scale.

This approach illustrated IEC's applicability through its concrete realization in an experiment to measure the dynamic range of emotional expression capability in schizophrenics and mentally healthy subjects. There has been no technique for measuring this kind of mental dynamic range so far, but

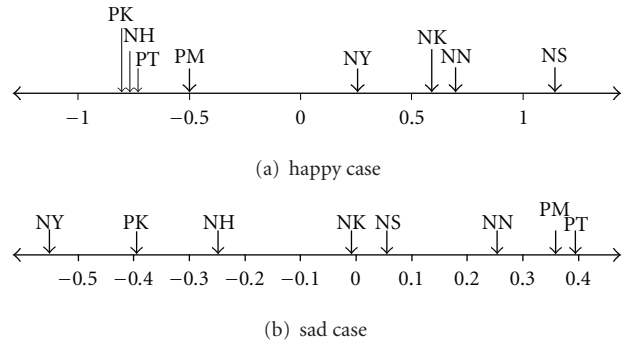

Figure 5: Psychological scales constructed using Scheffé's method of paired comparison and impression levels of the eight best lightings designed by the eight subjects. PM, PT, and PK are three schizophrenics, and NS, NN, NK, NY, and NH are five mentally healthy subjects.

IEC has the potential to be the technique. This point is the foundation for our expectation that IEC can be used as a tool for analyzing human awareness mechanisms.

IEC also has a potential for psychiatry. So far, schizophrenics' symptoms and daily-life functions have been diagnosed and measured using PANSS [5], LASMI, and other check lists of patient actions. These experimental results demonstrate the potential of IEC to be a new diagnostic tool in addition to or instead of these check lists.

3.2. Finding Unknown Auditory Facts. It is impossible to measure how users of hearing aids or cochlear implants really hear sounds, but they can report whether the sound is good or bad for them. This presents a typically adequate scenario equivalent to IEC application, and IEC was applied

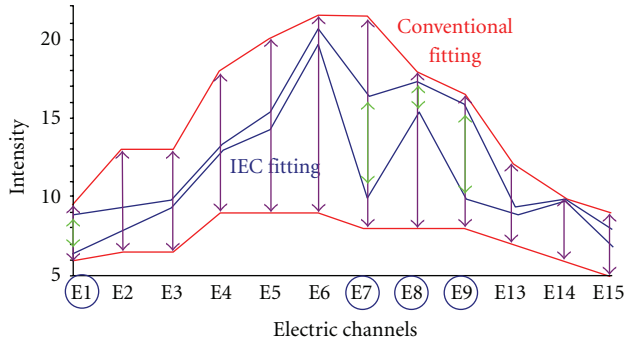

Figure 6: Cochlear implants fitting characteristics: those by conventional manual fitting (dot lines) and those by IEC fitting (solid lines). This figure was made based on Figure 5 in [7].

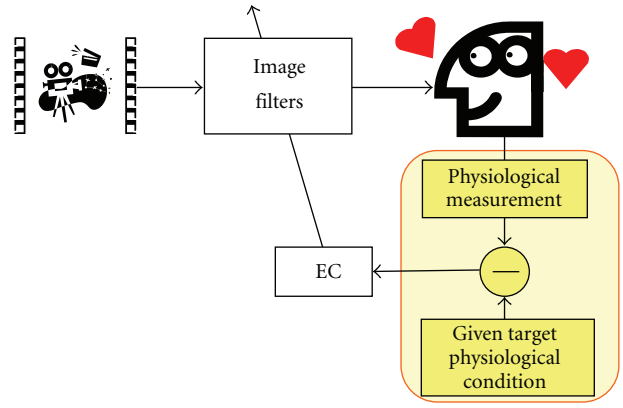

Figure 7: Example application of extended IEC for explaining its general framework. Filter evolves to bring the measured physiological data of an IEC user to the ideal condition of the targeted physiological condition.

to hearing aid fitting in [6] and cochlear implant fitting [7]. Since the IEC-based fitting is quite different from conventional fitting approaches, we were able to obtain new audio-psychophysiological facts from these new approaches.

There are two hypotheses for conventional cochlear implants fitting. One is "the more electric channels are, the better" to increase frequency resolution. Electric channels are set along the basilar membrane, which is equivalent to be set along the frequency axis. This is why the interval between neighbor electric channels corresponds to the minimum frequency difference that an audio-nerve system can distinguishes. Another is "the bigger dynamic range of electric stimulus, the better" to hear sounds from the minimum intensity level (threshold value: T-value) to the maximum one (comfort value: C-value). These two hypotheses look natural, and all existing current cochlear implants fittings are based on them.

A French team applied IEC to cochlear implant fitting and obtained quite strange results in [7], as shown in Figure 6. The dotted lines show C-values and T-values for the 15 electric channels obtained by a conventional fitting method, and the C–T ranges are set to be maximized. However, the fitting characteristics obtained by IEC-based fitting method are quite different from those of a conventional method as shown by solid lines.

Only 3 or 4 among 15 electric channels work and their dynamic ranges are quite a bit narrower than their C–T ranges. In spite of such poor fitting characteristics from the point of view of a conventional fitting, the IEC-based fitting characteristics showed higher word recognition rate; whereas a rate of around 50% was obtained with conventional fitting, more than 90% was achieved with the IEC fitting. This was a surprising result. The two hypotheses for cochlear implant fitting cannot explain these experimental results.

This fact implies that there must be unknown audio-psychophysiological facts. As it was found by IEC, IEC may be able to find other unknown facts in several areas of human science as an analysis tool. This is why we can expect to use IEC as a tool for analyzing awareness mechanisms.

We applied IEC to hearing aid fitting and showed that the IEC-based fitting realized almost equivalent performance to the fitting done by a fitter of the top level [6]. After this engineering approach, we analyzed the characteristics of the fitted hearing aids in different sound environments.

Some facts that we found thanks to the IEC approach are: (1) loudness characteristics obtained by conventional fitting method using band noises or pure tones differed from those obtained by IEC fitting using voices (2) differences among fitting characteristics obtained by IEC fitting using several different speeches with/without noise are small, while those are quite different from fitting characteristics obtained by IEC fitting using music with/without noise. These facts were not known until an IEC fitting method was applied.

3.3. IEC with Physiological Responses. IEC is an optimization system wherein an IEC user evaluates the outputs from a target system based on a priori knowledge, experiences, and/or preferences and makes the EC optimize the target system. Since the user evaluation is a subjective evaluation, we may say that IEC is a system applying psychological feedback to the EC.

Extended IEC is another IEC framework that extends the feedback from psychological data of an IEC user to his/her physiological data [8]. Let's explain it using the example in Figure 7. Let the goal or ideal physiological responses of relaxation or excitement, for example, be supervised data. The physiological responses of an IEC user who is watching movies are measured, and the EC optimizes the coefficients of an image filter to minimize the differences between the supervised data and the measured physiological responses. Thanks to this framework, we can extend IEC to handle human physiological data.

Extended IEC is an optimization framework rather than an analytical tool mentioned in Sections 3.1 and 3.2. However, note that we realized the IEC's potential as a tool for analyzing human awareness mechanism by changing our view from the IEC as an optimization tool to the IEC having psychological feedback as an analytic tool for human science. In the same way, we would like to expect that IEC becomes a tool for awareness science by extending it to the Extended IEC.

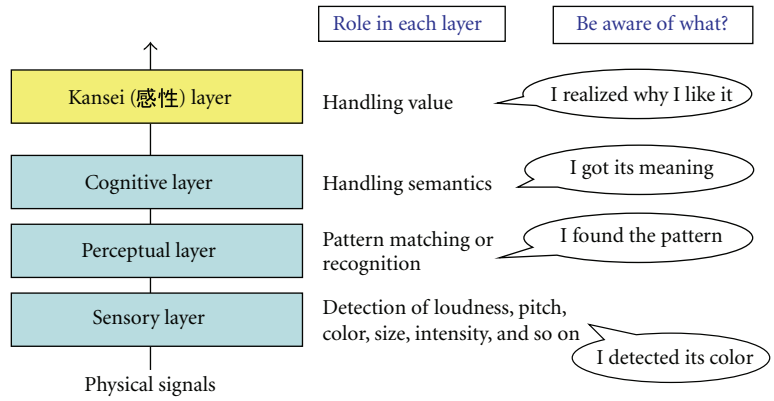

Figure 8: Awareness in different psychological layers.

4. IEC Approaches for Modeling Human Awareness Mechanism

4.1. Analysis of Human Awareness Mechanism. Psychological awareness is composed of three layers: the sensory layer, the perceptual layer, and the cognitive layer. In the case of sound, sound loudness or frequency is handled at the sensory layer; melody or word voice is distinguished and recognized at the perceptual layer; meaning of recognized sounds is handled at the cognitive layer. Accordingly, we can open a road for an ambulance when we hear its emergent siren.

Besides these three psychological layers, humans have a *KANSEI* layer that handles values of input stimuli, which are not handled by psychology. Although there are several definitions of this concept in the areas of information science, psychology, linguistics, design, and others, we define it as the capability or function of handling subjective values of inputs to a human being in this paper. For example, "I like this relaxed melody" is a process resulting from this KANSEI layer. There are many IEC applications using this layer, and we take account of the function in the layer in our analysis of the awareness mechanism.

We have different levels of awareness in different layers, as shown in Figure 8. As an example with the KANSEI layer, suppose you come to be attracted by a pot. The preferable feeling comes first, but it may be hard for you to immediately explain its reason. After a while, you may become *aware* of the reason, for example, the balance of color and its shape or its similarity to a toy that you played well with in your childhood, and thus you become able to explain the reason for your preference.

We should consider the mentioned psychological layers in our analysis of the awareness mechanism and the construction of awareness models.

One of the methods we can use to analyze the awareness mechanism would be to construct an input-output relationship for humans. Continuing with the above-mentioned example wherein a pot is evaluated, anyone can answer how they like the pot (z), when they see it. However, it is quite difficult to explain the relationship between a visual image of the pot (input x) and the evaluation (output z). After thinking for a while, they may become *aware* that the vertical-horizontal ratio of the pot (r) and the curvature of design pattern (c) are the points underpinning their evaluation and become able to explain the reason for their preference.

In other words, we may say that they *are aware of* the two hidden variables of the vertical-horizontal ration of the pot and the curvature of design pattern. That is, we can say that "$f()$ in $z = f(x)$ was so complex that the subject could not explain the input-output relationship at first glance. After that, they found the latent variables r and c and could interpret them as $z = f(x) = g_1(r) + g_2(c)$ and thus explained their evaluation, z, using a simple relationship involving $g_1()$ and $g_2()$."

4.2. Modeling of the Human Awareness Mechanism. We describe how we are proposed to realize the awareness mechanism in computer. We have not reached the stage of reporting our experimental results yet, but are rather at the stage of discussing our idea.

Step 1. The first step is to make an evaluation model of an IEC user. The IEC user inputs graphics, movies, sounds, and others from a target system and outputs his or her evaluations of them. We can make an IEC user model using a model system, a machine learning algorithm optimizing the system, and his/her inputs and outputs as training data (see Figure 9). The learning algorithm optimizes the model system using the training data. The model system includes neural networks (NN), fuzzy or crisp rule-based systems, neurofuzzy systems, and others; the machine learning algorithm includes a learning function of NN, and evolutionary computation.

The obtained user model can be analyzed and used to create an awareness model. When the relationship between the inputs to and outputs from the obtained evaluation model of an IEC user is simple, a computer may be able to easily explain why the IEC user evaluates the outputs from a target system from the obtained rule-based systems or neurofuzzy systems. The obtained rules are themselves the explanations. However, when the input-output relationship

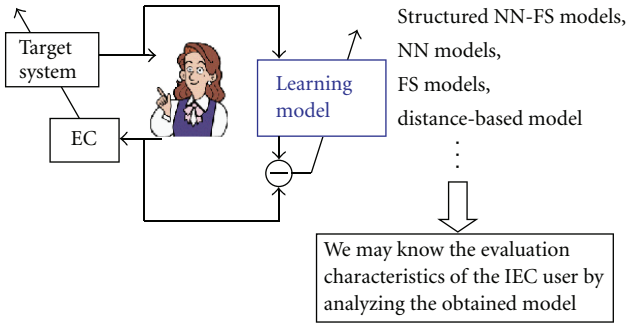

Figure 9: Making an evaluation model of an IEC user.

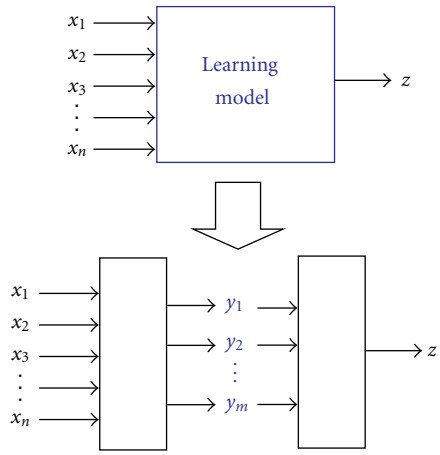

Figure 10: Finding out latent variables inside the model in the Step 1 that explains the output from the inputs.

is complex, we need to turn to a second step, which we will now describe.

Step 2. Resolving a complex system into several simpler systems is one of methods for decreasing complexity. We assume here that there are hidden variables or latent variables between the inputs and outputs as shown in Figure 10. Then, we can interpret that analyzing an awareness mechanism is to find the latent variables.

For example, when we imagine the type of students from their examination scores in mathematics, language, and physics, and from their heights and weights, maybe we can obtain an impression of the students with regards to their scholastic ability or their body builds.

Once a computer can extract the unknown latent variables between inputs to an IEC user, that is, outputs from a target system, and outputs from the IEC user, that is, his/her subjective evaluations, the computer may be able to come to explain why the IEC user evaluates the outputs from the target system with the subjective evaluations using the extracted latent variables. This is one example approach for constructing awareness mechanism.

The key point is how we can obtain such latent variables. Possible methods include (a) structure analysis of a structured NN-FS model, (b) introducing statistical methods for finding latent variables, (c) making a learning model by a math equation using genetic programming and analyzing the obtained equation, and (d) others.

When the relationship between inputs and outputs to/from an IEC user is complex, for example, it is strongly nonlinear, it must be difficult to find latent variables unlike the mentioned scholastic ability and body builds in the above. As the first stage of constructing awareness model, we should start the simple case using the discussed approach and then develop new method for finding latent variables.

5. Conclusions

We emphasize the applicability of IEC for awareness science, especially for analyzing the human mechanism of awareness, by showing its applicability for human science with some concrete approaches. At this time, we are at the start line in the process of using the proposed approach and must develop concrete routes towards the goal of analyzing the awareness mechanism and providing an awareness machine model for engineering applications. We hope that introducing the applicability of EC in this paper increases the interests of researchers in awareness computing into not only engineering aspects of the awareness computing but also its scientific aspects and helps them in their research.

Acknowledgments

This paper was made based on the author's keynote speech at the Second International Symposium on Aware Computing (ISAC 2010) held in Tainan, Taiwan. The author would like to thank Prof. Qiangfu Zhao of the University of Aizu and other committee members for giving the author the opportunity to present his views on approaches for awareness science at ISAC 2010. This work was supported in part by Grant-in-Aid for Scientific Research (23500279).

References

[1] R. Dawkins, *The Blind Watchmaker*, Longman, Essex, UK, 1986.
[2] H. Takagi, "Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation," *Proceedings of the IEEE*, vol. 89, no. 9, pp. 1275–1296, 2001.
[3] K. Aoki and H. Takagi, "3-D CG lighting with an interactive GA," in *Proceedings of the 1st International Conference on Conventional and Knowledge-based Intelligent Electronic Systems (KES '97)*, pp. 296–301, Adelaide, Australia, 197.
[4] K. Aoki and H. Takagi, "Interactive GA-based design support system for lighting design in 3D computer graphics," *Transactions of IEICE*, vol. 81, no. 7, pp. 1601–1608, 1.
[5] S. R. Kay, A. Fiszbein, and L. A. Opler, "The positive and negative syndrome scale (PANSS) for schizophrenia," *Schizophrenia Bulletin*, vol. 13, no. 2, pp. 261–276, 1987.
[6] H. Takagi and M. Ohsaki, "Interactive evolutionary computation-based hearing aid fitting," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 3, pp. 414–427, 2007.

[7] P. Legrand, C. Bourgeois-Republique, V. Péan et al., "Interactive evolution for cochlear implants fitting," *Genetic Programming and Evolvable Machines*, vol. 8, no. 4, pp. 319–354, 2007.

[8] H. Takagi, S. Wang, and S. Nakano, "Proposal for a framework for optimizing artificial environments based on physiological feedback," *Journal of Physiological Anthropology and Applied Human Science*, vol. 24, no. 1, pp. 77–80, 2005.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 525396, 6 pages
doi:10.1155/2012/525396

Research Article

Variance Entropy: A Method for Characterizing Perceptual Awareness of Visual Stimulus

Meng Hu and Hualou Liang

School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA 19104, USA

Correspondence should be addressed to Hualou Liang, hualou.liang@drexel.edu

Received 28 December 2011; Revised 22 March 2012; Accepted 23 March 2012

Academic Editor: Cheng-Hsiung Hsieh

Entropy, as a complexity measure, is a fundamental concept for time series analysis. Among many methods, sample entropy (SampEn) has emerged as a robust, powerful measure for quantifying complexity of time series due to its insensitivity to data length and its immunity to noise. Despite its popular use, SampEn is based on the standardized data where the variance is routinely discarded, which may nonetheless provide additional information for discriminant analysis. Here we designed a simple, yet efficient, complexity measure, namely variance entropy (VarEn), to integrate SampEn with variance to achieve effective discriminant analysis. We applied VarEn to analyze local field potential (LFP) collected from visual cortex of macaque monkey while performing a generalized flash suppression task, in which a visual stimulus was dissociated from perceptual experience, to study neural complexity of perceptual awareness. We evaluated the performance of VarEn in comparison with SampEn on LFP, at both single and multiple scales, in discriminating different perceptual conditions. Our results showed that perceptual visibility could be differentiated by VarEn, with significantly better discriminative performance than SampEn. Our findings demonstrate that VarEn is a sensitive measure of perceptual visibility, and thus can be used to probe perceptual awareness of a stimulus.

1. Introduction

Over the past decades, entropy [1] has been widely used for analysis of dynamic systems. Among many measures, sample entropy (SampEn) is thought of as an effective, robust method due to its insensitivity to data length and its immunity to noise [2]. Until now, SampEn has been successfully applied for discriminant analysis of cardiovascular data [3], electroencephalogram data [4], and many others [5]. In addition, SampEn has been used in multiscale analysis for computing entropy over multiple time scales inherent in time series. For example, multiscale entropy [6] and adaptive multiscale entropy (AME) [7] both use SampEn to estimate entropy over multiple scales of time series.

Despite its popularity, it is not well recognized that there is an inherit drawback of SampEn used for discriminant analysis, that is, the calculation of SampEn is routinely based on the normalized data where the variance of data that may provide additional information for discrimination is discarded [8]. The normalization is essentially to rescale the data, which is appropriate if the analysis is driven by the search for order in the dynamics, but is otherwise inappropriate for discriminant analysis of two data sets as the rescaling can make them appear identical when they clearly are not. In fact, the variance and SampEn represent different aspects of the data: the variance measures concentration only around the mean of the data, whereas the entropy measures diffuseness of the density irrespective of the location of concentration of the data. In this paper, we proposed a new complexity measure, variance entropy (VarEn), to take into account both SampEn and the variance for improved discriminant analysis. Not only can it be used as a single-scale measure, but it can also be adapted for studying nonstationary data over multiple time scales. We applied VarEn to analyze cortical local field potential (LFP) data collected from a macaque monkey while performing a generalized flash suppression task [9], in which physical stimulation is dissociated from perceptual experience, to probe perceptual awareness of a visual stimulus. We showed that VarEn performed better than SampEn for both the

whole time series (single-scale) and multiscale analysis of LFP data in terms of discriminative ability in distinguishing different perceptual conditions (Visible versus Invisible). Our results suggest that the proposed VarEn measure is a useful technique for discriminant analysis of neural data and can be used to uncover perceptual awareness of a stimulus.

2. Method

2.1. Sample Entropy. Entropy describes the complexity or irregularity of system, which can be used to classify systems. So far, many attempts have been made for estimation of entropy, such as Kolmogorov entropy [10] and Eckmann-Ruelle entropy [11]. However, these methods usually require very long time series that is not always available. Approximate entropy can be efficiently computed for short and noisy time series [1], but introduces a bias via counting self-match when calculating the pairs of similar epochs. Sample entropy (SampEn) provides a refined version of approximate entropy to reduce the bias [2]. It is defined as the negative natural logarithm of conditional probability that two sequences similar for m points remain similar at the next $m + 1$ point in the data set within a tolerance r, where self-matches are excluded in calculating the probability.

In order to compute SampEn, a time series $I = \{i(1), i(2), \ldots, i(N)\}$ is embedded in a delayed m-dimensional space, where the m-dimensional vectors are constructed as $x_m(k) = (i(k), i(k+1), \ldots, i(k+m-1))$, $k = 1 \sim N-m+1$. The match of two vectors in the embedded space is defined as their distance lower than the tolerance r. $B^m(r)$ is the probability that two sequences of m points match within r, whereas $A^m(r)$ is similarly defined for an embedded dimension of $m + 1$. The SampEn is then calculated as

$$\mathrm{SampEn}(I, m, r) = -\ln\left(\frac{A^m(r)}{B^m(r)}\right). \quad (1)$$

In practice, it is common to set the tolerance r as a fraction of the standard deviation of the data, which effectively rescales the data to have similar dynamic scales. As a result, the normalization process obscures the difference in the scales of data sets, thus rendering the analysis inappropriate if the goal is chiefly to discriminate between data sets. We therefore introduce a variance entropy measure in the following section to rectify this shortcoming.

2.2. Variance Entropy. Variance entropy (VarEn) measure is designed for discriminant analysis by combining SampEn and the variance of data. Specifically, VarEn can be treated as inverse-variance weighted entropy to represent system complexity. For a time series x, VarEn is defined as

$$\mathrm{VarEn}(x, m, r) = \frac{\sum_{i=1}^{p} \mathrm{SampEn}(x_i, m, r) \times w_i}{\sum_{i=1}^{p} w_i}, \quad (2)$$

where x_i is the ith segment of x, obtained with a window that slides over time, p is the number of sliding windows, w_i is inverse variance of x_i, m and r are the parameters for calculation of $\mathrm{SampEn}(x_i, m, r)$. Specifically, to calculate VarEn, we first apply a sliding window over time series x. For a given window (e.g., ith time window, x_i), we compute its SampEn, variance, and w_i. After computing all p time windows, we then apply above formula to obtain VarEn of this time series x. A schematic representation of the processing steps is shown in Figure 1. It is also straightforward to extend our VarEn measure to the analysis of multitrial neural data in which x_i is the ith trial of x, and p is the number of trials. In this study, the parameters m and r were chosen as 2 and 0.2 for minimizing the standard error of entropy estimation [3]. We note that there is no pronounced difference with automatic selection of r [12].

Similar to SampEn, VarEn can also be applied to nonstationary data by considering multiple time scales inherent in the data, on which the entropy can be calculated. By multiple-scale entropy analysis, we compute the VarEn in adaptive multiscale entropy (AME) [7] to demonstrate its performance in comparison with the use of SampEn. AME is a multiscale analysis method in which the scales are adaptively derived directly from the data by virtue of multivariate empirical mode decomposition [13], which is fully data driven, well suited for the analysis of nonlinear/nonstationary neural data. Depending on the consecutive removal of low-frequency or high-frequency components, AME can be estimated at either coarse-to-fine (preserving high-frequency oscillations by progressively removal of low-frequency components) or fine-to-coarse scales (preserving low-frequency oscillations by progressively removal of high-frequency components) over which the sample entropy is performed. The coarse-to-fine and fine-to-coarse AME can be used separately or used in tandem to reveal the underlying dynamics of complex neural data. In this study, we use the VarEn to replace SampEn to perform multiscale analysis.

3. Results

In this section, we apply VarEn to analyze local field potentials (LFPs) collected from visual cortex of a macaque monkey while performing a visual illusion task, to characterize neural dynamics of perceptual awareness of a visual stimulus.

The visual illusion task used here is called generalized flash suppression (GFS) task, where a salient visual stimulus can be rendered invisible despite continuous retinal input, thus providing a rare opportunity to study neural mechanisms directly related to perception [9]. In the task, as soon as a monkey gained fixation for about 300 msec, the target stimulus indicated by a red disk was presented. At 1400 msec after the target onset, small random-moving dots appeared as the surroundings. With the immediate presence of the surroundings, the red disk could be rendered subjectively invisible. If the target stimulus disappeared from perception, the monkey was trained to release a lever; otherwise, monkey was to hold the lever. Therefore, based on the responses of the animal, the trial was classified as either "Visible" or "Invisible." Note that the stimuli in these two conditions were physically identical. Multielectrode LFP recordings were simultaneously collected from multiple cortical areas V1, V2, and V4 while monkeys performed the GFS task [14]. The data were obtained by band-pass filtering the full bandwidth

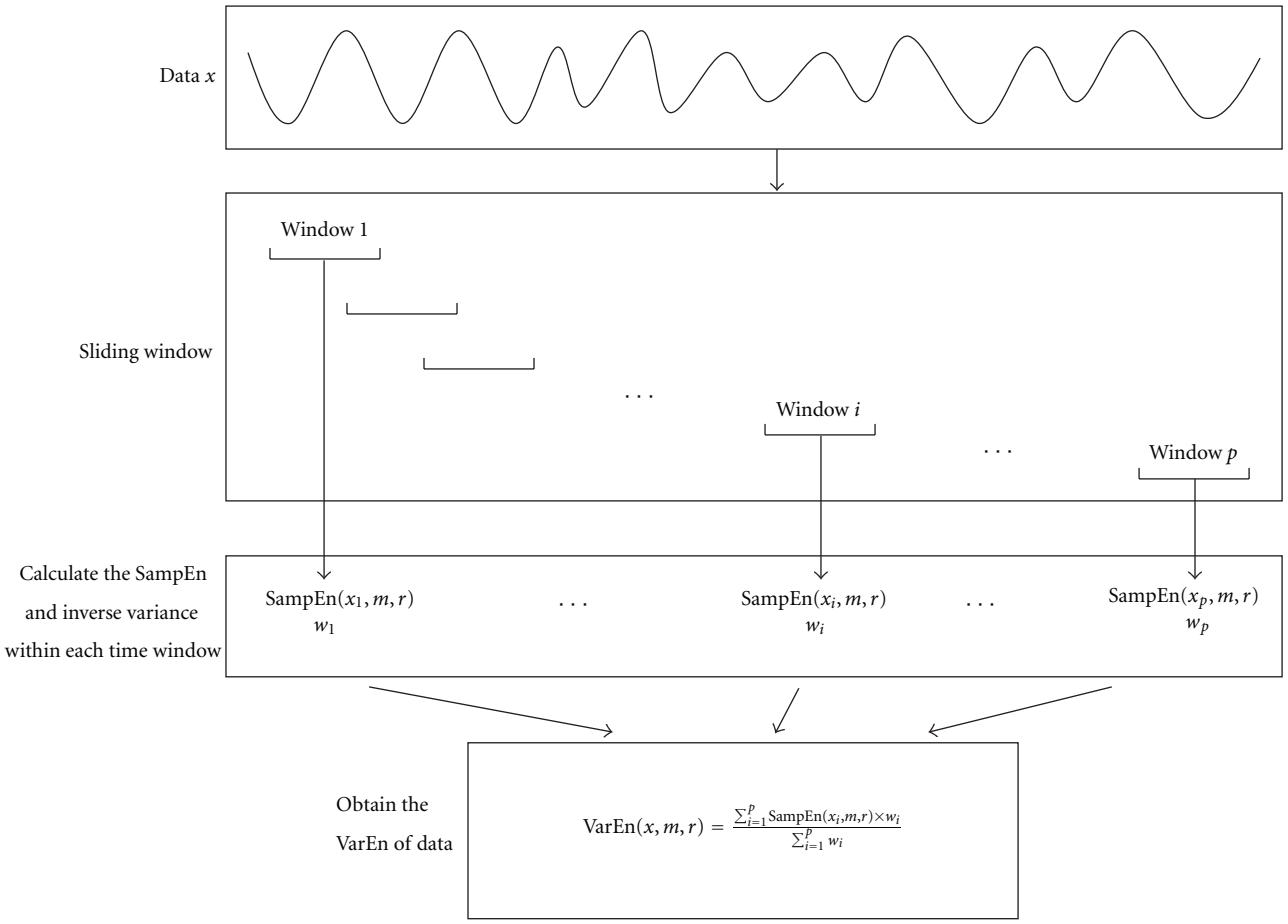

Figure 1: Schematic representation of the proposed variance entropy (VarEn).

signal between 1 and 500 Hz, and then resampled at 1 KHz. In this study, the LFPs of one second long after surrounding onset on a typical channel of area V4 over 87 trials were used in the analysis.

VarEn is first directly applied to the LFP data to discriminate two perceptual conditions: Visible versus Invisible. In Figure 2, we can see that the VarEn of the invisible condition is significantly greater than that of the visible condition ($P < 0.05$). As a comparison, SampEn is applied to the same LFP data. However, the result reveals that there is no significant difference between two perceptual conditions (Figure 2). This result suggests that VarEn carries more discriminative information by integrating SampEn and the variance of data.

Next, VarEn is applied to the LFP data for performing multiple-scale entropy analysis. As described in the Method, we herein apply the VarEn-based adaptive multiscale analysis (AME) to the LFP data for discriminating two perceptual conditions over different LFP scales. As a comparison, the original AME is also applied to the LFP data, where the SampEn is applied to calculate the AME entropy measure. Figure 3 shows the comparison of two methods at the coarse-to-fine scales. We can see from Figures 3(a) and 3(b) that both methods exhibit dominantly increasing trend

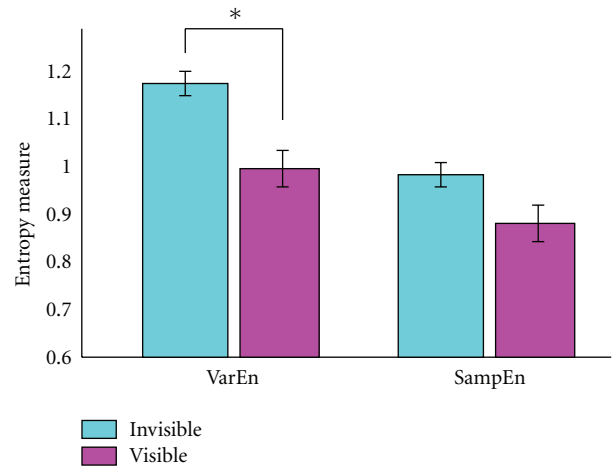

Figure 2: Comparison of the proposed variance entropy (VarEn) and sample entropy (SampEn) in discriminating the different perceptual conditions (Invisible versus Visible). Shown are the means and the standard errors of means of VarEn (left) and SampEn (right). As a result, the VarEn exhibits significant difference between two perceptual conditions ($p = 7.69e{-}4$, indicated by the sign "$*$"), but the SampEn does not ($P = 0.056$).

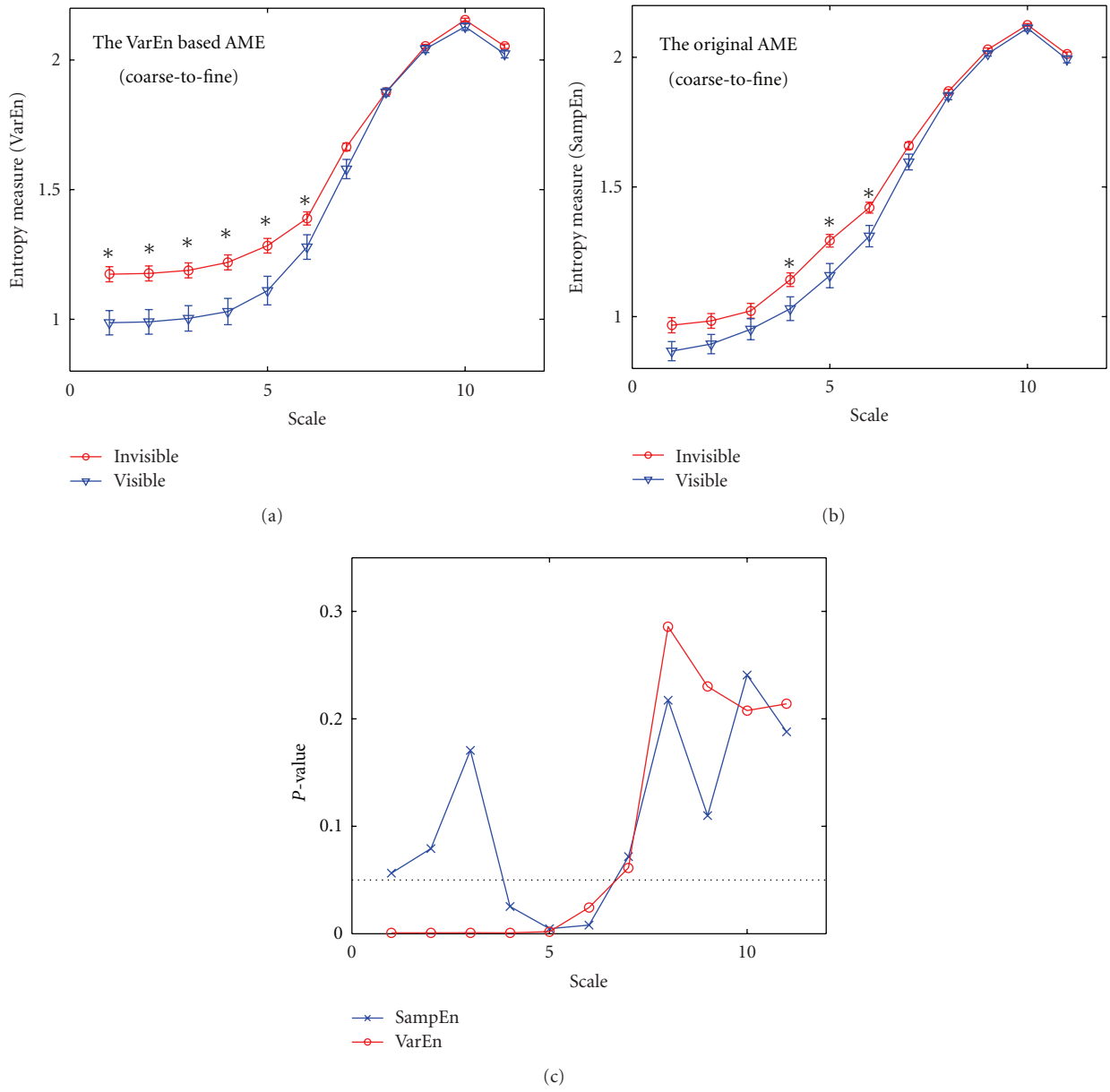

FIGURE 3: Coarse-to-fine comparison for the AME with VarEn (a) and the AME with SampEn (b) for discriminating the Invisible (red circle) and the Visible conditions (blue triangle) over multiple scales. The error bar refers to the standard error of mean. The leftmost red circle and blue triangle are the entropy measures of the raw data. For the AME with VarEn (a), significant differences occur at the 1st–6th scales, while the significant differences only occur at the 4th–6th scales for the AME with SampEn (b). The sign "$*$" refers to $P < 0.05$. The P values based on our VarEn and the SampEn along the coarse-to-fine scales are also compared (c), in which the horizontal black dotted line refers to the significance level ($P = 0.05$).

as the scale increases. However, the VarEn-based AME (Figure 3(a)) clearly provides significantly larger separation between two perceptual conditions than the original AME with SampEn (Figure 3(b)). Specifically, the VarEn-based AME at six significant scales (i.e., scale 1–6) shows significant differences between two perceptual conditions ($P < 0.05$) (Figure 3(a)), whereas the original AME differs only at three scales, that is, scales 4–6 (Figure 3(b)). A detailed comparison of P values between our VarEn and the SampEn along the coarse-to-fine scales is shown in Figure 3(c). These results indicate that VarEn is more sensitive than SampEn to detect perceptual difference between two conditions.

Similarly, the improvement of discrimination by VarEn occurs at the fine-to-coarse scales as well (Figure 4), in which AME with VarEn significant different at scales 1–3 (Figure 4(a)) while AME with SampEn only exhibits significant difference at scale 2 (Figure 4(b)). Detailed comparison of P values between our VarEn and the SampEn along the

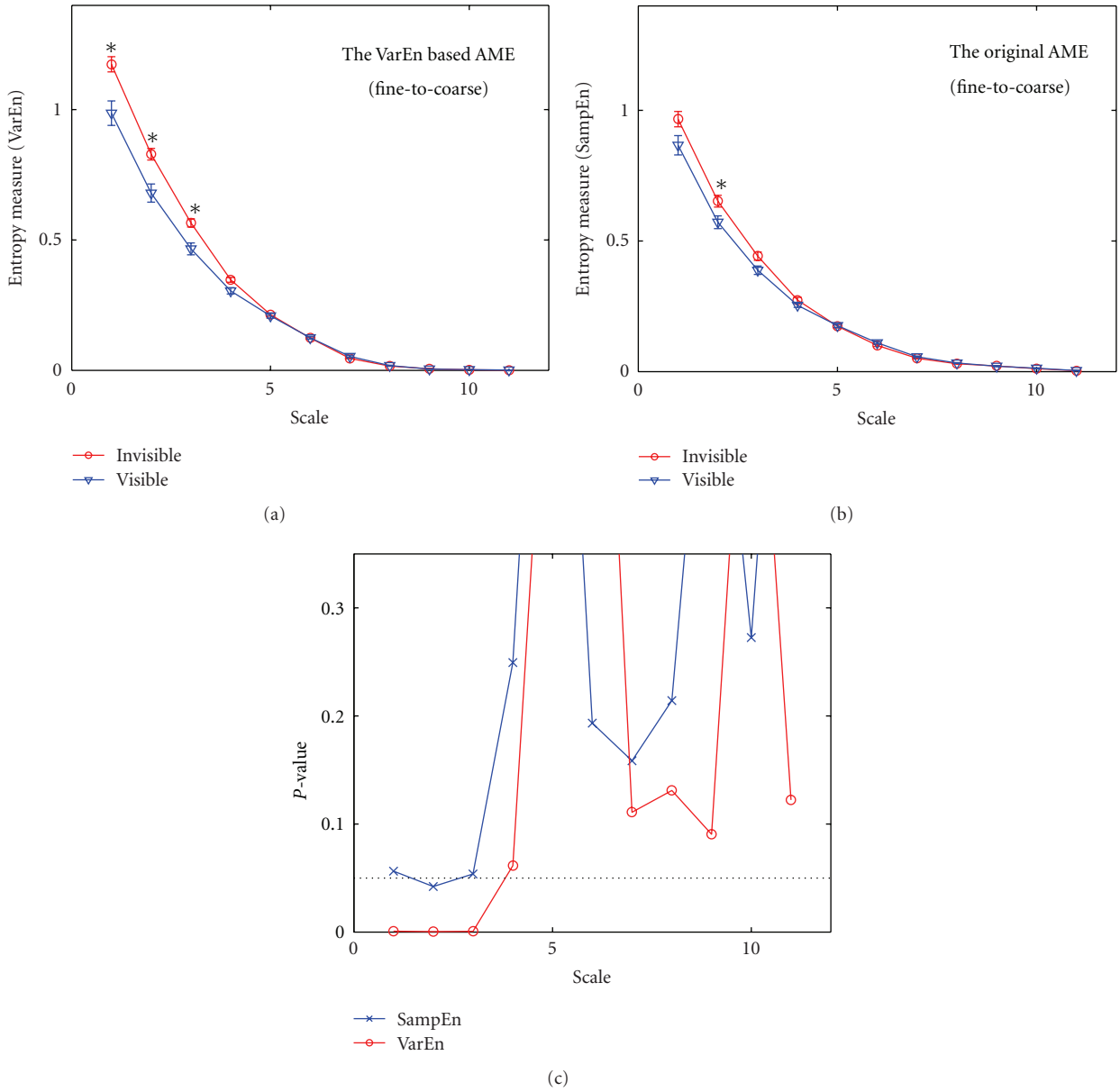

FIGURE 4: Fine-to-coarse comparison for the AME with VarEn (a) and the AME with SampEn (b) for discriminating the Invisible (red circle) and the Visible conditions (blue triangle) over multiple scales. For the AME with VarEn (a), significant differences occur at the 1st–3rd scales, whereas for the AME with SampEn (b) significant difference only occurs at the 2nd scale. Conventions are the same as in Figure 3. The P values based on our VarEn and the SampEn along the fine-to-coarse scales are also compared (c), in which the horizontal black dotted line refers to the significance level ($P = 0.05$). Note that the P values greater than 0.35 are truncated.

fine-to-coarse scales is also shown in Figure 4(c) to support our findings. These results, taken together, suggest that our VarEn can be used to obtain improved discriminative information for discriminant analysis of neural data.

In comparison of the coarse-to-fine AME (Figure 3) with the fine-to-coarse AME (Figure 4), the coarse-to-fine AME focuses on the high-frequency oscillations, whereas the fine-to-coarse AME emphasizes the low-frequency oscillations. When applied to the LFP data, we can see that the AME at coarse-to-fine scales exhibit more discriminative scales than those at the fine-to-coarse scales, indicating that the low-frequency scales may not contain as much discriminative information as the high-frequency scales. The low-frequency oscillations mainly correspond to evoked potentials, which are presumably the same as the stimuli in the GFS task are identical; this explains why the fine-to-coarse scales exhibit less discriminative ability for separating two perceptual conditions. Furthermore, among all the significant differences, the AME measure in the invisible condition is higher than that in the visible condition, which suggests that perceptual

suppression is likely to be related to more complex neural processes than the normal visible condition.

4. Discussion and Conclusion

In this paper, we proposed a simple complexity measure, variance entropy (VarEn), by combining SampEn and the variance to achieve improved discriminant analysis. Our measure was motivated by the observation that the calculation of SampEn is based on the normalized data where the variance is routinely discarded, which may otherwise provide the additional information for discrimination analysis. We applied VarEn to the analysis of cortical local field potential data collected from visual cortex of monkey performing a generalized flash suppression task. We showed that VarEn performed better than SampEn for both the whole time series (single-scale) and multiscale analysis of LFP data in terms of discriminative ability in distinguishing different perceptual conditions. The results suggest that our proposed VarEn measure is a useful measure for discriminant analysis of neural data and can be used to uncover perceptual awareness of a stimulus.

To quantify the complexity of a system, our proposed VarEn measure is defined as inverse-variance weighted SampEn. Inverse-variance weighting is typically used in statistical meta-analysis to combine several estimates of an unknown quantity to obtain an estimate of improved precision [15]. While other forms of weights (e.g., amplitude) could be used, such a choice of weight is optimal in providing the unbiased and minimum variance estimator. In comparison of VarEn to SampEn, the key difference is that SampEn discards the variance of the data, whereas VarEn combines the variance with SampEn via inverse-variance weighting. As such, if the main objective is for discriminant analysis, VarEn is preferred as it incorporates the variance information into its estimation. On the other hand, SampEn is appropriate if the analysis is driven by the search for order in the dynamics.

Acknowledgments

This work is partially supported by NIH. The authors thank Dr. Melanie Wilke for providing the data, which were collected at the laboratory of Dr. Nikos Logothetis at Max Planck Institute for Biological Cybernetics in Germany.

References

[1] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 6, pp. 2297–2301, 1991.

[2] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate and sample entropy," *American Journal of Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[3] D. E. Lake, J. S. Richman, M. Pamela Griffin, and J. Randall Moorman, "Sample entropy analysis of neonatal heart rate variability," *American Journal of Physiology*, vol. 283, no. 3, pp. R789–R797, 2002.

[4] E. N. Bruce, M. C. Bruce, and S. Vennelaganti, "Sample entropy tracks changes in electroencephalogram power spectrum with sleep state and aging," *Journal of Clinical Neurophysiology*, vol. 26, no. 4, pp. 257–266, 2009.

[5] S. Ramdani, B. Seigle, J. Lagarde, F. Bouchara, and P. L. Bernard, "On the use of sample entropy to analyze human postural sway data," *Medical Engineering and Physics*, vol. 31, no. 8, pp. 1023–1031, 2009.

[6] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, Article ID 068102, 4 pages, 2002.

[7] M. Hu and H. Liang, "Adaptive multiscale entropy analysis of multivariate neural data," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 12–15, 2012.

[8] J. S. Richman, D. E. Lake, and J. R. Moorman, "Sample entropy," *Methods in Enzymology*, vol. 384, pp. 172–184, 2004.

[9] M. Wilke, N. K. Logothetis, and D. A. Leopold, "Generalized flash suppression of salient visual targets," *Neuron*, vol. 39, no. 6, pp. 1043–1052, 2003.

[10] P. Grassberger and I. Procaccia, "Estimation of the Kolmogorov entropy from a chaotic signal," *Physical Review A*, vol. 28, no. 4, pp. 2591–2593, 1983.

[11] J. P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, no. 3, pp. 617–656, 1985.

[12] S. Lu, X. Chen, J. K. Kanters, I. C. Solomon, and K. H. Chon, "Automatic selection of the threshold value r for approximate entropy," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 8, pp. 1966–1972, 2008.

[13] N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proceedings of the Royal Society A*, vol. 466, no. 2117, pp. 1291–1302, 2010.

[14] M. Wilke, N. K. Logothetis, and D. A. Leopold, "Local field potential reflects perceptual suppression in monkey visual cortex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 46, pp. 17507–17512, 2006.

[15] L. V. Hedges and I. Olkin, *Statistical Methods for Meta-Analysis*, Academic Press, Orlando, Fla, USA, 1985.

Hindawi Publishing Corporation
Applied Computational Intelligence and Soft Computing
Volume 2012, Article ID 650818, 6 pages
doi:10.1155/2012/650818

Research Article

Environmental Sound Recognition Using Time-Frequency Intersection Patterns

Xuan Guo,[1] Yoshiyuki Toyoda,[1] Huankang Li,[2] Jie Huang,[1] Shuxue Ding,[1] and Yong Liu[1]

[1] *Graduate Department of Computer and Information Systems, Graduate School of Computer Science and Engineering, The University of Aizu, Aizu-Wakamatsu 965-8580, Japan*
[2] *Department of Computer Science and Engineering, Shanghai Jiaotong University, 200240 Shanghai, China*

Correspondence should be addressed to Jie Huang, j-huang@u-aizu.ac.jp

Received 13 January 2012; Accepted 27 February 2012

Academic Editor: Zhishun She

Environmental sound recognition is an important function of robots and intelligent computer systems. In this research, we use a multistage perceptron neural network system for environmental sound recognition. The input data is a combination of time-variance pattern of instantaneous powers and frequency-variance pattern with instantaneous spectrum at the power peak, referred to as a time-frequency intersection pattern. Spectra of many environmental sounds change more slowly than those of speech or voice, so the intersectional time-frequency pattern will preserve the major features of environmental sounds but with drastically reduced data requirements. Two experiments were conducted using an original database and an open database created by the RWCP project. The recognition rate for 20 kinds of environmental sounds was 92%. The recognition rate of the new method was about 12% higher than methods using only an instantaneous spectrum. The results are also comparable with HMM-based methods, although those methods need to treat the time variance of an input vector series with more complicated computations.

1. Introduction

Understanding environmental sounds is an essential function of human hearing. For example, people can recognize the beginning of a rain shower by the rain sound, be cautious when they hear footsteps coming from behind at night, and open the door to welcome visitors after the sound of the door-knocking. Environmental sound recognition is also important for intelligent robots and computer systems. An intelligent robot can be aware of the environments by the audition and use its hearing function to complement its vision [1].

In recent years, environmental sound recognition has received increasing attention, and we have seen some pioneering research in this field. An environmental sound database (RWCP-DB) has been created for research use [2]. The sounds in the database were recorded in an anechoic environment with durations of 250 to 500 ms. In total, there are 105 instances, with each instance including 100 samples. We reclassified this database into 12 types and 45 kinds as listed in Table 1. For many sounds, there are multiple instances with similar but different materials.

An environmental sound recognition method using the instantaneous spectrum at the power peak was proposed [3]. It was reported that the rate of recognition was about 80% for 20 instances of environmental sounds. In this research, the target sounds are limited to impact sounds that have a single power peak followed by exponential attenuation. The instantaneous spectrum $S_p(\omega_m)$ was calculated at the power peak, where ω_m $(m = 1, 2, \ldots, M)$ is the frequency. Since the input information was only based on the peak spectrum without time variance, it was not able to capture the environmental sounds and thus the recognition rate was low.

It is natural to consider using existing methods that have proven useful for speech recognition, for example, the hidden Markov Model (HMM) method and the time delay neural network (TDNN) method [4–6], since those methods deal with time variations of an input vector series. Miki and others achieved recognition rate of 95.4% using HMM method for 90 instances of RWCP-DB [5], and Sasou, and

Table 1: The RWCP environmental sound database.

Sound type	Kind of materials	Instances
Impact sound	Wood plates	12
	Metal cans, boxes, and so forth	10
	Plastic cases	3
	Glass cups, bottles, and so forth	8
	Bundle of paper	1
	Handclap/handclaps	4
Falling pieces	Grains	2
	Coin/coins	7
	Dice	3
Air jet	Small air pump	1
	Spray	1
	Firecracker	1
	Air bubbles	1
	Dryer	1
Friction sound	File	1
	Sand paper	2
	Saw	2
Musical instruments	Castanets	1
	Cymbals	1
	Drum	1
	Horn	1
	Kara	1
	Maracas	1
	Ring	1
	String	1
	Whistle	3
	Tambourine	1
Phone, buzzer	Buzzer	1
	Clock alarm	2
	Phone	4
	Toys	3
Open	Cap	2
Broken	Chopsticks	1
	Tearing paper	1
	Crumpling paper	1
Release	Clip	2
Shaking	Metal bell/bells	7
Rotation	Coffee mill	1
Others	Doorlock	1
	Leaf through a book	2
	Mech bell	1
	Padlock	1
	Punch	1
	Shaver	1
	Stapler	1

others reported the recognition rate for 59 instances of RWCP-DB using AR-HMM method was 83.0% [6].

The recognition rate of the HMM method was greater than that of the peak-spectrum method. Because the HMM method uses a time series of frequency-feature vectors $[S_n(\omega_m)]$ that includes the time-frequency variance of the signals, where ω_m $(m = 1, 2, \ldots, M)$ is the frequency and $S_n(\omega_m)$ indicates the spectrum (or cepstrum) for time frame n $(n = 1, 2, \ldots, N)$. However, HMM-based methods may not be the best choice for environmental sound recognition

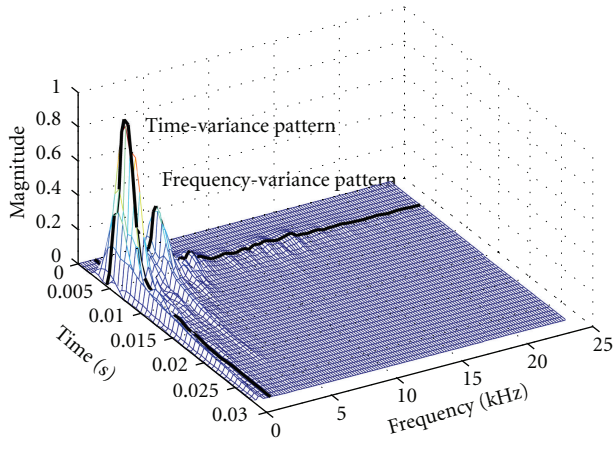

Figure 1: The time-frequency intersection pattern refers to the combination of the time-variance patter containing instantaneous powers (or their square roots) for all time frames and the frequency-variance patter with the instantaneous spectrum at power peak. (The time-variance pattern is illustrated as the line along with the spectrum-peaks.)

because environmental sounds differ from human speech. The frequency characteristics of most environmental sounds do not significantly change over time, and therefore it is not necessary to deal with state-transferring in many cases, as the HMM methods for speech signals require.

We can use a simpler method using the combination of a time-variance pattern containing the instantaneous powers (or their square roots) calculated by the sum-of-squares method for all time frames and a frequency-variance pattern with the instantaneous spectrum at the power peak as illustrated in Figure 1. Since this combination contains both time-variance and frequency variance of the signal, it incorporated almost the information needed for environmental sound recognition. We call this input data type a time frequency intersection pattern and refer to the time-variance patter of power as power-variance pattern. Thus, the information can be represented as $[S_p(\omega_m), P(t_n)]$, where ω_m $(m = 1, 2, \ldots, M)$ is the frequency, $S_p(\omega)$ indicates the spectrum at the time frame of power peak, and $P(t_n)$ indicates the power of sound for time frame t_n $(n = 1, 2, \ldots, N)$. The total information includes two vectors with sizes M and N (total $M + N$), which is less than that of HMM-based methods ($M \times N$ in total). This method can drastically reduce the input data while preserving the main time-frequency characteristics of environmental sounds.

We use perceptron NNs for environmental sound recognition. A multistage classification-recognition strategy is adopted to cover environment sounds with different time lengths. The first stage is the classification part, which classifies environmental sounds into three categories, single bursts, repeated sounds, and continuous sounds, based on their long-term power-variance patterns. The second stage is the recognition part, for individual recognition of each sound. In this stage, three different NN groups are used for different categories of environmental sounds. Two experiments were conducted using an original environment sound database recorded in an ordinary room and the RWCP database recorded in an anechoic chamber to verify the proposed new method.

2. Environmental Sound Database and Preprocessing

Since this research is concerned with a project that aims to develop a security patrol and home-helper robot capable of understanding environmental sounds, the target environmental sounds are chosen to be important for the robot to achieve its tasks. As seen in Table 3, 10 kinds of environmental sounds were selected and recorded in an ordinary room environment, with 30 samples of each kind. The original sampling frequency was 44.1 kHz.

For comparison with the previous methods, we selected 10 kinds of sounds and a total of 45 instances from the RWCP-DB as seen in Table 4.

Since there are unlimited kinds of environmental sounds, no database can cover all of them. Therefore, no system will be able to recognize all environmental sounds. Instead, for a practical system, the target sounds must be limited according to the practical environment and the purpose of tasks. That is, environmental sound recognition is task dependent.

At the preprocessing stage, the environmental sound data were downsampled to 8 kHz. The instantaneous power was calculated for each time frame of 128-point length. While the long-term power-variance patter contains the power data of 48 frames, the short-term power-variance patter is of 16 frames. The peak spectrum was calculated around power peak with a time frame of 64 points. All data were normalized to have a maximum value of one.

3. System Construction

In many cases, environmental sounds can be mainly classified into collision sounds, friction sounds, vibration sounds, electric sound, and other noises. Based on their power-variance patterns, environmental sounds can be roughly classified into single bursts, repeated bursts, continuous sounds, and other noises. It is reasonable to first classify the environmental sounds into different categories based on their long-term power-variance patterns in the classification stage. Recognition based on the combination of short-term power-variance patterns and frequency-variance patterns at the power peak will be performed in the second stage.

The data flow of the environmental sound recognition system is presented in Figure 2. The system consists of a classification part and a recognition part.

A three-layer perceptron NN is used for sound classification and recognition. The construction of the NN is described in Table 2.

3.1. Classification by Long-Term Power-Variance Patterns. The data needed for classification is the long-term power-variance patterns for each input sound. An example of the

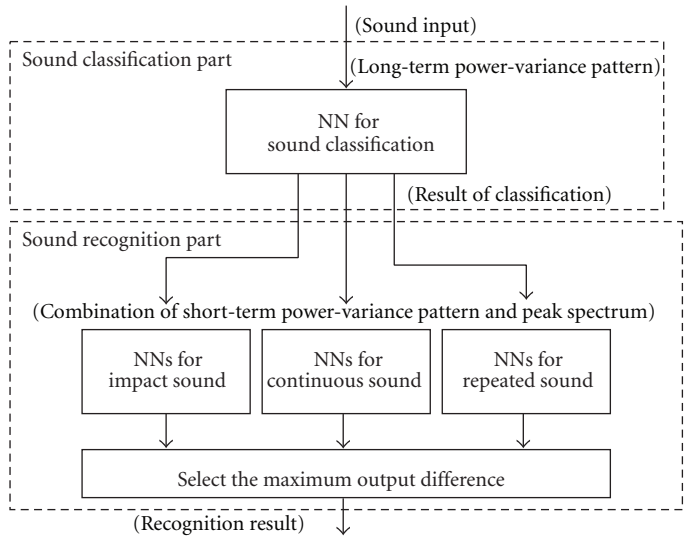

Figure 2: System data flow.

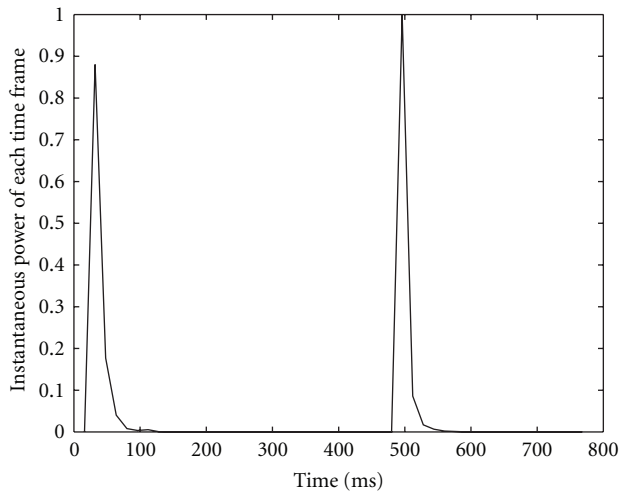

Figure 3: A sample of long-term power-variance pattern (a door-knocking sound).

Table 2: Construction of NNs for classification and recognition parts.

Input layer neuron	48
Intermid layer neuron	32
Output layer neuron	2

long-term power-variance pattern of a door-knocking sound is presented in Figure 3.

This classification stage classifies sounds with short impact sounds as single-impact sounds; sounds of friction, vibration, noises, and electric sounds like phone bells as continuous sounds; some sounds with repetition, for example, hand claps or knocks on a door, as repeated sounds.

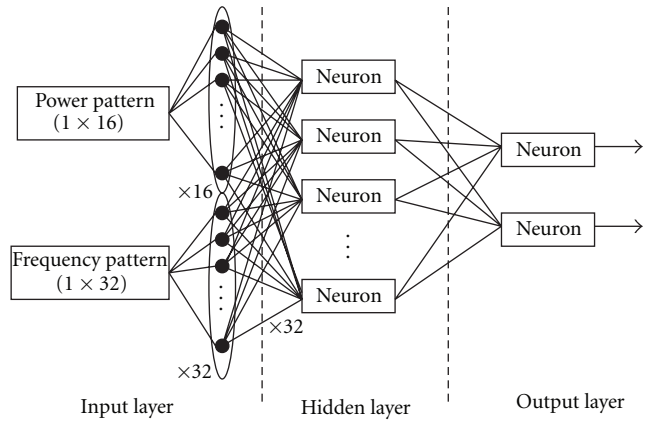

Figure 4: NN in the recognition part.

3.2. Construction of the Recognition Part. For almost all kinds of environmental sounds, the time variances of the frequency characteristics are usually rather stable and there are few marked changes during their period compared with speech sounds. The input data for the recognition part assigns the short-term power-variance pattern to the first 16 inputs and the instantaneous spectrum calculated at the power peak to the remaining 32 inputs, as seen in Figure 4. The output layer of each NN has two neurons that correspond to the results of correct and incorrect matching.

The three NNs in the recognition part correspond to the three target sound categories. Each NN, constructed by a three-layered perceptron, is trained for one target sound category. The final recognition result depends on the difference between the two output neurons of each NN. The NN that obtains the maximum difference of correct and incorrect output is dominant and gives the final recognition result (Figure 2).

Table 3: Results of recognition experiments for environmental sounds in the original database.

Sound kind	First stage rate	Final recognition rate
Boll impact	100%	100%
Metal impact	100%	95%
Door opening/closing	100%	85%
Lock	100%	95%
Switch on/off	100%	100%
Typing	100%	75%
Repeated typing	80%	80%
Knock	90%	90%
Telephone ringing	100%	100%
Japanese vowels	100%	100%
Average		92.0%

Table 4: Results of recognition experiments for environmental sounds in the RWCP database.

Sound kind	First stage rate	Final recognition rate
Wood impact	100%,	96.5%
Metal impact	99.5%,	92.5%
Clap	97.5%,	89.2%
Plastic impact	100%,	100%
Grains falling	100%,	80.0%
Telephone ringing	100%,	88.3%
Metal bell	99.2%,	98.3%
Spray	100%,	95.0%
Whistle	100%,	100%
Drier	100%,	86.0%
Average		92.7%

4. Recognition Experiments

Two experiments using the original prerecorded environmental sound database and the RWCP database were conducted. In all of the experiments, the computer system used was an MS-Windows PC with an Athlone 1600 XP CPU and 512 MB of memory. The NNs were implemented using the MATLAB programming language.

For the original database, 10 samples of each sound kind were used for NN training, and 10 samples of data were used for the recognition tests. The NN training time was about 1 hour in total, and the recognition time for each input data sample was less than 0.1 second. The results of the recognition are listed in Table 3. The average rate of recognition was 92.0%.

From the RWCP database, data for 10 kinds of sounds (total of 45 instances) were selected for the experiments. In the experiments, 10 samples of each sound kind were used for NN training and 20 samples were used for testing. Since there were not enough kinds of repeated sounds in this database, only single-impact and continuous sounds were tested. The required training time was 2 hours, and the recognition time for each data sample was less than 0.1 second. The recognition results are presented in Table 4. The average recognition rate was 92.7%.

5. Conclusion

In this research, we propose a multistage environmental sound recognition method. The method consists of a classification stage and a recognition stage. The classification stage classifies environmental sounds into three categories based on their long-term power-variance patterns, and the recognition stage recognizes the sound kind based on a combination of the short-term power-variance pattern and the instantaneous spectrum at the power peak.

The merit of this method is that it uses a one-dimensional intersectional time-frequency pattern that combines the power-variance pattern and the instantaneous spectrum at the power peak. The recognition rate of the new method was 12% higher than methods using only an instantaneous spectrum at the power peak. The results are also comparable with HMM-based methods, although those methods must accommodate the time variance of the input vector series with more complicated computations.

References

[1] J. Huang, N. Ohnishi, and N. Sugie, "Building ears for rbotos: Sound localization and separation," *Artificial Life and Robotics*, vol. 1, no. 4, pp. 157–163, 1997.

[2] S. Nakamura, K. Hiyane, F. Asano, and T. Endo, "Sound scene data collection in real acoustical environments," *Journal of the Acoustical Society of Japan*, vol. 20, no. 3, pp. 225–232, 1999.

[3] K. Hiyane and J. Iio, "Non-speech sound recognition with microphone array," in *Proceedings of the IEEE International Workshop Hands-Free Speech Communication*, 2001.

[4] K. J. Lang, A. H. Waibel, and G. E. Hinton, "A time-delay neural network architecture for isolated world recognition," *Neural Networks*, vol. 3, pp. 23–43, 1990.

[5] K. Miki, T. Nishiura, S. Nakamura, and G. Kashino, "Environmental sound recognition by HMM," in *Proceedings of the Spring Meet of The Acoustical Society of Japan*, no. 1-8-8, 2000.

[6] A. Sasou and K. Tanaka, "Environmental sound recognition based on AR-HMM," in *Proceedings of the Autumn Meet of The Acoustical Society of Japan*, no. 3-Q-7, 2002.