Research Article

A Novel Fusion Method by Static and Moving Facial Capture

Shuai Liu,^{1,2} Weina Fu,^{1,3} Wenshuo Zhao,¹ Jiantao Zhou,¹ and Qianzhong Li²

¹ College of Computer Science, Inner Mongolia University, Hohhot 010012, China

² School of Physical Science and Technology, Inner Mongolia University, Hohhot 010012, China

³ Department of Computer Science and Technology, Hohhot University of Nationalities, Hohhot 010012, China

Correspondence should be addressed to Shuai Liu; cs.liushuai@imu.edu.cn

Received 27 June 2013; Accepted 21 August 2013

Academic Editor: Su-Qun Cao

Copyright © 2013 Shuai Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For many years, face recognition has been one of the most important domains in pattern recognition. Nowadays, face recognition is more required to be used in video actually. So moving facial capture must be studied firstly because of performance requirement. Since classic facial capture method is not so suitable in a moving environment, in this paper, we present a novel facial capture method in a moving environment. Firstly, continuous frames are extracted from detecting videos by similar characteristics. Then, we present an algorithm to extract the moving object and restructure background. Meanwhile, with analysis of skin color in both moving and static areas, we use the classic faces capture method to catch all faces. Finally, experimental results show that this method has better robustness and accuracy.

1. Background

Today is the era of electronic public security, which uses electronic equipment and network in human security. So identity verification becomes the most important area in electronic social security. Meanwhile, face recognition becomes one of the most widely used biometric identification technologies because of its positive features such as being direct, friendly, convenient, difficult to counterfeit, and cost-effective [1]. Moreover, face recognition also has a wide range in biometric authentication, video surveillance, social security, and other application areas [2]. Generally, face recognition can be divided into three steps: (1) facial detection and segmentation from scene, (2) capture of facial feature, and (3) matching and recognition of human face [3].

Facial detection and capture is the first step in face recognition method. Detecting and tracking faces rapidly from frames in video is the basis of face recognition. Then, recognition will process when an image makes for identification.

However, there are many problems in face detection, like how to use information of facial time and space, how to overcome low resolution ratio and huge range of mutative scale, and how to recognize faces with intense transformation or

a hidden facial part. These are the emphases in studies nowadays [4]. Meantime, facial detection and capture in video need high quality in recognition time sometimes [5]. Thus, it needs low computational cost. So we have to use simpler model with higher accuracy. So, statistical models are used widely in facial detection today [6, 7].

In recent years, neural network, Support Vector Machine (SVM), and AdaBoost are the widest statistical models in facial detection and capture [8]. Thereinto, AdaBoost is used widely in dynamical detection for its fast speed and well stability [9]. Meantime, AdaBoost is also used in other classificatory areas just like classification of music and protein [10, 11]. Besides, AdaBoost is used in other biometric authentications like iris recognition and facial recognition [12].

Though AdaBoost is widely used, it has obvious deficiencies. The first deficiency is its less robustness with transformation of illumination and expression. Secondly, it cannot detect deflecting faces. So, in this paper, we ameliorate AdaBoost by moving object capture [13]; that is, we detect a moving object by using AdaBoost for a tiny time range. Admittedly, facial moving speed is slow. So, frames with tiny time range can be extracted by similar characteristics. We determine relevance of frames by area comparison of moving objects.

Meantime, we determine face when area ratio of facial image is registration more than a threshold. Then, with capture of moving object with restructured background, we analyze skin color in moving area and use it in static area. In this way, all faces are detected and extracted no matter whether these faces are front or deflecting.

The remainder of the paper is organized as follows. We present and analyze our method in Section 2. Then, we present a novel fusion algorithm in Section 3. Moreover, we experiment some data with this novel algorithm and classic AdaBoost algorithm in Section 4. Finally, Section 5 summarizes the main results of the paper.

2. Theory of Methods

There are many expressions of color set in computers and they form different color space. RGB, HLS, YCbCr, and YIQ are the widest used today. Admittedly, skin set gives expression to different cluster features when it is in different color spaces. Researchers discover that skin set shows better clustering results in HLS and YCbCr [14]. Generic terms of YCbCr are called YUV [5]. Y expresses luminance and U, V express chrominance signals.

Since Duan presents her hierarchical method of skin color for color spaces [15], we can analyze distribution of skin color in YIQ and YUV. Then, we use threshold of i and θ to detect skin colors where i is i th component of YIQ and θ is phase angle of YUV. So formula of θ is shown in

$$\theta = \tan^{-1} \left(\frac{|V|}{|U|} \right). \quad (1)$$

We revolve the component of chromatic aberration of YUV to form YIQ. So I contains color information from Orange to Ching and Q contains Green to Magenta. Then, skin tone is between Red and Yellow, and θ is basic in domain [100, 150].

In this paper, we use $\theta \in [100, 150]$ and $I \in [20, 90]$. Then we use (2) to transform YUV and YIQ to RGB:

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}, \quad (2)$$

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}.$$

Then, in order to increase speed of facial detection, we fuse skin color detection into face detection after transformation. We extract a frame as original image and detect possible skin color area. Moreover, we crop following images by using skin space model. Then, we cascade connect weak classifiers to strong classifiers. We have (3) to show a weak classifier $h_j(x)$ with feature f_j . In (3), θ_j is clustering threshold of minimum error in training samples and p_j expresses direction of inequality:

$$h_j(x) = \begin{cases} 1, & \text{if } p_j f_j < p_j \theta_j, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

In this way, we drop those weak classifiers whose clustering rate less than 50%. Then, in order to enhance accuracy of facial capture, we enhance weight for them with better performance and reduce weight for them with worse performance. Computation of weight is shown in

$$w_{1,i} = \begin{cases} \frac{1}{2m}, & x_i \text{ is negative sample, } m \text{ is the number of negative samples,} \\ \frac{1}{2l}, & x_i \text{ is positive sample, } l \text{ is the number of positive samples,} \end{cases} \quad (4)$$

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}, \quad w_{t,i} \in (0, 1) \text{ is probability distribution,} \quad (5)$$

$$h_t = h_k, \quad \text{whose } \varepsilon_k = \min \left\{ \varepsilon_j \mid \varepsilon_j = \sum_i w_t |h_j(x_i) - y_i| \right\}, \quad (6)$$

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}, \quad \text{where } \beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t}, \quad (7)$$

$$e_i = \begin{cases} 0, & \text{where } x_i \text{ classify right,} \\ 1, & \text{where } x_i \text{ classify wrong,} \end{cases}$$

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T a_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T a_t, \quad a_t = \log \frac{1}{\beta_t}, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Equation (4) is initialization of weight, and (5) is normalization of weight. Equation (6) is classifier h_t , which has minimum error rate ε_t for all classifiers h_j of feature j with weight w_t . Equation (7) shows transformation of all weights. Equation (8) presents the strong classifier, which is composed by T weak classifier.

3. A Novel Fusion Algorithm with Background Restructured

In this paper, we restructure background by (9). Without background, we compute function G to express difference of two next frames and function B to express whether there is a moving object in it. Then, applying $I_n(x)$ as pixel gray of foreframe and $T_n(x)$ as threshold, we consider a pixel z is moving when it fits in (10). In this way, we can find all background points by training. Figure 1 is a flow chart of background training. Consider

$$G_{i,i-1}(x, y) = |X_i(x, y) - X_{i-1}(x, y)|, \quad (9)$$

$$B_{i,i-1}(x, y) = \begin{cases} 1, & \text{if } G_{i,i-1}(x, y) \geq \text{threshold,} \\ 0, & \text{otherwise,} \end{cases}$$

$$\begin{aligned} |I_n(x) - I_{n-1}(x)| &> T_n(x), \\ |I_n(x) - I_{n-2}(x)| &> T_n(x). \end{aligned} \quad (10)$$

FIGURE 1: Restructure background with training.

After background restructure, we detect facial area from some continuous frames by (8) with skin color detection. We know (8) process well with front faces. Then, based on faces detected by (8), we search moving object around them. We trust a moving object is a face where registration rate of moving object and facial area is more than a threshold and its color belongs to skin color space. The novel fusion method can extract some deflecting faces, which cannot be found by a classic method.

The following steps are presented to show the novel fusion method.

Step 1. When the video plays, catch three continuous frames as original frames from video. Otherwise go to Step 8.

Step 2. If background is not complete, restructure it one time.

Step 3. Extract skin area with skin model and morphological operation.

Step 4. Extract detected areas from corresponding positions of original images, and then these areas are processed to connected rectangle or oval regions.

Step 5. Use (8) to detect faces.

Step 6. Detect moving faces from nearest facial area.

Step 7. Stamp results and go to Step 1.

Step 8. Procedure finished.

We have Figure 2 to show the flow chart of the novel fusion method.

4. Experimental Result and Its Analysis

In this paper, we validate our method from a video, which contains both single face frames and multifaces frames. We

FIGURE 2: Flow chart of our fusion extraction method.

process classic facial capture method and our method to detected faces. Then comparisons of time and rate are shown.

4.1. Single Face Detection. In frame sequence of the video, the face is moving with different kinds. In this paper, we extract the face image by using front, side, up, down, lean, and shaded face. As we know, classic method cannot detect faces when they are not frontal and can only detect a part of faces when they are shaded. Oppositely, the fusion method can detect faces accurately. We put them in Figure 3. Figures from upper two rows are capture result for classic method and figures from bottom two rows are for fusion method.

4.2. Multifaces Capture. In Section 4.1, we choose faces with different kinds and detect them in Figure 4. Figures from upper two rows are capture result for classic method and figures from bottom two rows are presented for fusion method.

4.3. Analysis of Experimental Result. We have (11) to define mean computational time of every image. With fames of single face, it costs 177.502 ms by using classic method and 127.887 ms by this fusion method. Moreover, it costs 229.631 ms by classic method and 153.963 ms by this fusion method with frames of multifaces (two faces in each image). Then, we have Table 1 to show these results. In Table 1, we find the same problem of these two methods is that it costs much time when detection fails. It is because we need to search the whole frame in this condition. Consider the following:

$$T = \frac{\sum_{i=1}^n T_i}{n}. \quad (11)$$

Then, we have (12) to define detection accuracy of every image. In (12), N_c is frame number with correct detection and N_t is total frame number. With fames of single face, the accuracy is 72.0% by using classic method and 96.8% by this fusion method. Moreover, the accuracy is 60.5% by classic method and 89.6% by this fusion method with frames

TABLE 1: Performance comparison between a single PC and cloud platform.

Mean detection time	Single face (success)	Single face (fail)	Multifaces (success)	Multifaces (fail)
Classic method	156.392	231.791	178.812	303.626
Mean time (ms)	177.502		229.631	
The novel fusion method	126.120	181.354	147.948	205.335
Mean time (ms)	127.887		153.963

FIGURE 3: Comparison of single face capture between two methods.

TABLE 2: Comparison of the face detection accuracy between two methods.

	Classic method	The novel fusion method
Single face	72.0%	96.8%
Multifaces	60.5%	89.6%

of multifaces (two faces in each image). We have Table 2 to show these results. In Table 2, we find accuracy of the fusion method is better than the classic one. We check fail frames of these two methods and find that the classic method shows more negative than the fusion one when deflection of faces is

large. The classic method cannot detect all deflecting faces, but the fusion one can detect most of them. Consider the following:

$$R = \frac{N_c}{N_t}. \quad (12)$$

5. Conclusions

In this paper, by fused skin color model, facial detection method, and moving object capture algorithm, we present a fusion facial detection method. This method takes full

FIGURE 4: Comparison of multifaces detection between two methods.

advantage of information in continuous frames of the detecting video and shows that it is positive in facial detection. Furthermore, we reach that the fusion method has well detecting effect when expressions and facial gestures change greatly. This fusion method remedies deficiencies of the classic method. Finally, we validate our method by using lots of experimental results. The experimental results indicate that the fusion method makes a good effect where faces are moving variously.

The deficiency of this method is that detecting accuracy is low when faces move quickly. In fact, the quickly moving faces lead to error of moving facial judgment. It is because the threshold of two continuous frames is too large that our method treats them as two faces.

Acknowledgments

This work is supported by Grants Programs of Higher-Level Talents of Inner Mongolia University (nos. 125126, 115117), Scientific Projects of Higher School of Inner Mongolia (NJZY13004), National Natural Science Foundation of China (nos. 61261019, 61262082), Key Project of Chinese Ministry

of Education (no. 212025), and Inner Mongolia Science Foundation for Distinguished Young Scholars (2012JQ03). The authors wish to thank the anonymous reviewers for their helpful comments in reviewing this paper.

References

- [1] Y. Yan and Y.-J. Zhang, "State-of-the-art on video-based face recognition," *Chinese Journal of Computers*, vol. 32, no. 5, pp. 878–886, 2009.
- [2] F. Pan, X. Wang, and B. Xiao, "Study on fast face detection of color image," *Chinese Journal of Scientific Instrument*, vol. 25, no. 5, pp. 561–564, 2004.
- [3] 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.
- [4] I. Dryden and K. V. Mardia, *The Statistical Analysis of Shape*, Wiley, London, UK, 1998.
- [5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

- [6] D. B. Graham and N. M. Allinson, "Characterising virtual eigensignatures for general purpose face recognition," *Face Recognition*, vol. 163, pp. 446–456, 1998.
- [7] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.
- [8] S. Zhou and R. Chellappa, "Beyond a single still image: face recognition from multiple still images and videos," in *Face Processing: Advanced Modeling and Methods*, Academic Press, New York, NY, USA, 2005.
- [9] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kégl, "Aggregate features and ADABOOST for music classification," *Machine Learning*, vol. 65, no. 2-3, pp. 473–484, 2006.
- [10] X. Jiang, R. Wei, Y. Zhao, and T. Zhang, "Using Chou's pseudo amino acid composition based on approximate entropy and an ensemble of AdaBoost classifiers to predict protein subnuclear location," *Amino Acids*, vol. 34, no. 4, pp. 669–675, 2008.
- [11] C. W. Hsieh, H. H. Hsu, and T. W. Pai, "Protein crystallization prediction with AdaBoost," *International Journal of Data Mining and Bioinformatics*, vol. 7, no. 2, pp. 214–227, 2013.
- [12] Q. Wang, X. Zhang, M. Li, X. Dong, Q. Zhou, and Y. Yin, "Adaboost and multi-orientation 2D Gabor-based noisy iris recognition," *Pattern Recognition Letters*, vol. 33, no. 8, pp. 978–983, 2012.
- [13] W. Fu, Z. Xu, S. Liu, X. Wang, and H. Ke, "The capture of moving object in video image," *Journal of Multimedia*, vol. 6, no. 6, pp. 518–525, 2011.
- [14] Z. Jin, Z. Lou, J. Yang, and Q. Sun, "Face detection using template matching and skin-color information," *Neurocomputing*, vol. 70, no. 4–6, pp. 794–800, 2007.
- [15] L. Duan, W. Gao, G. Cui, and H. Zhang, "A hierarchical method for nude image filtering," *Journal of Computer-Aided Design & Computer Graphics*, vol. 14, no. 5, pp. 404–409, 2002.