MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

Research on Far-Field Face Detection for Recognition

Zhifei Wang and Zhenjiang Miao
Institute of Information Science, Beijing Jiaotong University, 100044
Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing, China
Phone:(86)010-51683149, Email:{06112059, zjmiao}@bjtu.edu.cn

Abstract

Far-field face detection is the first step of automatic face recognition in video surveillance. In this paper, we propose a framework on extraction of high-resolution images in video that subjects are far from camera. To guarantee the tradeoff between accuracy and speed, our method uses four techniques including motion detection using Gaussian Mixture Models (GMMs), skin model detection and Adaboost-based detector with the expansion of skin region. Moreover, face hallucination based on eigentransformation is adopted to obtain high-resolution images for recognition. Experimental results on real data show that our system can acquire acceptable detection rate with promising speed.

1. Introduction

Face detection has drawn considerable attention in the past. Most of existing methods only work efficiently on high-resolution images. In video surveillance, the size of interested face is small because of distance between camera and subject. Resolution affects face detection and recognition badly.

In recent years, an interesting application domain is face detection on far-field unconstrained video [4]. The first low-resolution face detection technique is proposed by Hayashi [5] to detect 6*6 pixels faces. They try using upper-body images based on Haar features and improve face detection rate exhaustively for low-resolution images. However, some problems keep uncertain, such as, how to get upper-body image data and whether the upper-body detector is efficient.

However, it is necessary to guarantee the speed of detection system in video. In this case, some feature-based methods may be considered by taking face knowledge into account and combining cues such as color, motion and geometry. Many papers [6-8] have shown the success of detection by different skin models. Also, temporal relationship between frames is used to detect faces in video instead of each frame in [9-10]. In addition, the detected images in far-field conditions have poor resolutions. Therefore it is necessary to get high-resolution images by face hallucination. Hallucinating static images has been widely studied [12-13]. Nevertheless, few literatures [14] focus on video-based super-resolution. As the reconstructed faces can be directly used for recognition, it is very useful to get high quality images from low-resolution video sequences.

In this paper, we propose a novel framework for detecting and hallucinating face images from far-field videos, shown in Fig.1. We make a survey on the potential efficiency of those existing detection methods on far-field domain in section 2. Based on the investigation results, section 3 presents our method, which includes four techniques, “motion detection based on GMMs”, “skin GMMs detection based on $C'bC'r$ and H-SV space”, “Adaboost-based detector with the expansion of skin region”, and “face hallucination based on eigentransformation”. Experimental results are shown in Section 4 and Section 5 concludes this paper.

2. Survey on Existing Methods on Far-Field Face Detection

Most of appearance-based techniques aim to tackle the general problem of detection in still images achieving good performance. Sung [15] develops a detection system which assumes a range of working sizes (the minimum size is 20*20) and performs a multi-scale search on images. Based on the Sung’ s work, Rowley [16] uses a multi-layer neural network trained with multiple face and non-face prototypes at different scales, but it only detects upright faces. The same main limitation of them is large computation. Viola [1] proposes a rapid general object detector, which is based on the idea of a boosted cascade of weak classifiers with Haar features. The system reduces dramatically latency at high levels of accuracy and achieves a good tradeoff between accuracy and speed. Adaboost-based algorithm is considered as a standard for

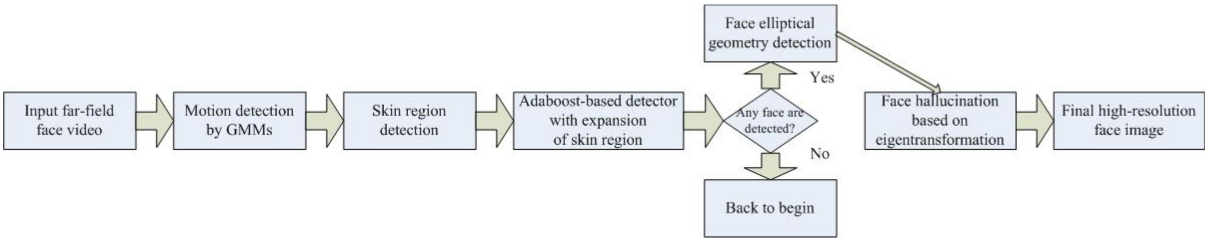

Fig.1. Architecture of our approach

face detection. For that reason, we use Adaboost-based detector as the basis of our method. However, the minimum detectable face size of conventional Adaboost-based detector in OpenCV is 18*18. We try decreasing training data to the size of 12*12 and 8*8. However, experimental results indicate that detection rate decreases in the low-resolution. Then, we do the similar experiment "Expansion of input images" as [5] with MIT+CMU frontal face set, performance is improved certainly.

Compared with Appearance-based techniques, Feature-based ones provide faster performance in constrained conditions. Feature-based techniques, such as color and geometry information, may be used as a supplement to optimize the Appearance-based ones by means of restricting the search area. Yow [7] proposes an organic-feature-based method to achieve pose-robust but time-consuming and failure in low-resolution. Color information is a powerful descriptor in detecting faces. You [8] proposes a new model (H-SV-V), which can efficiently decrease lighting effects. In addition, Hsu [6] finds that skin tone of different people form a compact area in the CbCr plane. The $C^{'}bC^{'}r$ space is more efficient for detecting faces.

Recently, the problem of real-time detection in context of video has been focused, which integrates the temporal coherence during video frames. [9-10] integrate facial probabilities and temporal information to form a method which is superior to frame-based ones. Video-based methods can gain good performance in far-field detection and may be the future direction. In this work, we will use background GMMs [2-3] for motion detection as a supplement to face detection.

3. Proposed Method

Our method will combine the advantages of appearance with feature-based techniques. First, background model is built using GMMs to gain the location of subject. Then, feature-based technique based on skin color is performed to find skin region. Next, appearance-based technique is integrated with Adaboost by expansion of skin region. Finally, eigentransformation based 2DPCA will be adopted for hallucination to get high-resolution images. The architecture of our approach is shown in Fig.1.

3.1 Motion Detection Based on GMMs

We first perform motion detection based on Gaussian Mixture Models (GMMs) [2-3], which can achieve real-time motion detection. Moreover, the parameters of GMMs can be updated for motion detection. The details are shown in our previous work [3]. Some experimental results of motion detection are shown in Fig.5 and Fig.6.

3.2 Skin Model Based Face Detection

It is necessary to choose an appropriate color space for building skin model. Hsu [6] transform nonlinearly YCbCr to $C^{'}bC^{'}r$ for making skin cluster luminance independent. A new space H-SV-V [8] is also robust to lighting change. Based on $C^{'}bC^{'}r$ and H-SV space, GMMs with order (K=2) is utilized for representing skin model, which is described in (1).

$$P(m,\Theta)=\sum_{i=1}^{K}\alpha_i\frac{1}{\sqrt{(2\pi)^d\,|\Sigma_i|}}\exp\left\{-\frac{1}{2}(m-\mu_i)^T\Sigma^{-1}(m-\mu_i)\right\} \quad (1)$$

Let $(\mu_{C^{'}bC^{'}r},\sigma_{C^{'}bC^{'}r})$ and $(\mu_{H-SV},\sigma_{H-SV})$ to be the mean and variance of skin Gaussian model based on $C^{'}bC^{'}r$ and H-SV.

3.3 Adaboost-based Detector with Expansion of Skin Region

Based on Viola and Jones's milestone work [1], Hayashi [5] do research on the low-resolution images with Haar features. They use some psychological results that human can recognize faces in low-resolution using upper-body images better than solely face images. Although they gain the detection rate from 39% to 73% for 6*6 pixel faces, some technologies keep questionable, such as "How to combine face detector with upper-body detector?" However, expansion of input image helps to improve the system, and experiments have verified it.

Then, we expand skin region by bi-cubic with the scaling factor of two, and apply 24*24 pixel face detector aim to detect faces with the minimum size of 12*12 pixel. Here, a new detector is produced using frequency-band limitation of features and conditions, such as $W \geq 2, H \geq 2$ (2)

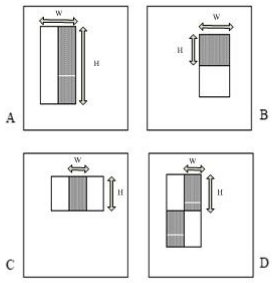

Fig.3. W and H of Haar features

3.4 Recognition Oriented Face Hallucination

3.4.1 2DPCA Analysis

2DPCA [11] is used to decompose images into a weighted combination of eigen-faces. Suppose that there are M samples in total, the j th training image is denoted by $L_j(m \times n, j = 1,2,..., M)$, and the average image is denoted by $\bar{L}$. Then the image covariance matrix G_t can be evaluated by

$$G_t = \frac{1}{M}\sum_{j=1}^{M}(L_j - \bar{L})^T (L_j - \bar{L}) \quad (3)$$

Alternatively, the *criterion* $J(X)$ can be expressed by

$$J(X) = X^T G_t X \quad (4)$$

The optimal projection axes are the orthonormal eigenvetors X_k of G_t corresponding to the first d largest eigenvalues. For a face image f_L, a family of principal component vectors Y_k can be computed by projecting it onto X_k:

$$Y_k = f_L X_k, k = 1,2,...,d. \quad (5)$$

Then we can get each Y of M training image samples:

$$\begin{cases} Y_j = L_j X, j = 1,2,..., M \\ X^T = \frac{1}{M}\sum_{j}^{M} Y_j^T L_j \end{cases} \quad (6)$$

3.4.2 Eigentransformation for Face Hallucination

Based on eigentransformation for hallucination in [12], we apply 2DPCA to the input low-resolution image f_{IL}. From (5), the input face image f_{IL} can be represented by

$$f_{IL} = Y_{IL} X^T \quad \Rightarrow \quad X^T = Y_{IL}^T f_{IL} \quad (7)$$

From (6) and (7), equation (7) can be rewritten as

$$f_{IL} = \frac{1}{M}\sum_{j}^{M} Y_{IL} Y_j^T L_j = \bar{r}L \qquad \bar{r} = Y_{IL} Y^T = [r_1, r_2, ..., r_M] \quad (8)$$

Here, $\bar{r}$ describes the weight that each training face contributes in representing the input image. Replacing each low-resolution image L_j by its corresponding high-resolution one H_j, we have

$$f_H^{ET} = \frac{1}{M}\sum_{j}^{M} Y_{IL} Y_j^T H_j = \bar{r}H \quad (9)$$

f_H^{ET} is expected to be an approximation to original image.

3.4.3 Face Residue Compensation

Face residue compensation is then used to enhance hallucinated images. Coupled PCA [13] is adopted to infer high-resolution residue (H) from low-resolution one (L). The relation between them is described as $H = B_H B_L^T L$, where B_L and B_H are orthogonal matrices, which are optimized by the error energy function $E(B_L, B_H) = \sum_{i=1}^{n} \left\| H_i - B_H B_L^T L_i \right\|^2$. The final high-resolution result is obtained by adding high-resolution residue to hallucinated high-resolution face. The diagram of hallucination algorithm is shown in Fig.4.

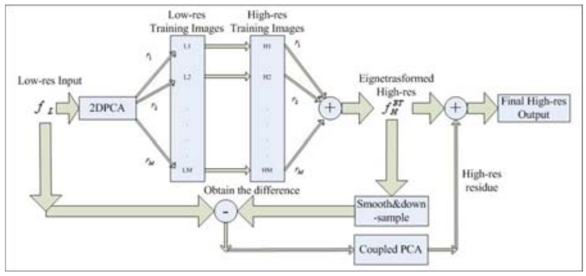

Fig.4. Block diagram for face hallucination

4. Experimental Results

We design experiments on outdoor and indoor videos with size of 320*240. 24*24 Adaboost-based face detector is made using 1586 face images (from Lab), along with 3144 non-face images (from Internet) as training data. Fig.5 and Fig.6 show the detected results of our method on outdoor and indoor respectively. Although we merge detected results of several frames, false positives that appeared in Fig.5 are observed. Compared with outdoor results, indoor performance is much better. Nevertheless, we can acquire real-time with the speed reaching at about 13 frames per second.

After face detection, the focus of face hallucination is to get high resolution frontal faces with 100*100. So we build several training subsets with various low-resolution images and the corresponding high-resolution ones. Every training subset is based on 20 persons and contains 80 images (four low and one high) for each person, and sizes of low-resolution images are 24*24, 18*18, 12*12, 8*8, respectively. Fig.6 shows the obvious difference between our method and Cubic B-Spline.

Fig.5. Face detection results on outdoor cases

5. Conclusions and Future Work

In this paper, we pay more attention to build a novel framework for far-field face detection to achieve a tradeoff between accuracy and speed. Motion detection and expansion of skin region technique are introduced into the system improving accuracy. Skin model is used to optimize the speed by means of restricting search area. Face hallucination by eigentransformation can be used to get high-resolution images for recognition. However, there exist some problems to be solved in the future, such as more robust motion detection, face detection, and online face hallucination algorithms et al.

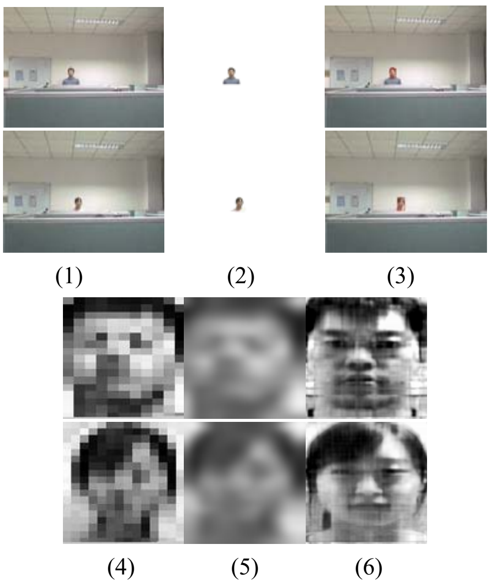

Fig.6. Our experimental results. (1)Input video streams. (2) Motion detection. (3) Face detection. (4)Up-sample the original results and histogram equalization on them. (5) and (6) show the results of Cubic B-Spline and our method respectively.

6. Acknowledgements

This work is supported by National 973 Key Research Program of China (2011CB302203), National Nature Science Foundation of China (60973061), Ph.D. Programs Foundation of Ministry of Education of China (20100009110004), and Ph.D. Technology Innovation Fund of BJTU (141040522).

References

[1]P.Viola, M.J.Jones,"Robust real-time face detection," IJCV, Vol.57, No.2, pp.151-173, 2004.

[2]Stauffer C, Grimson W. E. L, "Adaptive background mixture models for real-time tracking," CVPR'99, Vol.2, pp.246-252.

[3]Y.D Sun, B.Z Yuan, Z.J Miao, and C.K Wan, "Better Foreground Segmentation for Static Cameras via New Energy Form and Dynamic Graph-cut," ICPR'06,Vol.4, pp.49-52.

[4]A.Pnevmatikakis, L.Polymenakos, "Far-Field, Multi-Camera, Video-to-Video Face Recognition," In Journal of Advanced Robotics Systems, I-Tech, Vienna, Austria, pp.467-486, 2007.

[5]S.Hayashi, O.Hasegawa, "A detection technique for degraded face images," CVPR'06, Vol.2, pp.1506-1512.

[6]R.L.Hsu, M.Abdel-Mottaleb, and A.K.Jain, "Face Detection in Color Images," IEEE TPAMI, Vol.24, No.5, pp.696-706, 2002.

[7]K.C.Yow, R.Cipolla, "Feature-based Human Face Detection," Image and Vision Computing, Vol.15, No.9, pp.713-735, 1997.

[8]Y.P You, B.Z Yuan, "Illumination Independent Face Detection," Chinese Journal, Signal Processing, Vol.3, 2004.

[9]K.Mikolajczyk, R.Choudhury, and C.Schmid, "Face detection in a video sequence-a temporal approach,"CVPR'01,Vol.2, pp.96-101.

[10]R.C Verma, C.Schmid, and K.Mikolajczyk, "Face detection and tracking in a video by propagating detection probabilities," IEEE TPAMI, Vol.25, No.10, pp.1215-1228, 2003.

[11]J.Yang, D.Zhang, A.F.Frangi, and J.Y.Yang, "Two-Dismensional PCA: A New Approach to Appearance-based Face Representation and Recognition," IEEE TPAMI, Vol.26, No.1, pp.131-137, 2004.

[12]X.Wang, X.Tang, "Hallucinating face by eigentransformation," IEEE TSMC-C, Vol.35, No.3, pp.425-434, 2005.

[13]W.Liu, D.Lin, and X.Tang, "Hallucinating Faces: TensorPatch Super-Resolution and Coupled Residue Compensation," CVPR'05, Vol.2, pp.478-484.

[14]M.Elad, A.Feuer, "Super-Resolution Reconstruction of Image Sequences," IEEE TPAMI, Vol.21, No.9, 1999.

[15]K-K.Sung, T.Poggio, "Example-Based Learning for View-Based Human Face Detection," IEEE TPAMI, Vol.20, No.1, pp.39-51. 1998.

[16]H.A.Rowley, S.Baluja, and T.Kanade, "Neural Network-Based Face Detection," IEEE TPAMI, Vol.20, No.1, pp.23-38, 1998.