Hindawi
Journal of Robotics
Volume 2018, Article ID 5368624, 9 pages
https://doi.org/10.1155/2018/5368624

Research Article

A Computing Model of Selective Attention for Service Robot Based on Spatial Data Fusion

Huanzhao Chen and Guohui Tian

School of Control Science and Engineering, Shandong University, Jinan, Shandong 250061, China

Correspondence should be addressed to Guohui Tian; g.h.tian@sdu.edu.cn

Received 23 March 2018; Accepted 14 May 2018; Published 2 July 2018

Academic Editor: L. Fortuna

Robots and humans are facing the same problem: they all need to face a lot of perceptual information and choose valuable information. Before the robots provide services, they need to complete a robust real-time selective attention process in the domestic environment. Visual attention mechanism is an important part of human perception, which enables humans to select the visual focus on the most potential interesting information. It also could dominate the allocation of computing resource. It also could focus human's attention on valuable objects in the home environment. Therefore we are trying to transfer visual attention selection mechanism to the scene analysis of service robots. This will greatly improve the robot's efficiency in perception and processing information. We proposed a computing model of selective attention which is biologically inspired by visual attention mechanism, which aims at predicting focus of attention (FOA) in a domestic environment. Both static features and dynamic features are composed in attention selection computing process. Information from sensor networks is transformed and incorporated into the model. FOA is selected based on a winner-take-all (WTA) network and rotated by inhibition of return (IOR) principle. The experimental results showed that this approach is robust to the partial occlusions, scale-change illumination, and variations. The result demonstrates the effectiveness of this approach with available literature on biological evidence. Some specific domestic service tasks are also tailored to this model.

1. Introduction

Humans are able to choose the salient part of the visual scene and focus on this area. Building a computing model to detect such salient has become a popular approach. Its advantage lies in the allocation of computing resources in the following cognition and analysis operations. The most popular definition for selective attention has been proposed by the psychologist J. William. The potential of an area or a location in an image to appear distinct from the rest in the scene is roughly referred to as its visual salience. It is believed that the saliency is affected by top-down knowledge and bottom-up stimulus-driven processes [1]. Saliency computing and selective attention mechanism are widely used in many applications including object of interest image segmentation, visual scene analysis, and object recognition [2].

Here a question is raised about how to choose a specific focus of attention based on multiple formations that describe the visual scene. Based on this theory, most popular approaches of bottom-up attention, Itti model, hypothesized that many kinds of feature maps on image feed into a unique representation called saliency map. The saliency map is a two-dimensional scalar map representation while it topographically represents activity saliency of image and is used to choose the salient location. Specifically when this location is salient, it is referred to as an active location of a saliency map, no matter the facts that it corresponds to a thing of an obvious colour in a dark background or to a moving object in a still background. Switching attention based on this scalar topographical representation to focus on the most salient location could draw attention on the area of the highest activity in the saliency map[3].

The concept of "salient" is usually mentioned in the process of bottom-up computations. The term "saliency" characterizes specific parts in a scene that might be users or objects that are assumed to be relative to the neighbouring area [4].

Figure 1: Illustration of intelligent space system.

The representation of saliency in a specific scene is computationally introduced in the process of choosing spatial area. For selection and convergence process, when one wins the spatial competition in one or more feature maps at different spatial scales, the related location is defined as salient. After that, detailed measurement of saliency integrating multiple kinds of feature maps is encoded by saliency map. In that way it provides potential heuristic information for choosing attention on salient locations, without consideration of the detailed feature responses that made those locations salient. The saliency map is where the attention spotlight should be placed as well as the most salient location in the scene [5]. A proper neural architecture to find the area needing user's attention from the generated saliency map is winner-take-all network, which implements a neural distributed maximum detector [4]. Using this mechanism, however, brings another computing problem: how to lay attention on the different saliency location or winning area in a saliency map. One plausible computing strategy has already got biological experimental support. It contains transiently inhibiting neurons for current focus of attention in presenting saliency map. The currently attended location is then suppressed. This progress will be repeated and generates attention recording. This kind of inhibitory strategy aims at transferring the FOA. This phenomenon has been named as "inhibition of return" (IOR) and became a popular way of simulating selective attention mechanism [5]. By using IOR process, we are able to rapidly find and shift the focus of attention on different salient locations with decreasing saliency, rather than being bound to attend only to the location of maximal saliency at a given time.

We also use intelligent space to obtain information about the users in a comprehensive way. Intelligent space is capable of improving both perception and executive skills of service robots. Figure 1 illustrates the overview of intelligent space. Users are capable of communicating with both robots and machines, with the help of intelligent space. Many useful services could be provided by actuators discretely installed in the space. The main suggestion of the intelligent space is to install sensors and actuators separately in domestic space, so that service robot could provide timely, stably, and intelligent service with limited configuration [6]. The sensor network can capture human motion information as the user moves in everyday life scenes. In the meantime, intelligent space can offer some service to users by its actuators. It is important to fuse the data from different resources as the sensor network of intelligent space is discretely distributed in target environment. Cellular neural network is an effective way to deal with spatiotemporal temporal processing [7]. Here we choose spatial transformation as a mathematical way in data fusion operation. This mathematical approach is relatively simple and straightforward comparing with neural network. In conclusion, intelligent space is utilized in our research aiming to develop both perception skills and executing skills for robots.

In this paper, we proposed a computing model of visual attention based on data fusion using intelligent space. The model is turned out to select visual attention accurately and robustly by a set of experiments.

2. Related Work

The popular attention selection model is inspired by the biological structure and the behaviour of human visual system demonstrations [1, 8]. Usually, saliency map is proposed based on topographical combination with multiscale features in colour images. Gabor filtering and center-surround operator are used to process the features. Center-surround mechanism is inspired by photoreceptor interactions and visual receptive field organization. It consists of the center and the surround as the feature values at the corresponding image pixels derived from responses to stimuli [9]. A dynamical neural network based is built on WTA network and IOR mechanism, to choose focus of attention and to balance the saliency map. A few other saliency computing models based on Itti model for salience computation are presented [10–13].

In recent years, some models are proposed by new approaches other than Itti's approach. Zhao proposed a multi-context deep learning framework for salient object detection where local context and global context are all concerned [14]. Wang presented a deep neural network based saliency detection mechanism by integrating global search and local estimation [15, 16]. Liu presented saliency tree as a novel saliency detection framework which could provide a hierarchical representation of saliency for generating high-quality regional and pixel-wise saliency maps [17]. Zhang proposed a saliency-guided unsupervised feature learning framework for scene analysis [18]. Wei proposed an algorithm for simultaneously localizing objects and discovering object classes based on bottom-up multiple class learning [19]. Lin proposed a novel salient object detection algorithm by integrating multi-level features including local contrast, global contrast, and background priors [20]. Chen proposed a protoobjects computational model by predicting fixations using a combination of saliency and mid-level representations of shape [21].

Most models of fixation prediction operation at the feature level are like the classical Itti model. Others emphasize that salient object is more important while defining it needs researchers annotation. Most of the existing models consider low-level visual features and locations of objects. The major difference among them is the method for dealing with incoming sensor information, feature types, and computing saliency. For domestic service robot, there are two main shortages. At first, considering the complex environment of domestic environment, blocking is a big trouble. Then human behaviour is a key component of robot attention for service

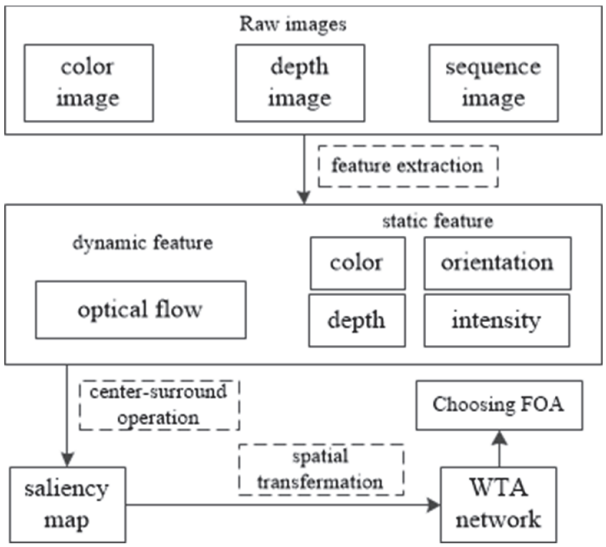

Figure 2: Illustration of selective attention model procedure.

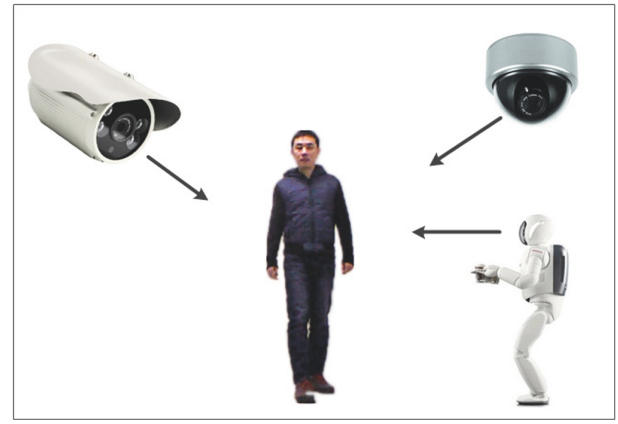

Figure 3: Illustration of multiview situation from sensor network in intelligent space.

task, so that human behaviour deserves more interests in the saliency computing process. Unlike other studies which have emphasize elaborately designed low-level feature descriptors, our approach proposed a saliency model with the help of intelligent space platform. So more information from different cameras is mixed and computed. This has the potential to improve the accuracy and robustness of the saliency computing.

3. A Saliency Computing Model of Selective Attention

The general procedure of saliency computing model for selective attention is illustrated in Figure 2. Raw images (colour image, time ordered image sequences, and depth image) are obtained from cameras distribution from intelligent space. At first, features are extracted using Gabor filter and difference of Gaussian filter. Besides, dynamic features are calculated by computing of optical flow. Then the homogeneous transformation matrix is applied for representing geometric relations between cameras from different coordinate system. After that a saliency map is proposed for nonlinearly combining the features and feeds them into a WTA network. By integrating with inhibition of return method, the FOA is finally selected and rotated.

3.1. Feature Extraction and Processing. For this computing model, the input is captured as colour images and depth images. Colour images are obtained, being digitized at a specific resolution. Based on dyadic Gaussian pyramids and being created in few spatial scales, this is considered as progressively low-pass filter. The colour images are subsampled to produce horizontal and vertical image-reduction factors. Depth images are obtained by infrared sensor of Microsoft Kinect and are capable of detecting user movement in key areas.

Static features could be extracted from one single image and represent some static character (like outline and chromatic aberration). Colour images are obtained from ceiling camera from intelligent space platform, as shown in Figure 3. As b, g, and r refer to the blue, green, and red from the colour images, then we build intensity feature map$I = (r + g + b)/3$. This measures the intensity value of all colour channels of an image. Four broadly tuned colour channels are created as follows:

$$B = b - \frac{(g + r)}{2} \tag{1}$$

$$R = r - \frac{(g + b)}{2} \tag{2}$$

$$G = g - \frac{(b + r)}{2} \tag{3}$$

$$Y = \frac{(b + r)}{2} - \frac{|r - g|}{2} - b \tag{4}$$

for red, green, red, and yellow. Based on these colour channels, four Gaussian pyramids are created as $R(\sigma)$, $G(\sigma)$, $B(\sigma)$, and $Y(\sigma)$, while the scale is $\sigma \in [0, 1, 2, 3, ..., 8]$. By using Gaussian pyramid in image processing to extract multiscale features and downsampling, we defined the difference of cross scale as $\ominus$ between two feature maps based on center-surround operation. This operation is generated based on point-by-point subtraction and interpolation. This computing process is highly relative to the sensitive features which is shown in a series of maps $I(c, s)$, with $s = c + \delta$, $c \in [2, 3, 4]$ and $s \in [3, 4]$.

$$I(c, s) = |I(c) \ominus I(s)| \tag{5}$$

By simulating the colour-double-opponent system in parts of human brain, chromatic and spatial opponency is constructed for the red/green, green/red, blue/yellow, and yellow/blue colour pairs in our model.

$$RG(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \tag{6}$$

$$BY(c, s) = |(B(c) - Y(c)) \ominus (B(s) - Y(s))| \tag{7}$$

Gabor filters are the product of a cosine grating and a 2D Gaussian envelope, approximating the receptive field sensitivity profile (impulse response) of orientation selective neurons in primary visual cortex. Next, we extracted information of local orientation from based on Gabor pyramids $O(\sigma, \theta)$, where $\sigma \in [0, 1, 2, ..., 8]$ represents the scale and $\theta \in [0^\circ, 45^\circ, 90^\circ, 135^\circ]$ is the preferred orientation.

$$G(x, y, \theta) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\delta^2}\right)\cos\left(2\pi\frac{x'^2}{\lambda} + \psi\right) \quad (8)$$

Orientation feature maps $O(c, s, \theta)$ are encoded as a group local orientation contrast between the center and surround scales:

$$O(c, s, \theta) = |O(c, \theta) \ominus (s, \theta)| \quad (9)$$

Here we obtain a depth image based on infrared camera from Microsoft Kinect. At first, Kinect captures Speckle Pattern in different scales and it is saved as Primary or Reference Pattern. Then, the infrared camera detects the objects and gets its corresponding Speckle Pattern. The depth information is obtained by matching Primary or Reference Pattern and triangulation method of ground resistance test. Finally it generates a depth image to describe the spatial information by using 3D Reconstruction and Local offset. After a normalization operation, we can get depth map m_d to describe the spatial relationship of the environment.

Next the dynamic features from image sequences are extracted by computing optical flow. The mobile objects (normally users and robots) are salient part and deserve more attention from service robots. Suppose the image intensity function $I(x, t)$ is

$$I(x, t) \approx I(x + \delta x, t + \delta t) \quad (10)$$

where δx is the displacement of the local image region at (x, t) after time δt. The left side of the equation is expanded based on a Taylor series:

$$I(x, t) = I(x, t) + \nabla I \cdot \delta x + \delta t \cdot I_t + O^2 \quad (11)$$

where $\nabla I = (I_x, I_y)$ and I_t are the first-order partial derivatives of $I(x, t)$ and O^2 is the second- and higher-order terms, which are assumed negligible. Subtract $I(x, t)$ on both sides while ignoring O^2 and dividing by δt:

$$\nabla I \cdot (v) + I_t = 0 \quad (12)$$

This equation is called optical flow constraint equation. Suppose (u, v) is the image velocity, while $\nabla I = (I_x, I_y)$ is the spatial intensity gradient. Optical flow is defined as the distribution of the velocity in given part of the image. The optical flow at given time is described as feature map m_{of}.

3.2. Space Transformation Using Coordination Transform. In this chapter, we proposed a systematic and generalized method about three-dimensional coordinate transform. That is because the service robot and visual devices under real circumstances are from different area, with respect to different reference coordinates. We use a 3D Euclidean space-homogeneous coordinate representation to develop matrix transformations. This operation includes scaling, rotating, translating, and perspective transformation. So features can be flexibly selected to generate a saliency map and focus the attention accurately. In this way, some common problems like pattern noise can be successfully solved.

The homogeneous coordinate representation implies that the n-dimensional vector could be represented by a $n+1$ component vector. Thus in a 3D space a position vector and an augmented vector $\mathbf{w}$ can be used to represent that a position vector $\mathbf{p}$ is in homogeneous coordinate representation.

$$\mathbf{w} \times \mathbf{p} = \left(wp_x, wp_y, wp_z, w\right)^t \quad (13)$$

In physical world the coordinates corresponded to homogeneous coordinate as follows:

$$p_x = \frac{wp_x}{w}, \quad p_y = \frac{wp_y}{w}, \quad p_z = \frac{wp_z}{w} \quad (14)$$

The component w represents the scale factor of the homogeneous coordinate. The homogeneous transformation matrix is a 4×4 matrix. This matrix is used to map a homogenous coordinate position vector from one coordinate system to the other coordinate system. The homogeneous transformation matrix T is as follows:

$$T = \begin{bmatrix} \mathbf{R}_{3\times3} & \mathbf{p}_{3\times1} \\ \mathbf{f}_{1\times3} & w_{1\times1} \end{bmatrix} = \begin{bmatrix} RotationMatrix & PositionVector \\ PerspectiveTransf & ScalingFactor \end{bmatrix} \quad (15)$$

The two components, respectively, are the 3×3 rotation matrix and the 3×1 position vector. The vector is the rotated coordinate system origin with respect to the reference system. For the second row, one component is a 4×4 homogeneous transformation matrix and the other component of T represents *ScalingFactor*. The matrix maps a vector expressed in homogeneous coordinates with respect of the $OUVW$ coordinate system to the reference coordinate $OXYZ$ system. That is,

$$\hat{p}_{xyz} = T\hat{p}_{uvw} \quad (16)$$

$$T = \begin{bmatrix} n_x & s_x & a_x & p_x \\ n_y & s_y & a_y & p_y \\ n_z & s_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{n} & \mathbf{s} & \mathbf{a} & \mathbf{p} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (17)$$

As is shown, (17) represents a homogeneous transformation matrix in a 3D space. The space transfer operation is used to

composite an overall saliency map based on the saliency maps from different cameras. In conclusion, the homogeneous transformation matrix geometrically represents the location of a fixed coordinate system with respect to a reference coordinate system from other visual resources.

3.3. Building Saliency Map and Finding FOA. In this chapter, we compute saliency by building a saliency map referring to biological principles to describe the value of saliency in the given image. The most salient location in the specific coordinate is referred to as FOA. It represents the maximum area of the space that the robot should focus on at given time. Next we need to propose a FOA selection and inhibition mechanism to switch FOA among saliency maps. There are two major steps which are biologically plausible. At first, winner-take-all network is used to model the saliency score all over the environment and identify FOA. Then Argmax function is used to identify of the saliency map against the highest contributing score of the saliency map activity at the winning location. The neurons of WTA network include a single capacitance, a leakage conductance, and a voltage threshold. A prototypical spike will be produced as the charge delivered by synaptic input reaches the threshold of the neuron. Finally the saliency became the neurons feed into a winner-take-all (WTA) neural network. There is one and only one neuron that could be activated by the WTA network at one time. Other neurons are inhibited at the same time to ensure that the only activated exclusively neuron represents the FOA in space.

We can build four "feature maps" corresponding to the features extracted by methods proposed in previous chapter. Intensity map m_i is obtained in (1). Colour map m_c describes colour features obtained by (2) and (3). Orientation feature map is $m_c = \mathrm{O}(c, s, \theta)$. Depth map m_d is extracted from sensors in Kinect. Moreover, the dynamic feature map m_{of} obtained from the computing of optical flow in (8). Here we use entropy method to choose the weigh for every part of saliency amount. The entropy can be obtained by the following equation:

$$e_s = \sum_{k=1}^{4} e_k \tag{18}$$

where e_k, $k \in [1, 2, 3, 4]$, means the entropy for m_c,m_i,m_o and m_d. e_s is the summation of e_k.The final saliency map S is generated by accumulation static and the former feature maps as follows:

$$S = \frac{e_1}{e_s} m_c + \frac{e_2}{e_s} m_i + \frac{e_3}{e_s} m_o + \frac{e_4}{e_s} m_d + m_{of} \tag{19}$$

In this way, features are nonlinearly accumulated by their entropy contribution in normalization coefficient so that the result will be more precise and comprehensive. Then the final saliency map S from one specific resource is generated.

The neurons in the saliency map S could receive activation signals from previous process and remain independent of each other. Every neuron of the saliency should activate the corresponding WTA network neuron. Based on biological mechanism, the neuron could be fast activated according to saliency map on salient place of space. Here some relative process of choosing FOA is included as follows. First, the FOA in original state is not fixed. Another winner neuron with salient place will be excited. Then, all other neurons are suppressed to make the FOA exclusively and wait for next signal. At last IOR mechanism is activated in saliency computing in order to switch the FOA dynamically based on saliency map [22]. The FOA could be unconstrained transferred from the previous FOA to the next FOA while other areas are suppressed. In order to locate the FOA, the winning position is

$$FOA = \operatorname{argmax} S \tag{20}$$

In this way, the FOA which is highly relative to service task can be picked out. This mechanism can be verified by experiments in the following chapter.

Figure 4: Illustration of intelligent space system.

4. Experiments

In this chapter we present a few experiments in real images to demonstrate the computing model of selective attention model. In order to stimulate indoor environment for service robot, we build a room and interconnected corridor experiment environment based on intelligent space, as is shown in Figure 4.

It is a simple environment by contrast as there are fewer obstacles and a simple background in the corridor. However the room is a sophisticated environment as there are various kinds of furniture and electronics. All these objects will make it hard for attention selection. We built an intelligent space platform. There are a few cameras, including ceiling camera and robot sensors, as is shown in Figure 5. This sensor network could be able to observe the target environment in a more comprehensive way. We can obtain colour images, depth images, and time sequence images based on our

Figure 5: Illustration of cameras in intelligent space system.

Figure 6: Original colour image and depth image in a corridor.

approach. Based on this configuration, a series of experiments are performed in this real scenario.

General Performance. The first experiment is deployed in the corridor. Figure 6 illustrates the original images (colour images and depth images) both from a room and a corridor. There are one user and a service robot in the experiment. This is in accordance with the scenes of daily life. The first two rows are colour images from ceiling camera and robot. The bottom image is depth image from robot. We also capture a series of time ordered image sequences. It implies that colour image contains the global colour features from users and objects. Depth image emphasizes the saliency about users located in specific area. Then we calculated the optical flow

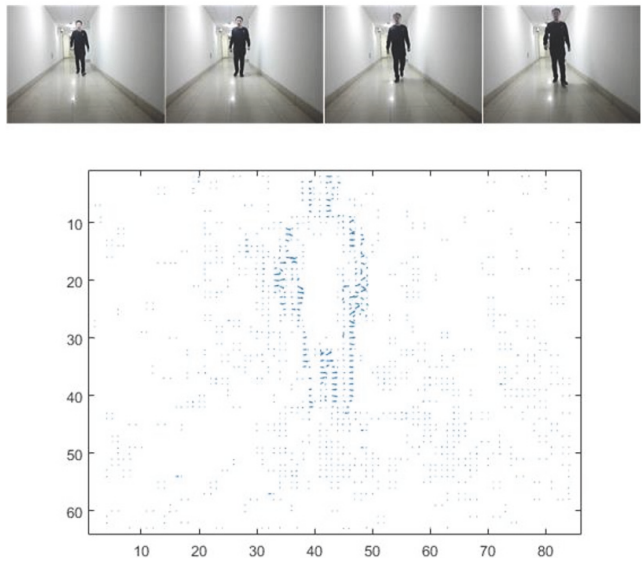

Figure 7: Illustration of cameras in intelligent space system.

using the image sequence. The result of saliency computing experiment based on our approach is illustrated in Figure 7. We can see that as user is walking in the corridor, the outline of the user is clearly shown in the chart. As depth image is static and captured in fixed area, the user's movement as dynamic feature is more accurate to reveal the saliency of the target environment. It turns out to be a good supplementary method.

After that, the final result of the saliency computing result is shown in Figure 8. The first row is saliency maps while the bottom picture visualizes the FOA in the environment. From Figure 8 we can see the model can find the user from the corridor accurately. In the common sense, a human waiter should also focus on the user and prepare to offer some service. So our result is in agreement with biological sense. As for the indoor environment, despite the fact that the background is complex, we can see that our attention is in the user. By common sense, the model chooses the FOA accurately. Due to the existing of depth information, the user can be more attractive than the background.

Comparison Experiment with Existing Model. Then we make two comparison experiments with traditional computing approach as is shown in Figure 9. The first row is the original colour images while one is from a corridor and others are from a room. The second row is result of traditional model, while the last row is result of our model. By comparison, we can see that the result of Itti model is coarse and inaccurate. Some other parts which are less salient are chosen (like desks and chairs). That is because traditional approaches are mainly focused on the saliency parts in image. Normally the human user is of great importance and should be the focus of the service tasks. On the contrary, our model is able to choose FOA more precisely and more reasonably, due to the process of depth image and event item which emphasize

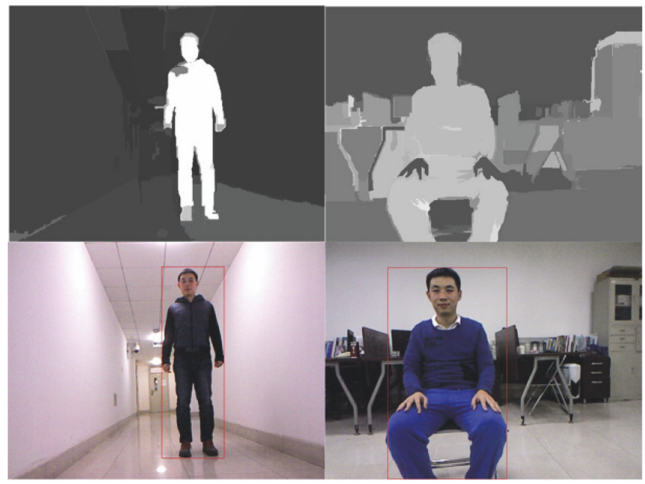

Figure 8: General performance experiment.

Figure 9: Comparison experiment with existing model.

discrete information about intelligent space device or human behaviour.

Influence of Noise. Figure 10 is an example of influence of noise. The top-left is an image with salt and pepper noise while the top-right image is a depth image. The bottom images are the result of saliency map and corresponding FOA. As is shown in the figure, the user and service robot are chosen as saliency part. The influence of noise has been eliminated and the model chosen is right FOA by common sense. In such images, our model was also in better performance as our subjective perception of saliency. Due to the depth information, noise in colour images can be excluded from the result. This operation improves the robustness of the saliency computing model.

5. Discussion

Selective visual attention is capable of finding attention timely on salient objects or users in the domestic environment. This kind of skill enables service robots to locate salient things in a sophisticated visual scene. As it makes human able to find possible food, predator, or mate in the outer world. This paper focuses on developing computational models for service robots that describe the detailed process on attention selection in a specific scene. This has been an important challenge for both robotic and computational neuroscience.

We proposed a computing model for visual attention based on spatial data fusion. Using images spatially independently from the sensor network is applied for visualizing static occlusion culling. We also captured dynamic features

Figure 10: Illustration of cameras in intelligent space system.

in order to reflect human motion states. This is a key component of potentially interesting information for a service robot. When performing in a real environment both quantitatively and qualitatively, the application of our approach has been found to correspond more accurately to locations and saliency maps of eye fixations in visual environment than traditional approaches. Our computing method approach is inspired by the framework of Itti model and has also been found to be more useful in performing better performance than any other same state-of-the-art existing same saliency computing approaches. The better performance of the proposed model has been confirmed through a series of experiments. Most traditional approaches on saliency computing merely are based on low-level features from colour image. This paper covers spatial information and dynamic features which are important to be concerned in scene analysis for service robot. Our model shows great improvement compared with traditional methods both in task-related accuracy and in noise resistance capability. What is more, compared with traditional approaches, it produces better results for attention selection. To some extent, it is observed that the result using our saliency computing approach is consistent manual annotations.

At present, there are a few shortages in our research. Some low-level visual features in multiscales are considered in the research such as depth, hues, orientation, and intensities. While biological principle (like the comprehensive shape features) derived is extensible and could be used to more feature dimension. In the future, new approaches with less computing cost, incurred through multiplicative top-down weight on bottom-up saliency maps, which will combine both goal-driven and stimulus-driven attention should be concerned. Also it should optimize the efficiency of guiding on potential target locations, simultaneously being productive to change visual stimulus situations. Furthermore, when robot operates in unconstrained environments where unexpected events such as accidents may occur, the attention system based on saliency computing needs to bridge the attention selection to further application in further research. This has great potentials to improve its guiding capabilities for more visual computing tasks.

6. Conclusions

In this paper, we proposed a novel saliency computing model for selective attention based on human visual system and traditional bottom-up saliency computing approach. This new computing model with feature fusion is built inspired by the framework of Itti model. It performs salience computing based on multiple information fusions. The main contribution of this paper is to build a computing model for visual attention selection for service robots based on spatial data fusion. By introducing colour images and depth information from different locations, our model solved the problem of image matching under complex condition like noise corruption. The result of our model is consistent with the biological evidence of study of visual attention in human psychophysics and optimized target detection speed, tailored to a specific domestic service task.

Visual attention mechanism plays an important role in the human visual system and for robots and other machines [1]. There are two main goals for developing the selective attention model. At first, selective attention model can be used to cope with massive complexity perception information. This model can prioritize saliency areas in visual images. More potential operations exist, especially for large image information, that need to be processed or if real-time processing is needed for service robots. The other goal is the ability to support task plans. This is always necessary especially for service task performs in a domestic and sophisticated unknown environment. As service robot usually works in the same environment with human. It is reasonable to transfer the human attention mechanism to service robots in order to perform service tasks.

The novel computing model of selective attention proposed in the paper is a biological plausible approach. Furthermore, it could extend to other applications not only in saliency computing related approaches. A large bunch of potential opportunity stays in modifying and updating the saliency computing framework of Itti model and referring to recent research in biology. In the future, it would be interesting in exploiting performance of the saliency computing model built on the application of the proposed computing model. Besides service task inducing in our research, more application of the extracted saliency computing model can be present to segmentation, bionic cognition, and so on.

Data Availability

The data used to support the findings of this study have not been made available because some of technical reasons.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] F. Baluch and L. Itti, "Mechanisms of top-down attention," *Trends in Neurosciences*, vol. 34, no. 4, pp. 210–224, 2011.

[2] M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[3] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[4] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[5] H.-W. Kwak and H. Egeth, "Consequences of allocating attention to locations and to other attributes," *Perception & Psychophysics*, vol. 51, no. 5, pp. 455–464, 1992.

[6] H. Hashimoto, "Intelligent space - how to make spaces intelligent by using DIND?" in *Proceedings of the SMC2002: IEEE International Conference on Systems, Man and Cybernetics*, pp. 14–19, Yasmine Hammamet, Tunisia, 2002.

[7] L. Fortuna, P. Arena, D. Balya, and A. Zarandy, "Cellular neural networks: a paradigm for nonlinear spatio-temporal processing," *IEEE Circuits Systems Magazine*, vol. 1, no. 4, pp. 6–21, 2001.

[8] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[9] D. Sen and M. Kankanhalli, "A bio-inspired center-surround model for salience computation in images," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 277–288, 2015.

[10] X. Hou, J. Harel, and C. Koch, "Image signature: highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, 2012.

[11] R. Achanta and S. Süsstrunk, "Saliency detection using maximum symmetric surround," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 2653–2656, September 2010.

[12] 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.

[13] C. Yang, L. H. Zhang, H. C. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 3166–3173, Portland, Ore, USA, June 2013.

[14] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015*, pp. 1265–1274, IEEE, Massachusetts, Mass, USA, June 2015.

[15] T. Chen, L. Lin, L. Liu, X. Luo, and X. Li, "DISC: deep image saliency computing via progressive representation learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1135–1149, 2016.

[16] J. Pan, E. Sayrol, X. Giro-I-Nieto, K. McGuinness, and N. E. O'Connor, "Shallow and deep convolutional networks for saliency prediction," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 598–606, usa, July 2016.

[17] Z. Liu, W. Zou, and O. Le Meur, "Saliency tree: a novel saliency detection framework," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 1937–1952, 2014.

[18] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2175–2184, 2015.

[19] X. H. Shen, Y. Wu, and Evanston, "A unified approach to salient object detection via low rank matrix recovery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 853–860, Providence, RI, USA, June 2012.

[20] M. Lin, C. Zhang, and Z. Chen, "Global feature integration based salient region detection," *Neurocomputing*, vol. 159, no. 1, pp. 1–8, 2015.

[21] Y. Chen and G. Zelinsky, "Computing Saliency over Proto-Objects Predicts Fixations During Scene Viewing," *Journal of Vision*, vol. 17, no. 10, p. 209, 2017.

[22] M. I. Posner, "Components of visual orienting," *Attention Performance*, vol. 32, no. 4, pp. 531–556, 1984.