Real-time Embedded Face Recognition for Smart Home

Fei Zuo, Student Member, IEEE, Peter H. N. de With, Senior Member, IEEE

Abstract — *We propose a near real-time face recognition system for embedding in consumer applications. The system is embedded in a networked home environment and enables personalized services by automatic identification of users. The aim of our research is to design and build a face recognition system that is robust for natural consumer environments and can be executed on low-cost hardware. For enabling distribution of computations, we propose a processing pipeline for face recognition, which consists of (1) face detection by stepwise pruning; (2) coarse-to-fine facial feature extraction for face normalization; and (3) face identification by cascaded discriminant analysis. The system has been applied in varying environments, such as an experimental home network, and achieves over 95% recognition rate and 3-4 frames/s processing speed*[1].

Index Terms — Biometrics, face detection, face recognition, facial feature extraction, smart home.

I. Introduction

DURING the past few years, the demand for 'smart' devices in consumer electronics is increasing. This is motivated by e.g. the widespread use of low-cost embedded electronics in all kinds of environments, and the need for systems that understand the user such that easy interfaces can be designed. Consequently, it is desirable that electronic devices can consciously sense and understand their surroundings and adapt their services according to the contexts (e.g. environments). One important aspect in this context-awareness is the user identification, in which machines can actively identify users present in their environments and customize services accordingly. Face recognition provides a natural visual interface for such context-aware applications, endowing devices perceptual abilities to communicate naturally with people, e.g. to recognize people and remember their preferences and peculiarities [1].

A. Face recognition: perceptual interface for home devices

Face recognition can be used as a natural interface to facilitate devices to be aware of their surroundings and especially their customers. Compared with other existing identification technologies such as fingerprint and iris recognition, face recognition has several characteristics that are advantageous for consumer applications.

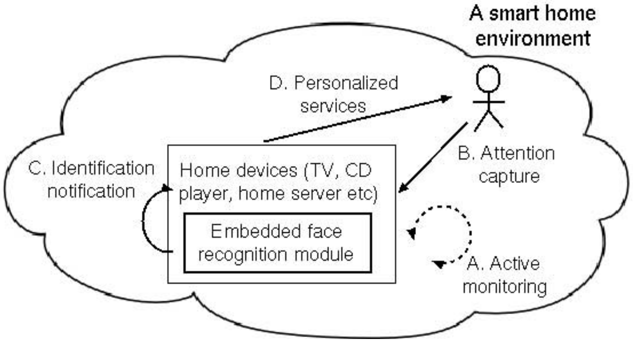

A smart home environment
D. Personalized services
B. Attention capture
C. Identification notification
Home devices (TV, CD player, home server etc)
Embedded face recognition module
A. Active monitoring

Fig. 1. Face recognition embedded in a smart home environment.

1. ***Nonintrusive and user-friendly interfaces***: In traditional fingerprint and iris identification, the users are required to undergo a 'pause-and-declare' procedure to authenticate themselves [1]. This can cause great inconveniences for users in consumer applications. Face recognition, on the other hand, provides a natural means for perceptual interface without special requirements for user actions.
2. ***Low-cost sensors and easy setup:*** Face recognition can make use of a wide range of inexpensive consumer cameras, such as DV cameras, web cameras and embedded cameras in mobile devices. No special setup is required and the identification procedure can be transparent to the user.
3. ***Active identification:*** In a smart home environment, a face recognition system can *actively* monitor its surroundings and identify users. This can be carried out without any human intervention or even awareness of the user.

In Fig. 1, we give an abstract description of the perceptual interface of home devices, which is employed in this paper. A face recognition module is embedded into home devices (e.g. televisions, video and music players), which continuously monitors its surroundings and captures active user presence. By automatic identification of the user's face, home devices can tailor services (e.g. selection of favorite TV or music programs) for the individual user.

B. Face recognition: the challenges

As stated in the previous subsection, face recognition in consumer applications offers several benefits compared with other biometrics, such as unobtrusiveness and wide applicability. This has aroused great interest in face recognition research in recent years [3]. However, although many face processing techniques have been proposed thus far, face recognition for 'real-life' applications (e.g. in a consumer environment) still remains to be a difficult problem. The

[1] This research was funded by the ITEA (Information Technology for European Advancement) project HomeNet2Run.

Fei Zuo is with Eindhoven University of Technology, Den Dolech 2, 5600MB, Eindhoven, The Netherlands (e-mail: F.Zuo@tue.nl).

Peter H. N. de With is with Eindhoven University of Technology, Signal Processing Systems Group. He is also with LogicaCMG Eindhoven, 5605JB Eindhoven, The Netherlands (e-mail: P.H.N.de.With@tue.nl).

Manuscript received January 7, 2005

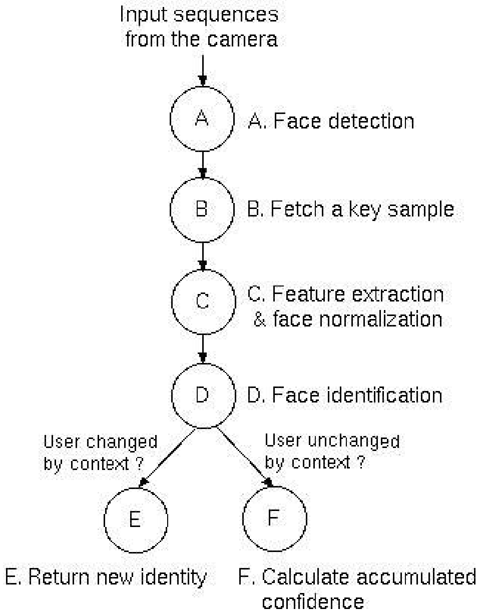

Input sequences from the camera
A
A. Face detection
B
B. Fetch a key sample
C
C. Feature extraction & face normalization
D
D. Face identification
User changed by context ?
User unchanged by context ?
E
F
E. Return new identity
F. Calculate accumulated confidence

Fig. 2. Face recognition operation flow.

relatively unconstrained operating environment imposes more challenges compared to conventional 'smug-shot' face recognition such as used in security applications. These extra challenges include the following:

1. Large variability of operating environments (e.g. illumination and backgrounds).
2. Processing efficiency with low-cost consumer hardware.
3. Nearly unconstrained capturing of facial images.

In this paper, we present a consumer-oriented face recognition system *HomeFace*, with special emphasis on the above challenges. The system is embedded in a smart home environment for user identification as illustrated in Fig. 1. More specifically, the experimental environment consists of a home network using e.g. Ethernet and IEEE-1394 connections. A prototype of our recognition system was integrated in this network and successfully demonstrated at the IFA consumer electronics exhibition [2].

The remainder of the paper is organized as follows. In Section II, we will give an overview of the *HomeFace* system architecture. The core of the system is a pipeline of three processing modules, namely, face detection, facial feature extraction for face normalization, and finally, face identification. Sections III-V illustrate each processing module in more detail. In Section VI, we provide the experimental results and analyze the performance of the complete system. Section VII presents discussions and conclusions.

II. Architecture Overview of *HomeFace*

The embedded *HomeFace* system consists of a kernel component performing the face detection/recognition and an interfacing component providing a uniform interface to different hosting devices. The kernel component employs a processing pipeline consisting of face detection (attention capture), feature extraction for face normalization (preprocessing for classification), and face identification.

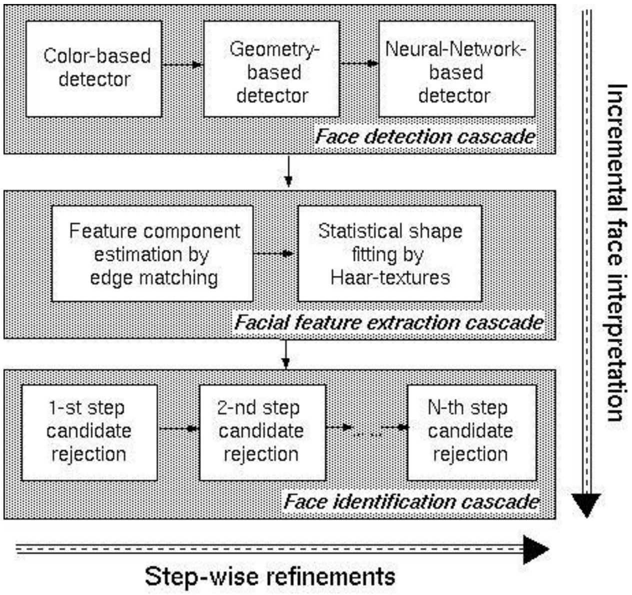

Color-based detector
Geometry-based detector
Neural-Network-based detector
Face detection cascade
Feature component estimation by edge matching
Statistical shape fitting by Haar-textures
Facial feature extraction cascade
1-st step candidate rejection
2-nd step candidate rejection
N-th step candidate rejection
Face identification cascade
Incremental face interpretation
Step-wise refinements

Fig. 3. Algorithm architecture.

A. Processing flow

Fig. 2 portrays the operation flow of the *HomeFace* system. The system operates on live-feed video from embedded cameras and performs the identification in near real time. For each frame, the system detects existing faces and keeps track of the located faces (Fig. 2, stage A). Motions of faces from the same user form a smooth trajectory, and a new user can be easily identified by detecting a 'break' of this trajectory. Note that in both cases, the subsequent facial feature extraction and face identification are performed. From all detected faces, those faces with large rotations (in-plane or out-of-plane rotation) or low resolutions are discarded. The remaining face images are selected as key samples for further processing (Fig. 2, stage B). To accommodate for varying face scale, position and pose, an accurate extraction of salient facial features is required for normalization purposes (Fig. 2, stage C).

The final face identification (Fig. 2, stage D) searches for a database of reference faces and returns the 'best-match'. For a user coming into the scene for the first time (Fig. 2, stage E), a minimum of *k* key samples need to be processed before an identification decision is made. For users already present in the scene, the identification result for the current frame will only contribute to the accumulated recognition confidence (Fig. 2, stage F). In this way, more stable identification can be achieved w.r.t. a sequence of face images.

The processing pipeline can be executed on one single processing unit (*the centralized mode*) or distributed over a number of processing units connected by a home network (*the distributed mode*). In the distributed mode, different processing tasks (e.g. face detection, facial feature extraction

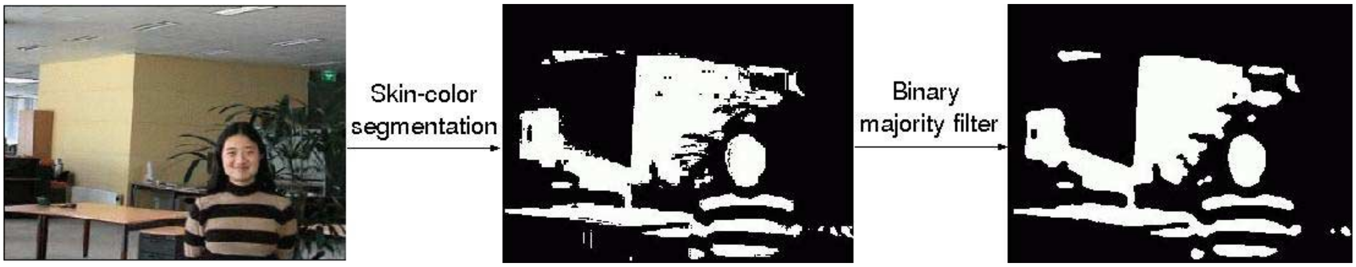


Fig. 4. The color-based face detector at the first stage.

and face identification) can be carried out at various locations in a home network. The task distribution enables the execution of the software on multiple low-end hardware platforms, thereby reducing the total cost. Furthermore, since the chain of processing tasks form a pipeline, the distribution of tasks can enable parallel processing and/or high processing efficiency.

B. Image processing algorithm architecture

Fig. 3 shows that three key steps are involved in the face processing pipeline, namely, face detection, facial feature extraction for face normalization and face identification. These key steps interpret human faces with incremental semantic levels. The three key processing steps sequentially answer the following questions: if a human face is present in the scene, where the key facial features are (e.g. eyes and mouth), and what identity the face represents. Note that the answer to each question is built up upon the understanding of the previous questions (see the vertical *incremental face interpretation* arrow in Fig. 3).

Orthogonally, within each processing step, a set of component algorithms are cascaded into a chain, where the outputs from the previous algorithms act as the inputs for the subsequent algorithms. The subsequent algorithms operate at finer granularity and further improve the results generated by the previous algorithms. By *step-wise refinements* (Fig. 3), high system performance can be achieved in terms of both robustness and processing efficiency.

In the following Sections (III to V), we present each stage in more detail.

III. Face Detection

Tremendous effort has been spent on designing face detection techniques in the past decade [4]. A variety of image features are adopted, such as color, edges, image pixels in the spatial domain and transformed image signals in DCT/Wavelet domain. Based on these features, various classifiers are built to discriminate face regions from non-face background regions. However, it is difficult for one single technique to achieve both robustness and training/processing efficiency simultaneously. For example, face detectors based on skin-color segmentation (e.g. [5]) are relatively fast, but color-features alone are insufficient for the accurate discrimination between faces and non-faces. Conversely, learning-based techniques (e.g. Support Vector Machine [6]) offer more robust detection performance, however, they usually involve high training/processing complexity, which can be too costly for consumer use.

To build a face detector that is highly efficient and robust, we use a detector cascade [7] that employs successive face detectors with incremental complexity and detection capability. The detectors are cascaded in such a way that each detector progressively restricts the possible face candidates into fewer areas (see Fig. 3), thereby improving the overall performance in a step-wise manner. The detector cascade consists of a skin-color-based face detector, a geometry-based face detector and a Neural-Network-based (NN) face detector. The detector cascade offers the following advantages:

1. Various image features are used including color, geometry and direct image pixels in the spatial domain. These features complement each other and capture different properties of the face images.
2. The initial detectors are capable of quickly pruning large areas of non-face background regions, which significantly reduces the number of candidate regions for the final learning-based detector. This largely reduces the overall computation cost, while still retaining high detection accuracy.

A. Color-based face detector

Color information has been proven to be an effective image feature for coarsely locating potential facial regions. We analyzed a sample set of facial pixels taken from pictures of people of different race, gender, age and under various lighting conditions. We have found that the chrominance components (C_b, C_r) in the YC_bC_r color space form a condensed cluster. We have fit a convex hull to the skin-color cluster in the $C_b - C_r$ plane. Skin-regions in an input image can be quickly determined by checking a look-up table (see Fig. 4). In addition to pixel-based color verification, we apply a binary majority filter as a post-processor to smooth the segmentation result.

Notice that the color pruning criterion used in this stage is quite relaxed, resulting in a higher false acceptance rate than a more intricate color-based detection approach. However, for an initial detector in the whole detector cascade, we intend to put more emphasis on speed than accuracy and the overall detection performance is reinforced by the succeeding detectors.

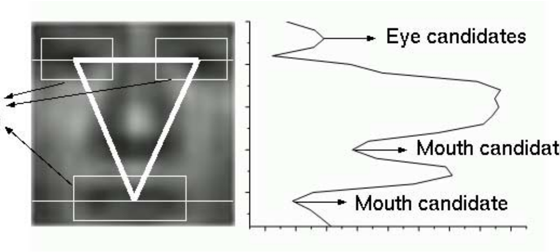


Fig. 5. The geometry-based face detector, left: sliding windows for eyes and mouth; right: vertical intensity profiles of the face window.

B. Geometry-based face detector

In order to select potential face areas present in the skin blobs acquired in the first detector, we enumerate all small windows (20×20 pixels) extracted from each skin blob at every position and scale. This is similar to the image pyramid architecture used by Rowley *et al.* [8]. For each window, we apply a geometry-based face structure detector which examines the triangle-relationship between eye and mouth regions. The algorithm of the geometry-based face detector is summarized as follows.

Algorithm 1: ***The geometry-based face verification.***

1. *Generate the vertical profile of the candidate region (Fig. 5 right side);*
2. *Select local minima of the profile as candidate vertical locations of eyes and mouth. If no proper minima are found, return non-face;*
3. *For each candidate vertical location, a sliding window is applied to search horizontally for the most probable eye-pair (or mouth). The average region intensity is used as a fast evaluation criterion. If the lowest average intensity is above a threshold, return non-face;*
4. *Check whether the selected feature group (eyes + mouth) forms an approximate equilateral triangle. If not, return non-face.*

C. Learning-based detector

For the final detector in the cascade, a neural-network-based (NN) detector is used. Its purpose is the final verification of facial regions and its principle is quite similar to Rowley's approach [8]. Since the previous detectors in the cascade already discard a large portion of background regions, only a small number of ambiguous candidates are passed to the NN detector. This significantly reduces the potential high computation cost of the NN detector. In the following, we provide quantitative results of our proposed face detector cascade.

D. Evaluation

To evaluate the performance of our proposed detector cascade, we use a test data set composed of color images downloaded from the Internet (most of them from *Yahoo News* photo gallery), randomly selected samples from the color FERET face database [9], and 10 self-recorded video sequences simulating consumer home environments. We obtained an average of 92% percent detection rate. For the test on video sequences, the detector achieves approximately 0.1 second/frame processing speed, with a frame resolution of 320×240 pixels, using a Pentium IV PC (1.7 GHz).

IV. Facial Feature Extraction and Face Normalization

The outputs from the face detector (see the previous subsection) contain faces with varying scale, position and pose. The direct use of unaligned facial images will potentially lead to identification failures. To this end, we need an accurate extraction of salient facial features for normalization purposes.

One earlier approach for feature extraction is proposed by Yuille *et al.* [10]. In this approach, parameterized deformable templates are used to minimize a predefined fitting function. However, it is too computationally expensive for consumer use. More recently, the Active Shape Model (ASM) and Active Appearance Model (AAM) [11] were proposed as two effective techniques for feature extraction. In those techniques, a constrained statistical shape model is adopted, which offers fast and stable performance. However, both ASM and AAM suffer from wrong convergence when the model is not initialized sufficiently close to the real face position.

We propose a two-step coarse-to-fine feature extractor [12] as shown in Fig. 3. First, we use Edge Orientation Matching (EOM) [13] for a fast determination of approximate facial feature positions within the detected face region. This ensures a 'close' initialization of the shape model used by the second step, alleviating the wrong convergence problem caused by improper model initializations. Second, H-ASM (an enhanced version of ASM) is adopted to iteratively fit a statistical shape model to the input image. Note that in contrast with ASM, we use 2-D Haar-wavelet-based texture attributes in H-ASM to guide the deformation process, which facilitates increased robustness and fast processing.

A. Feature estimation by EOM

Since prominent facial features like eyes and mouth are often characterized by relatively pronounced edges, edge information (with both strength and orientation) proved to be effective in face processing [13]. The edge map of a given image region is obtained by convolving the image with a horizontal and vertical 3×3 Sobel edge filter. Based on the convolution results ($S_X(x,y)$ and $S_Y(x,y)$), the Edge-Strength image (ES) and the Edge-Orientation image (EO) are generated, given by

$$\begin{aligned} ES(x,y) &= \sqrt{S_X^2(x,y)+S_Y^2(x,y)}, \\ EO(x,y) &= \arctan\frac{S_Y(x,y)}{S_X(x,y)}. \end{aligned} \tag{1}$$

Using the average ES and EO as a template, a multi-resolution search is performed over the detected facial region for the position and scale yielding the best match (Fig. 6). The matching function between two image regions P_1 and P_2 is defined as [13]:

$$M(P_1,P_2)=\sum_x\sum_y d_e(x,y), \text{ and} \tag{2}$$

$$d_e(x,y)=\begin{cases}\sin(EO_1(x,y)-EO_2(x,y)),\\ \quad \text{if } ES_1(x,y),\ ES_2(x,y)>T;\\ 1, \quad \text{otherwise,}\end{cases} \tag{3}$$

where ES_i (EO_i) represents the ES (EO) image of P_i ($i=1,2$), respectively. The parameter T is a threshold.

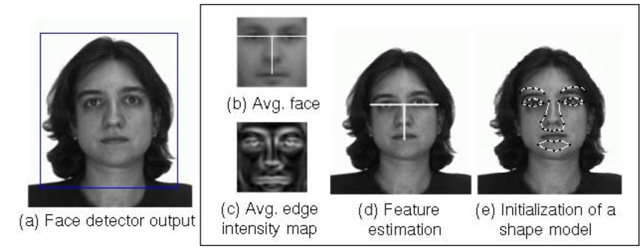

(a) Face detector output
(b) Avg. face
(c) Avg. edge intensity map
(d) Feature estimation
(e) Initialization of a shape model

Fig. 6. Feature estimation by EOM.

B. Deformable shape fitting by H-ASM

1) Facial feature model with enhanced Haar textures

In this stage, we build a facial feature model as an ordered set of N_f feature points, as shown in Fig. 7. For each feature point F_i ($1\le i\le N_f$), we define an attribute pair $\langle\vec{P}_i,\vec{T}_i\rangle$, where $\vec{P}_i$ specifies the coordinates of F_i and $\vec{T}_i$ describes the local 2-D texture structure of F_i. We employ 2-D local textures because they contain richer and more reliable pattern information than 1-D profiles normal to the contour. The latter is widely used in gradient-based deformable models, such as ASM. In our case, we model $\vec{T}_i$ by extracting an $M\times M$ block around F_i from the image, which is subsequently transformed using the Haar-transform for robustness and processing efficiency. A closer examination of the local feature patterns in face images shows that they usually contain relatively simple patterns with strong contrast. The 2-D basis images of the Haar-transform match quite well with these patterns, so that it is attractive to exploit them for efficient signal representation. Furthermore, the simplicity of the Haar-transform supports the requirement of a real-time implementation.

2) Fast computation of Haar textures

Haar decomposition mainly involves summations of pixel sub-blocks (see the example Haar basis images shown in Fig. 7), which can be efficiently computed by using two auxiliary 'integral images' [14]. Furthermore, the illumination normalization for each block (by zero mean and one standard deviation) can be conveniently integrated into the computation of Haar coefficients. Interested readers are referred to [12] for a more detailed explanation of the procedure for the fast computation of Haar coefficients.

3) Haar feature selection

From the above, each local feature attribute vector $\vec{T}_i$ can be represented by a transformed data vector

$$\vec{T}_i=(H_2(F_i),H_3(F_i),...,H_c(F_i))^T, \tag{4}$$

where $H_k(F_i)$ gives the k-th (in zigzag order) Haar coefficient[2] of the $M\times M$ block surrounding F_i. Only a part, say c where c □ $M\times M$, of the total coefficients are selected, due to the highly compact description of the local feature structure after Haar transformation. For most feature points, less than 4% of the total coefficients are required to retain up to 95% of the signal energy. Note that c can be variable to adapt to different feature structures. This significantly increases the efficiency of the algorithm and reduces the interference of the image noise.

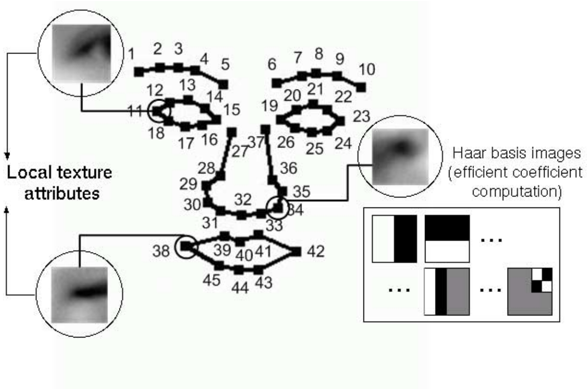


Fig. 7. Topological shape with local textures modelled by Haar wavelets.

4) Haar feature weighting

The selected coefficients $H_k(F_i)$ ($1\le k\le c$) may not be of equal importance to characterize F_i. Coefficients with large variances for different face images may contain 'noisy' features and should be suppressed. Therefore, a *feature weighting scheme* is adopted, which assigns more weights (higher importance) to those coefficients with less variance over a set of training samples. To this end, Equation (4) is rewritten to:

$$\vec{T}_i=(H_2(F_i)/\sigma_2,H_3(F_i)/\sigma_3,...,H_c(F_i)/\sigma_c)^T, \tag{5}$$

where σ_k ($1\le k\le c$) represents the standard deviation of $H_k(F_i)$. In this sense, more stable coefficients are emphasized, which exert more influence in the subsequent feature matching (see the section below).

5) Feature extraction by H-ASM

From the initial estimation of the feature position, an initial shape model can be overlayed to the real image (see Fig. 6(e)). To extract N_f key feature points, the shape model is deformed to fit to the real image in an alternating two-step procedure.

[2] Note that $H_1(F_i)=0$, due to the illumination normalization.

1. For each feature point F_i, a local optimal position is searched along a trace composed of eight radial lines originating from the current model point. The search is based on evaluating the Euclidean distance between the current local feature attribute $\vec{T}_i$ (Equation (5)) and an average $\overline{\overline{T}}_i$ computed over a set of manually labeled training face images.
2. The shape is updated and regulated to maintain a plausible face shape based on the PCA-based (Principal Component Analysis) shape statistics, which is similar to ASM.

C. Face normalization

Assuming all feature points are coplanar and the face remains essentially rigid, we use an affine transformation to warp an input face with varying scale, position and pose (with slightly non-frontal view) to a standard frame. Suppose the extracted feature locations are $\{\vec{P}_i\}$, where $\vec{P}_i = (x_i, y_i)^T$ ($1 \le i \le N_f$), and the feature locations of a standard face are $\{\overline{\overline{P}}_i\}$. By establishing correspondences between $\vec{P}_i$ and $\overline{\overline{P}}_i$, we can quickly estimate the six affine parameters a_j ($1 \le j \le 6$) by minimizing the following:

$$\sum_i \left\| \mathrm{F}(\vec{P}_i, a_1, a_2, ..., a_6) - \overline{\overline{P}}_i \right\|^2 , \qquad (6)$$

where $\mathrm{F}(\vec{P}_i, a_1, a_2, ..., a_6) = \vec{P}_i\,'$ represents an affine transformation that maps $\vec{P}_i$ to $\vec{P}_i\,'$ with parameters a_j ($1 \le j \le 6$). This can be solved as a standard least-squares-fitting problem. Based on the estimated parameters, the affine warping is performed on the original detected face to obtain a geometrically normalized face.

Furthermore, a photometric normalization (by zero mean and one standard deviation) is performed on the geometrically normalized face for illumination correction.

D. Evaluation

We have tested our facial feature extraction algorithm on a subset of FERET and AR face databases [9] [15]. Our proposed H-ASM has achieved doubled convergence area compared with the conventional ASM, while retaining a similar low computation cost (30-70 ms on a Pentium-IV PC, 1.7GHz). The initial feature estimation step in Section IV-A is able to bring the initial shape into the convergence area of H-ASM, thereby further increasing the stability of the feature extraction. For converged cases, our feature extraction approach achieves an average point-to-point error of 4-7 pixels.

V. Face Identification

The normalized face image (Section IV-C) is then supplied to an image-based classifier, where it is compared with a database of *N* faces. Due to large variations in illumination, expression and pose, face images from one person form a highly complex distribution in the image space, causing great difficulties in discriminating different persons. Existing subspace-based face recognition approaches (e.g. PCA (Principal Component Analysis) [16] and LDA (Linear Discriminant Analysis) [17] concentrate on deriving a transformed face representation space Φ, in which face clusters from different persons are better separated. Based on a metric d_Φ, the nearest-neighbor (NN) rule is widely applied to find the closest match to a probe face image *p* from a face database. However, we have found that using one single (distance) metric is insufficient for completely discriminating all persons. Therefore, we propose a face classifier based on a multistage rejection-based LDA (Fig. 3), where unlikely candidates are gradually rejected until the best-match is found.

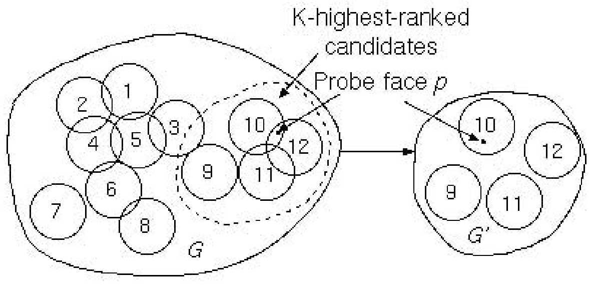


Fig. 8. Face identification by cascaded rejection.

The multistage rejection-based face identification algorithm is summarized as follows.

Algorithm 2: ***Face identification by cascaded rejection.***

Input*: A face database* G_0 *of* N_0 *persons, each having* T_i *(*$1 \le i \le N_0$*) sample images. A probe face image p.*

Training*: Perform a linear discriminant analysis on* N_0 *persons and derive a classifier* A_0.

Identification*:*

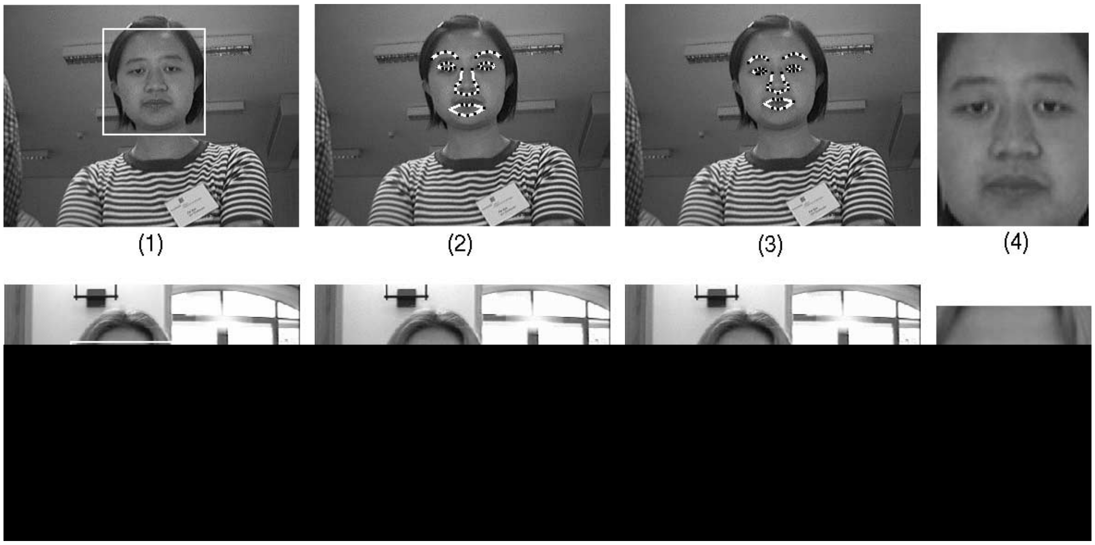
1. *Let* $A = A_0$, $G = G_0$ *and* $N = N_0$.
2. *Compare p with N candidates under A and select K candidates with highest rankings. Reject the rest N – K candidates.*
3. *Train a new classifier A' using a local one-versus-all linear discriminant analysis on set G'* ($G' \subset G$) *containing only the selected K persons.*
4. *If the best-match is found, stop; else let A = A', N = N - K, and G = G', and goto 2) in the identification stage.*

In the above identification framework, refined classifications are performed over gradually reduced candidate sets. For each stage, a different metric is defined for increased discrimination of the remaining candidates, thereby increasing the overall identification performance (see also Fig. 8).

VI. Experimental Results

A. Experiments with the HomeFace system

To evaluate the performance of *HomeFace* system, we have



[redacted]

Fig. 9. Processing stages for face recognition. (1) Face detection. (2) Feature estimation. (3) Extracted features. (4) Normalized face for recognition.

performed real-life tests in a variety of indoor environments, capturing a wide range of home situations.

Fig. 9 shows some intermediate face processing results prior to the final identification. The processing gives good performance with respect to different environments and complex backgrounds. For the identification, we set up a face database composed of 25 people. Each person has 4 to 8 sample pictures. For tests under a variety of environments, the integrated *HomeFace* system achieves a total recognition rate of 95% percent. Furthermore, the operation is computationally efficient, with a 3-4 frames/second processing speed on a P-IV PC in the centralized mode. Fig. 10 shows some examples of the face identification results in two typical testing environments.

To accommodate for large variances in illumination and face appearances over time, the *HomeFace* system has a 'self-developing' nature, where samples that are not correctly identified are progressively added to the training set. In this way, the performance of the system can be gradually improved by learning from past experiences.

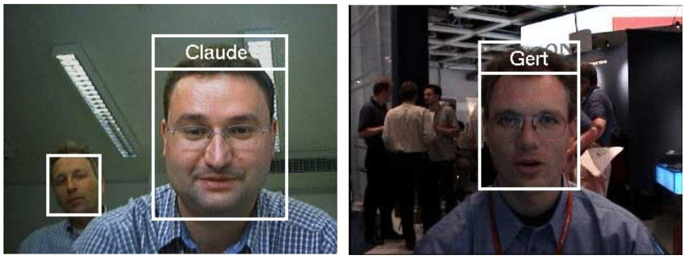
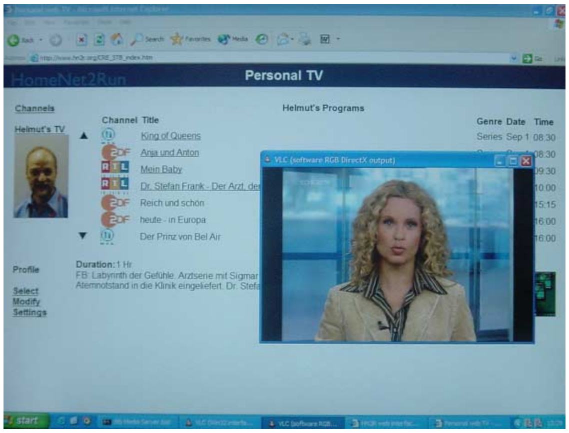


Fig. 10. Sample results from *HomeFace*.

B. Application embedding

The *HomeFace* system was embedded in a number of home applications to facilitate 'smart' user interactions. Specifically, we have integrated *HomeFace* into a user management and authentication software running on a home server, which actively identifies home users and 'pushes' user identity information to a number of home devices such as televisions and music players. Accordingly, TV and music programs can be automatically selected on a personalized basis. In Fig. 11, we give an example web interface for smart TV program selection. A successful demonstration was presented at the IFA consumer electronics exhibition [2].



Personal TV

Channels

Helmut's TV

Helmut's Programs

Channel	Title	Genre	Date	Time
	King of Queens	Series	Sep 1	08:30
ZDF	Anja und Anton			08:30
RTL	Mein Baby			09:30
RTL	Dr. Stefan Frank - Der Arzt, der			10:00
ZDF	Reich und schön			15:15
ZDF	heute - in Europa			16:00
	Der Prinz von Bel Air			16:00

Duration: 1 Hr
FB: Labyrinth der Gefühle. Arztserie mit Sigmar
Atemnotstand in die Klinik eingeliefert. Dr. Stefa

Profile

Select
Modify
Settings

Fig. 11. Web interface for smart TV program selection by face recognition.

We have also implemented the *HomeFace* system in the distributed mode, where the face detection and feature extraction/face identification are carried out at different processing units in a home. The task distribution enables the execution of the software on multiple low-end hardware platforms, thereby reducing the hardware cost. The only extra cost is the information exchange between different processing units. By using MPEG-4 object-based coding of extracted faces, we achieved a very low bit-rate network transmission of less than 5KBps.

VII. CONCLUSION

In this paper, we have presented a fast embedded face recognition system for smart home applications. In contrast with highly constrained face recognition under consistent environments, our proposal concentrates on robust real-life recognition under unconstrained environments. Furthermore, processing efficiency and low-cost are our major concerns. To meet these requirements, we developed a pipeline of face processing algorithms, namely, face detection, facial feature extraction for face normalization, and face identification. During each processing stage, we use step-wise refinements for improved robustness and efficiency. The system has achieved over 95% recognition rate and 3-4 frames/sec. processing speed.

Our proposed face recognition system has a relatively unconstrained operation environment and offers high processing efficiency using low-cost hardware. This is especially attractive for consumer home applications, where easy user interfaces and low expenses are of most interest. In a distributed version of the face recognition system, the face processing pipeline is distributed over a number of different processing units. The distribution may be such that expensive computations are performed by units that have ample computation power, so that the overall processing cost is reduced considerably. For example, in a surveillance application, face detection is done at the home entrance, and the feature extraction and face identification is carried out on a media server in the home.

The current *HomeFace* system is able to handle (near) front-view faces with out-of-plane rotation within ±20 degrees. Larger head-pose variations are not considered in the recognition. Future research is still required to handle large head rotations to facilitate more natural user interfaces.

ACKNOWLEDGMENT

The authors would like to thank Frank van Tuijl and Peter van Grinsven from Philips Research Labs. Eindhoven for their support for the integration of the *HomeFace* face recognition system into the HomeNet2Run demonstrator.

REFERENCES

[1] A. Pentland and T. Choudhury, “Face recognition for smart environments’”, *IEEE Computer*, vol. 33, no. 2, pp. 50-55, Feb. 2000.

[2] HomeNet2Run project website, http://www.extra.research.philips.com/euprojects/hn2r/.

[3] W. Zhao, R. Chellappa, P.J. Philips and A. Rosenfeld, “Face recognition: A literature survey”, *ACM Computing Surveys*, Vol. 35, no. 4, pp. 399-458, Dec. 2003.

[4] M.-H. Yang, D. Kriegman, and N. Ahuja, “Detecting faces in images: A Survey”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, 2002.

[5] J. Yang, W. Lu and A. Waibel, “Skin-color modelling and adaptation”, *Proc. ACCV*, pp. 687-694, 1998.

[6] Y. Ma and X. Ding, “Face detection based on hierarchical Support Vector Machines”, *Proc. ICPR*, pp. 222-225, 2002.

[7] F. Zuo and P.H.N. de With, “Fast human face detection using successive face detectors with incremental detection capability”, *Proc. SPIE, Electronic Imaging 2003*, vol. 5022, pp. 831-841, Jan. 2003.

[8] H. Rowley, S. Balujua and T. Kanade, “Neural network-based face detection”, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-28, 1998.

[9] P.J. Phillips, H. Moon, S.R. Rizvi and P.J. Rauss, “The FERET evaluation methodology for face recognition algorithms”, *IEEE Trans. PAMI*, Vol. 22, pp. 1090-1104, 2000.

[10] A. Yuille and P.W. Hallinan and D. Cohen, “Feature extraction from faces using deformable templates”, *Int. Journal Computer Vision*, Vol. 8, pp. 99-112, 1992.

[11] T. Cootes, “Statistical models of appearance for computer vision”, *Tech. Rep.*, Univ. of Manchester, 2001.

[12] F. Zuo and P.H.N. de With, “Fast facial feature extraction using a deformable shape model with Haar-wavelet based local texture attributes”, *Proc. ICIP*, pp. 1425-1428, 2004.

[13] B. Froba and C. Kublbeck, “Real-time face detection using Edge-Orientation Matching”, *Proc. AVBPA*, pp. 78-83, 2001.

[14] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features”, *Proc. CVPR*, pp. 511-518, 2001.

[15] A.M. Martinez and R. Benavente, “The AR Face Database”, *CVC Tech. Rep.* #24, 1998.

[16] M. Turk and A.P. Pentland, “Face recognition using Eigenfaces”, *Proc. CVPR*, pp. 586-591, 1991.

[17] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class-specific linear projection”, *IEEE Trans. PAMI*, vol. 19, No. 7, pp. 711-720, July 1997.

Fei Zuo received both her B.Eng. and M.Sc. in computer science from Xi'an Jiaotong University, P. R. China. In 2001, she became a Ph.D student in the VCA (Video Coding and Architectures) group at Electric Engineering Depart., Eindhoven University of Technology, The Netherlands. Her research interests include face detection and recognition, object extraction and classification. From 2001-2003, she was involved in the ITEA project HomeNet2Run and worked on embedded face recognition. Since 2004, she is also cooperating with System Identification Technology group at Philips Research Labs. At Eindhoven and contributes to biometric template protection on face recognition.

Peter H.N. de With graduated in electrical engineering from the University of Technology in Eindhoven. In 1992, he received his Ph.D. degree from the University of Technology Delft, The Netherlands, for his work on video bit-rate reduction for recording applications. He joined Philips Research Labs Eindhoven in 1984, where he became a member of the Magnetic Recording Systems Department. From 1985 to 1993, he was involved in several European projects on SDTV and HDTV recording. In this period, he contributed as a principal coding expert to the DV standardization for digital camcording. In 1994, he became a member of the TV Systems group at Philips Research Eindhoven, where he was leading the design of advanced programmable video architectures. In 1996, he became senior TV systems architect and in 1997, he was appointed as full professor at the University of Mannheim, Germany, at the faculty Computer Engineering. In Mannheim he was heading the chair on Digital Circuitry and Simulation with the emphasis on video systems. Since 2000, he is with CMG Eindhoven (now LogicaCMG) as a principal consultant and he is professor at the University of Technology Eindhoven, at the faculty of Electrical Engineering. He has written and co-authored numerous papers on video coding, architectures and their realization. Regularly, he is a teacher of the Philips Technical Training Centre and for other post-academic courses. In 1995 and 2000, he co-authored papers that received the IEEE CES Transactions Paper Award, and in 2004, the SPIE Investigators Award. In 1996, he obtained a company Invention Award. In 1997, Philips received the ITVA Award for its contributions to the DV standard. Mr. de With is a senior member of the IEEE, program committee member of the IEEE CES and ICIP, chairman of the Benelux community for Information and Communication Theory, co-editor of the historical book of this community, former scientific board member of CMG, scientific advisor of the Dutch Imaging school ASCII, IEEE ISCE and board member of various working groups.