Video Parsing Based on Head Tracking and Face Recognition

Pengxu Li[§], Haizhou Ai[†], Yuan Li[§] and Chang Huang[§]
Department of Computer Science and Technology,
Tsinghua University, Beijing, 100084, China
+86-10-62795495
[§]{lpx02, yuan-li, huangc99}@mails.tsinghua.edu.cn, [†]ahz@mail.tsinghua.edu.cn

ABSTRACT
In this paper, we describe a fully automatic video retrieval prototype system that uses an image or a video sequence of an interested identity as probe. The system is based on face vision techniques including face detection and tracking, face alignment and recognition. Given a film or TV sitcom, first face trajectories are extracted in video by head tracking that decompose the video into segments corresponding to certain identity, then frames containing faces of higher quality are selected and normalized according to face alignment results, and finally different segments are associated by face recognition. Experiments are carried out on news video, feature length film video and TV sitcom to show its effectiveness. Potential usage of our system includes intelligent DVD/VCD browsing, video database retrieval, meeting record browsing, etc.

Categories and Subject Descriptors
I.5.4 [**Pattern Recognition**]: Applications – Computer vision; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis – *Object recognition*

General Terms
Algorithms, Design, Experiment

Keywords
Video content retrieval, video parsing, face recognition, face vision.

1. INTRODUCTION
After over a decade intensive research in computer vision society, face vision has been somewhat matured that makes it possible to apply face vision to video analysis and content extraction. In this paper, we discuss the problem of parsing the video based on face information facilitated by face detection [10], face and head tracking [9] and face alignment [22] tools. First face trajectories are extracted in video that decompose the video into segments corresponding to faces of certain identity. And then frames containing faces of higher quality are selected and normalized according to face alignment results. Finally different segments are associated by face recognition.

CIVR'07, July 9-11, 2007, Amsterdam, The Netherlands.

What we propose here is a fully automatic video retrieval prototype system. It uses an image or a video sequence of an interested identity as probe. Given a film or TV sitcom it retrieves shots containing this identity. Potential usage of our system includes intelligent DVD/VCD browsing, video database retrieval, meeting record browsing, etc.

The challenges in face recognition are that faces of a same person under different poses, expressions and illuminations differ a lot. Further in videos like films or television sitcoms, problems become even more complicated as they usually contain occlusions, motion blurs and videos are usually in lower resolution than still images. While in films or TV sitcom the retrieval task is a little easier as the gallery set usually consists tens of people which is not quite large and people themselves don't change a lot in a video.

The system we propose consists of three parts: (i) Face trajectories are extracted by a multi-state particle filter which decompose the video into segments corresponding to certain identities. (ii) Face alignment based on local texture classifiers are done on each frames. Frames with large alignment confidence are believed to have high quality i.e. frontal, non occlusion, etc. Thus those frames are selected as representatives of a segment. (iii) Shape free textures are obtained by warping faces to a referenced mean shape to remove pose and expression variation. Multi Feature MRC-Boosting classifiers are trained to measure distance between those textures.

1.1 RELATED WORKS
One approach to handle pose and illumination issue in video is building a 3D face model. The 3D point distribution data is usually acquired indirectly via stereo vision [4] or directly from laser scanner [1]. Then a parametric 3D deformable model consists of a mean 3D shape and a set of linear deformations of the shape derived by principal component analysis. A probe image is fitted to the model via analysis by synthesis method.

The other approach is to treat a person's appearance under different conditions as a manifold. In [1] images of a person in the same shot are modeled as a linear subspace obtained by PCA. Those manifolds are clustered according to a distance measure of Constraint Mutual Subspace method (CMSM). In [1] non-linear manifolds are formed as Gaussian Mixture Model (GMM). In [6] stable locally linear manifold patches are found using Mixtures of Probabilistic PCA (PPCA) and use boosting to learn a distance measure named Boosted Manifold Principal Angles (BoMPA).

3D model based methods usually achieves good recognition results especially on faces with large variance in position and

illumination. However, in order to build a 3D model they require 3D scanned face points or images taken at different poses. There is not a publicly available 3D dataset yet and meanwhile the largely available 2D image datasets cannot be directly utilized in those methods. Another disadvantage of 3D methods is that fitting a probe image to a 3D model usually involves an analysis by synthesis procedure which is computationally expensive. Manifold based methods are somewhat data dependent and needs careful tuning of parameters.

Our approach is a compromise between manifold based method and pure image based method since it employed a head tracker for extracting character trajectories and hence measures distance on the video sequence level. According to face alignment results frames contain non-frontal face or occlusions are automatically skipped while with head tracking frames before and after those frames are still associated. So it outperforms pure image based algorithm and is more robust than manifold based method.

Compared to video based approaches, face recognition algorithms on still images especially for frontal faces have been studied for much longer time and thus more mature. In our system, we select one of those algorithms as a basic distance measure. For a close examination, see the following discussions.

There are two different types of models, generative models for similarity or distance measure, and discriminative models for classification. Generative models include subspace models [11] such as Eigenface, Fisherface, ICA etc., graph models such as Elastic Bunch Graph Matching (EBGM) [14], while discriminative models include Bayesian model [11], other classification models such as MRC Boosting [19], Gabor AdaBoosting [21] etc.

Recently, discriminative models attract more attentions. As a representative approach of discriminative models, the Bayesian method outperforms many conventional methods. MRC-Boosting, which is a work following the Bayesian intra/extra personal difference framework, further treats this problem as a so-called target detection, where a target class should be separated from the surrounding clutter class. In each boosting iteration, it obtains a weak classifier through finding a projection vector by minimizing the target scatter matrix and maximizing the clutter scatter matrix like in LDA method. Without involving searching a large pool of candidate weak classifiers, it is very computational effective.

In our system, we select MRC-Boosting for face recognition [19] and make an extension to this method called Multi-Feature (MF) MRC-Boosting [7] in which we use wavelet features in MRC-Boosting for face recognition and further in boosting learning we use confidence-rated domain partition based weak classifiers under the framework of Real AdaBoost [14] instead of its original binary ones under that of Discrete AdaBoost.

A work similar to ours is that in [16] the author also employed a tracker to extract face sets. However, they use part based generative model to describe a face set which is different from our discriminative model. The head tracker and face alignment algorithm also make our system distinctive from theirs. In [20] single feature MRC-Boosting has been applied to record meeting video analysis. Their system is pure image based method and requires a certain amount of human interaction for labeling face positions.

The rest of this paper is organized as follows: Section 2 presents the proposed prototype system. Section 3 is about the head tracker and face alignment algorithm. Section 4 introduces the multi feature MRC-Boosting algorithm for distance measurement. Experimental results on a news video, a feature length film 'City of Angels', a TV sitcom 'Friends' are reported in section 5. And finally Section 6 draws the conclusion.

2. SYSTEM FRAMEWORK

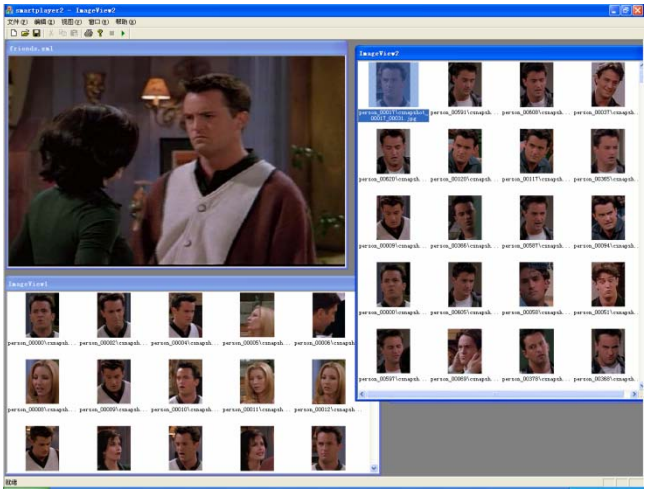


Figure 1 Interface of our retrieval system

Figure 1 demonstrates the interface of our retrieval system. The system works as follows: Given a video file, it automatically decomposes the video into segments corresponding to different individuals. The extracted sequences are listed as thumbnails in the 'Probe Window' (lower left sub window in figure1). Select a sequence from the 'Probe Window', all sequences of this individual are retrieved and shown in the 'Result Window' (right sub window in figure1).

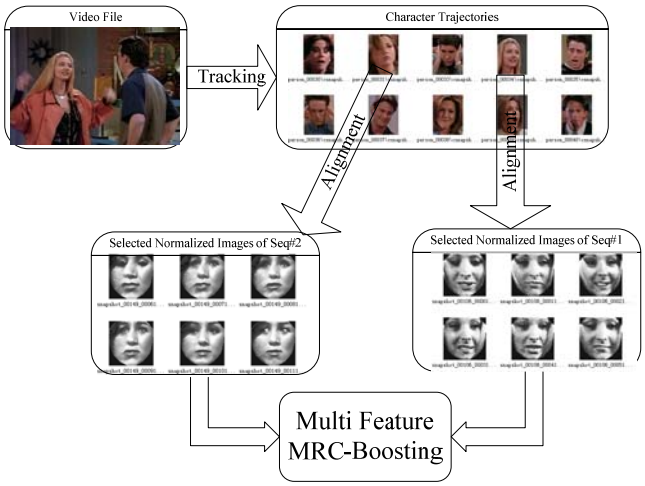


Figure 2 Framework of our retrieval system

The framework of our system is shown in figure 2. A video is analyzed in following steps:

By head tracking [9] a video file V is decomposed into several sequences q_i. Each sequence is a set of head images I_{ij} corresponding to the same individual.

$$\text{Track}(V) = \{q_i, i = 1...n\}, q_i = \{I_{ij}\} \quad (1)$$

Face alignment [22] is done on each head image which gives a face shape s_{ij} together with a confidence c_{ij}.

$$\text{Align}(I_{ij}) = (s_{ij}, c_{ij}) \quad (2)$$

According to the alignment confidence a subset $\{k_{i,1}, k_{i,2}, \cdots, k_{i,m_i}\}$ is selected from each q_i to remove hard faces. Then shape free texture t_{ij} is obtained by warping the representatives to a mean shape.

$$\text{Warp}(I_{i,k_{ij}}, s_{i,k_{ij}}) = t_{ij} \quad (3)$$

Similarities between face textures are computed according to a classifier obtained by MF MRC-Boosting. Then sequence similarity $S(q_i, q_j)$ is computed by a matching between representative faces.

3. HEAD TRACKING AND FACE ALIGNMENT

3.1 HEAD TRACKING

We use the head tracking algorithm proposed in [9] to extract face trajectories in video that decompose the video into segments corresponding to certain identity. Head tracker is superior to face tracker in capturing the trajectories of certain identity since in many cases although face may be invisible yet head still can be discriminated that makes face can be tracked over longer segments via head tracking. With the development of multiview face detection algorithm [5] [10], it can be integrated in head tracking as a powerful observation cue. The head tracker we used [9] utilizes a such state-of-the-art face detector [10] together with two general image cues, under the framework of multi-state particle filter. For detail, please see [9].

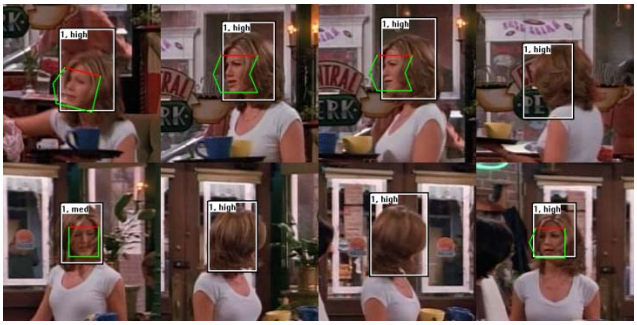


Figure 3 Tracking result

3.2 FACE ALIGNMENT

We use the face alignment algorithm proposed in [22] to extract face shape that consists of 88 landmark points. This algorithm differs from the classical ASM method in that it uses boosted local texture classifiers as local texture models rather than the conventional descriptive PCA models, in which classifiers as discriminative local texture models are trained over large data sets that greatly improves the accuracy and robustness of the ASM method. For detail, please see [22].

To remove small angle pose variations and expression variations, we warp faces to a reference mean shape. First a triangulation is formed from the mean shape of frontal view faces. Then according to the face shape extracted, each face is warped by affine transformations between each pair of corresponding triangles. This procedure effectively removes small angle pose variations. This procedure is demonstrated below in Figure 4.

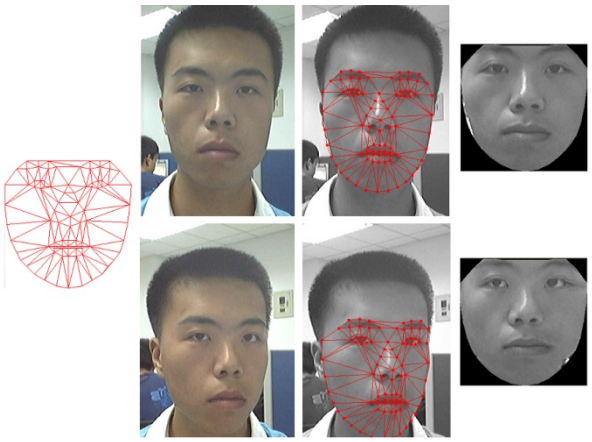


Figure 4 Extracting shape free textures (from left to right: triangulated mean shape, sample faces, aligned shapes of sample faces, normalized shape free textures, respectively)

3.3 FRAME SELECTION BASED ON ALIGNMENT CONFIDENCE

As described in Section 2, we select representative faces from a tracked sequence according to face alignment confidence. The confidence *c* is a real number between 0 and 1. When factors like occlusion, motion blur, or large angle pose variation presents the confidence value will be close to 0, otherwise *c* will be close to 1.

We assume that in most sequences there exists at least one frontal view face. So we select top N faces from all the faces with a confidence above a threshold value α. Figure 5 demonstrates this process.

4. MULTI FEATURE MRC-BOOSTING

We propose an extension of MRC-Boosting algorithm in [7]. As stated in [7], MF MRC-Boosting can effectively handle lighting and expression variations. So it is used as an image similarity measure in the video retrieval system.

MRC-Boosting was proposed by Xu, et al in [19]. Given a particular feature space, in each boosting iteration it tries to find the most discriminative feature by computing the projection vector which minimizes intra class difference scatter matrix while maximizes extra class difference scatter matrix. In each iteration, given samples $x_1, \ldots, x_n$ with weights carried by differences of sample i, j, $D_t(i, j)$ the projection vector is computed as follows

Figure 5 Frame selection. (The first six images in each row are key frames from tracked trajectories. The green curve indicates the alignment confidence of each image. The last four images are first two representative faces and extracted shape free textures.)

$$S_I = \sum_{i,j:y_i=y_j} D_t(i,j)(x_i - x_j)(x_i - x_j)^T$$
$$S_E = \sum_{i,j:y_i\neq y_j} D_t(i,j)(x_i - x_j)(x_i - x_j)^T \quad (4)$$

$$w = \arg\max_w \frac{w^T S_E w}{w^T S_I w} \quad (5)$$

4.1 REAL ADABOOST with LUT WEAK CLASSIFIERS

The original MRC-Boosting algorithm employed a Discrete AdaBoost in which weak classifiers are threshold functions with Boolean valued output. However, Real AdaBoost [14] whose weak classifiers are real-valued confidence rated function can give more accurate predictions and hence faster convergence. So here we use Real AdaBoost algorithm with LUT-type domain partition weak classifiers [18].

Usually a LUT based classifier partitions the feature domain uniformly [18], as there is no knowledge of a priori probability distribution function (pdf). Whereas in MRC-Boosting, since the differences are more or less centralized distributed, so we should partition the bins of LUT according to the specific type of distribution in way of using small bins at the dense part of its feature domain and large bins at the sparse part. This way of domain partition provides better approximation of sample distribution. One example is given in Figure 6.

Let $f_I(x)$ and $f_E(x)$ be the pdf of two kinds of differences,

$f(x) = \frac{1}{2}(f_I(x) + f_E(x))$ be the total pdf. Partition the feature space into p bins:

$B_0 = [c_0, c_1), B_1 = [c_1, c_2), \ldots, B_{p-1} = [c_{p-1}, +\infty)$, where

$c_0, \ldots, c_p$ are chosen to satisfy

$$c_0 = 0, c_p = \sup\{x : f(x) > 0\}$$
$$\int_{B_i} f(x)\,dx = \frac{1}{p}, i = 0, \cdots, p-1 \quad (6)$$

The sum of sample weights in each bin is:

$$W_I^i = \int_{B_i} f_I(x)\,dx, W_E^i = \int_{B_i} f_E(x)\,dx, i = 0, \ldots, p-1 \quad (7)$$

According to [14] the (smoothed) optimal weak classifier is

$$h(x) = \sum_{i=0}^{p-1} \frac{1}{2} \ln \frac{W_I^i + \varepsilon}{W_E^i + \varepsilon} I_i(x)$$
$$I_i(x) = \begin{cases} 1, x \in B_i \\ 0, x \notin B_i \end{cases} \quad (8)$$

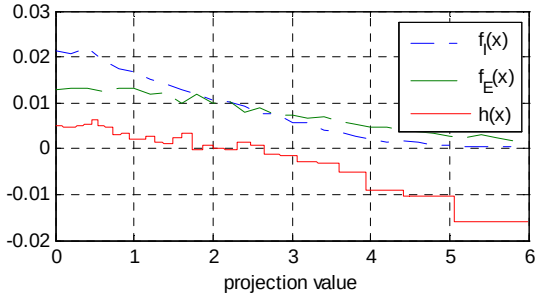


Figure 6 A LUT-type weak classifier with adaptive domain partition

4.2 EMPLOY WAVELET FEATURES BY MF MRC-BOOSTING

Instead of a single gray scale image, we use Gabor wavelet at five scales and eight orientations to represent a face. Thus each face is represented by 80 images (both magnitude and phase image are utilized); each image constitutes a specific feature in the feature

space of a particular scale and orientation. In order to take all these 80 feature spaces into consideration, we propose to extend the MRC-Boosting algorithm as follows: in each boosting iteration, the optimal projection direction for every feature space is computed, and among all the 80 corresponding features selected, the algorithm chooses the feature on which the optimal classifier has the smallest normalization error Z.

So the optimal weak classifier is computed analytically inside a feature space like MRC-Boosting; while statistically among different feature spaces like AdaBoost. Hence the algorithm inherits both MRC-Boosting's efficiency in computation and AdaBoost's adaptability of the training samples.

We call this algorithm Multi-Feature MRC-Boosting (MF MRC-Boosting), for details, see Figure 7.

Input

$S = \{(X_1, y_1), \ldots, (X_m, y_m)\}$

where $X_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,n}), x_{i,j} \in \Re^{D_j}, y_i \in \aleph$

i.e. training dataset with m samples, every sample contains n features and an identity

Compute

1) Initialize $D_1(i,j) = \begin{cases} 1/2N_I, y_i = y_j \\ 1/2N_E, y_i \neq y_j \end{cases}$ where N_I and N_E a re the number of intra difference and extra difference respectively.

2) For $t = 1, \ldots, T$

   a) For $p = 1, \ldots, n$

      i. Under distribution D_t, on feature p compute the scatter matrix S_I,S_E and the optimal projection vector $w_{t,p}$, via (4)(5)

      ii. Obtain a weak classifier $h_{t,p}(\Delta)$ on projection space $\Delta_{ij} = (x_{i,p} - x_{j,p}) \cdot w_{t,p}$ with Real AdaBoost via (6)(7)(8)

      iii. Compute normalization factors

$$Z^I_{t,p} = \sum_{i,j:y_i = y_j} D_t(i,j) \exp\left(-h_{t,p}(\Delta_{ij})\right)$$

$$Z^E_{t,p} = \sum_{i,j:y_i \neq y_j} D_t(i,j) \exp\left(h_{t,p}(\Delta_{ij})\right)$$

$$Z_{t,p} = Z^I_{t,p} + Z^E_{t,p}$$

   b) Find the weak classifier with the minimal normalization factor $k_t = \arg\min_p Z_{t,p}$ and let

$$w_t \leftarrow w_{t,k_t}, h_t \leftarrow h_{t,k_t}$$

   c) Update sample weights[1]

$$D_{t+1}(i,j) = \begin{cases} \dfrac{D_t(i,j) \exp\left(-h_t\left(\left(x_{i,k_t} - x_{j,k_t}\right) \cdot w_t\right)\right)}{Z^I_{t,k_t}}, y_i = y_j \\ \dfrac{D_t(i,j) \exp\left(h_t\left(\left(x_{i,k_t} - x_{j,k_t}\right) \cdot w_t\right)\right)}{Z^E_{t,k_t}}, y_i \neq y_j \end{cases}$$

Output

A strong classifier

$$H(X_i, X_j) = sign\left[\sum_{t=1}^{T} h_t\left(\left(x_{i,k_t} - x_{j,k_t}\right) \cdot w_t\right)\right]$$

Figure 7 Multi Feature MRC-Boosting algorithm

5. EXPERIMENTS

We evaluate the proposed system on three typical types of videos: news videos, feature length films and TV sitcom. A video is first decomposed into sequences by head tracking. Sequences are manually labeled with identity. For each main character, use one sequence as probe to retrieve other sequences from all decomposed sequences. Then precision/recall curve is computed following the standard definition.

Since face recognition is performed on frontal view faces, only sequences contain at least one frontal view face are considered. And the tracker is set to only track faces of size at least 48 by 48 pixels.

5.1 TRAINING

We train the MF MRC-Boosting algorithm on CMU-PIE dataset [15] to obtain a general recognition model. CMU-PIE dataset contains more than 40,000 images of 68 individuals. Images of the same individual comprise large variations in lighting condition, expression and poses. So we use this dataset as an approximation of faces in videos.

We select 90 images of each individual, 60 for training, 25 for probe and 5 for gallery. Images are warped to a reference mean shape as described in Section 3.2 then resized to 64 by 64. Gabor wavelets of five scales and eight orientations are extracted. Both phase and magnitude components are down sampled to 16 by 16. The training procedure converges at about 300 weak classifiers and the classifier we used is composed of 500 weak classifiers. On test set, the rank-1 recognition accuracy reaches 98%.

5.2 RECOGNITION

Sequence similarity is computed as weighted average similarities between matched images.

[1] The weights are normalized with different normalization factors here to balance the impact of intro and extra personal differences.

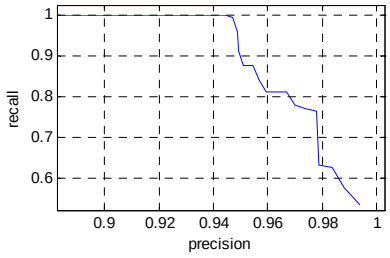
$$S(q_a, q_b) = \frac{\sum_i c_{a,i} c_{b,k_i} \mathrm{H}(t_{a,i}, t_{b,k_i}) + \sum_j c_{a,l_j} c_{b,j} \mathrm{H}(t_{a,l_j}, t_{b,j})}{\sum_i c_{a,i} c_{b,k_i} + \sum_j c_{a,l_j} c_{b,j}} \quad (9)$$

$$k_i = \arg\max_k \mathrm{H}(t_{a,i}, t_{a,k}), l_j = \arg\max_l \mathrm{H}(t_{a,l}, t_{b,j})$$

5.3 NEWS VIDEO

Our first experiment is performed on the TREC-VID video *20051213_145800_CCTV_DAILY_CHN.mpg* (Figure 8a) [23]. This is a news video so only the announcer is considered to be the main character.

(a)



(b)

Figure 8 Results on news video (a) sample faces from the video (announcer is labeled with red rectangle) (b) Precision-Recall curve

The video is decomposed into 226 sequences, which constitutes of 10 incorrectly tracked sequences, 29 sequences don't contain any frontal-view face, 17 sequences of the announcer and 170 sequences of other individuals. The precision recall curve is shown in Figure 8b. The accuracy is high because the announcer is always appeared in fontal-view, good lighting conditions and no occlusion.

5.4 FILM

Our second experiment is on the film 'City of Angels'. We use first one hour of this film. There are two main characters, Seth and Dr. Maggie Rice.

The tracker gives 800 sequences including 30 incorrectly tracked sequences, 101 sequences don't contain any frontal-view face, 200 sequences of Seth, 267 of Dr. Maggie Rice and 202 of other characters.

Some of the retrieval results are shown in Figure 9. Each sequence is represented by a frame with medium alignment confidence.

Figure 9 Example retrieval results on 'City of Angels'. Each row contains rank-0(query image), 5, 10, 15…35 sequences.

The quantitative results are shown as blue curves in Figure 10.

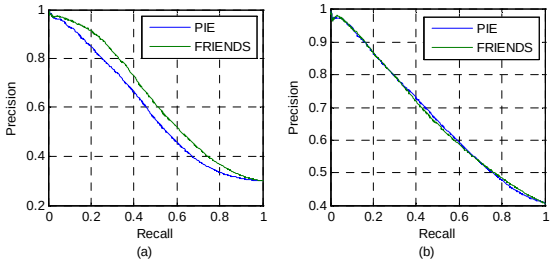


Figure 10 Precision/recall curves for (a) Seth (b) Dr. Maggie Rice

5.5 TV SITCOM

The third experiment is on episode 'The One Where Ross Finds Out' of TV sitcom 'Friends'. There are six main characters, Joey, Chandler, Ross, Rachel, Phoebe and Monica.

The tracker gives 622 sequences including 21 incorrectly tracked sequences, 97 sequences don't contain frontal-view face, 74 sequences of Chandler, 88 of Ross, 41 of Phoebe, 51 of Monica, 128 of Rachel, 43 of Joey and 79 of other characters.

Example retrieval results are shown in Figure 11. Precision/recall curves are shown in blue in Figure 13. The performance is not very satisfactory compared to previous two experiments. This might because in 'Friends' the intra personal differences are more complex and thus are not properly covered by the CMU-PIE training dataset we used.

Figure 11 Example retrieval results on 'Friends' (recognition model is trained on CMU-PIE). Each row contains rank-0(query image), 5, 10, 15…35 sequences.

Figure 12 Example retrieval results on 'Friends' (recognition model is trained on another episode of 'Friends'). Each row contains rank-0(query image), 5, 10, 15…35 sequences.

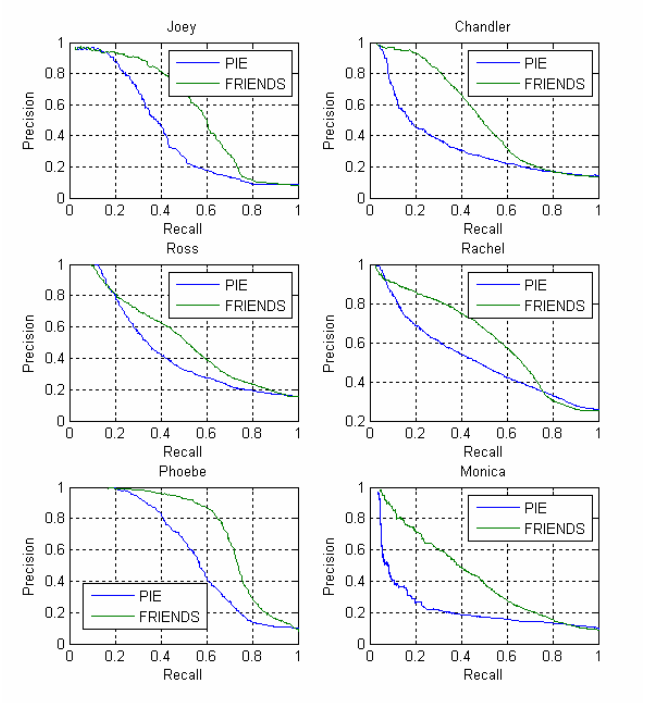


Figure 13 Precision/recall curves for 'Friends'

5.5.1 ADDITIONAL TRAINING

We further labeled the ground-truth of an adjacent episode 'The One with the Baby on the Bus', as additional training data. The new training set is composed of 2126 images of the six afore mentioned characters from this episode as well as 2000 images from CMU-PIE dataset. As described in Section 5.1, we trained a classifier composed of 500 weak classifiers.

The new example retrieval results are shown in Figure 12. Precision/recall curves are shown in green in Figure 13. We can see that the accuracy is significantly improved. This improvement may come from two factors: the learning procedure gains some general knowledge about the sample distribution in video; and some specific knowledge about those six characters in those two episodes. The former case means that this is a proper augmentation to the training set while the latter case may lead to an over fit.

To see which the main factor is, we evaluate the new model on the film 'City of Angels'. All conditions are same as in Section 5.4 and the new precision/recall curves are shown as green curves in Figure 10. The little improvement of accuracy verifies that the additional training data brings mainly general knowledge and little specific knowledge. This also indicates that given more proper training data the retrieval accuracy in a general video may be further improved.

6. CONCLUSIONS

We have proposed a fully automatic video retrieval system. It integrates the state-of-the-art head tracking, face alignment and face recognition algorithms.

In our system, head tracking algorithm links frames containing not so good faces to adjacent frames containing good faces. According to face alignment confidence we effectively select

good faces as representatives for recognition. The MF MRC-Boosting algorithm we use successfully handles intra personal lighting and expression variations. Satisfactory experimental results are achieved on news videos, films and TV sitcom.

7. ACKNOWLEDGMENTS

This work is supported in part by National Science Foundation of China under grant No.60332010, No.60673107, National Basic Research Program of China under grant No.2006CB303100, and it is also supported by a grant from Intel Corporation.

8. REFERENCES

[1] O Arandjelovic, G Shakhnarovich, et al. Face recognition with image sets using manifold density divergence, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 pp. 581-588

[2] Ognjen Arandjelovic, Roberto Cipolla. Automatic Cast Listing in Feature-Length Films with Anisotropic Manifold Space, 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06) pp. 1513-1520

[3] Zhao, Chellappa, Rosenfeld & Phillips, Face Recognition: A Literature Survey, UMD CS-TR-4167, 2000

[4] Mark Everingham and Andrew Zisserman, Identifying Individuals in Video by Combining Generative' and Discriminative Head Models, ICCV 2005.

[5] M. Jones and P. Viola, Fast Multi-View Face Detection. MERL-TR2003-96, July 2003.

[6] Tae-Kyun Kim, Ognjen Arandjelovic and Roberto Cipolla, Boosted Manifold Principal Angles for Image Set-Based Recognition, Pattern Recognition, accepted for publication conditioned on minor revision, 2006.

[7] Pengxu Li, Haizhou Ai, Face Recognition Using Wavelet Features via MRC Boosting. Submitted to 2nd International Conference on Biometrics (ICB2007), Seoul, Korea, August 27-29, 2007

[8] Y. Li, S. Gong, and H. Liddell. Constructing facial identity surfaces for recognition. International Journal of Computer Vision, 53(1):71–92, 2003.

[9] Yuan Li, Haizhou Ai, , et.al, Robust Head Tracking Based on a Multi-State Particle Filter, 7th IEEE International Conference, Automatic Face and Gesture Recognition, AFG2006, pp.335-340, Southampton, UK, April 10-12 2006.

[10] Chang Huang, Haizhou Ai, et.al, Vector Boosting for Rotation Invariant Multi-View Face Detection, The IEEE International Conference on Computer Vision (ICCV-05), pp.446-453, Beijing, China, Oct 17-20, 2005.

[11] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian Face Recognition. Pattern Recognition, vol.33, no.11, November 2000

[12] B. Moghaddam，Face Recognition by Humans and Machines，A Tutorial Survey，CVPR'01 Short Course，CVPR2001, Dec. 2001.

[13] S. Satoh, Y. Nakamura, and T. Kanade, Name-It: Naming and Detecting Faces in News Videos, IEEE MultiMedia, 1999, 6(1):22-35.

[14] R. E. Schapire, Y. Singer. Improved boosting algorithms using confidence-rated predictions. Machine Learning, 1999, 37(3). 297-336.

[15] T. Sim, S. Baker, and M. Bsat, The CMU Pose, Illumination, and Expression (PIE) Database. Technical Report CMU-RI-TR-01-02, 2001.

[16] Josef Sivic, Mark Everingham, and Andrew Zisserman, Person spotting: video shot retrieval for face sets. International Conference on Image and Video Retrieval (CIVR 2005), Singapore 2005

[17] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face Recognition by Elastic Bunch Graph Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, no.7, 1997.

[18] Bo Wu, Haizhou Ai, Chang Huang, LUT-Based AdaBoost for Gender Classification, In LNCS, Vol.2688, pp.104-110, Springer-Verlag, 2003.

[19] Xun Xu and Thomas S. Huang, Face Recognition with MRC-Boosting, ICCV 2005, Beijing, China.

[20] Xun Xu Yong Rui Huang, T.S Recognizing Faces in Recorded Meetings via MRC-Boosting, Multimedia and Expo, 2006 IEEE International Conference on, Toronto, ON, Canada

[21] Lei Zhang, Stan Z. Li, et.al. Boosting Local Feature Based Classifiers for Face Recognition. First IEEE Workshop on Face Processing in Video. 2004, Washington, USA.

[22] Li Zhang, Haizhou Ai, et.al, Robust Face Alignment Based on Local Texture Classifiers, The IEEE International Conference on Image Processing (ICIP-05), Genoa, Italy, September 11-14, 2005.

[23] http://www-nlpir.nist.gov/projects/trecvid