Integrated Framework for Expression and Identity Recognition in 3D Faces

Chao Li, Armando Barreto and Malek Adjouadi
Electrical and Computer Engineering, Florida International University
{cli007, barretoa, adjouadi}@fiu.edu

Abstract

Interest in face recognition technology has recently increased in academia and industry because of its wide potential application and its importance to meet the security needs of today's world. This paper proposes a method to tackle an important problem in 3D face recognition: the deformation of facial geometry that results from the expression changes of a subject. A framework composed of three subsystems: expression recognition system, expressional face recognition system and neutral face recognition system is proposed and implemented. The recognition of faces that were neutral or exhibited one expression, i.e. smiling, was tested on a database of 30 subjects. The results proved the feasibility of this framework.

1. 3D face recognition in Biometrics

Biometrics is the science of identifying, or verifying the identity of, a person based on physiological or behavioral characteristics [1].

Face recognition is a particularly compelling biometric approach because it is the one used every day by nearly everyone as the primary means for recognition of other humans. Face recognition has a wide range of potential applications for commercial, security, and forensic purposes.

Our work focuses specifically on the Face Identification Problem: ("Who am I?"). This involves a one-to-many matching process that compares a query face image against all the gallery images in a face database to determine the identity of the query face. In the identification task, it is assumed that the person is in the database. The identification of the query image is done by choosing the image in the database that has the highest similarity score with the query image.

Most of the face recognition methods developed until recently use 2D intensity images obtained by photographic cameras as the data format for processing. Although 2D face recognition has achieved considerable success, certain problems still exist, because the 2D face images used not only depend on the face of a subject, but also depend on the imaging factors, such as the environmental illumination and the orientation of the subject. These two sources of variability in the face image often make the 2D face recognition system fail. That is the reason why 3D face recognition is believed to have an advantage over 2D face recognition.

Recently, increased attention has been directed to 3D face recognition efforts. However, there are still several challenges that this technology must overcome. Most of the 3D face recognition systems treat the 3D face surface as a rigid surface. But in reality the face surface is deformed by different expressions of the subject. So, systems which treat the face as a rigid surface are significantly challenged when dealing with faces with expressions. In [2] Chang et al. found that the performance of Iterative Closest Point (ICP) and PCA methods decline when dealing with faces displaying expressions. In [3], Bronstein et al. present a 3D face recognition approach based on a representation of the facial surface, invariant to isometric deformation by facial expression.

We attempt to tackle the expression challenge in 3D face recognition from a different point of view. Because the deformation of the face surface is always related with different expressions, we have proposed an integrated expression recognition and face recognition system. The following sections outline the rationale for our proposed system, its organization, and the algorithms used for the implementation of its key modules in a prototype meant to deal with neutral and smiling faces. The results of testing this prototype are presented and discussed.

2. Expression and identity recognition

Some psychological models propose that a connection exists between the recognition of expression and identity from faces [4]. Etcoff and Magee [5] found that people are slower in identifying happy and angry faces than they are in identifying faces with neutral expression. Other experiments show that people are slower in identifying pictures of familiar faces when they exhibit uncommon expressions [6].

In our proposed framework the incoming 3D range image is first processed by an expression recognition system to find the most appropriate expression label for it. The expression label could be one of the six prototypical expressions of the faces, which are happiness, sadness, anger, fear, surprise and disgust [7]. In addition, the face could also be labeled as 'neutral'. Therefore, the expression recognition system will yield one of seven possible expression labels. Our framework proposes that a different face recognition approach be used for each type of expression. So, a query face initially classified as 'neutral' will be processed by a recognition module optimized for neutral face identification, whereas an "expressional" face (i.e., one that displays an expression different from neutral), such as a smiling face, will be routed to an identification module for that type of face expression (e.g., a "smiling face" recognition module), where the system will find the right face by modeling the variations of the features between the neutral face and the expressional face.

Figure 1 shows an outline of this framework, which only deals with happy (smiling) expressions in addition to neutral. Smiling is the most common (non-neutral) expression displayed by people in public.

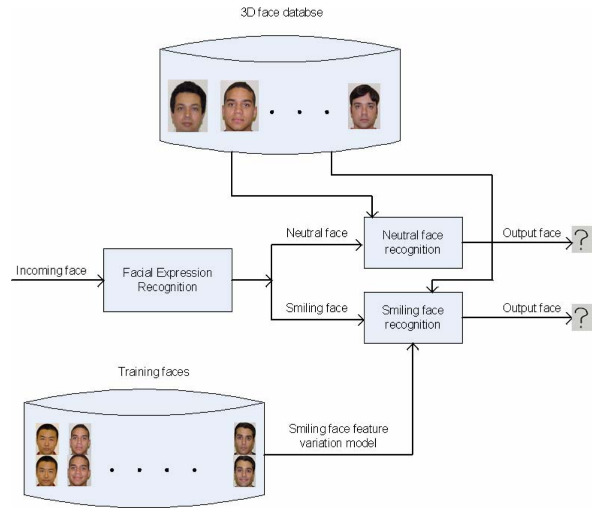

Figure 1. Simplified 3D face expression and identity recognition framework

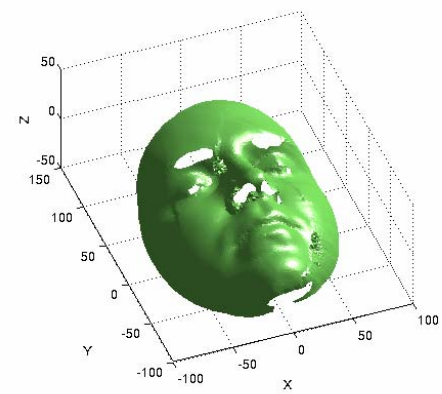

Figure 2. 3D face surface acquired by the scanner

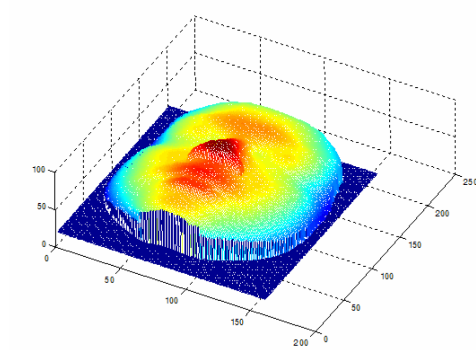

Figure 3. Mesh plot of the converted range image

3. Data acquisition and preprocessing

To test the performance of our framework, we developed a database with 3D face scans from 30 subjects, including faces with the most common expression i.e., smiling, as well as neutral faces from the same subjects. Each subject participated in two data acquisition sessions, which took place in two different days. In each session, two 3D scans were acquired using a Fastscan 3D laser scanner (Polhemus, Inc.) One was a neutral expression; the other was a smiling (happiness) expression. Figure 2 shows an example of the 3D scans obtained using this scanner.

In our experiment, a method based on the symmetric property of the face is used to register the face image. In converting the 3D scan from triangulated mesh format to a range image with a sampling interval of 1.8mm (e.g., Fig 2), trilinear interpolation was used [8]. A cubic spline interpolation method was used to patch the unavoidable holes that appear in hairy areas (e.g., eye brows). An example of the resulting 3D range image is shown in Fig 3.

4. 3D expression recognition

The first block required to implement the framework shown in Figure 1 is a "Facial Expression Recognition" module that could separate neutral faces from faces displaying other expressions. Facial expressions are generated by contractions of facial muscles, which result in temporally deformed facial features such as eye lids, eye brows, nose, lips and skin textures, often revealed by wrinkles and bulges. Typical changes of muscular activities for spontaneous expressions are brief, usually between 250 ms and 5 s. Three stages have been defined for each expression, which are onset (attack), apex (sustain) and offset (relaxation). In contrast to these spontaneous expressions, posed or deliberate expressions can be found very commonly in social interactions. These expressions typically last longer than spontaneous expressions.

In our experiment, we sought to recognize social smiles, which were posed by each subject, in their apex period. Smiling is the easiest of all expressions to find in photographs and is readily produced by people on demand. The most distinctive features associated with a smile are the bulge of the cheek muscle and the uplift of the corner of the mouth, as can be seen in Fig 4. The line on the face generated by a smiling expression is called the nasal-labial fold ("smile line").

Figure 4. Illustration of the features of a smiling face

The following steps are followed to extract the features from the face scan:

1. An algorithm is developed to obtain the coordinates of five characteristic points A, B, C, D and E in the face range image as shown in Figure 4. A and D are at the extreme points of the base of the nose. B and E are the points defined by the corners of the mouth. C is in the middle of the lower lip.
2. The first feature is the width of the mouth BE normalized by the length of AD. Obviously, while smiling the mouth becomes wider. The first feature is represented by *mw*.
3. The second feature is the depth of the mouth (The difference between the Z coordinates of points B-C and E-C) normalized by the height of the nose to capture the fact that the smiling expression pulls back the mouth. This second feature is represented by *md*.
4. The third feature is the uplift of the corner of the mouth, compared with the middle of the lower lip d1 and d2, as shown in the figure, normalized by the difference of the Y coordinates of points A-B and D-E, respectively and represented by *lc*.
5. The fourth feature is the angle of AB and DE with the central vertical profile, represented by *ag*.
6. The last two features are extracted from the semicircular areas shown, which are defined by using AB and DE as diameters. The histograms of the range (Z coordinates) of all the points within these two semicircles are calculated (Figure 5). This figure shows that the smiling face tends to have large values at the high end of the histogram because the bulge of the cheek muscle. On the other hand, a neutral face has large values at the low end of the histogram distribution.

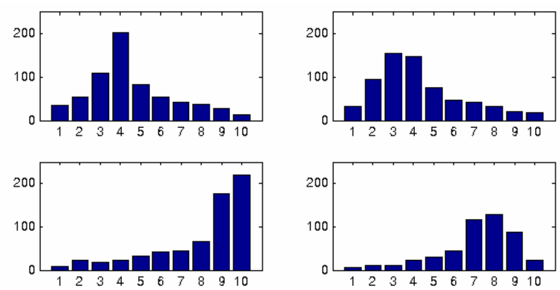

Figure 5. Histogram of range of left and right cheeks for neutral (top row) and smiling (bottom row) face

Therefore, two features can be obtained from the histograms: One is called the 'histogram ratio', *hr*, and the other is called the 'histogram maximum', *hm*:

$$hr = \frac{h6 + h7 + h8 + h9 + h10}{h1 + h2 + h3 + h4 + h5} \quad (1)$$

$$hm = i \; ; \quad i = \arg\{\max(h(i))\} \quad (2)$$

After all six features (*mw*, *md*, *lc*, *ag*, *hr* and *hm*) have been extracted this becomes a general classification problem. Two pattern classification methods were explored to recognize the expression of the incoming faces:

1) Linear discriminant classifier: (Linear Discriminant Analysis-LDA) - LDA tries to find the subspace that best discriminates different classes by maximizing the between-class scatter matrix S_b , while minimizing the within-class scatter matrix S_w in the projective subspace. S_w and S_b are defined as follows,

$$S_w = \sum_{i=1}^{L} \sum_{\vec{x}_k \in X_i} (\vec{x}_k - \vec{m}_i)(\vec{x}_k - \vec{m}_i)^T \quad (3)$$

$$S_b = \sum_{i=1}^{L} n_i (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T \quad (4)$$

where m_i is the mean vector for the individual class X_i, n_i is the number of samples in class X_i and m is the mean vector of all the samples. L is the number of classes. The LDA subspace is spanned by a set of vectors W, satisfying

$$W = \arg\max \left| \frac{W^T S_b W}{W^T S_w W} \right| \tag{5}$$

2) Support Vector Machine (SVM) - SVM relies on preprocessing the data to represent patterns in a high dimension, typically much higher than the original feature space. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two categories can always be separated by a hyperplane [9]. In our research, the LIBSVM package (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) was used to implement the support vector machine.

5. 3D face recognition

5.1. Neutral face recognition

In previous related work, we have found that the central vertical profile and the face contour are both discriminant features for every person [8]. Therefore, for neutral face recognition, the same method as in [10] is used. In this approach, the results of central vertical profile matching and contour matching are combined. The combination of the two classifiers improves the performance noticeably. The final similarity score for the probe image is the product of ranks for each of the two classifiers. The image which has the smallest score in the gallery will be chosen as the matching face for the probe image.

5.2. Smiling face recognition

For the recognition of faces labeled as 'smiling' by the expression recognition module, the probabilistic subspace method proposed by Moghaddam and Pentland [11] is used. The following paragraphs provide an outline of this method and the related principal component analysis (PCA).

Subspace methods are commonly used in computer vision, including face recognition. For example, a raw 2D image can be represented as a vector in a high dimensional space. In most cases, however, the information which needs to be extracted has a much lower dimension. That is where subspace methods such as *principal component analysis* (PCA), or the previously introduced *linear discriminant analysis* (LDA), can be applied to cope with the problem of reducing excessive dimensionality in the data to be analyzed.

5.2.1. PCA. Unlike LDA, which seeks a set of features that results in the best separation of each class, PCA seeks a projection that best represents the data in a least-squares sense. In PCA, a set of vectors are computed from the eigenvectors of the sample covariance matrix C,

$$C = \sum_{i=1}^{M} (\vec{x}_i - \vec{m})(\vec{x}_i - \vec{m})^T \tag{6}$$

where m is the mean vector of the sample set. The eigen space Y is spanned by k eigenvectors u_1, u_2, ….. u_k, corresponding to the k largest eigenvalues of the covariance matrix C.

$$\vec{y}_i = (\vec{x}_i - \vec{m})^T \left[\vec{u}_1 \vec{u}_2 \ldots \vec{u}_k \right] \tag{7}$$

The dimensionality of vector $\vec{y}_i$ is K (K<<M).

5.2.2. Probabilistic subspace method. In [11][12] Moghaddam et al. presented an unsupervised technique for visual learning, which is based on density estimation in high dimensional spaces using an eigen decomposition. The probability density is used to formulate a maximum-likelihood estimation framework for visual search, target detection and automatic object recognition. Using the probabilistic subspace method, a multi-class classification problem can be converted into a binary classification problem.

Let Δ represent the difference between two vectors in a high dimensional subspace.

$$\Delta = I1 - I2 \tag{8}$$

Δ belongs to the intrapersonal space in the high dimensional subspace if I1 and I2 are two different instances of the same subject; Δ belongs to the interpersonal or extrapersonal space if I1 and I2 are instances from different subjects. *S(Δ)* is defined as the similarity between I1 and I2. Using Bayes Rule,

$$S(\Delta) = P(\Omega_I \mid \Delta) = \frac{P(\Delta \mid \Omega_I) P(\Omega_I)}{P(\Delta \mid \Omega_I) P(\Omega_I) + P(\Delta \mid \Omega_E) P(\Omega_E)} \tag{9}$$

$P(\Delta|\Omega_I)$ and $P(\Delta|\Omega_E)$ are the likelihoods of intrapersonal space and extrapersonal space. The likelihood function can be estimated by traditional means, i.e. maximum likelihood estimation or Parzen window estimation if there are enough data available. In most cases, because of the high dimensionality of the subspace, training data are not sufficient. Subspace density estimation is another choice, which is the case in our experiment. $P(\Omega_I)$ and $P(\Omega_E)$ are *a priori* probabilities for intrapersonal and extrapersonal subspace. Thus, according to the maximum *a posteriori* (MAP) rule, if $P(\Omega_I|\Delta)$ is greater than $P(\Omega_E|\Delta)$, the two images are considered to be different instances of the same subject, otherwise, they belong to two subjects.

A method based only on Ω_I can be used to simplify the computation. This maximum-likelihood (ML) similarity measure ignores extrapersonal variations.

$$S'(\Delta) = P(\Delta \mid \Omega_I) \qquad (10)$$

In [11], it was found that the Ω_I density in (10) carries greater weight in modeling the posterior similarity used for MAP recognition. The extrapersonal Ω_E , serves a secondary role and its accurate modeling is less critical. By dropping the Ω_E likelihood in favor of an ML similarity, the results typically suffer only a minor deficit in accuracy as compared to $S(\Delta)$.

5.2.3. Subspace density estimation. Given the high dimensionality of Δ, traditional methods are not suitable for the purpose of probability density estimation. An efficient subspace density estimation method proposed in [11][12] was used. The vector space of R^N is divided into two complementary subspaces: DIFS (Difference In Feature Space), F , and DFFS (Difference From Feature Space), $\overline{F}$, as show in Figure 6. F is spanned by the first M (M<<N) eigenvectors corresponding to the largest M eigenvalues of principal component decomposition results. As derived in [11], the complete likelihood estimate can be written as the product of two independent marginal Gaussian densities.

$$\hat{P}(\Delta \mid \Omega) = \left[\frac{\exp\left(-\frac{1}{2} \sum_{i=1}^{M} \frac{y_i^2}{\lambda_i} \right)}{(2\pi)^{\frac{M}{2}} \prod_{i=1}^{M} \lambda_i^{1/2}} \right] \left[\frac{\exp\left(-\frac{\varepsilon^2(\Delta)}{2\rho} \right)}{2\pi\rho^{(N-M)/2}} \right] = P_F(\Delta \mid \Omega) \hat{P}_{\overline{F}}(\Delta \mid \Omega; \rho) \qquad (11)$$

where $P_F(\Delta \mid \Omega)$ is the true marginal density in F , $\hat{P}_{\overline{F}}(\Delta \mid \Omega; \rho)$ is the estimated marginal density in the orthogonal complement $\overline{F}$, y_i are the principal components and $\varepsilon^2(\Delta)$ is the PCA residual. From [11], the optimal value for ρ is the average of the $\overline{F}$ eigen values. In the experiment for smiling face expression recognition, because of the limited number of subjects (30), the central vertical profile and the contour are not used directly as vectors in a high dimensional subspace. Instead, they are down sampled to a dimension of 17 for the analysis.

The dimension of subspace F is set to be 10, which contains approximately 97% of the total variance. The dimension of complementary subspace $\overline{F}$ is 7.

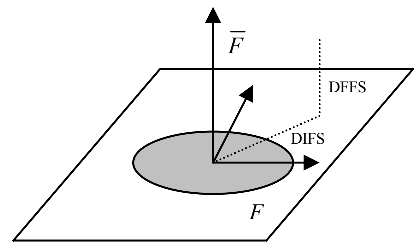

Figure 6. The principal subspace F and its orthogonal complement $\overline{F}$ for a Gaussian density

In this case also, independent ranks are computed for the central profile and the contour of each gallery face. The overall rank is found by sorting the product of these two ranks and is used to determine the final recognition result.

6. Experiments and Results

One gallery and three probe databases were formed for the evaluation of our methods. The gallery database had 30 neutral faces, one for each subject, acquired in the first data acquisition session. Three probe sets were constituted as follows and used in experiments 2 and 3.

Probe set 1: 30 neutral faces acquired in session 2.
Probe set 2: 30 smiling faces acquired in session 2.
Probe set 3: 60 faces (probe set 1 and probe set 2).

The following experiments were undertaken:

Experiment 1: Expression recognition module. The leave-one-out cross validation method was used to test the expression recognition classifier. Every time, the faces collected from 29 subjects in both data acquisition sessions were used to train the classifier and the four faces of the remaining subject collected in both sessions were used to test the classifier. The results are the average of the 30 recognition rates. Both classifiers used resulted in similarly high expression recognition rates: LDA 90.8% and SVM 92.5%.

Experiment 2: Testing the neutral and smiling recognition modules separately. In the first two sub experiments, probe faces were directly fed to the neutral face recognition module. In the third sub experiment leave-one-out cross validation was used to verify the performance of the smiling face recognition module. In each cycle, 29 subjects' faces from both acquisition sessions were used for the training and the remaining subject's smiling face from the second session was used as testing face. The probe sets used and the recognition rates achieved in each sub experiment (2-1, 2-2 and 2-3) are as follows:

2-1 Neutral face recognition (Probe set 1), neutral face recognition module: Rank 1 = 97%, Rank 3 = 97%
2-2 Smiling face recognition (Probe set 2), neutral face recognition module: Rank 1 = 57%, Rank 3 = 83%
2-3 Smiling face recognition (Probe set 2), smiling face recognition module: Rank 1 = 80%, Rank 3 = 97%

From these results, it can be seen that when the incoming faces are all neutral, the algorithm which treats all the faces as neutral achieves a very high recognition rate. On the other hand, if the incoming faces are smiling faces, then the neutral face recognition algorithm does not perform well, since only a 57% rank-one recognition rate is obtained. In contrast, when the smiling face recognition algorithm is used to deal with smiling faces, the recognition rate can go back as high as 80%.

Experiment 3: Testing a realistic scenario. This experiment emulated a realistic situation in which a mixture of neutral and smiling faces (probe set 3) must be recognized. Sub experiment 1 investigated the performance obtained if the expression recognition front-end is bypassed, and the recognition of all the probe faces is attempted with the neutral face recognition module alone. The last two sub experiments implemented the full framework shown in Figure 1. (Faces were first sorted according to expression and then routed to the appropriate recognition module.) In 3-2 the expression recognition was performed with the LDA classifier, while in 3-3 it was implemented through the SVM approach.

3-1 Neutral face recognition module only: Probe 3 used: Rank 1 = 77%, Rank 3 = 90%
3-2 Integrated expression and face recognition: Probe 3 used. (using LDA): Rank 1 = 87%, Rank 3 = 97%
3-3 Integrated expression and face recognition: Probe 3 used. (using SVM): Rank 1 = 85%, Rank 3 = 97%

It can been from these results that, if the incoming faces include both neutral faces and smiling faces the recognition rate can be improved about 10 percent by using the integrated framework proposed here.

7. Discussion

Experiment 1 was aimed at determining the level of performance of the Facial Expression Recognition Module, by itself. The average success rate in identifying the expressions of the face belonging to a subject not used for training, in each case, was 90.8% with LDA and 92.5% when SVM was used. This confirms the capability of this module to successfully sort these two types of faces (neutral vs. smiling).

The results from Experiment 2 confirmed one of the basic assumptions behind the framework proposed. That is, a system meant to recognize neutral faces may be successful with faces that are indeed neutral (Rank 1 = 97%), but may have much less success when dealing with faces displaying an expression, e.g., smiling faces (Rank 1 = 57%). On the other hand, in the third sub experiment we confirmed that a module that has been specifically developed for the identification of individuals from smiling probe images (Probe set 2) is clearly more successful in this task (Rank 1 = 80%).

Finally, Experiment 3 was meant to simulate a more practical scenario, in which the generation of probe images does not control the expression of the subject. Therefore for all three sub experiments in Experiment 3 we used the comprehensive Probe set 3, including one neutral range image and one smiling range image from each of the subjects. In the first sub experiment we observe the kind of results that could be expected when these 60 probe images are processed by a "standard" Neutral Face Recognition Module alone. This yielded a 77% rank-one face recognition, which highlights the need to account for the possibility of a non-neutral expression in 3D face recognition systems. On the other hand, in sub experiments two and three the same mixed set of images (Probe set 3) was applied to the complete framework we propose. The difference between these last two sub experiments is only the algorithm used in the Facial Expression Recognition Module. The results indicate that, regardless of the method used in the first module of the framework, this more comprehensive approach to expression and identity recognition yields improved results (87% and 85% rank-one recognition, when using LDA and SVM, respectively.)

8. Conclusions

In this paper we have presented an alternative framework proposed to enhance the performance of 3D face recognition algorithms, by acknowledging the fact that the face of a subject is a deformable surface that undergoes significant changes when the subject displays common expressions. Our main proposition is that, instead of ignoring the possibility of significant facial changes due to expressions, 3D face recognition systems should account for the potential presence of an expression in their probe images. Our suggested framework requires the development of two new functional modules, in addition to a “standard” face recognition module for neutral faces:

- A Facial Expression Recognition Module, capable to “tag” an incoming probe image with an appropriate expression label, and route it to an appropriate “specialized” face recognition classifier (matched to the expression found in the probe face), where the identity of the subject will be estimated.
- “Specialized” Face Recognition Classifiers that are trained to identify faces with expressions other than “neutral”.

In this work we have developed the framework for the simplest case in which we consider only neutral and “smiling” faces, as one very common form of expression, frequently displayed by people in public.

Our implementation of the framework, using 3 sets of probe images revealed that:

- It is possible to implement an appropriate module for the sorting of neutral vs. smiling 3D face images, based on classification of six facial features we have defined, and utilizing LDA or SVM approaches (Experiment 1).
- While a contemporary “neutral” face classifier is capable of achieving a good performance, (97% rank-one recognition), when identifying neutral 3D faces, the performance of this same classifier is much weaker (57% rank-one recognition) when dealing with otherwise comparable “smiling” faces. (Experiment 2).
- It is feasible to develop a “specialized” classifier that will identify “smiling” faces with a reasonable level of success (80% rank-one recognition), which is clearly higher than the performance of the “neutral” face classifier for the same scenario (Experiment 2).
- The framework proposed (Figure 1) is better able to identify subjects from a mixture of neutral and smiling 3D faces (87% and 85% rank-one recognition) than a standard 3D face recognition system (77% one-rank recognition) that relies on the assumption that the subjects are expressionless during the capture of the probe images (Experiment 3).

The work reported in this paper represents an attempt to acknowledge and account for the presence of expression on 3D face images, towards their improved identification. In comparison with other methods that pursue similar goals, the method introduced here is computationally efficient. Further, this method also yields as a secondary result the information of the expression found in the faces.

9. Acknowledgement

This work was sponsored by NSF grants IIS-0308155, CNS-0520811, HRD-0317692 and CNS-0426125.

10. References

[1] Bolle, R., Connell, J., Pankanti, S., Ratha, N., Senior, A., 2003. Guide to Biometrics, Springer.

[2] Chang, K., Bowyer, K., Flynn, P., 2005. Effects of facial expression in 3D face recognition. SPIE 5779: Biometric Technology for Human Identification II.

[3] Bronstein, A., Bronstein, M., Kimmel, R., 2003. Expression-invariant 3D face recognition. Proceedings of Audio & Video-based Biometric Person Authentication (AVBPA), pp. 62-69.

[4] Hansch, E., Pirozzola, F., 1980. Task Relevant Effects on the Assessment of Cerebral Specialization for Facial Emotions. Brain and Language 10: 51-59.

[5] Etcoff, N., Magee, J., 1992. Categorical perception of facial expressions. Cognition 44: 227-240.

[6] Hay, D., Young A., Ellis, A.,1991. Routes through the face recognition system. Q. J. Exp. Psychol A-Human Exp. Psy. 43: 761-791.

[7] Ekman, P., Friesen, W., 1971. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology 17(2): 124-129.

[8] Li, C., Barreto, A., 2005. Profile-Based 3D Face Registration and Recognition. Lecture Notes on Computer Science 3506: 484-494.

[9] Duda, R., Hart, P., Stork D., 2001, Pattern Classification.

[10] Li, C., Barreto, A., Zhai, J., and Chin, C., Exploring Face Recognition by Combining 3D Profiles and Contours. Proceedings of the 2005 IEEE SoutheastCon, pp. 576 - 579.

[11] Moghaddam, B., Pentland, A., 1995. Probabilistic Visual Learning for Object Detection. Proceedings of International Conferrence of Computer Vision (ICCV' 95), pp.786-793.

[12] Moghaddam, B., Pentland, A., 1997. Probabilistic Visual Learning for Object Representation. IEEE Trans. on Pattern Analysis & Machine Intelligence 19(7): pp. 696-710.