28

A Feature-level Fusion of Appearance and Passive Depth Information for Face Recognition

Jian-Gang Wang[1], Kar-Ann Toh[2], Eric Sung[3] and Wei-Yun Yau[1]
[1]Institute for Infocomm Research, [2]School of Electrical & Electronic Engineering, Yonsei University, Seoul, [3] School of Electrical & Electronic Engineering, Nanyang Technological University
[1,3]Singapore, [2]Korea

1. Introduction

Face recognition using 2D intensity/colour images have been extensively researched over the past two decades (Zhao et al., 2003). More recently, some in-roads into 3D recognition have been investigated by others (Bowyer et al., 2006). However, both the 2D and 3D face recognition paradigm have their respective strengths and weaknesses. 2D face recognition methods suffer from variability in pose and illumination. Intuitively, a 3-D representation provides an added dimension to the useful information for the description of the face. This is because 3D information is relatively insensitive to illumination, skin-color, pose and makeup, and this can be used to compensate the intrinsic weakness of 2D information. However, 3D face lacks texture information. On the other hand, 2D image complements well 3D information. They are localized in hair, eyebrows, eyes, nose, mouth, facial hairs and skin color precisely where 3D capture is difficult and not accurate. A robust identification system may require fusion of 2D and 3D. Ambiguities in one modality like lighting problem may be compensated by another modality like depth features. Multi-modal identification system hence usually performs better than any one of its individual components (Choudhury et al., 1999).

There is a rich literature on fusing multiple modalities for identity verification, e.g. combining face and fingerprint (Hong and Jain, 1998), voice and face biometrics (Bruneli, 1995; Choudhury et al. 1999) and visible and thermal imagery (Socolinsky et al., 2003). The fusion can be done at feature level, matching score level or decision level with different fusion models. The fusion algorithm is critical part to obtain a high recognition rate. (Kittler et al., 1998) considered the task of combining classifiers in a probabilistic Bayesian framework. Several ways (sum, product, max, min, major voting) to combine the individual scores (normalized to range [0, l]) were investigated, based on the Bayesian theorem and certain hypothesis, from which the Sum Rule (adding the individual scores) is shown to be the best in the experimental comparison in a multilevel biometric fusion problem. Appearance and depth were fused at matching score level for face recognition by min, sum and product in (Chang et al., 2004; Tsalakanidou et al., 2003), by weighted sum in (Beumier & Acheroy, 2001; Wang et al., 2004a, 2005, 2006). There are some limitations in the existing decision fusion models. Statistical models (e.g. kNN, Bayesian) rely heavily on prior

Source: Face Recognition, Book edited by:Kresimir Delac and Mislav Grgic, ISBN978-3-902613-03-5, pp.558, I-Tech,Vienna, Austria, June2007

statistical assumptions which can depart from reality; Linear models (e.g. weighted sum, LDA) are limited to linear decision hyper-surfaces; Nonlinear models (e.g. Neural Networks, RBF, SVM) involves nonlinear optimization. Moreover, the learning process could be very tedious and time consuming.

Multivariate Polynomial (MP) provides an effective way to describe complex nonlinear input-output relationship since it is tractable for optimization, sensitivity analysis, and predication of confidence intervals. With appropriate incorporation of certain decision criteria into the model output, MP can be used for pattern analysis and could be a fusion model to overcome the limitations of the existing decision fusion models. However, the full MP has dimension explosion problem for large dimension and high order system. The MP model can be considered a special example of kernel ridge regression (KRR) (Taylor & Cristianini, 2004). Instead of using the kernel trick to handle the computational difficulty of MP, we consider the use of a reduced multivariate polynomial model.

In this chapter, we proposed to use an extended Reduced Multivariate Polynomial Model (RMPM) (Toh et al., 2004; Tran et al., 2004) to fuse appearance and depth information for face recognition where simplicity and ease of use are our major concerns. RMPM is found to be particullary suitable for problems with small number of features and large number of examples. In order to apply RMPM to face recognition problem, principal component analysis (PCA) is used for dimension reduction and feature extraction and a two-stage PCA+RMPM is proposed for face recognition. Furthermore, the RMPM was extended in order to cater for the new-user registration problem. We report a stage of development on fusing the 2D and 3D information, catering for on-line new user registration. This issue of new user registration is non-trivial since current available techniques require large computing effort on static database. Based on a recent work by (Toh et al., 2004), a recursive formulation for on-line learning of new-user parameters is presented in this chapter (Tran et al., 2004). The performance of the face recognition system where appearance and depth images are fused will be reported.

There are three main techniques for 3D facial surface capture. The first is by passive stereo using at least two cameras to capture a facial image and using a computational matching method. The second is based on structured lighting, in which a pattern is projected on a face and the 3D facial surface is calculated. Finally, the third is based on the use of laser range-finding systems to capture the 3D facial surface. The third technique has the best reliability and resolution while the first has relatively poor robustness and accuracy. Existing 3D or 3D plus 2D (Lu & Jain, 2005; Chang et al., 2003, 2005; Tsalakanidou et al., 2003; Wang et al. 2004a) face recognition techniques assume the use of active 3D measurement for 3D face image capture. However, the active methods employ structured illumination (structure projection, phase shift, gray-code demodulation, etc) or laser scanning, which are not desirable in many applications. The attractiveness of passive stereoscopy is its non-intrusive nature which is important in many real-life applications. Moreover, it is low cost. This serves as our motivation to use passive stereovision as one of the modalities of fusion and to ascertain if it can be sufficiently useful in face recognition (Wang et al., 2005, 2006). Our experiments to be described later will justify its use.

(Gordon, 1996) presented a template-based recognition method involving curvature calculation from range data. (Beumier C. & Acheroy M., 1998, 2001) proposed two 3D difference methods based on surface matching and profile matching. (Beumier & Acheroy, 1998) extended the method proposed in (Gordon, 1996) by performing face recognition

using the similarity of the central and lateral profiles from the 3D facial surface. The system is designed for security applications in which the individuals are cooperative. Structured light was used to obtain the facial surface. However, the drawback of a structured light system is its bulkiness and its limited field of depth constrained by the capabilities of the camera and projector lens. Both (Gordon, 1996) and (Beumier & Acheroy, 2001) realized that the performance of the 3D facial features based face recognition depends on the 3D resolution. (Lee & Milios, 1990) proposed a method based on extend Gaussian image for matching graph of range images. A method to label different components of human faces for recognition was proposed by (Yacoob & Davis, 1994). (Chua et al., 2000) described a technique based on point signature, a representation for free form surfaces. (Blanz & Vetter, 2003, 1999) used a 3D morphable model to tackle variation of pose and illumination in face recognition, in which the input was a 2D face image.

3D face recognition is one of the three main contenders for improving face recognition algorithms in "The Face Recognition Grand Challenge (FRGC)" (WWWc). While 3D face recognition research dates back to before 1990, algorithms that combine results from 3D and 2D data did not appear until about 2000. (Beumier & Acheroy, 2001) also investigated the improvement of the recognition rate by fusing 3D and 2D information. The error rate was 2.5% by fusing 3D and gray level using a database of 26 subjects. Recently (Pan et. al., 2003) used the Hausdorff distance for feature alignment and matching for 3D recognition. More recently, (Chang et. al., 2003, 2004, 2005) applied PCA with 3D range data along with 2D image for face recognition. A Minolta Vivid 900 range scanner, which employs laser-beam light sectioning technology to scan workpieces using a slit beam, was used for obtaining 2D and 3D images. (Chang et. al. 2004) investigated the comparison and combination of 2D, 3D and IR data for face recognition. Based on PCA representations of face images, they reported 100% recognition rate when the three modalities are combined on a database of 191 subjects. (Tsalakanidou et al., 2003) developed a system to verify the improvement of the face recognition rate by fusing depth and colour eigenfaces on XM2VTS database, PCA is adopted to extract features. The 3D models in XM2VTS database are built using an active stereo system provided by Turing Institute (WWWa). By fusing the appearance and depth Fisherfaces, (Wang et al., 2004a) showed the gain in the recognition rate when the depth information is used. Recently, (Wang et al., 2006) developed a face recognition system by fusing 2D and passive 3D information based on a novel Bilateral Two-dimensional Linear Discriminant Analysis (B2DLDA). A good survey on 3D, 3D plus 2D face recognition can be found in (Bowyer et al., 2006).

Thanks to the technical progress in 3D capture/computing, an affordable real-time passive stereo system has become available. In this paper, we set out to find if present-day passive stereovision in combination with 2D appearance images can match up to other methods that rely on active depth data. Our main objective is to investigate into combining appearance and depth face images to improve the recognition rate. We show that present-day passive stereoscopy, though less robust and accurate, is a viable alternative to 3D face recognition. The SRI Stereo engine (WWWb) that outputs a high range resolution (≤ 0.33 mm) was used in our applications. The entire face detection, tracking, pose estimation and face recognition steps are described and investigated. A hybrid face and facial feature detection/tracking approach is proposed that collects near-frontal views for face recognition. Our face detection/tracking approach automatically initializes without user intervention and can be re-initialized automatically if tracking of the 3D face pose is lost. Comparisons with the

existing methods (SVM, KNN) are also provided in this chapter. The proposed RMPM method can yield compariable results with SVM. It is clear that computation load of the RMPM is much lower than SVM.
The rest of the chapter is organized as follows. Fusion of appearance and depth information is discussed in Section 2. Section 3 presents some issues related to stereo face recognition system. Section 4 discusses the experiment on XM2VTs and the implementation of the algorithm on a stereo vision system. Section 5 concludes the work with recommendation for future enhancements to the system.

2. Fusing Appearance and Depth Information

As discussed in section 1, we aim to improve the recognition rate by combining appearance and depth information. The manner of the combination is crucial to the performance of the system. The criteria for this kind of combination is to fully make use of the advantages of the two sources of information to optimize the discriminant power of the entire system. The degree to which the results improve performance is dependent on the degree of correlation among individual decisions. Fusion of decisions with low correlation can dramatically improve the performance.
In this chapter, a novel method using the RMPM has been developed for face recognition. A new feature is formed by concatenting the feature of an appearance image and the feature of a depth/disparity image. A RMPM is trained by using the combined 2D and 3D features. We will show that The RMPM can be easily formulated into recursive learning fashion for online applications.
In Section 2.1 and 2.2, we briefly discuss the RMPM and the auto-update algorithm of the RMPM for new user registration in face recognition. The face recognition based on the RMPM will be discussed in section 2.3. The new-user registration is discussed in Section 2.4.

2.1 Multivariate polynomial regression

The general multivariate polynomial model can be expressed as

$$g(\alpha, \mathbf{x}) = \sum_{i=1}^{K} \alpha_i x_1^{n_1} x_2^{n_2} \cdots x_l^{n_l} \tag{1}$$

where the summation is taken over all nonnegative integers $n_1, n_2, \ldots, n_l$ for which $n_1 + n_2 + \ldots + n_l \leq r$ with r being the order of approximation. $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_K]^T$ is the parameter vector to be estimated and $\mathbf{x}$ denotes the regressor vector $[x_1, x_2, \ldots, x_l]^T$ containing l inputs. K is the total number of terms in $g(\boldsymbol{\alpha}, \mathbf{x})$.
Without loss of generality, consider a second-order bivariate polynomial model (r = 2 and l = 2) given by

$$g(\boldsymbol{\alpha},\mathbf{x})=\boldsymbol{\alpha}^T\mathbf{p}(\mathbf{x}) \tag{2}$$

where $\boldsymbol{\alpha} = [\alpha_1\ \alpha_2\ \alpha_3\ \alpha_4\ \alpha_5\ \alpha_6]^T$ and $\mathbf{p}(\mathbf{x}) = [1\ x_1\ x_2\ x_1^2\ x_1x_2\ x_2^2]^T$. Given m data points with $m>K$ (K = 6 here) and using the least-squares error minimization objective given by

$$s(\boldsymbol{\alpha}, \mathbf{x}) = \sum_{i=1}^{m} [y_i - g(\boldsymbol{\alpha}, x_i)]^2 = [\mathbf{y} - \mathbf{P}\boldsymbol{\alpha}]^T [\mathbf{y} - \mathbf{P}\boldsymbol{\alpha}] \tag{3}$$

the parameter vector $\boldsymbol{\alpha}$ can be estimated from

$$\boldsymbol{\alpha} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{y} \tag{4}$$

where $\mathbf{P} \in \Re^{m\times K}$ denotes the Jacobian matrix of $\mathbf{p}$(x):

$$\mathbf{P} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & x_{1,1}^2 & x_{1,1}x_{2,1} & x_{2,1}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,m} & x_{2,m} & x_{1,m}^2 & x_{1,m}x_{2,m} & x_{2,m}^2 \end{bmatrix} \tag{5}$$

and $\mathbf{y} = [y_1, y_2, \ldots, y_m]^T$ is the known interface vector from training data. In (5), the first and second subscripts of the matrix elements, $x_{j,k}$ (j = 1,2, k = 1,…m) indicate the number of the inputs and the number of instances, respectively.

It is noted here that (4) involves computation of the inverse of a matrix, the problem of multicollinearity may arise if some linear dependence among the elements of x are present. A simple way to improve numerical stability is to perform a weight decay regularization using the following error objective:

$$s(\boldsymbol{\alpha}, x) = \sum_{i=1}^{m} [y_i - g(\boldsymbol{\alpha}, x_i)]^2 + b\|\boldsymbol{\alpha}\|^2 = [y - P\boldsymbol{\alpha}]^T [y - P\boldsymbol{\alpha}] + b\boldsymbol{\alpha}^T\boldsymbol{\alpha} \tag{6}$$

where $\|\cdot\|^2$ denotes the l_2-norm and b is a regularization constant.

Minimizing the new objective function (6) results in

$$\boldsymbol{\alpha} = (\mathbf{P}^T\mathbf{P} + b\mathbf{I})^{-1}\mathbf{P}^T\mathbf{y} \tag{7}$$

where $\mathbf{P} \in \Re^{m\times K}$, $y \in \Re^{m\times 1}$ and $\mathbf{I}$ is a (K×K) identity matrix. This addition of a bias term into the least-squares regression model is also termed as ridge regression (Neter et al., 1996).

2.2 Reduced Multivariate Polynomial Model

To significantly reduce the huge number of terms in the above multivariate polynomials, a reduced model (Toh et al., 2004) was proposed as:

$$g_{RM}(\boldsymbol{\alpha}, \mathbf{x}) = \alpha_0 + \sum_{k=1}^{r}\sum_{j=1}^{l} \alpha_{k,j} x_j^k + \sum_{k=1}^{r} \alpha_{rl+k} \left(\sum_{j=1}^{l} x_j\right)^k + \sum_{k=2}^{r} (\alpha_k^T \cdot x)\left(\sum_{j=1}^{l} x_j\right)^{k-1} \quad l, r \geq 2 \tag{8}$$

where x_j, j = 1, 2,…l, are the polynomial inputs, $\{\alpha\}$ are the weighting coefficients to be estimated, and l, r correspond to input-dimension and order of system respectively. The number of terms in this model is: $k = 1+r+l(2r-1)$.
Comparing with existing classifiers, RMPM has some advantages as follows: (1) Number of parameters (polynomial coefficients) increases linearly with model-order and input-dimension, i.e. no dimension explosion as in the case of full multivariate polynomials; (2) Nonlinear decision hyper-surface mapping; (3) Fast single-step least-squares optimal computation which is linear in parameter space, tractable for optimization, sensitivity analysis, and prediction of confidence intervals; (4) Good classification accuracy: comparable to SVM, Neural Networks, RBF, Nearest-Neighbor, Decision Trees (Toh et al. 2004).

2.3 Face recognition

RMPM is found to be particullary suitable for problems with small number of features and large number of examples (Toh et al., 2004). It is known that the face space is very large. In order to apply RMPM to face recognition problem, dimension reduction is necessary. In this paper, principal component analysis (PCA) is used for dimension reduction and feature extraction and a two-stage PCA+RMPM is proposed for face recognition.
Learning. PCA is applied to appearance and depth images, respectively. In this chapter, the fusion of the appearance and depth information is completed at feature level, this can be done by concatenating the eigenface features of the appearance and depth images. The learning algorithm of a RMPM can be expressed as

$$\mathbf{P} = RM(r, [W_{Eigen\_appearance} \quad W_{Eigen\_depth}]) \tag{9}$$

where r is the order of the RMPM, $W_{\text{Eigen\_appearance}}$ and $W_{\text{Eigen\_depth}}$ are eigenface features of the appearance and depth/disparity images respectively. The parameters of the RMPM can then be learned from the training samples using (7).
Testing. A probe face, F_T, is identified as a face of the gallery if the output element of the reduced model classifier (appearance and depth), $\mathbf{P}^T\boldsymbol{\alpha}$, is the maximum (and ≥0.5) among the all faces in the training gallery.

2.4 New user registration

New user registion is an important problem for an online biometric authentication system. Although it can be done offline where the system are re-trained on the new training set, an automatic online user registration is an interesting research topic. The efficiency is the major concern of online registration. In this chapter, we extend the RMPM by adding an efficient user registration capability. We will discuss this extension as follows.
We set the learning matrix, $\mathbf{P}^T$, as follows: the initial value of an element in learning matrix is set to be 1 if the sample corresponds to the indicated subject, else it is set to be 0. According to the definition of RMPM classifier, a face F_T is determined to be a new user (not a previously registered user) if

$$\text{Maximum}(\mathbf{P}^T(F_T)\boldsymbol{\alpha}) < 0.5. \tag{10}$$

Assume we have n face images in the original training set. A new training set is formed by adding a new user, f_{new} which is detected based on the above rule, to the original database.

Assum S_{new} and S_{old} are the sum of the faces of the training samples in the new traing set and the original training set respectively, we have

$$S_{new} = S_{old} + f_{new} \tag{11}$$

and the (small sample) mean of the new training set will be

$$m_{new} = S_{new}/(n+1) \tag{12}$$

The new eigenfaces can be computed using m_{new}.
Let $\mathbf{f}_i$ be the vector of all polynomial terms in (8) which is applied to the i-th samples. Assuming the parameters of RM are represented as t when a new user is registered and are represented as t-1 when the new user is not resigtred. $\mathbf{F_T} = [\mathbf{f}_1, \mathbf{f}_2 \ldots, \mathbf{f}_t]^T$. Let $\mathbf{M}_t = (\mathbf{P}^T\mathbf{P} + b\mathbf{I})$, then (7) becomes

$$\boldsymbol{\alpha} = \mathbf{M}_t^{-1}\mathbf{P}_t^T\mathbf{y}_t \tag{13}$$

Next we have (Tran et al., 2004) :

$$\mathbf{M}_t = \mathbf{M}_{t-1} + \mathbf{f}_t\mathbf{f}_t^T \tag{14}$$

$$\mathbf{P}_t^T\mathbf{y}_t = \mathbf{P}_{t-1}^T\mathbf{y}_{t-1} + \mathbf{f}_t\mathbf{y}_t \tag{15}$$

Finally, the new estimate $\boldsymbol{\alpha}_t$ can be calculated using the previous estimate $\boldsymbol{\alpha}_{t-1}$ the inversion of $\mathbf{M}_{t-1}$ and the new training data $\{\mathbf{f}_t, \mathbf{y}_t\}$, we have

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \lambda\mathbf{M}_t^{-1}\mathbf{f}_t(\mathbf{y}_t - \mathbf{f}_t^T\boldsymbol{\alpha}_{t-1}) \tag{16}$$

Based on (14) to (16), the parameters of RMPM can be computed automatically when a new user arrives.

3. Stereo Face Recognition System

The proposed algorithm is evaluated using XM2VTS database and an inhouse database which was collected by a stereo vision system. In this section, we discuss the stereo vision system.
We proposed a hybrid approach to detect and track head/face in real-time. The signal flow diagram is shown in Fig. 1. The output of a stereo vision system (WWWb) is a set containing three images: left image, right image and disparity image. It should be noted that the left image and the disparity image are fully registered. This means that we can detect facial features from either left image or disparity image depending upon which is more easy. For instance, we can detect nose tip from the disparity image and eye corners from the left image. In our approach, by combining disparity and intensity images, either the head or the face/3D pose is tracked automatically. The head is tracked if face features are not available, e.g. when the person is far away from the stereo head or when the face is in the profile view. The face are tracked once the facial features, such as the nostrils, eyebrows, eyes and mouth are found. Disparity maps of the face are obtained at frame rate using commercially available stereo software, e.g. SRI International Small Vision System (SVS) (WWWb). The range data of a person is extracted from the disparity map by assuming the person of interest

is the nearest object to the camera. A novel method is proposed where the head is separated from the connected head-and-shoulder component using morphological watersheds. The head contour is modeled as an ellipse, which can be least-squares fitted to points obtained in the watershed segmentation. The eye corners, mouth corners are extracted using SUSAN corner detector; the nose tip is detected in the disparity image by a template metching. Using the calibrated parameters of the vision system, the head pose can be estimated using a EM enhanced vanishing point, formed by the eye lines and mouth line, based method (Wang & Sung, 2007).

3.1 Face detection

Proper face detection is important for accurate face recognition. The task includes locating the face, extracting facial features and consistent image normalization. Although the face detection and tracking with a single camera is a well explored topic, the use of the stereo technology for this purpose has now become an important interest (Morency et al. 2002; Daniel 2002; Rafael et al., 2005). The availability of commercial hardware to resolve low-level problems with stereoscopic cameras, as well as lower prices for these types of systems, turns them into an appealing sensor with which intelligent systems could be developed. The use of stereo vision provides a higher grade of information that bring several advantages when developing face recognition system. On one hand, the information regarding disparities becomes more invariable to illumination changes than the images provided by a single camera, this being a very advantageous factor for the background segmentation. Furthermore, the possibility to know the distance to the person could be of great assistance for the tracking as well as for a better analysis of their gestures.

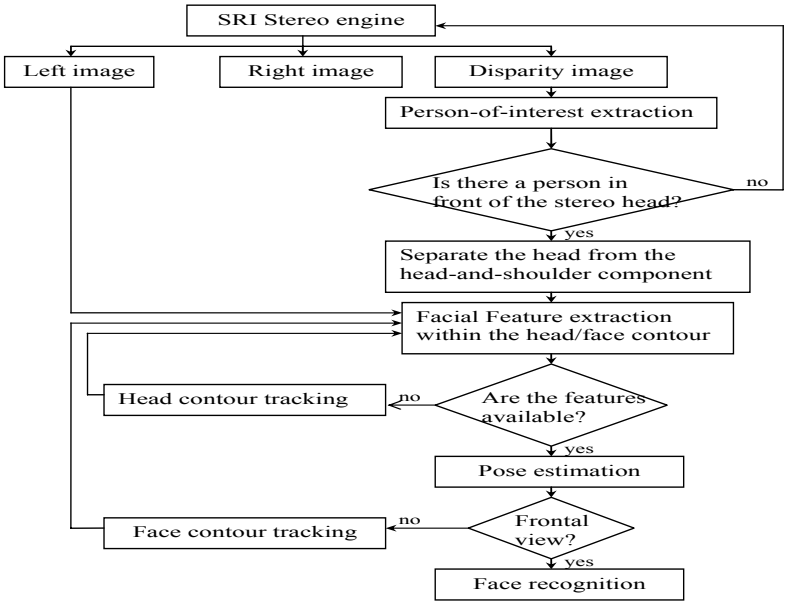


Figure 1. Flowchart of the head/face tracking algorithm

In our system, face detection becomes much easier with available 3D information. An object-oriented segmentation is applied to obtain the person-of-interest, who in our case is the one closest to the camera. The nearer the object to the camera, the brighter the pixel in the disparity image. Hence histogram-based segmentation can be applied effectively.
Subject-of-interest can be segmented out by thresholding their distances from the stereo head. The thresholds are selected based on the peak analysis of the disparity histogram. This will help in tracking the objects efficiently. Two persons at different distances in front of the camera are separated using the disparity map as shown in Fig. 2.

(a) (b) (c) (d) (e)

Figure 2. Extraction of a person-of-interest in a disparity image. (a) left image (b) right image (c) disparity image (d) person-of-interest (near face) (e) far face and background

3.2 Head location using morphological watersheds

We have discussed the segmentation of a disparity image. By observing the shape of the segmented image, we can see that the head can be located by blob analysis. Here, a novel method, which employs morphological watersheds transform in conjunction with the distance transform to separate head from shoulder, is proposed. Watershed is an efficient tool to detect touching objects. To minimize the number of valleys found by the watershed transform, one need to maximize the contrast of our objects of interest. Distance transform determines the shortest distance between each blob pixel and the blob's background, and assigns this distance value to the pixel. Here, we apply a distance transform to the head-and-shoulder image to produce a distance image. In the distance image, there is a maximum in the head and shoulder blobs, respectively. In addition, the head and the shoulder have touching zones of influence. Applying watershed operation to the distance image, the head can be separated from the head-and-shoulder component by the watershed line. Then the segmented head contour is least-squares fitted to an ellipse. We have tested this method on a face database built in our laboratory. There are 3000 face images of 100 student volunteers with 10 different head poses where the subjects turn their head from left to right. At each pose, the database includes two intensity face images (stereo) and one disparity image computed using SVS. The experimental results are satisfactory, the heads can be located with a successful rate 99% (Wang et al., 2004b). An example is shown in Fig. 3. Some of the resulting frames of the sequences of a person are shown in Fig. 4. Frame 2, 20, 40, 60 and 80 are shown respectively from left to right.

(a) disparity segmentation (b) distance transform of (a) (c) watershed line of (b) (d) overlay (c) to (a)

Figure 3. Separate the head from the head-and-shoulder component using distance transform and morphological watershed

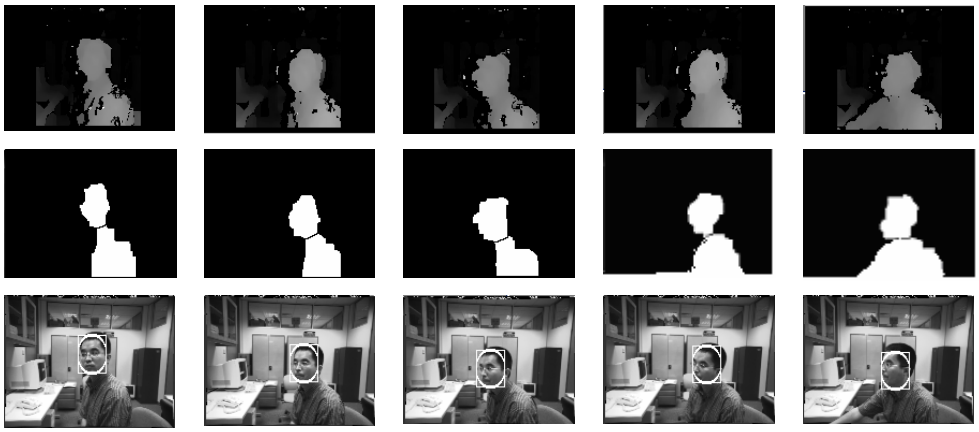


Figure 4. Head contours of a sequence face images; first row: disparity images; second row: the heads are located by the distance transform and watershed operation; third row: the elliptical head contours. From left to right: frame 2, 20, 40, 60, 80 respectively.

Although the algorithm works well, the watershed line may not incidentally correspond to the neck. Fortunately, this can be detected by checking the eccentricity of the elliptical head. The eccentricity is assumed 1.2 in our experiment. If the watershed is found not to correspond to the neck, i.e. the difference between the ratio of the elliptical head and the assumed one is significant, a candidate head will be placed below the vertically maxima of the silhouette, in a manner similar to (Darrell et al., 2000), and will be refined in the feature tracking stage. (Daniel 2002) proposed a alternative stereo head detection approah. A simple human-torso model is used besides the depth.

3.3 Feature extraction

Many of the existing feature extraction methods are based on artificial template matching. An artificial template is a small rectangular intensity image that contains, for example, an eye corner, where the corner is located in the centre of the template. The image region, which best matches the artificial template is extracted from current image. The problem with artificial template matching is that it fails when it is applied to images that are different from those that were used to generate the artificial image.

Recently, the SUSAN (Smallest Univalue Segment Assimilating Nucleus) operator (Smith & Brady, 1997) was found to be an efficient facial feature extraction tool (Hess & Martinez, 2004; Wu et al., 2001; Gu et al., 2001). We adopt the method in our approach where the eye and mouth corners can be located using the SUSAN corner detector. In order to apply the corner detector, we need to establish a rectangular search region for the mouth and two rectangular regions for the left and right eyes respectively. These initial search regions for eye and mouth are found by our method in (Wang & Sung, 1999, 2000). The mouth corners and eye corners were extracted with a reliability of 90% and an average position error of 2.25 pixels. The size of the head is about 50×60 in a 154×114 image. The ground truth of the positions of the eye and mouth corners are found by manually measuring them.

3.4 Head/Face tracking

(Birchfield, 1998) presented an algorithm for tracking a person's head. The head's projection onto the image plane is modeled as an ellipse whose position and size are continually updated by a local search combining the output of a module concentrating on the intensity gradient around the ellipse's perimeter with that of another module focusing on the color histogram of the ellipse's interior. Since these two modules have roughly orthogonal failure modes, they serve to complement one another. Extensive experimentation shows the algorithm's robustness with respect to full 360-degree out-of-plane rotation, up to 90-degree tilting, severe but brief occlusion, arbitrary camera movement, and multiple moving people in the background. We adopted Birchfield's method to track head/face.

3.5 Pose estimation

As we are using stereo camera, we can compute the pose from the 3D coordinates of the feature points directly, like that by (Matsumoto & Zelinsky, 2000). However, it is found that the pose is quite sensitive to the coordinates' errors. The head pose can be computed using three feature points, e.g. two eye corners and one mouth corners. Hence, we can get different pose estimations based on different three-point feature groups. We found the result is not stable even if the mean pose of some different poses (e.g. 5 poses can be computed from two eye corners and two mouth corners) is computed. Instead of computing the pose directly based on the 3D coordinates of the feature points, we adopted a robust EM enhanced vanishing point based pose estimetion method (Wang & Sung, 2007) in this chapter. The novel approach assumes the full perspective projection camera model. Our approach employs general prior knowledge of face structure and the corresponding geometrical constraints provided by the location of a certain vanishing point to determine the pose of human faces. To achieve this, eye-lines, formed from the far and near eye corners, and mouth-line of the mouth corners are assumed parallel in 3D space. Then the vanishing point of these parallel lines found by the intersection of the eye-line and mouth-line in the image can be used to infer the 3D orientation and location of the human face. In order to deal with the variance of the facial model parameters, e.g. the ratio between the eye-line and the mouth-line, an EM framework is applied to update the parameters. We first compute the 3D pose using some initially learnt parameters (such as ratio and length) and then adapt the parameters statistically for individual persons and their facial expressions by minimising the residual errors between the projection of the model features points and the actual features on the image. In doing so, we assume every facial feature point can be associated to each of features points in 3D model with some *a posteriori* probability. The expectation step of the EM algorithm provides an iterative framework for computing the *a posterori* probabilities using Gaussian mixtures defined over the parameters.

3.6 Normalizations of appearance and disparity images

A face is detected and tracked using stereo vision as the person moves in front of the stereo camera. The frontal face pose is automatically searched and detected from the captured appearance and depth images. These images are subsequently used for face recognition.

Using the image coordinates of the two eye centers the image is rotated and scaled to occupy a fixed size array of pixels (88×64). In the stereo vision system, the coordinates of pixel are consistent with the coordinates in the left image. The feature points, two eye centers can hence be located in the disparity image. The tip of the nose can be detected in the

disparity image using template matching of (Gordon, 1996). From coplanar stereo vision principle, we have,

$$d' = Bf / d \tag{17}$$

where d' represents the depth, d is the disparity, B is the baseline and f is the focal length of the calibrated stereo camera. Hence we can compute the depth image from a disparity image with (17). The depth image can be normalized using the depth of the nose tip, i.e. the nose tip of every subject is translated to the same point in 3D relative to the sensor. After that, the depth image is further normalized by the two eye centers.

Problems with the 3D data are alleviated to some degree by preprocessing to fill in holes (a region where there is missing 3D data during sensing) and spikes. We adopt the method in (Chang et al., 2004) to detect the spike, and then remove the holes by linear interpolation of missing values from good values around the edges of the hole.

4. Experiments

We evaluate our algorithm on XM2VTS database (Masser et al., 1999) and a large face detabase collected using stereo vision system described in section 3. The purpose is to show the gain in the recognition rate when the depth information is used.

4.1 Experiment on XM2VTS database

The XM2VTS database consists of color frontal, color profile and 3D VRML models of 295 subjects (Masser et al., 1999). The following tests were conducted for performance evaluation. The main reason for adopting this database is that 3D VRML model of those subjects are provided on top of the 2D face images, and this 3D model can be used to generate the depth map for our algorithm.

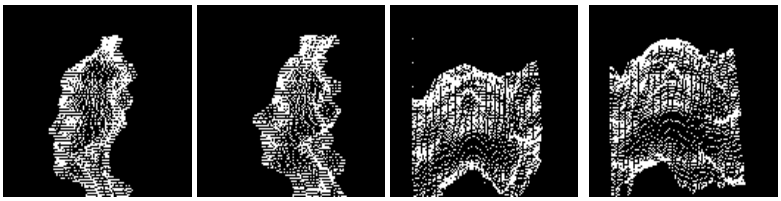


Figure 5. 3D VRML face models consisting of triangles

Generation of depth images. Depth image is an image where the intensity of a pixel represents the depth of the correspondent point with respect to the 3D VRML model coordinate system. 3D VRML model of a face in XM2VTS database is displayed in Fig 5. There are about 4000 points in a 3D face model to represent the face. The face surface is triangulated with these points. In order to generate a depth image, a virtual camera is placed in front of the 3D VRML model, see Fig. 6. The coordinate system of the camera is defined as follows: the image plane is defined as the X-Y plane and the Z-axis is along the optical axis of the camera and pointing toward the frontal object. The initial plane of Y_c-Z_c is positioned parallel to Y_m-X_m plane of the 3D VRML model. The projective image can be obtained using the perspective transform matrix of the camera. Z_c coincides with Z_m; however in reverse directions. X_c is parallel to X_m and Y_c parallel is to Y_m; however they are with reverse directions.

The intrinsic parameters of the camera must be properly defined in order to generate depth image from the 3D VRML model. The parameters include (u_0, v_0), the coordinates of the image-center point (principle point), and f_u and f_v, scale factors of the camera along the u- and v-axis respectively. The position of the origin of the camera system, (x_0, y_0, z_0), under the 3D VRML model coordinate system is also set.

Perspective projection is assumed, i.e. for a point $P(x_m, y_m, z_m)$ in a 3D VRML model of a subject, the 2D coordinates of $\boldsymbol{P}$ in its depth image is computed as follows:

$$u = u_0 + f_u\,(x_m/(z_0\text{-}z_m)) \tag{18}$$

$$v = v_0 - f_v\,(y_m/(z_0\text{-}z_m)) \tag{19}$$

In our approach, z-buffer algorithm is applied to handle the face-self occlusion for generating the depth images.

We used the frontal views in XM2VTS database (CDS001, CDS006 and CDS008 darkened frontal view). CDS001 dataset contains 1 frontal view for each of the 295 subjects and each of four sessions. This image was taken at the beginning of the head rotation shot. So there are a total of 1,180 colour images, each with a resolution of 720 × 576 pixels. CDS006 dataset contains 1 frontal view for each of the 295 subjects and each of the four sessions. This image was taken from the middle of the head rotation shot when the subject had returned his/her head to the middle. They are different from those contained in CDS001. There are a total of 1,180 colour images. The images are at resolution 720 × 576 pixels. CDS008 contains 4 frontal views for each of the 295 subjects taken from the final session. In two of the images, the studio light illuminating the left side of the face was turned off. In the other two images, the light illuminating the right side of the face was turned off. There are a total of 1,180 colour images. The images are at resolution 720 × 576 pixels. We used the 3D VRML-Model (CDS005) of the XM2VTSDB to generate 3D depth images corresponding to the appearance images mentioned above. The models were obtained with a high-precision 3D stereo camera developed by Turing Institute (WWWa). The models were then converted from their proprietary format into VRML.

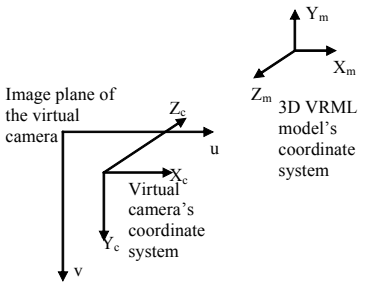
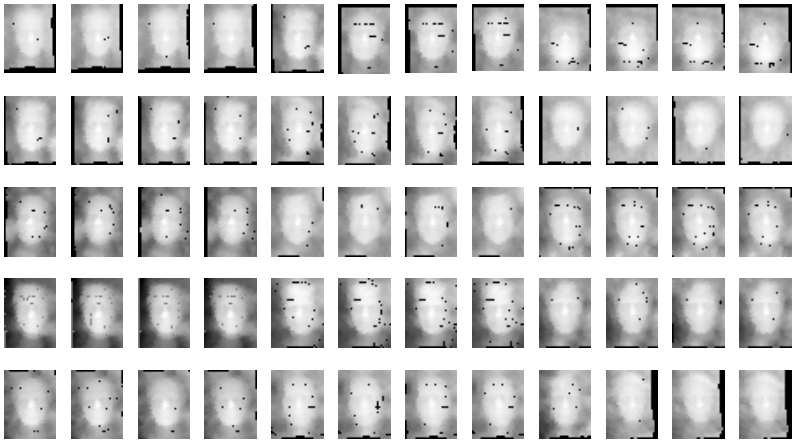


Figure 6. The relationship between the virtual camera and the 3D VRML face model

Therefore, a total of 3540 pairs of frontal views (appearance and depth pair) of 295 subjects in X2MVTS database are used. There are 12 pairs of images for each subject. We pick randomly any two of them for the learning gallery, while the remainder ten pairs per subject are used as probes. The average recognition rate was obtained over 66 runs. In the XM2VTS database, there is only one 3D model for each subject. In order to generate more than one view for learning and testing, some new views are obtained by rotating the 3D coordinates

of VRML model away the frontal (about the Y_m axes) by some degrees. In our experiments, the new views obtained at ±1°, ±3°, ±5°, ±7°,±9°,±11°.
Some of the normalized face image samples in XM2VTS database are shown in Fig. 7, where appearance face images are shown in Fig. 7(a) and the correspondent depth images are shown in Fig. 7(b). The resolution of the images is 88 × 64. We can see significant changes in illumination, expressions, hair, and eye glasses/no eyeglasses due to longer time lapse (four months) in photograph taking. The first 40 Eigenfaces of 2D and 3D training samples are shown in Fig. 8.

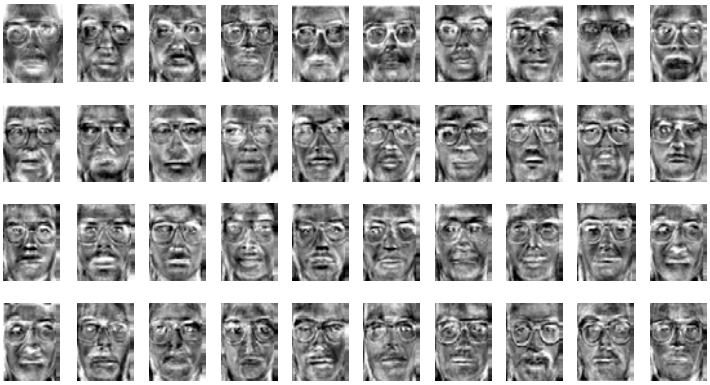
(a) Normalized appearance face images, the column 1-4: images in CDS001; column 5-8: images in CDS006; column 9-12: images in CDS008



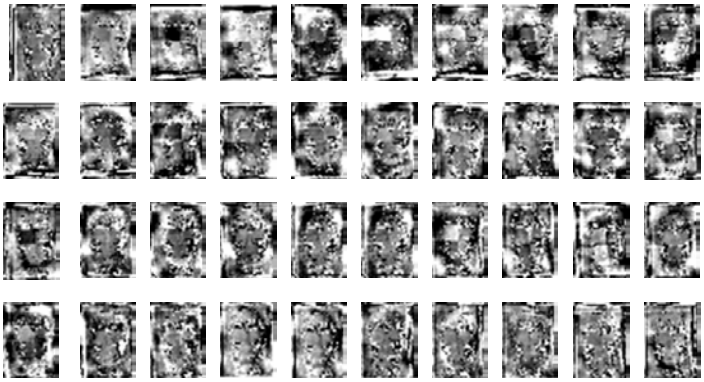
(b) Normalized depth images corresponding to (a)
Figure 7. Some normalized samples (appearance and corresponding depth images) from XM2VTS database

Recognition. Using the gallery and probe described above, the evaluation of the recognition algorithm has been done, including the recognition when the number of the eigenfaces varies from 20 to 80 with a step increment of 10. The order r of the RMPM is set to be 2 while b is set to be (10^{-4}). The experimental results support our hypothesis that the combined modality outperforms the individual modalities of appearance and depth. It also shows that each contains independent information from the other. In this experiment, the recognition rate for 3D alone is nearly the same with the one on 2D.

In order to compare the proposed RMPM method with the existing methods, e.g. SVM, KNN, we evaluate the performance of SVM, KNN and the proposed RMPM using the same database. The results for 2D plus 3D, 3D alone and 2D alone by the proposed RMPM, a RBF-kernel SVM (OSU SVM package, WWWd) and KNN on XM2VTS databse are shown in Fig. 9. to Fig. 11. respectively. We can see that the recognition rate has been improved by fusing appearance and depth. By comparing the results in Fig. 9 to Fig. 11, we can see that the RMPM method can yield results compariable with SVM. Both the RMPM and SVM get better results than the one by the KNN.



(a) The first 40 Eigenfaces of the appearance gallery



(b) The first 40 Eigenfaces of the depth gallery

Figure 8. Eigenface features of the appearance and depth training samples

4.2 Experiment on stereo vision system

Encouraged by the good performance of the recognition algorithm on XM2VTS database, we implemented the algorithms on a stereo vision system. We aim at identifying a face by fusing disparity/depth and intensity information from a binocular stereo vision system (WWWb), which outputs the disparity/range information automatically in real-time. The existing 3D face recognition techniques assume the use of active 3D measurement for 3D face image capture. However, active methods employ structured illumination (structure projection, phase shift, etc) or laser scanning, which is not desirable in many applications. In this paper, we use passive stereo to obtain 3D face images and a face recognition using appearance and depth is presented. A major problem of using passive stereo is its low accuracy, and thus no passive method for 3D face recognition has been reported. Thanks to the technical progress in 3D capture/computing, an affordable real-time stereo system is now available by which one can get a comparable resolution of 3D data in real-time. In this paper, we used SRI stereo head (WWWb), in which the stereo process interpolates disparities to 1/16 pixels. Both internal and external parameters are calibrated by an automatic calibration procedure. The disparity change, Δd is $(1/16)\times 7.5\ \mu m = 0.46875\ \mu m$. Here a pixel size of 7.5 μm. We used MAGA-D stereo head, where the baseline, B, is 9cm, the focus length, f, is 16mm, Hence, when the distance from the subject to the stereo head, s, is 1m, the range resolution, i.e. the smallest change in range that is discernable by the stereo geometry,

$$\Delta r = (s^2/Bf)\Delta d = (1m^2/(90mm\times 16mm))\times 0.46875\mu m\times 10^{-3} = 0.3255\ mm$$

The range resolution is high enough for our face recognition applications, a fact verified by our experiments. The SRI Small Vision System outputs the disparity/range information automatically in real-time. The size of the left image, right image and disparity image is 320 × 240. In our experiments, the distance between the person-of-interest to the camera is about 1m to 1.5m. At this distance, the size of the face region is big enough and the disparity image is good with 16 mm lens. We calibrate the camera within this distance range in order that a good disparity image can be obtained as the distance between the person to the camera is within the distance range. In our experiments, the entire face detection, tracking, feature extraction, pose estimation and recognition can be performed in real time at 15 frames per second on a P4 3.46Ghz, 1G memory PC. When a subject is found to be a new user by the system, i.e. when the output of the reduced model is less than 0.5, the user will be registered by the system automatically.

Using the above-mentioned system, a database is built to include 116 subjects. There are 12 pairs of images (appearance and depth) for each subject. The face images are captured over a period of six months. Some normalized appearance and disparity images are shown in Fig. 12. We pick randomly any two of them for the learning gallery, while the remainder ten pairs per subject are used as probes. The average recognition rate was obtained over 66 runs. We use the same parameters for RMPM on XM2VTS database, i.e. the order of the RMPM, r, is set to be 2. b is set to be (10^{-4}). The new users that are not included in the database can be registered automatically using the method described in Section 2.4.

The recognition results for 2D plus 3D, 3D alone and 2D alone by using the proposed RMPM, SVM and KNN are given in Fig. 13 to Fig. 15 respectively. We can see that the RMPM method can yield results comparable with SVM.

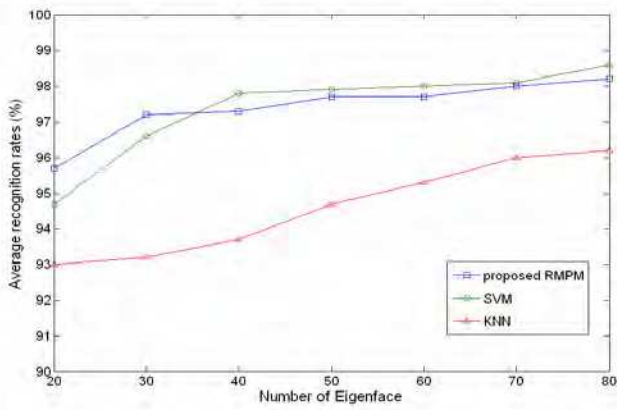


Figure 9. Recognition rates for **2D+3D** vs. number of eigenfaces on XM2VTS database

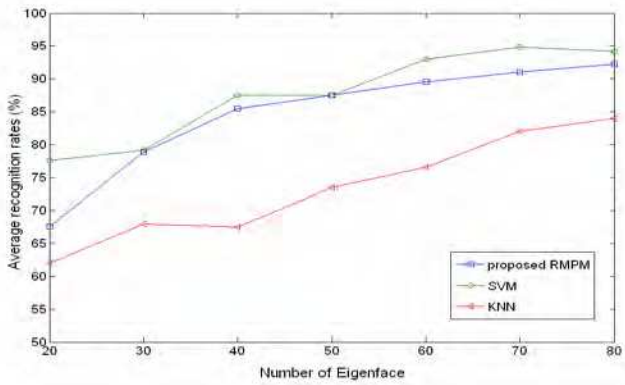


Figure 10. Recognition rates for **3D** vs. number of eigenfaces on XM2VTS database

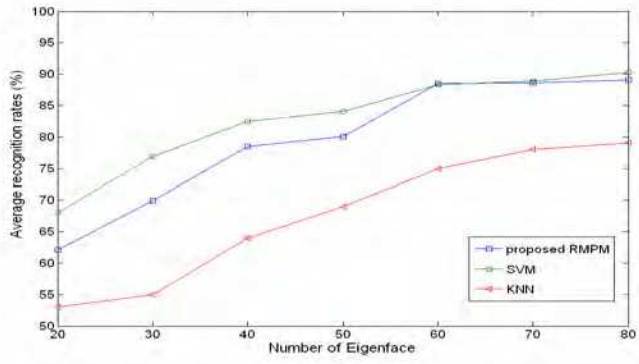


Figure 11. Recognition rates for **2D** vs. number of eigenfaces on XM2VTS database

5. Conclusion

In this paper, we contributed a stereo face recognition formulation which combines appearance and disparity/depth at feature level. We showed that the present-day passive stereovision in combination with 2D appearance images can match up to other methods which rely on active depth data. A Reduced Multivariate Polynomial Model was adopted to fuse the appearance and disparity images. RMPM is extended so that the problem of new-user registration can be overcome. We evaluated the performance of such fusion on XM2VTS face database. The evaluation results, which included results from appearance alone, depth alone and fusion of them respectively, using XM2VTS database, showed improvement of recognition rate from combining 3D information and 2D information. The performance using fused depth and appearance was found to be the best among the three tests. Furthermore, we implemented the algorithm on a real-time stereo vision system where near-frontal views were selected from stereo sequence for recognition. The evaluation results, which included results from appearance alone, depth alone and fusion of them respectively, using a database collected by the stereo vision system also showed improvement of the recognition rate by combining 3D information and 2D information. The RMPM can yield comparable results with SVM while the computation load of the RMPM is much lower than SVM.

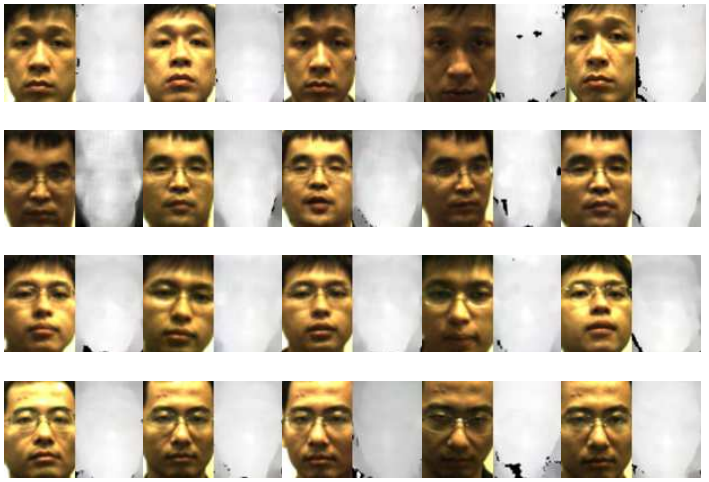


Figure 12. Normalized appearance and disparity images from the stereo vision system

The face recognition approach is anticipated to be useful for some on-line applications, such as visitor identification, ATM, and HCI. Prior to such implementations in physical systems, the performance of the system should be investigated on data with larger pose variance in terms of the verification accuracy. This is our future work. In addition, new dimension reduction method could be investigated in the future work.

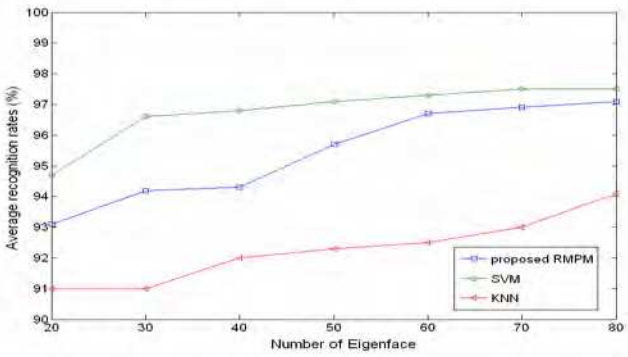


Figure 13. Recognition rates for **2D+3D** vs. number of eigenfaces on our database

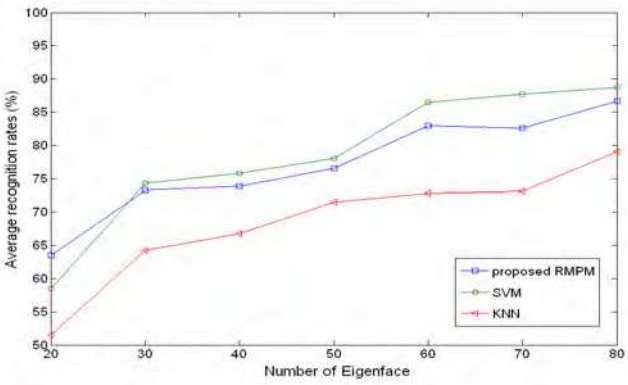


Figure 14. Recognition rates for **3D** vs. number of eigenfaces on our database

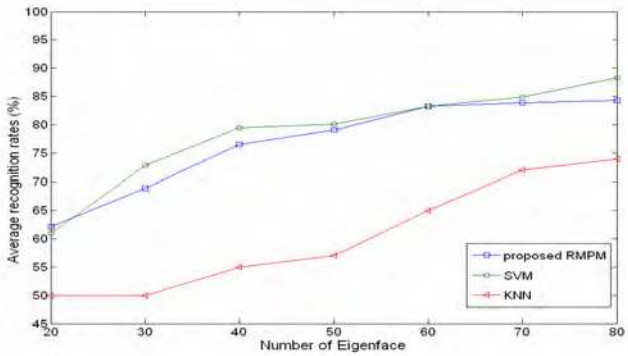


Figure 15. Recognition rates for **2D** vs. number of eigenfaces on our database

6. References

Beumier C. & Acheroy M. (1998). Automatic face verification from 3D surface, *Proceedings of British Machine Vision Conference*, pp. 449-458.

Beumier C. & Acheroy M. (2001). Face verification from 3D and grey level clues, *Pattern Recognition Letters*, Vol. 22, 1321-1329.

Birchfield, S. (1998). An elliptical head tracking using intensity gradients and color histograms, *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition*, Santa Barbara, California, pp. 232-237.

Blanz V. & Vetter T. (1999). A Morphable Model for the Synthesis of 3D Faces, *Proceedings of SIGGRAPH,* pp. 187-194.

Blanz V. & Vetter T. (2003). Face Recognition Based on Fitting a 3D Morphable Model, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 25, No. 9, 1063-1074.

Bowyer K. W.; Chang K. & P. Flynn (2006). A survey of approaches and challenges in 3D and multimodal 3D + 2D face recognition, *Computer Vision and Image Understanding*, Vol. 101, 1-15.

Bruneli R. & Falavingna D. (1995). Person identification using multiple cues, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 17, No. 10, 955-966.

Chang K.; Bowyer K.; Flynn P. (2003). Face Recognition Using 2D and 3D Facial Data, *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Nice, France, pp. 187-194.

Chang K.; Bowyer K. & Flynn P., (2004). Multi-biometrics using facial appearance, shape and temperature, *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 43-48.

Chang K.; Bowyer K. & Flynn P. (2005). An Evaluation of Multimodal 2D+3D Face Biometrics, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 27, No. 4, 619-624.

Chellappa R.; Wilson C.L. & Sirohey S. (1995). Human and machine recognition of faces, *Proceedings of the IEEE*, Vol. 83, No. 5, 705-740.

Choudhury T.; Clarkson B.; Jebara T. & Pentland A. (1999). Multimodal person recognition using unconstrained audio and video, *Proceedings International Conference on Audio and Video Based Person Authentication*, pp.176-181.

Chua C. S.; Han F. & Ho Y. K. (2000). 3D face recognition using point signature, *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 233-238.

Daniel B. R. & Martin H. (2002).Head Tracking Using Stereo Machine Vision and Applications, Vol. 13, 164–173

Darrell T.; Gordon G.; Harville M. & Woodell J. (2000). Integrated person tracking using stereo, color and pattern detection, *International Journal of Computer Vision*, Vol. 37, No. 2, 175-185.

Gordon G. (1996). Face Recognition Based on Depth Maps and Surface Curvature, *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pp.176-181.

Gu H.; Su G. & Du C. (2001). Feature points extraction from face, *Proceedings of Conference on Image and Vision Computing*, pp. 154-158.

Hess M. & Martinez M. (2004). Facial feature extraction based on the smallest univalue segment assimilating nucles (susan) algorithm, *Proceedings of Picture Coding Symposium*, San Franscisco, California.

Hong L. & Jain A.(1998). Integrating faces and fingerprints for person identification, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 12, 1295–1307.

Kittler J.; Hatef M.; Duin R. P.W. & Matas J. (1998). On combining classifiers, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, 226-239.

Lee J. C. & Milios E. (1990). Matching range Images of human faces, *Proceedings of. IEEE International Conference on Computer Vision*, pp. 722-726.

Lu X. & Jain A. K. (2005). Deformable analysis for 3D face matching, *Proceedings of 7th IEEE Workshop on Applications of Computer Vision*, pp. 99-104.

Masser K.; Matas J.; Kittler J.; Luettin J. & Maitre G. (1999). XM2VTSDB: The extended M2VTS Database, *Proceedings of AVBPA*, pp. 72-77.

Matsumoto Y. & Zelinsky A. (2000). An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. *Proceedings IEEE International Conference on Automatic Face and Gesture Recognition*, pp.499-505.

Morency L.-P.; Rahimi A.; Checka N. & Darrell T. (2002). Fast Stereo-Based Head Tracking for Interactive Environments, *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 375-380.

Neter J.; Kutner M. H.; Nachtsheim C. J. & Wasserman W. (1996). *Applied Linear Regression Models*, third ed.

Pan G.; Wu Y. & Wu Z. (2003). Investigating profile extraction from range data for 3D recognition, *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pp. 1396-1399.

Rafael M.-S.; Eugenio A.; Miguel G.-S. & Antonio G. (2005). People detection and tracking through stereo vision for human-robot interaction. *Proceedings of the Mexican International Conference on Artificial Intelligence*, *Lectures Notes on Artificial Intelligence* 3789, Ed. Springer, pp. 337-346.

Smith S. M. & Brady J. M. (1997). SUSAN-A new approach to low level image processing, *International Journal of Computer Vision*, Vol. 23, 45-78.

Socolinsky D.A.; Selinger A. & Neuheisel J.D. (2003). Face recognition with visible and thermal infrared imagery, *Computer Vision and Image Understanding*, Volumn 91, Issue 1-2, 72 – 114.

Taylor S. & Cristianini N. (2004). *Kernel methods for pattern analysis*, Cambridge University Press.

Toh K.-A. ; Tran Q.-L. & Srinivasan D. (2004). Benchmarking a reduced multivariate polynomial pattern classifier, *IEEE Transactions on PAMI*, Vol. 26, No. 6, 740-755.

Tran Q.-L.; Toh K.-A. & Srinivasan D. (2004). Adaptation to changes in multimodal biometric authentication, *Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems*, pp.981-985.

Tsalakanidou F.; Tzovaras D. & StrintzisAfzal M. G. (2003). Use of depth and colour eigenfaces for face recognition, *Pattern Recognition Letters,* Volume 24, Issues 9-10, 1427-1435

Wang J.-G. & Sung E. (1999). Frontal-view face detection and facial feature extraction using color and morphological operations, *Pattern Recognition Letters*, Vol. 20, 1053-1068.

Wang J.-G. & Sung E. (2000). Morphology-based front-view facial contour detection, In: *Proceedings of IEEE International Conference on Systems, Man and Cybernetics,* Nashville, Tennessee, USA, 8-11 October 2000.

Wang J.-G. ; Kong H. & Venkateswarlu R. (2004a). Improving face recognition rates by combining color and depth Fisherfaces," *Proceedings of 6th Asian Conference on Computer Vision*, pp.126-131.

Wang J.-G.; Lim E. T. & Venkateswarlu R. (2004b). Stereo face detection/tracking and recognition, *Proceedings of IEEE International Conference on Image Processing*, pp. 638-644.

Wang J.-G.; Toh K.-A. & Venkateswarlu R. (2005). Fusion of appearance and depth information for face recognition, *Proceedings of 5th International Conference on Audio- and Video–Based Biometric Person Authentication*, pp. 919-928.

Wang J.-G.; Kong H. & Yau W.-Y. (2006). Bilateral Two Dimensional Linear Discriminant Analysis for Stereo Face Recognition. *Proceedings of IEEE International Conference on Pattern Recognition*, pp.429-432

Wang J.-G.; Sung E. (2007). EM Enhancement of 3D Head Pose Estimated by Point at Infinity, Image and Vision Computing, *Special issue of HCI'04 Workshop on Computer Vision in Human-Computer Interaction*, Lew Michael S. et al. (eds), to appear.

Wu H.; Inada J.; Shioyama T.; Chen Q. & Simada T. (2001). Automatic Facial Feature Points Detection with SUSAN Operator, *Proceedings of Scandinavian Conference on Image Analysis,* pp. 257-263.

WWWa. Turing Institute: http://www.turing.gla.ac.uk

WWWb. Videre Design, MEGA-D Megapixel Digital Stereo Head, http://users.rcn.com/mclaughl.dnai/sthmdcs.htm

WWWc. The Face Recognition Grand Challenge, http://www.frvt.org/FRGC/

WWWd. OSU SVM package, http://svm.sourceforge.net/download.shtml

Yacoob Y. & Davis L. S. (1994). Labeling of human faces components from range data, *CVGIP: Image Understanding*, Vol. 60, No. 2, 168-178.

Zhao W.; Chellappa R.; Rosenfeld A. & Phillips P. J. (2003). Face Recognition: A Literature Survey, *ACM Computing Surveys*, 399-458.

Face Recognition

Edited by Kresimir Delac and Mislav Grgic

ISBN 978-3-902613-03-5
Hard cover, 558 pages
Publisher I-Tech Education and Publishing
Published online 01, July, 2007
Published in print edition July, 2007

This book will serve as a handbook for students, researchers and practitioners in the area of automatic (computer) face recognition and inspire some future research ideas by identifying potential research directions. The book consists of 28 chapters, each focusing on a certain aspect of the problem. Within every chapter the reader will be given an overview of background information on the subject at hand and in many cases a description of the authors' original proposed solution. The chapters in this book are sorted alphabetically, according to the first author's surname. They should give the reader a general idea where the current research efforts are heading, both within the face recognition area itself and in interdisciplinary approaches.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821