

## Multi-Subregion Based Probabilistic Approach Toward Pose-Invariant Face Recognition

Takeo Kanade<sup>1</sup>, Akihiko Yamada<sup>2</sup>

<sup>1</sup>Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213

<sup>2</sup>SANYO Electric Co., Ltd., Osaka, Japan 573-8534

E-mail: tk@ri.cmu.edu, aki-yamada@rd.sanyo.co.jp,

### Abstract

*Current automatic facial recognition systems are not robust against changes in illumination, pose, facial expression and occlusion. In this paper, we propose an algorithm based on a probabilistic approach for face recognition to address the problem of pose change by a probabilistic approach that takes into account the pose difference between probe and gallery images. By using a large facial image database called CMU PIE database, which contains images of the same set of people taken from many different angles, we have developed a probabilistic model of how facial features change as the pose changes. This model enables us to make our face recognition system more robust to the change of poses in the probe image. The experimental results show that this approach achieves a better recognition rate than conventional face recognition methods over a much larger range of pose. For example, when the gallery contains only images of a frontal face and the probe image varies its pose orientation, the recognition rate remains within a less than 10% difference until the probe pose begins to differ more than 45 degrees, whereas the recognition rate of a PCA-based method begins to drop at a difference as small as 10 degrees, and a representative commercial system at 30 degrees.*

### 1 Introduction

Many face recognition algorithms have been developed and some have been commercialized for applications such as access control and surveillance. Several studies have been reported in recent years [1, 2, 5] that compare those algorithms and evaluate the state-of-the-art of face recognition technology. These studies show that current algorithms are not robust against changes in illumination, pose, facial expression and occlusion.

Of these, pose change is one of the most important and difficult issues for the practical use of automatic face recognition. For example, the Face Recognition Vendor

Test 2000 [1], sponsored by the Department of Defense and the National Institute of Justice, reports that the recognition rate by representative face recognition programs drops by 20 percent under different illumination conditions, and as much as 75 percent for different poses.

Most algorithms [3,7,10,11] proposed so far for pose invariant face recognition need several images of each subject. We propose an approach that can recognize faces in a variety of poses even if a gallery database includes images of only one pose per person.

The method works as follows. When a probe image is given, the face region in the image is detected and its landmarks, such as the eyes, are localized. The resulting probe face region is registered with that of the face in the gallery. The face region is divided into a set of small subregions, and each subregion is compared with the corresponding subregion of the face in the gallery. To compare the two, a similarity value for each subregion, defined by the sum of squared difference (SSD) after image normalization (so effectively the same as normalized correlation), is computed after finer alignment is done in order to compensate for the potential error in registration and the local deformation due to pose and other variations. The total similarity value between the probe face and the gallery face is then obtained by combining the similarity values of all subregions.

The key idea of our approach is that in combining those similarity values of subregions we take into account how the similarity value of each subregion, and thus its utility, changes as the pose of the face changes.

We have developed a probabilistic model of that change by using a large set of training images from the CMU PIE database [4], which consists of face images of a set of people from many viewing angles. In a face recognition task of different poses, it was shown that our algorithm outperformed a baseline algorithm (PCA) and a commercial product for face recognition.

---

<sup>1</sup> This work was done while the second author was visiting the Robotics Institute of Carnegie Mellon.

## 2 Modeling Change of Local Appearance across Poses

Our approach is categorically that of appearance-based template matching. In template matching, if we use the whole face region for comparison, it is not easy to take into account changes in appearance due to pose differences, because the appearance in a different part of a face changes in a different manner due to its complicated three-dimensional shape. Instead, one can compare several subregions of the face separately, such as the eyes, nose and mouth [7,9]. It is not understood, however, which subregions provide stable and discriminative information, in particular, with respect to pose changes.

Generally, when using a similarity value, such as SSD, the value varies by three factors: the differences in the identity, the poses, and the location in the face. We decided to perform a systematic study by computing similarity values of several subregions of a face for a large number of combinations of the same and different identities and poses.

### 2.1 The CMU PIE Database

The CMU PIE database [4] consists of face images of 68 subjects  $\times$  13 poses  $\times$  21 different illumination conditions  $\times$  2 occasions. We will use part of this database in this paper. We will use only those images with frontal illumination; thus 13 images per person  $\times$  13 poses for 68 people. So, each image  $I$  in the study is labeled by  $(i, \phi)$ , where  $i$  is identity of the person ( $|\{i\}|=68$ ), and  $\phi$  is pose ( $|\{\phi\}|=13$ ). Figure 1 shows a sample set of images of different poses for one person  $i$ =Yamada. Poses  $\phi$  are denoted by symbols, like c34, c14, etc, where c27 is the frontal view and c37 and c11 are the views of about 45°.

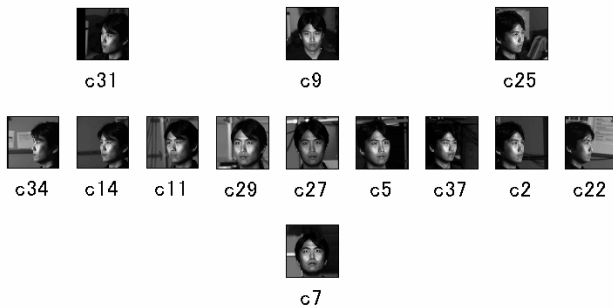


Figure 1. An example set of face images in the CMU PIE database. The database has 68 subjects with 13 poses per person, taken almost simultaneously [4]. The 13 poses cover from left profile (c34) to right profile (c22), and slightly up or down with c27 is the frontal view.

### 2.2 Local Subregions in Face and Their Similarity Values

For our study, three facial landmark points, i.e., the pupils of both eyes and the midpoint of the mouth, are manually located. The image is rotated and resized in plane so that the line that connects left and right pupils is horizontal and its length is nominal. The face region is then cropped to the size of 128 x 128 pixels.

As shown in Figure 2, a 7-by-3 lattice is placed on the face, whose position and orientation are defined by the three landmarks. Finally, we create a 9 x 15 pixel subregion centering at each of the lattice points, resulting in 21 subregions in total. For each subregion the intensity values are normalized to have zero mean and a unit variance.

As the similarity measure, the SSD (sum of squared differences) values  $s_j$  between corresponding  $j$ -th subregions for all the pair of images  $I_k = (i_k, \phi_k)$  vs.  $I_m = (i_m, \phi_m)$  in the training dataset were calculated. Note that since we compute the SSD after image normalization for each subregion, it contains effectively the same information as normalized correlation.

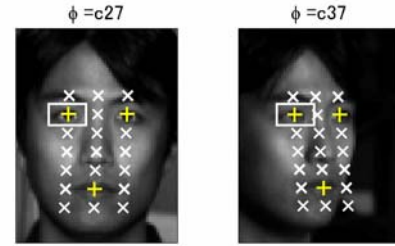


Figure 2. The facial landmark points are hand-labeled and the 7x3 lattice points are placed on the face based on their positions. The size of each subregion at the lattice points is 9 x 15 pixels.

### 2.3 Map of Similarity Values

In order to comprehend how these similarity values vary with identity and pose of the face, we plot them in two-dimensional maps. The leftmost map in Figure 3 shows the similarity values of the subregion at the right eye plotted as a two-dimensional image with gallery's identity  $i_k$  (all with  $\phi_k=c27$ ) as the horizontal axis and probe's identity  $i_m$  (all with  $\phi_m=c5$ ) as the vertical axis. Pose c27 is the frontal and pose c5 is slightly left. The darker (the smaller SSD value) the "pixel" is, the more similar are the two corresponding regions in the gallery face and probe face. Naturally, along the diagonal of the map, that is, when  $i_k=i_m$ , the map is dark, meaning the similarity is high.

The other similarity maps are for  $(i_k, \phi_k=c27)$  vs.  $(i_m, \phi_m=c37)$ ,  $(i_k, \phi_k=c27)$  vs.  $(i_m, \phi_m=c2)$ ,  $(i_k, \phi_k=c27)$  vs.  $(i_m, \phi_m=c22)$ . They correspond to cases where, while the

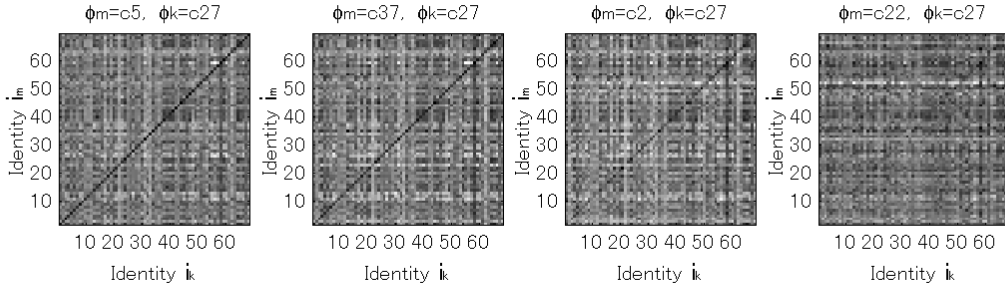


Figure 3. The two-dimensional maps of similarity values of the subregion around an eye in Figure 2. The four maps from left to right correspond to the case where the pose of one of the images  $\phi_k$  is c27 and the other pose  $\phi_m$  is c5, c37, c2 or c22, respectively. The horizontal and vertical axes are the subject's identity numbers of  $I_k$  and  $I_m$ , respectively. The darker the pixel in the maps is, the more similar the corresponding subregion is in the two images.

gallery remains to consist of frontal faces, the probe poses move from gradually left to all the way to the left profile (c22). It is clear that the similarity decreases even for the same identity (i.e., at the diagonal) as the pose moves away from frontal.

### 2.3 Prior Distributions of Similarity Values

From each similarity map like in Figure 3, we compute two histograms of similarity values. One is for the diagonal part. It represents the distribution of similarity values between face images of the same person. The other is for the non-diagonal part, which is the distribution of similarity values between faces of different people. Figure 4 shows these histograms, each one for the corresponding map in Figure 3. The histograms of the first type are shown by solid curves, and the second type by broken curves.

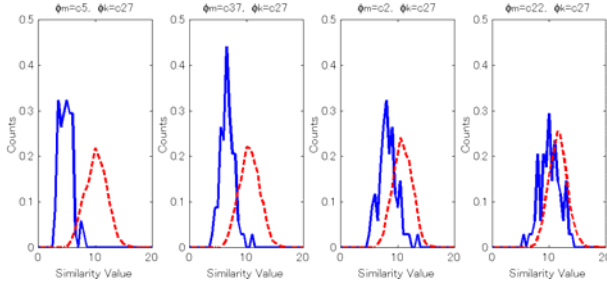


Figure 4. Each graph contains the two histograms of similarity values: their distribution for the same identity (solid curves) and that for the different identity (broken curves). The four graphs are for the combinations that corresponds to those in Figure 3..

The favorable situation is that the two histograms be as separate as possible, because that means that the similarity values of that subregion have the discriminative power to tell whether two faces are of the same person or not. It is clear that for the frontal (c27) gallery, the discriminative power of the eye subregion decreases as the pose of the probe moves from slightly left (c5), more left (c37), further left (c2), and all the way to profile (c22).

From the histograms, we create  $P(s_j | same, \phi_k, \phi_m)$ , the conditional probability density of the  $j$ -th similarity value  $s_j$  given that the images are of the class *same* identity and the poses of the two images are  $\phi_k$  and  $\phi_m$ , respectively. Likewise we also create  $P(s_j | diff, \phi_k, \phi_m)$  for the class of *different*.

In this paper, we approximate these distributions by a Gaussian distribution. Accordingly,

$$P(s_j | same, \phi_k, \phi_m) = \frac{1}{\sqrt{2\pi}\sigma_j^{same}} \exp\left[-\frac{1}{2}\left(\frac{s_j - \mu_j^{same}}{\sigma_j^{same}}\right)^2\right]$$

$$P(s_j | diff, \phi_k, \phi_m) = \frac{1}{\sqrt{2\pi}\sigma_j^{diff}} \exp\left[-\frac{1}{2}\left(\frac{s_j - \mu_j^{diff}}{\sigma_j^{diff}}\right)^2\right] \quad (1)$$

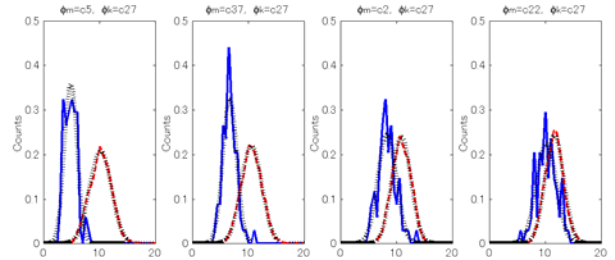


Figure 5. Histograms (solid and broken curves) and Gaussian fits (dotted curves) to them.

where  $\mu_j^{same}$  and  $\mu_j^{diff}$ ,  $\sigma_j^{same}$  and  $\sigma_j^{diff}$  are the means and standard deviations of class *same* and *diff*, respectively, which are obtained from the histograms. Figure 5 shows how these Gaussian models fit to the histograms.

## 3 Recognition

Imagine that we have a gallery images  $\{I_g=(i_g, \phi_g)\}$  and a probe image  $I_p=(i_p, \phi_p)$ . What we want to know is the probe identity  $i_p$ .

### 3.1 Posteriori Probability

For the probe image  $I_p$  and an image  $I_g$  in the gallery we compute the similarity values of all subregions, i.e.,  $\{s_1, s_2, \dots, s_J\}$ . Once we have developed probabilistic

models of similarity values of each subregion, they enable us to properly combine these similarity values, one computed for each subregion, to reach to the total decision for recognition whether the two faces are from the same identity or the different identity.

The posteriori probability that the probe image and the gallery image are of the same identity, given the  $j$ -th similarity value and their poses, is

$$P(\text{same} | s_j, \phi_p, \phi_g) = \frac{P(s_j | \text{same}, \phi_p, \phi_g)P(\text{same})}{P(s_j | \text{same}, \phi_p, \phi_g)P(\text{same}) + P(s_j | \text{diff}, \phi_p, \phi_g)P(\text{diff})} \quad (2)$$

The values  $P(\text{same})$  and  $P(\text{diff})$  are a priori probability of identify and non-identity, respectively, and the conditional densities,  $P(s_j | \text{same}, \phi_p, \phi_g)$  and  $P(s_j | \text{diff}, \phi_g, \phi_p)$ , are from the models obtained beforehand in (1).

### 3.2 Marginal Distribution for Unknown Pose of Probe

It is reasonable to assume that we have good knowledge of gallery pose  $\phi_g$ , but we may not have that for the probe pose  $\phi_p$ . In that case we cannot use (2).

One of the ways to deal with this is to determine the pose by using some pose estimation algorithms. However, the pose estimation may not be as easy.

Another way is to compute the marginal distributions of (1) over  $\phi_p$  as

$$P(s_j | \text{same}, \phi_g) = \sum_p P(\phi_p)P(s_j | \text{same}, \phi_g, \phi_p) \\ P(s_j | \text{diff}, \phi_g) = \sum_p P(\phi_p)P(s_j | \text{diff}, \phi_g, \phi_p) \quad (3)$$

Then, we can develop a posterior probability similar to (2), given the  $j$ -th similarity values and the pose of the gallery image (but not that of the probe):

$$P(\text{same} | s_j, \phi_g) = \frac{P(s_j | \text{same}, \phi_g)P(\text{same})}{P(s_j | \text{same}, \phi_g)P(\text{same}) + P(s_j | \text{diff}, \phi_g)P(\text{diff})} \quad (4)$$

### 3.3 Combining Similarity Values for Identification

Given a probe image  $I_p$ , we compute for each image  $I_g$  in the gallery (whose pose  $\phi_g$  is known) the similarity values of all subregions, i.e.,  $\mathbf{S}_{pg} = \{s_1, s_2, \dots, s_J\}$ . Then, for each of the similarity values in  $\mathbf{S}_{pg}$ , we compute (2) or (3), depending on whether the pose information of the probe is available or not. Let's denote the resultant value as  $h(\text{same} | s_j, I_g, I_p)$ .

The total similarity between  $I_p$  and  $I_g$  is now ready to be computed. Since we have not yet modeled the

probabilistic dependency among  $s_j$ 's, we chose to use the sum rule [6] in order to obtain the total similarity value,

$$H(\text{same} | \mathbf{S}_{pg}, I_g, I_p) = \sum_{j \in \mathbf{S}_{pg}} h(\text{same} | s_j, I_g, I_p) \quad (5)$$

The identity  $i_p$  is determined to be the identity  $i_g$  of the gallery image that gives the highest value of  $H$  above.

## 4 Recognition Results

We evaluate our algorithm by comparing its performance with a standard PCA-based method and a commercial product.

### 4.1 Training and Test Dataset

We used half of the subjects (34 subjects) in the CMU PIE Database as a training dataset and obtained the statistical model described above. The images of the remaining 34 subjects were used as probe images in the recognition test. The test dataset, therefore, consists of the images of 13 poses x 34 subjects. As the gallery images, we use frontal (c27) images of all 68 subjects. This makes this recognition task a little more difficult than otherwise since the gallery includes images of 34 subjects that are not included in the testset, which no probe image should match.

In this experiment below, we use  $P(\text{same}) \ll 1$ ,  $P(\text{diff}) = 1 - P(\text{same})$ ,  $P(\phi_k) = 1/13$ , as there is no better prior knowledge.

### 4.2 Case 1: When the Probe Pose is Known

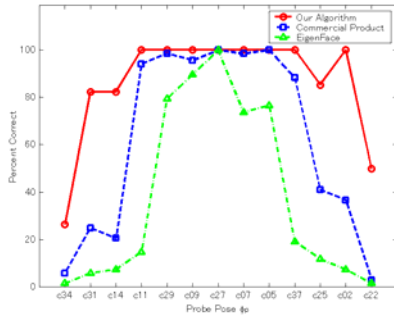
Let's assume that the probe pose was known in advance, and thus we can use equation (2). Figure 6(a) plots the recognition scores with respect to the pose of the probe. Scores of three algorithms are shown: our algorithm, a PCA algorithm, and a commercial face recognition program.

When a probe is at a frontal pose (c27), the scores are 100%, since exactly the same images are included in the gallery for the frontal pose. As the probe pose moves away from the frontal, the scores deteriorate. The greater the width at which the scores remain high, the more pose invariant the algorithm is.

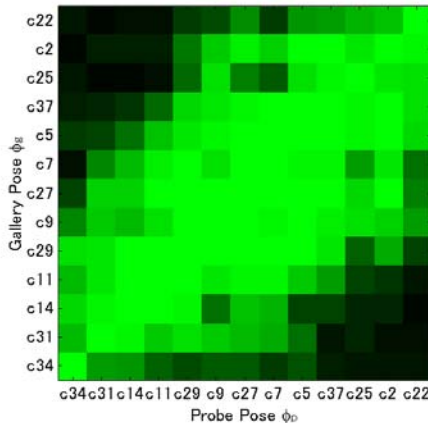
The PCA algorithm drops as soon as the probe pose moves 15° away from frontal. The commercial program maintains its high performance till 30°. Our algorithm's score stays high till 45° {c34, c31, c14, c25, c02 and c22}.

The difference in the scores by algorithms become larger as the probe poses pull away from the gallery pose c27. Especially, at poses such as  $\phi_p = \{c34, c31, c14, c25, c02 \text{ and } c22\}$ .

Figure 6(b) is a two-dimensional plot where the gallery's pose is also varied. The width of the region with high scores along the diagonal indicates the degree with which the algorithm can accommodate the difference of the pose between the gallery and the pose. When compared with 7(a) and 7(b), which are the corresponding plots for a PCA algorithm and a commercial program, respectively, our algorithm clearly outperforms them.



(a)



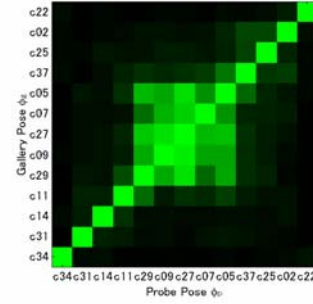
(b)

Figure 6. (a) Recognition scores with respect to probe poses, when the probe pose is known, by our algorithm, an eigenface (PCA) method, and a commercial product measured by using the CMU PIE database. (b) Recognition scores of our algorithm for all the combinations of gallery pose and probe pose.

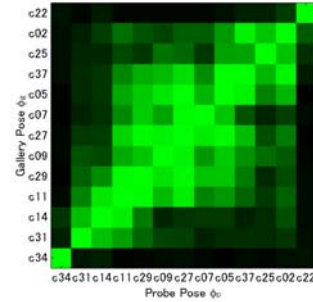
#### 4.3 Case 2: When the Probe Pose is Unknown

When the probe pose is unknown, we must use the marginal distribution method in Equation (3).

The results for the case of unknown probes pose are shown in Figure 8. The two plots correspond to those in Figure 6. They indicate that even when the probe pose is not known at all, the recognition scores are not much lower than for the case of known probe poses.

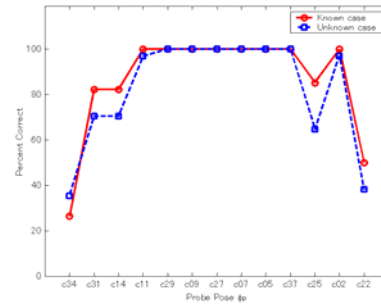


(a)

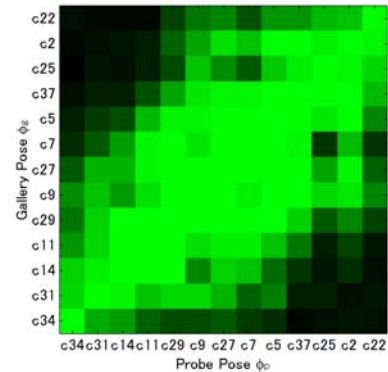


(b)

Figure 7 (b) Recognition scores for all the combinations of gallery pose and probe by an eigenface (PCA) algorithm (a) and by a commercial program (b), when probe poses are known. These should be compared to Figure 6(b).



(a)



(b)

Figure 8. (a) Comparison of recognition scores by our algorithm between the case of known probe pose and the case of unknown probe pose. (b) Recognition scores by our algorithm for combinations of probe and gallery poses when the probe pose is unknown. Compared with Figure 6(a), one can observe that the performance remains relatively the same.

#### 4.4 Discriminativeness of Subregions

An interesting question is which subregion (or which part) of a face has the most discriminating power for recognizing faces. We performed the recognition task for the gallery at the frontal c27 pose and the probe at one pose by using only one subregion at a time. This gives us 7x3 recognition scores, which we can think indicate the discriminating power of each subregion for that particular combination of gallery and probe poses. We repeat this for all the probe poses.

Figure 9 shows the results as a 7x3 “image” for each pose of probe; the brighter the “pixel” is, the more powerful is the corresponding subregion. As the probe pose changes from central to left, as shown in {c29, c11, c14...}, the right side of the face becomes more discriminating than the left side, which is very natural.

On the contrary, the left side of the face becomes more discriminating at {c5, c37, c2, c22}. For the nodding faces, such as {c9, c7}, the right and left sides of the face have almost the same discriminating power.

It is interesting to notice that the nose subregion and cheek subregion become less discriminating rather quickly, probably because the former is three-dimensional and thus changes its appearance quickly, and the latter is uniform and thus is less useful from the beginning.

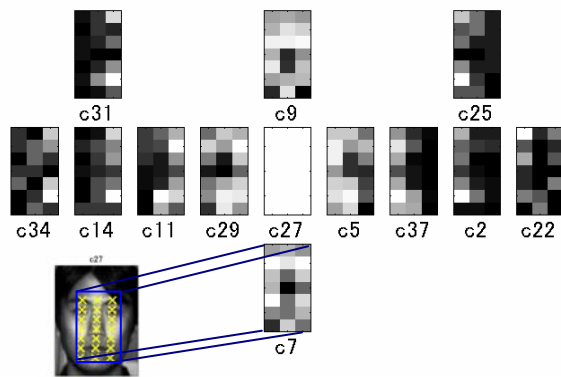


Figure 9. Maps of discriminating powers of subregions for various probe poses when the gallery pose is frontal (c27).

#### 5 Conclusion

We have proposed a face recognition method based on a probabilistic model of how the each local subregion of a face changes its appearance (thus its utility for the task of face recognition, as the pose changes). The algorithm outperformed a PCA method and a commercial product.

#### 6 References

- [1] D. Blackburn, M. Bone, and P. Phillips. Facial Recognition Vendor Test 2000: Evaluation Report, 2000.
- [2] P. J. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET Evaluation Methodology for Face Recognition Algorithms: IEEE Trans. On PAMI, 22(10): 1090-1103, 2000.
- [3] Ralph Gross, Iain Matthews and Simon Baker. Eigen Light-Fields and Face Recognition across Pose. In Proceedings of the 5<sup>th</sup> International Conference on Automatic Face and Gesture Recognition, 2002.
- [4] T. Sim, S. Baker and M. Bsat. The CMU Pose Illumination and Expression (PIE) Database. In Proceedings of the 5<sup>th</sup> International Conference on Automatic Face and Gesture Recognition, 2002.
- [5] R. Gross, J. Shi, and J. Cohn. Quo Vadis Face Recognition?: Third Workshop on Empirical Evaluation Methods in Computer Vision, December, 2001.
- [6] J. Kittler, M. Hatef, R. Duin and J. Matas. On Combining Classifiers: IEEE Trans. On PAMI. 20 (3): 226-239, 1998.
- [7] A. Pentland, B. Moghaddam, T. Starner. View-based and Modular Eigenspaces for Face Recognition. M.I.T Media Laboratory Perceptual Computing Section Technical Report No.245.
- [8] P. S. Penev and J. J. Atick. Local Feature Analysis: A General Statistical Theory for Object Representation. Network: Computation in Neural Systems 7(3), 477-500.
- [9] R. Brunelli and T. Poggio. Face Recognition: Features versus Templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.15, No.10, October 1993.
- [10] D. Beymer. Pose-Invariant Face Recognition Using Real and Virtual Views. M.I.T., A.I. Technical Report No.1574, March 1996.
- [11] T. Vetter and T. Poggio. Linear Object Classes and Image Synthesis from a Single Example Image. A.I. Memo No.1531 (C.B.C.L. Paper No.119), March 1995