Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition Letters 26 (2005) 1292–1303
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Identity verification by relative 3-D structure using multiple facial images ☆

Jau Hong Kao *, Yen Heng Chen, Jen Hui Chuang

Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan, Province of China

Received 3 December 2003; received in revised form 27 October 2004
Available online 15 December 2004

Abstract

Identity verification is one of the critical issues in the sector of security and has been emerging as an active research area. In recent years, technologies using biological features to address problems of identity verification have attracted numerous research interests. For examples, fingerprint recognition, voice recognition and pattern of blood vessels in the retina have spanned many commercial applications. However, special and expensive equipments such as fingerprint readers and iris scanners are often required and people have to be in unpleasant poses occasionally. This paper presents a study on computer vision technique and its application in face recognition to achieve identity verification. With multiple facial images taken from different view angles, relative affine structures are computed and are used as measurements. To that end, the explicit relationship between relative affine structure and the cross ratio which is a view-invariant under perspective projection is also addressed. The proposed method neither requires camera calibration nor reconstructs 3D models. According to simulation results, the developed approach can achieve satisfactory results given the feature points of facial images.

Keywords: Identity verification; Relative affine structure; Cross ratio; Perspective projection

1. Introduction

Machine recognition of faces has been a very active research topic in recent years (Belhhumeur et al., 1997; Chellappa et al., 1995; Samal and Iyengar, 1992; Zhang et al., 1997). Face recognition technology for still and video images has potentially numerous commercial and law enforcement applications. These applications range from

☆ This work is partly supported by National Science Council of Taiwan, Republic of China, under grant no. NSC92-2213E009006 and by Ministry of Economic Affairs, Taiwan, Republic of China, under grant no. g3-EC17A02S1032.
* Corresponding author.
E-mail addresses: gis88804@cis.nctu.edu.tw (J.H. Kao), jchuang@cis.nctu.edu.tw (J.H. Chuang).

0167-8655/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2004.11.008

static matching of well-formatted photographs such as passports, credit cards, driver's licenses, and mug shots, to real-time matching of surveillance video images presenting different constraints in terms of various processing requirements. Although humans seem to recognize faces in cluttered scene with relative ease, machine recognition which often spans several disciplines such as image processing, pattern recognition, computer vision, and neural networks is a much more daunting task. In particular, the problem can be formulated as follows: Given still or video images of a scene, identify one or more persons in the scene using a stored database of faces. A complete face recognition system generally includes two main stages. The first stage is the face detection stage that determines the existence of one or more faces in an image. Techniques used in this stage involve segmentation of faces from cluttered scenes and extraction of features from the face region. The challenges are mainly due to the fact that the position, orientation and size of face regions in an arbitrary image are usually unknown (Rowley et al., 1998; Yang and Huang, 1994; Jeng et al., 1998). A survey of face detection techniques can be found in (Kriegman et al., 2002). The second stage is the recognition stage which deals with the identification and matching problems. The goal is to determine the identities of the target faces obtained in the first stage. Considering important works developed so far in the recognition stage in the engineering literature, a brief survey on the face recognition researches in recent years is provided in what follows.

Most of existing face recognition algorithms are 2D-based. In terms of the nature of the facial features utilized, these 2D algorithms can generally be divided into two major categories: structure-based approaches and statistics-based approaches. The class of structure-based ones uses structural facial features, which are mostly local structures, e.g., the shapes of mouth, nose, and eyes (Mirhosseini and Yan, 1998; Lades et al., 1994; Kanade, 1974; Phillips, 1998). In (Kanade, 1974), an automated recognition system that uses a top-down control strategy directed by a generic model of expected feature characteristics is developed. They proposed an elastic graph matching model which extracts the feature vectors from image lattices based on a set of 2D Gabor filters. The main advantage of a structure-based face recognition method is the low sensibility to irrelevant data, e.g., moving hair or background, since it only handles data of interest instead of using all image data indiscriminately. The main disadvantage of such approaches is the high complexity in feature extraction.

The statistics-based approaches basically use the whole 2D image as facial features (Belhhumeur et al., 1997; Bichsel and Pentland, 1994; Lin et al., 1997; Liao et al., 1988). In this category of approaches, the principal component analysis (PCA) exhibits particular importance (Hotta, 2003). The principal components, e.g., Eigenface (Turk and Pentland, 1991; Pentland and Turk, 1991), of training face images are calculated and then used as a set of orthonormal basis. The complete space can be represented effectively by a significant small subset of these orthonormal facial images and the dimension of the feature space of facial images is thus reduced. Moreover, theoretical neuroscience has contributed to account for the view-invariance perception, which is also the underlying idea of our work for identify verification, of universals such as the explicit perception of featural parts and wholes in visual scenes. A survey of recent developments in theoretical neuroscience for machine vision can be found in (Colombe, 2003). These unsupervised learning methods are used to make predictive perceptual models of the spatial and temporal statistical structure in natural visual scenes. In particular, given the spatio-temporal continuity of the statistics of sensory input, invariant object recognition might be implemented using a learning rule that uses a trace of previous neural activity capturing the same object under different transforms in the short time scale. By first relating a modified Hebbian rule to error correction rules and exploring a number of error correction rules that can be applied to invariant pattern recognition, Rolls and Stringer (2001) developed learning rules related to temporal difference learning. The analysis of temporal difference learning provides a theoretical framework for better understanding the operation and convergence properties of rules useful for learning invariant representations. In contrast to

structure-based approaches, statistics-based ones are more straightforward and simple. However, it happens that important local features are used with small factor of importance. As for theoretical neuroscience, it is not yet obvious whether the full power of learning rules is expressed in the brain, and the practical applications in face recognition are needed for the understanding of the performance. The work in (Rolls and Stringer, 2001) provides suggestions about how they might be implemented. Although the above 2D-based face recognition approaches produce satisfactory results under normal conditions, their performance can deteriorate quickly by varying lighting condition or large change of the viewing geometry.

As the face recognition technology is an essential tool for law enforcement agencies' efforts to combat crime, fake or duplicated facial images which can easily cheat the 2D-based facial recognition systems raise problems of interest (Chellappa et al., 1995). To avoid such problems, a few 3D model-based face recognition are proposed wherein 3D feature points are reconstructed which provide important information for facial recognition. In (Atick et al., 1995) a method based on Karhonen-Loeve expansion is developed to reconstruct 3D face features. The method is claimed to be independent on lighting conditions. In (Ya and Zhang, 1998), the reconstruction of face surface is made rotation-invariant. A similar approach based on a depth map obtained from stereo images to perform face segmentation and recognition can be found in (Lengagne et al., 1996). In (Eriksson and Weber, 1999), a model-matching approach is provided to reduce the computational cost of 3D-based facial recognition algorithms.

In this paper, we propose a novel approach to identify a person with facial images using 3D information of facial feature points. Three reference points are first extracted to construct a reference plane in every image. By calculating a view-invariant relative depth, i.e., relative affine structure with respect to the obtained reference plane introduced in (Shashua and Navab, 1996), for each relevant feature point, an efficient face recognition algorithm is developed using the robust measurement. Compared with other 3D approaches that require specific structures in Euclidean space (Atick et al., 1995; Ya and Zhang, 1998), the proposed method uses only a few facial feature points and requires no camera calibration. In addition, iterative training is not required which leads to the issue of convergence in the neural network approaches. Experimental results show that the developed approach performs satisfactorily with an experimental facial image database.

In the following sections, we first introduce related projection geometry for one and two cameras. The geometrical relationships between two cameras such as parallax and relative affine structure are discussed in Section 3, together with the geometrical meaning of such a structure which is expressed in terms of the invariant under perspective projection, i.e., cross ratio. Algorithms for face recognition using relative affine structure are presented in Section 4. Simulation results for an experimental facial image database are given in Section 5. Finally, conclusion is given in Section 6.

2. Projective geometry for one and two cameras

The basic procedure of projecting 3D points onto an image by a perspective camera can be described as

$$m \propto PM, \tag{1}$$

where $\propto$ denotes the equality up to a scaling factor, P is the 3×4 projection matrix, $M = [X\ Y\ Z\ 1]^{\mathrm{T}}$ and $m = [x\ y\ 1]^{\mathrm{T}}$ represent the homogeneous coordinates of a 3D world point and the corresponding image point, respectively. In general, the image coordinate system is defined in terms of image pixels. The general form of the projection matrix can be represented as

$$P_{\mathrm{euc}} \propto KP_0T = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} [I|0] \begin{bmatrix} R & t \\ 0_3^{\mathrm{T}} & 1 \end{bmatrix}. \tag{2}$$

In (2), K gives the intrinsic parameters of the camera, the imaging system. As for T, it describes the location and orientation of the camera with respect to the world coordinate system. It is a 4×4 matrix describing the pose of the camera in terms of a rotation R and a translation t, which give

the extrinsic parameters. For an ideal camera model, both K and T are identity matrices and (2) becomes

$$m \triangleq P_0 M. \tag{3}$$

Consider two cameras taking pictures of an object, as illustrated in Fig. 1, wherein C and C' are the two optical centers of the two cameras and v and v' are their associated image planes, respectively. The projection of C' on v, $e = PC'$, observed from C and the projection of C on v' observed from C', $e' = P'C$, are defined as the epipoles of the two cameras, respectively. Without loss of generality, we assume that the world coordinate system is aligned with the image coordinate system of camera C, thus the projection matrices for C and C' become

$$P = K_{3\times3}[I_{3\times3}|0] = [K|0], \tag{4}$$

$$P' = K'_{3\times3}[R_{3\times3}|t_{3\times1}] = [K'R|K't]. \tag{5}$$

In addition, we have, by definition,

$$PC = K_{3\times3}[I_{3\times3}|0_{3\times1}]C_{4\times1} = 0$$

or

$$C \propto [0 \quad 0 \quad 0 \quad 1]^{\mathrm{T}}.$$

Since e' is the projection of C on v'

$$e' = P'C = K't. \tag{6}$$

Consider a 3D point M whose depth is z with respect to the camera coordinate system of camera C. Its projection on the image plane v, from (3), is equal to

$$m \propto PM = K\tilde{M}$$

with

$$M = \begin{bmatrix} \tilde{M} \\ 1 \end{bmatrix} = \begin{bmatrix} zK^{-1}m \\ 1 \end{bmatrix}$$

if m is normalized as $(x, y, 1)^{\mathrm{T}}$. The projection on image plane v' is then

$$m' \propto P'M \propto K'RK^{-1}m + \frac{1}{Z}K't. \tag{7}$$

With the above geometrical relationships and coordinate transformations between two cameras, Shashua and Navab (1996) derived the view invariant relative affine structure. The following section provides a brief review, together with its explicit geometric meaning.

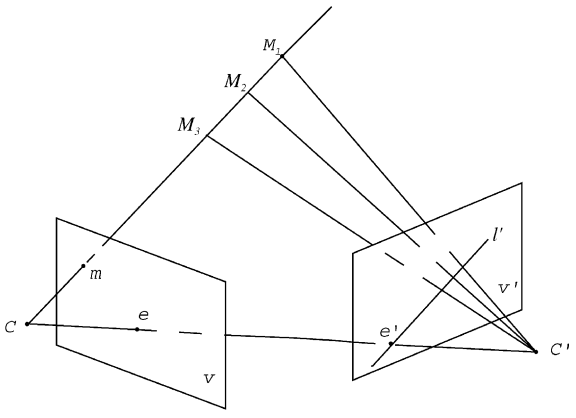

Fig. 1. A scene with two cameras and three 3D points.

3. Relative affine structure and its geometric meaning

In (Shashua and Navab, 1996), an affine framework for perspective views is proposed which is captured by a simple equation based on an invariant called relative affine structure. It is shown in (Shashua and Navab, 1996) that the framework unifies projection tasks including Euclidean, projective and affine in a natural and simple way. While the algebraic form of the relative affine structure is given clearly in (Shashua and Navab, 1996), as reviewed next, the direct relationship between the relative affine structure and a view-invariant cross ratio under perspective projection, is derived at the end of this section.

Given a reference plane π where the image points m and m' are projections of a 3D point $M_\pi \in \pi$ on image planes v and v', respectively. The homography induced by π can be obtained by $M_\pi = H_1 m$ and $M_\pi = H_2 m'$ as follows:

$$m' = H_2^{-1}M_\pi = H_2^{-1}H_1 m = H_\pi m. \tag{8}$$

Since H_π has eight entries (nine minus a scale factor), H_π can be determined uniquely by solving a system of linear equations obtained from three

point correspondences in general positions on π and the relationship $e' = H_\pi e$. Moreover, once H_π is computed we can use it to determine positions of points on π from a singe image.

The homogeneous coordinates of π can be written as

$$\pi = \begin{bmatrix} n_{3\times 1} \\ d_\pi \end{bmatrix}, \tag{9}$$

where n and d_π describe the normal vector and the depth of π, respectively. For the projection m of M_π on the image plane v, we have

$$m = PM_\pi = [K|0]M_\pi.$$

Since the depth of M_π is unknown, we can assume that

$$M_\pi = \begin{bmatrix} (K^{-1}m)_{3\times 1} \\ \rho \end{bmatrix}. \tag{10}$$

On the other hand, since M_π is on π, we have

$$\rho = \frac{-1}{d_\pi} n^{\mathrm{T}} K^{-1} m. \tag{11}$$

Now, by projecting M_π on v', we have

$$m' = H_\pi m \propto P' M_\pi = K' \left(R - \frac{t n^{\mathrm{T}}}{d_\pi} \right) K^{-1} m. \tag{12}$$

For more general scenes wherein not all of the 3D points are co-planar, parallax will be produced. For instance, M is a 3D point which is not on the plane π in Fig. 2. m'' and $H_\pi m$ are projections of M and M_π on v', respectively. From (7), (8), (12) and $e' = K't$, we have

$$\begin{aligned} m'' &\propto K'RK^{-1}m + \frac{1}{z}K't \\ &= H_\pi m + \left(\frac{z n^{\mathrm{T}} K^{-1} m + d_\pi}{d_\pi z} \right) e'. \end{aligned} \tag{13}$$

For a point $M = [zK^{-1}m \;\; 1]^{\mathrm{T}}$ which is not on the reference plane π, the distance from M to π is equal to

$$d = \pi^{\mathrm{T}} M = z n^{\mathrm{T}} K^{-1} m + d_\pi. \tag{14}$$

Substituting (14) into (13), we have

$$m'' \propto H_\pi m + \left(\frac{d}{d_\pi z} \right) e' = H_\pi m + \beta e'. \tag{15}$$

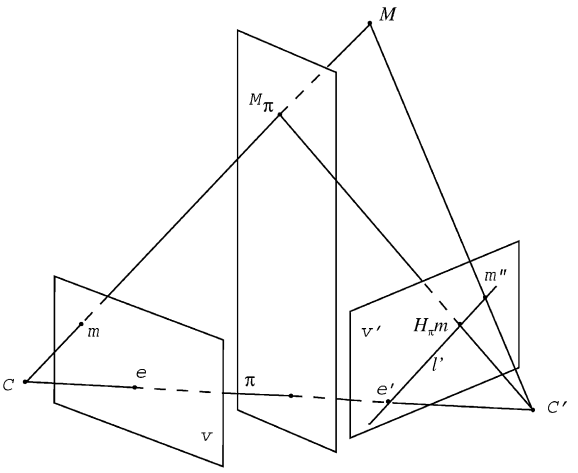

Fig. 2. An example of parallax. M is a point which is not on the reference plane π.

Since the value of the parallax term β in (15) is normalized, d_π can be dropped out, as stated in (Shashua and Navab, 1996). If we let $\beta_0 = 1$ for a reference point M_0 which is not on the reference plane (see Fig. 3), we are left with

$$d_\pi = \frac{d_0}{z_0}$$

and (15) can be rewritten as

$$m'' \cong H_\pi m + \left(\frac{z_0}{z} \frac{d}{d_0} \right) e' = H_\pi m + \lambda e' \tag{16}$$

with λ being the relative affine structure. In the following paragraph, we will investigate a different of

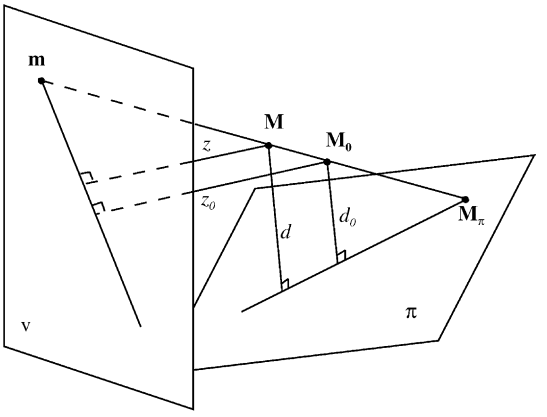

Fig. 3. The geometry of the relative affine structure. z and z_0 are depths of M and M_0 with respect to v, respectively.

relative affine structure and its relationship with cross ratio.

In (16), it is not difficult to see that λ is an invariant quantity since the variables z_0, z, d_0, d are governed by camera C only. Consider Fig. 3, by extending $\overline{MM_0}$, we can obtain two intersection points m and M_π, which are on v and the reference plane π, respectively. By triangular similarity, we have

$$\lambda = \frac{z_0}{z}\frac{d}{d_0} = \frac{\overline{mM_0}}{\overline{mM}}\frac{\overline{MM_\pi}}{\overline{M_0M_\pi}} = CR(m, M, M_0, M_\pi). \quad (17)$$

This leads to a conclusion that relative affine structure is in fact a measure of cross ratio.

Algorithm 1. Computation of relative affine structure for n pairs of image points

(1) Calculate the fundamental matrix F with 8 pairs of correspondences.
(2) Derive the epipoles e and e' using $F^{\mathrm{T}}e' = 0$ and $Fe = 0$.
(3) Derive the homography H_π of the reference plane with an epipole and 3 pairs of point correspondences.
(4) Choose a pair of correspondence m_0 and m_0' where m_0 and m_0' are image points on the left image and the right image, respectively.
(5) Scale H_π such that $m_0' \cong H_\pi m_0 + e'(\lambda_0 = 1)$.
(6) Obtain λ_i with $m_i' \cong H_\pi m_i + \lambda_i e'$, $1 \leqslant i \leqslant n-1$.

Since it is view-invariant, λ can be used as a useful feature to describe object points. Algorithm 1 summarizes the process to calculate the relative affine structure for n pairs of image points. By calculating relative affine structures of facial features of persons, we have developed an identity verification system based on face recognition using λ, as discussed next.

4. Face recognition using relative affine structures

With the properties of the view-invariant relative affine structure investigated in the previous section, this section presents the proposed approach to face recognition using such invariants. Recall that the relative affine structure of an object point is only dependent on the configuration of the first camera C, the position of the reference plane π and the reference point M_0. So, two facial images are used first to derive the relative affine structure for each feature point. The first image is denoted as the reference image and the extracted facial features are stored together with the obtained relative affine structures. To verify the identity of a new facial image, a new set of relative affine structures are obtained by the reference facial image and the new image. The similarity between the stored relative affine structures and the new set of relative affine structures is evaluated. Finally, the identity is verified by checking whether the similarity is higher than some specified thresholds.

The extraction of essential features of two facial images and the procedure adopted in this paper to obtain relative affine structures using the extracted features are explained here. To focus on the correctness of the theory, feature points are obtained manually from facial images taken from different points of view. On each given face image, fifteen feature points including eye and mouth corners, nose tip, ear lobes, etc. are extracted as shown in Fig. 4(b). The image of the front view of person A is labeled as A_f while the upward and downward looking facial images are labeled as A_u and A_d, respectively. In the same manner, three images of each of other persons are also taken. For example, Fig. 5 shows the images obtained for person B.

Table 1 shows the relative affine structures obtained for persons A and B with A_u and B_u being the reference images, respectively. Since the reference plane is defined by right ear lobe (point 14), right ear lobe (point 13) and chin (point 15), as illustrated in Fig. 6, the relative affine structure values of these three points are all zeros. The value of the relative affine structure of the nose tip (point 12), which is the reference point M_0, is defined as 1 for normalization. Since the depth from the camera to a person is usually several meters, $\frac{z_0}{z}$ in (17) is close to 1. Thus, the values of other relative affine structures given in (17) are close to $\frac{d}{d_0}$. From Fig. 6, we can see that the ratio for the eye corner is close to unity, the ratio for the mouth corner is

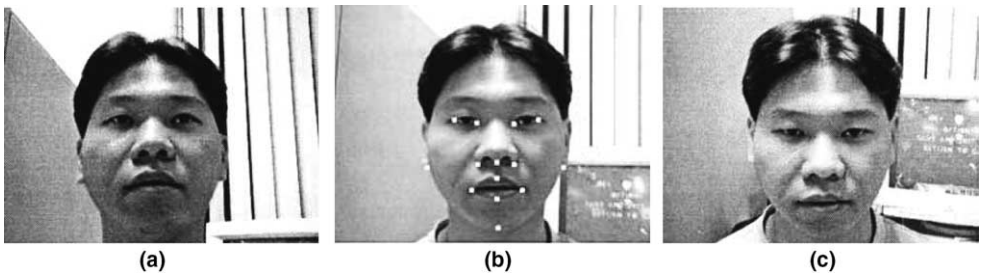

Fig. 4. Face images of person A. From left to right side, the images are labeled as A_u, A_f, and A_d, respectively.

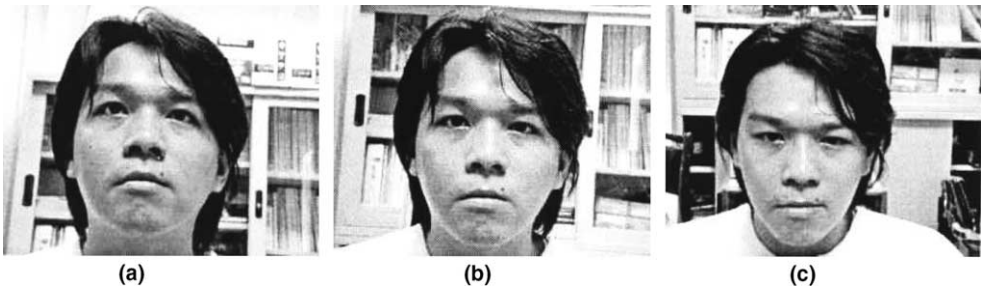

Fig. 5. Face images of person B. From left to right side, the images are labeled as B_u, B_f, and B_d, respectively.

Table 1
Relative affine structures obtained for persons A (λ_{1i}) and B (λ_{2i}) using $A_u\_A_f$ and $B_u\_B_f$, respectively

i	Feature point	λ_{1i}	λ_{2i}
1	Right eye corner (outer)	0.9951	1.0111
2	Right eye corner (inner)	0.9050	1.0391
3	Left eye corner (inner)	0.8112	1.0400
4	Left eye corner (outer)	0.7242	1.0590
5	Mouth corner (right)	0.4358	0.4594
6	Mouth corner (left)	0.4228	0.3430
7	Upper lip	0.6663	0.6598
8	Lower lip	0.4748	0.4518
9	Nose (right)	0.7256	0.7436
10	Nose (left)	0.6808	0.7281
11	Nose (center)	0.7734	0.8849
12	Nose (tip)	1.0000	1.0000
13	Ear lobe (right)	0.0000	0.0000
14	Ear lobe (left)	0.0000	0.0000
15	Chin	0.0000	0.0000

about 0.4, while the ratios for the upper and lower lips are about 0.65 and 0.45, respectively.

In our experiments, we use six groups of facial images for persons A through F (see Fig. 7 for facial images C_f through F_f). Each group consists of three images from three different points of view. With a personal computer equipped with a 333 MHz PentiumII processor and memory of 128 MB, the program implemented with MATLAB 6.1 under Microsoft Windows 2000 spends 0.1 second to obtain the relative affine structure for each data set, e.g., $A_u\_A_f$ with A_u being the reference image. A database is used to store such information obtained from the facial images. The details

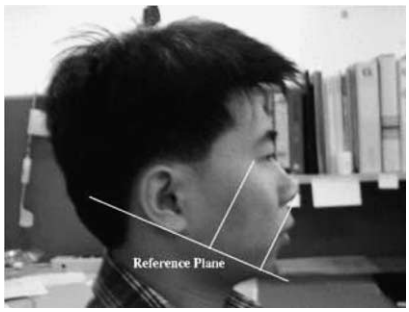

Fig. 6. Face image of side view of person F. The reference plane is defined by the two ear lobes and the chin. The 2D projections on images of these three feature points are used to calculate relative affine structures.

of the verification stage using this database and a method to improve the performance of the verification are given in the next section.

5. Experimental results

This section gives the experimental results of face recognition. For example, given the relative affine structures previously stored in the database for $X_u\_X_f$ and a facial image of an unknown person Y, we can investigate the identity of Y by evaluating the similarity between the relative affine structures for $X_u\_X_f$ and that for X_u_Y. The result of the comparison is then transformed into a score of matching error. If the score exceeds a threshold, the unknown person Y is identified not being the person X.

Table 2 shows the relative affine structure values for the fifteen facial features calculated for $A_u\_A_f$ and $A_u\_A_d$. Here, the dissimilarity between two corresponding relative affine structures, say λ_{1i} and λ_{2i}, is calculated as $Ds_i = \max(\lambda_{1i}/\lambda_{2i}, \lambda_{2i}/\lambda_{1i})$. For feature points lie on the reference plane, the relative affine structures are 0's by definition and the dissimilarity values are set to 1 (not shown). Eventually, the overall dissimilarity between these two set of relative affine structures are defined as the product of all Ds_i's. For this example, person with facial image A_d will be identified as person A since the overall dissimilarity, denoted as $Ds\_A_u\_A_f\_A_d$, is very close to 1.

Table 3 gives results similar to that in Table 2 but using facial image B_d of person B instead of A_d. It is readily observable that there are major differences between quite a few corresponding relative affine structure pairs. In particular, if $\lambda_{i1} * \lambda_{i2} < 0$, that means the feature points are not on the same side of the reference plane in the 3D

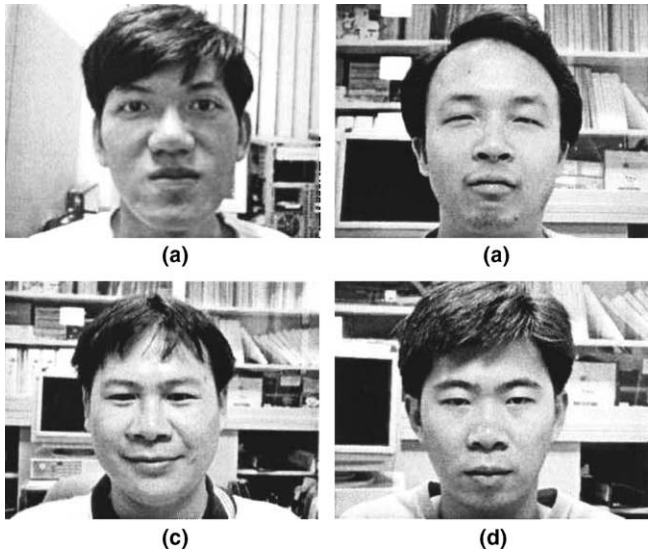

Fig. 7. Facial images (a) C_f, (b) D_f, (c) E_f, (d) F_f.

Table 2
Relative affine structures for $A_u\_A_f$ (λ_{1i}) and $A_u\_A_d$ (λ_{2i}), and their dissimilarity $Ds_i = \max(\lambda_{1i}/\lambda_{2i}, \lambda_{2i}/\lambda_{1i})$

i	Feature point	λ_{1i}	λ_{2i}	Ds_i
1	Right eye corner (outer)	0.9951	0.9510	1.0463
2	Right eye corner (inner)	0.9050	0.8961	1.0100
3	Left eye corner (inner)	0.8112	0.8183	1.0087
4	Left eye corner (outer)	0.7242	0.7189	1.0073
5	Mouth corner (right)	0.4358	0.4271	1.0204
6	Mouth corner (left)	0.4228	0.4409	1.0428
7	Upper lip	0.6663	0.6719	1.0084
8	Lower lip	0.4748	0.4871	1.0259
9	Nose (right)	0.7256	0.7243	1.0018
10	Nose (left)	0.6808	0.6853	1.0066
11	Nose (center)	0.7734	0.7593	1.0185
12	Nose (tip)	1.0000	1.0000	1.0000
13	Ear lobe (right)	0.0000	0.0000	1.0000
14	Ear lobe (left)	0.0000	0.0000	1.0000
15	Chin	0.0000	0.0000	1.0000
Overall dissimilarity				1.2141

Table 3
Relative affine structures for $A_u\_A_f$ (λ_{1i}) and $A_u\_B_d$ (λ_{2i}), and their dissimilarity $Ds_i = \max(\lambda_{1i}/\lambda_{2i}, \lambda_{2i}/\lambda_{1i})$

i	Feature point	λ_{1i}	λ_{2i}	Ds_i
1	Right eye corner (outer)	0.9951	1.0243	1.0293
2	Right eye corner (inner)	0.9050	2.3284	2.5727
3	Left eye corner (inner)	0.8112	3.5765	4.4088
4	Left eye corner (outer)	0.7242	4.5082	6.2254
5	Mouth corner (right)	0.4358	43.423	99.632
6	Mouth corner (left)	0.4228	−2.701	2.0000
7	Upper lip	0.6663	−3.049	2.0000
8	Lower lip	0.4748	−6.584	2.0000
9	Nose (right)	0.7256	−2.721	2.0000
10	Nose (left)	0.6808	−0.186	2.0000
11	Nose (center)	0.7734	−0.711	2.0000
12	Nose (tip)	1.0000	0.9999	1.0000
13	Ear lobe (right)	0.0000	0.0000	1.0000
14	Ear lobe (left)	0.0000	0.0000	1.0000
15	Chin	0.0000	0.0000	1.0000
Overall dissimilarity				4.63E+05

space, the dissimilarity value are set to 2 which leads to a big contribution to the overall dissimilarity. Since the overall dissimilarity of this example exceeds the threshold, person B is not identified as person A.

To further improve the stability of the verification system, each facial image can be used as the reference image and a composite dissimilarity measure can be obtained, which is the geometric mean of individual results. Table 4 shows the result of the verification of A_f using A_u and A_d while Table 5 shows similar results by using B_f instead of A_f.

Table 4
Verification of A_f using A_u and A_d

	Overall dissimilarity
$Ds\_A_u\_A_f\_A_d$	1.2141
$Ds\_A_f\_A_u\_A_d$	1.9270
$Ds\_A_d\_A_u\_A_f$	1.6926
Composite dissimilarity	1.5821

Table 5
Verification of B_f using A_u and A_d

	Overall dissimilarity
$Ds\_A_u\_B_f\_A_d$	499057.33
$Ds\_B_f\_A_u\_A_d$	631.10
$Ds\_A_d\_A_u\_B_f$	1955.54
Composite dissimilarity	8508.22

The composite dissimilarity 1.5821 in Table 4 indicates that the facial image A_f can be verified to be of person A. On the other hand, it is obvious that B_f is not a facial image of person A since the composite dissimilarity in Table 5 is too high.

By using the composite dissimilarity measure, a more robust identity verification system is developed and more experimental results are obtained. Table 6 shows the composite dissimilarity for the verifications of facial images A_u through F_u based on relative affine structure established using front and downward looking facial images. Similarly, Table 7 verifies facial images A_f through F_f and Table 8 verifies facial images A_d through F_d, respectively. It can be seen from these results that the threshold for similarity can be set comfortably at 2.5 for the composite dissimilarity that every person in our database can be correctly verified with the proposed approach. We can see easily that the developed identity verification system successfully performs the verification of our experimental database of facial images.

Table 6
Composite dissimilarities for the verification of facial images A_u through F_u

	A_f, A_d	B_f, B_d	C_f, C_d	D_f, D_d	E_f, E_d	F_f, F_d
A_u	**1.58**	22925.14	1509.41	16456.43	110.70	1995.84
B_u	95.19	**1.87**	439.52	1.48E+05	1.50E+06	1.23E+05
C_u	12.17	211.89	**1.94**	1.22E+05	351.99	86389.45
D_u	4.08E+05	122.75	3055.30	**1.61**	50602.31	10.73
E_u	7.85E+06	2861.12	38943.89	62857.37	**1.70**	810.17
F_u	2.89E+05	2348.07	206.42	32115.82	1.29E+07	**1.89**

Table 7
Composite dissimilarities for verification of facial images A_f through F_f

	A_u, A_d	B_u, B_d	C_u, C_d	D_u, D_d	E_u, E_d	F_u, F_d
A_f	**1.58**	13738.93	9625.68	26941.43	8.06E+06	1.02E+06
B_f	8508.22	**1.87**	372.59	18755.60	1.15E+06	8985.88
C_f	11998.25	14725.11	**1.94**	2791.98	2013.52	5650.15
D_f	3042.51	310.53	51170.67	**1.61**	425738	195.46
E_f	6979.54	1.90E+06	171595	3329.84	**1.70**	105005
F_f	151.60	2.03E+06	1146.79	1262.61	134511	**1.89**

Table 8
Composite dissimilarity for verification of facial images A_d through F_d

	A_u, A_f	B_u, B_f	C_u, C_f	D_u, D_f	E_u, E_f	F_u, F_f
A_d	**1.58**	149.81	11.38	2654.73	1336.72	162.89
B_d	705.13	**1.87**	422.78	3.18	4536.51	16885.11
C_d	10160.21	24.72	**1.94**	34498.32	6919.54	762.24
D_d	8318.70	1.09E+07	5.61E+05	**1.61**	1.75E+05	20261.21
E_d	14.15	3155.22	24.01	781.89	**1.70**	423.75
F_d	726.86	219.91	8222.20	4.88	5292.08	**1.89**

As for the sensitivity of the proposed algorithm, the relative affine structure is actually cross ratio in a form which is quite stable numerically. [1] This can be seen from Fig. 3 that the error of feature detection, in terms of variance of image pixels on the image plane, will results in minor change in the depth of the spatial structure, e.g., z and z_0, associated with a face. From above simulation results, it seems that differences among face structures of different individuals are much more significant than the differences due to the error of feature detection of facial images of the same person, which gives the robustness of the proposed approach.

6. Conclusion

This paper presents a study on computer vision technique and its application in face recognition to achieve identity verification. The explicit relationship between the relative affine structure and the cross ratio—an invariant under perspective projection, is addressed. Subsequently, relative affine structures derived from multiple images are used for face recognition. The proposed method neither requires camera calibration nor reconstructs 3D models. Moreover, as long as feature points of facial images are located accurately, the orientation and depth of the face are allowed to very more freely. As shown in our preliminary experiments, the proposed approach does achieve satisfactory results given the feature points of facial images. Slightly large scale of face database can be established for further investigation of the performance.

References

Atick, J.J., Griffin, P.A., Redlich, N.A., 1995. Face recognition from live video. Advance Imaging 10 (5), 58–62.

Belhhumeur, P., Hespanha, J., Kriegman, D., 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. 19 (7), 711–720.

Bichsel, M., Pentland, A.P., 1994. Human face recognition and the face image set's topology. CVGIP: Image Understanding 59 (2).

Chellappa, R., Wilson, C., Sirohey, S., 1995. Human and machine recognition of faces: A survey. Proc. IEEE 83 (5).

Colombe, J.B., 2003. A survey of recent developments in theoretical neuroscience and machine vision. In: Proc. 32nd Applied Imagery Pattern Recognition Workshop, pp. 205–213.

Eriksson, A., Weber, D., 1999. Towards 3-dimensional face recognition, vol. 1. In: Proc. IEEE Conf. on Africon, pp. 401–406.

Hotta, K., 2003. View-invariant face detection method based on local pca cells. In: Proc. 12th Internat. Conf. on Image Analysis and Processing, pp. 57–62.

Jeng, S.H., Liao, H.Y.M., Han, C.C., Chern, M.Y., Liu, Y.T., 1998. Facial feature detection using geometrical face model: An efficient approach. Pattern Recognition 31 (3), 273–282.

Kanade, T., 1974. Picture processing system by computer complex and recognition of human faces. Ph.D. dissertation, Robotics Institute, Carnegie Mellon University.

Kriegman, D.J., Yang, M.H., Ahuja, N., 2002. Detecting faces in images: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 24 (1), 34–58.

Lades, M., Vorbruggen, J.C., Buhmann, J., Lage, J., von der Malsburg, C., Wurtz, R.P., Konen, W., 1994. Distortion invariant object recognition in the dynamic link architecture. IEEE Trans. Comput. 42.

Lengagne, R., Tarel, J.P., Monga, O., 1996. From 2D images to 3D face geometry. In: Proc. IEEE Conf. on Automatic Face and Gesture Recognition, pp. 301–306.

Liao, H.Y.M., Han, C.C., Yu, G.J., Tyan, H.R., Chen, M.C., Chen, L.H., 1988. Face recognition using a face-only database: A new approach. In: Proc. Asian Conf. on Computer Vision, p. 1352.

Lin, S.H., Kung, S.Y., Lin, L.J., 1997. Face recognition/detection by probabilistic decision-based neural network. IEEE Trans. Neural Networks 8 (1).

Liu, J.S., Chuang, J.H., 2002. A geometry-based error estimation of cross-ratios. Pattern Recognition 35 (12), 155–167.

Mirhosseini, A.R., Yan, H., 1998. Human face image recognition: An evidence aggregation approach. Computer Vision and Image Understanding 71 (2), 213–230.

Pentland, A., Turk, M., 1991. Face recognition using eigenfaces. In: Proc. CVPR.

Phillips, P.J., 1998. Matching pursuit filters applied to face recognition. IEEE Trans. Image Process. 7 (8), 1150–1164.

Rolls, E.T., Stringer, S.M., 2001. Invariant object recognition in the visual system with error correction and temporal difference learning. Network: Comput. Neural Syst. 12, 111–129.

Rowley, H.A., Baluja, S., Kanade, T., 1998. Neural network-based face detection. IEEE Trans. Pattern Anal. Mach. Intell. 20 (1).

[1] Please see Liu and Chuang (2002) for a comprehensive error analysis of cross ratio in different forms.

Samal, A., Iyengar, P., 1992. Automatic recognition and analysis of human faces and facial expressions: A survey. Pattern Recognition 25, 65–77.

Shashua, A., Navab, N., 1996. Relative affine structure: Canonical model for 3D from 2D geometry and applications. IEEE Trans. Pattern Anal. Mach. Intell. 18 (9), 873–883.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. J. Cognitive Neurosci. 3 (1).

Ya, Y., Zhang, J., 1998. Rotation-invariant 3D reconstruction for face recognition. In: Proc. IEEE Conf. on Image Processing, vol. 1, pp. 156–160.

Yang, G., Huang, T.S., 1994. Human face detection in a complex background. Pattern Recognition 27 (1), 53–63.

Zhang, J., Yan, Y., Lades, M., 1997. Face recognition: Eigenface, elastic matching and neural nets. Proc. IEEE 85 (9).