Chapter 9

STEREO-BASED 3D FACE RECOGNITION SYSTEM FOR AMI

S.Ramalingam and D.Ambaye
Department of Computing, Middlesex University, UK
{s.ramalingam,d.ambaye}@mdx.ac.uk

Keywords: Ambient Intelligence, AmI, Non-Intrusive Verification and Authentication, NIVA, computer vision, face-recognition, Linear/Fisher Discriminant Analysis (LDA/FDA), Image Indexing, Stereo-Imaging.

1. Introduction

Ambient Intelligence (AmI) is the notion of technology embedded into our surroundings at work, home and leisure contexts that can be designed to make our lives more safe, more effective, less arduous and more enjoyable.

A key attribute of AmI technologies is that they should be both ubiquitous and innocuous. They are technologies embedded into our natural surroundings, present whenever needed, enabled by simple and effortless interactions, attuned to all our senses, adaptive to users and context and autonomously acting. Leading commentators suggest that the near future will see AmI applications in every day situations ranging from safe driving systems, smart buildings and home security to smart fabrics or e-textiles. It is also becoming of interest to a wide range of commercial and law enforcement applications. The key point being that they are poised to become part of every day life as we will know it.

Undoubtedly, one of the most important enablers of AmI is intelligent vision. The ability to detect, and recognise people or objects in the environment is a key pre-requisite for many AmI applications. The range of possible applications for intelligent vision is steadily expanding as advances in Ambient Intelligence (AmI) is the notion of technology embedded into our surroundings at work, home and leisure contexts that can be designed to make our lives more safe, more effective, less arduous and more enjoyable.

Although very reliable biometric methods exist, such as fingerprint analysis, retinal or iris scans, these methods are highly intrusive in respect to the overall capture to recognition cycle. For instance, such systems require users to be subjected to unnatural and repetitive identification processes. In some instances the additional process time overheads placed on the user can also be an issue. However, a personal identification system based on face images that

are frontal or partial in view can be less intrusive, more natural in environment [21] and faster. An appropriately designed face recognition technology system can enable user-friendly and fast access to an ATM machine or a computer, to control entry into restricted areas, to recognize individuals in specific areas (banks, stores).

A common issue for most face recognition systems is achieving consistent performance levels under non-standard conditions. Differing applications can bring varying levels and types of constraints which the system must cope with. It must operate under a variety of conditions, such as varying illuminations and facial expressions, it must be able to handle non-frontal facial images regardless of gender, age or race. This can have an impact on relative performances in terms of speed and accuracy. As would be expected, differing face recognition approaches currently available satisfy such constraints to varying degrees.

The Non-Intrusive Verification and Authentication (NIVA) project, of which the work described in this paper is a part, explores these and other issues in relation to face recognition in an AmI context. The project has the primary objective of establishing best practices on how to develop and implement face recognition systems that are more user-friendly and effective. It is currently investigating a range of approaches to face recognition and verification. The approach described in this paper, stereo-based 3D face recognition, is of interest because it offers particular characteristics which, taken together, could improve the usability and effectiveness of AmI applications. For instance, some potential advantages include:

- Intrinsic high levels of accuracy in recognition.
- Ease of adaptability to varying conditions in the environment (eg. pose, intensity).
- Authentication/verification in real-time with free movement to enable a more naturalistic user interaction.

Nevertheless, stereo-based face recognition systems suffer from the following disadvantages that they have not yet gained the usability in applications:

- Calibration of the stereo camera is a research problem in building robust recognition systems.
- 3D algorithms have been developed for tracking but not recognition, possibly due to the difficulty in discriminating face objects from non-face objects.
- Being 3D, any system has to cope with the handling of volumetric data and is a serious problem for large sized databases.
- Not being cost-effective as a hardware component.

We attempt to overcome these issues through the use of a commercially available stereo-camera system that is easy to set up and integrate with the vision system. We have demonstrated an easy means of handling 3D data in feature space, and used for recognition that has real-time and robust performance.

2. Face Recognition: Review

Face recognition deals with the identification or verification of one or more persons in the database. Identification verifies an unknown query image as existing in the database or not. Verification authenticates claimed identity. Face recognition systems fall into two groups, namely those that use static images and others that use video sequences. The actual technique of image processing, interpretation and recognition will actually depend on the application for which it is used. Hence a third degree of classification exists based on the application itself.

In the following sections, we review some of the recent face recognition systems based on the use of still images, 3D images as well those that employ retrieval and indexing mechanisms as part of the recognition process. We then briefly describe the proposed system that employs 3D imaging with indexing mechanisms.

2.1 Face Recognition from Still Images

A recent review paper [21] gives a thorough survey of face recognition that exist both as research and commercial systems. In most applications the images are available only in the form of single or multiple views of 2D intensity data, so that the inputs to the computer face recognition algorithms are visual only. For this reason, the literature reviewed in [21] is restricted to studies of human visual perception of faces. A summary of this detailed review is described in the following section.

Using Principal Component Analysis (PCA) many techniques have been developed:

- Eigenfaces which uses a nearest neighbour classifier.
- Feature-line-based methods, which replace the point-to-point distance with the distance between a point and the feature line linking two stored sample points.
- Fisherfaces which use linear/Fisher discriminant analysis.
- Bayesian methods, which use a probabilistic distance metric and SVM methods, which use a support vector machine as the classifier.

In face recognition applications utilising higher order statistics, independent component analysis (ICA) is argued to have more representative power than PCA. ICA is a generalisation of PCA, which de-correlates the higher order moments of the input in addition to the second order moments.

More recent methods have been concerned with both representation and recognition, so a robust system with good generalisation capability can be built. Key lessons learned in FRVT[11] were:

1 Given reasonable controlled indoor lighting, the current state of the art in face recognition is 90 per cent verification at a 1per cent false accept rate.

2 Face Recognition in outdoor images is a research problem.

3 The use of morphable models can significantly improve nonfrontal face recognition.

4 Identification performance decreases linearly in the logarithm of the size of the gallery.

5 In face recognition applications, accommodations should be made for demographic information since characteristics such as age and sex can significantly affect performance.

2.2 Face Recognition from Image Retrievals

In recent years, there has been an interesting trend in the application of relational database indexing and retrieval mechanisms being applied to image recognition. This has risen from the ever-increasing need for managing large set of images. In particular, vision applications such as face recognition and verification typically use a large database of face images as part of the system.

Automatic image retrieval systems are feasible in very limited domains [10]. Typically image retrieval mechanisms require human assisted and knowledge based assisted schemes for image understanding. We need to perform image understanding process before we perform indexing process. Current image database systems have been classified into two general groups : 1) Databases with no image understanding capabilities or 2)Vision systems with image repositories. The first type relies on alphanumeric descriptions of images stored in a database. It was pointed out in current research that these types of systems lack the ability to accurately interpret and retrieve complex images [2]. On the other hand traditional computer vision systems have addressed the latter problem, but rely on image repositories with little concern about efficient insertion, indexing and querying common in database optimization.

Indexing is vital for response critical real-time applications, be it banking or identifying a criminal during the short span of time spent in the que to the check-in desk of an air-port. With vision applications, indexing has so far been an implicit mechanism embodied within the vision system [7]. With the application of database indexing and query mechanisms, it is now possible to treat the recognition part as being separate from the retrieval part. However what makes such a system difficult is dealing with the understanding- and complexity of multi-dimensional image feature data that will be selected and used for indexing. This is a challenge for a conventional database system.

In [13] syntactic descriptions of appearances is made use of for retrieval of face images. Images are indexed based on the feature vectors, which are constructed using derivative filter outputs. The image database is a general purpose consisting of 1561 images. Given a query, the system retrieves a subset of classified objects. The system does not rank a retrieved object. The feature vectors are of fixed length. Matching is performed in two stages: in the first stage the user marks selected regions within the query image for which invariant vectors are calculated. The second stage involves spatial fitting mechanism to

derive corresponding query points. Image interpretation in this case is very computationally intensive and only possible to be performed off line. As well as the previous system, human intervention makes the system more error-prone in the query description process or an expert in the domain is required for query interpretation. The results do not inspire confidence.

In [20] face image retrieval system is presented for criminal identification. Once again the system employs human interface for image and text processing through description and visual browsing. The system relies on the presence of a domain expert during the process of interpretation. A special neural network in combination with a Kohonen's [6] MAP and Wu's LEP [19] model is used to generate a self-organized index tree. This is followed by an abstract facial image icon for retrieval. The system is tested on 100 facial images, but no results were reported.

An image retrieval mechanism involving subjectivity imprecision and uncertainty is attempted in [4]. Queries are mapped on to the database associated with statistical qualitative features, such as long, short, oval, square chin. The features once again are described in interactive manner. Elicitation of semantic attributes is a human assisted process.

Effective indexing of images [8] include R-Trees, inverted lists, weighted centre of mass, hash table indexing, two -level signature files, etc. In [9] face image retrieval technique using HMMs is presented for indexing video images. In particular it addresses the issues of illumination related to principle component analysis. The technique uses a set of local features and creates an HMM model for each local feature. To index a new image the system first needs to identify the cluster to which it belongs using the Viterbi algorithm [12]. This technique is tested on a face database of 100 depths images with various facial expressions, illuminations and occlusions. In addition 30 video sequences consisting of 25 images each of frontal faces are tested. HMM is used to find a match between the query and the database, the recognition is performed in the Eigen space. The paper reports high recognition rate based on individual features.

2.3 3D Face Recognition

True non-3D face recognition algorithms deal with 2D intensity images or 3D images derived from 2D projections in effect. Face recognition based on still images or captured frames in a video stream can be viewed as 2D image matching and recogniton; range images are not available in most commercial/law enforcement applications. Face recognition based on other sensing modalitites such as sketches and infrared images is also possible. Recently there has been a few succesful applications developed for robot navigation and face tracking with stereo imaging [5].

In [3], a system called PersonSpotter using stereo imaging is described. This system is able to capture, track, and recognize a person walking toward or passing a stereo CCD camera. It has several modules, including head a head tracker, pre-selector, landmark finder, and identifier. The head tracker determines the image regions that change due to object motion based on simple

image differences. A stereo algorithm then determines the stereo disparities of these moving pixels. The disparity values are used to compute histograms for image regions. Regions within a certain disparity interval are selected and referred to as silhouettes. Two types of detectors, skin color based and convex region based, are applied to these silhouettes. The outputs of these detectors are clustered to from regions of interest which usually correspond to heads. To track a head robustly, temporal continuity is exploited in the form of the thresholds used to initiate, track, and delete an object.

To find the face region in an image, the preselector uses a generic sparse graph consisting of 16 nodes learned from eight example face images. The landmark finder uses a dense graph consisting of 48 nodes learned from 25 example images to find landmarks such as the eyes and the nose tip. Finally, an elastic graph matching scheme is employed to identify the face. A recognition rate of about 90 per cent was achieved; the size of the database is not known.

2.4 NIVA System Overview

With the backdrop of the issues described in previous sections, this paper proposes an algorithm of indexing and retrieval that is an integral part of the recognition system and is based on multi-dimensional feature vectors . Further, the system is suitable for both 3-D and 2-D set of face images. Section 3 below describes important aspects of the proposed system. The NIVA Vision System incorporates an automatic mechanism that will enable easy indexing and retrieval from a face database. What is proposed is simplifying the features to produce a suitable index tree that narrows down search to a smaller subset of the original database. The indexed database can now be effectively used for recognition. This is a two step process that uses scores of match for retrieval and recognition. Any ambiguities of indexing are resolved through the recognition process by using appropriate distance metrics.

3. NIVA 3D Vision System

The proposed system comprises of two major modules namely, the Small Vision System (SVS) and the proposed vision system. The NIVA vision system architecture is shown in Fig. 9.1. The system is developed in Matlab6 integrated with the SVS stereo camera system. The SVS takes care of image capture, pre-processing, calibration, rectification, disparity estimation, and filtering. The database is assumed to be normalized with respect to lighting given that all images were captured under normal lighting conditions over an 8 week period, and pose variations were restricted to 2-D. Normal expressions were allowed during the capture. The NIVA VISION vision system consists of the modules classified as follows:

1 **Feature Space Representation:** The module transforms the database of face images into feature space. The query face is projected onto the feature space for further matching and indexing. The outputs from this module become the input to the next module, image indexing.

2 **Image Indexing:** This module applies the standard LDA and performs a classification. A rough set is produced with a score of match between the query and the database.

3 **Face Recognition:** The FDA is now applied on the rough set. Face is recognized with a very high accuracy.

3.1 NIVA 3D Stereo-based Face Database

The face recognition system takes as input two images from two slightly different views of the same face from the stereo camera and produces a three-dimensional image called the disparity map. The database considered here is a part of the student face database developed earlier[16]. The images were captured under normal room lighting conditions. The commercial SVS stereo camera was used to capture images and determine the disparity maps. This database was developed over a period of 2 months under normal room lighting conditions. The size of the database is 10 images of 200 student individuals from the Asian community. Each student as a subject was asked to sit before a PC on top of which was mounted the stereo camera. The subject was asked to turn from left to right going through marked dots along the wall at 10 equidistant points. This works out to approximately 18 deg bpose variation. Normal expressions were allowed. The captured image outputs appear as shown in Figure 9.2.

In the current work, a stereo imaging based 3D face database of 40 student subjects with 10 views of each is used. Stereo image pairs and disparity information are stored as 8-bit bitmaps. The image sizes vary but a typical size would be 80x80 pixels. The varying size of the images for a specific lens parameter is an advantage in discriminating one individual's features from the other.

4. Face Recognition in NIVA

Face recognition in NIVA VISION is a two step process. The first step is to obtain the candidate set of images. This step filters out the images to avoid unnecessary searches. The candidate set is processed for further matching and images are ranked according to their degree of similarity. In this paper, we link the retrieval mechanism to a vision system for face recognition. Work in progress of a database indexing and retrieval mechanism as applied to a vision system is described here. In particular, we are considering a 3D stereo-based face image retrieval system. Stereo-vision provides depth perception by merging information captured from multiple images at different viewpoints.

This work is an extension to the work on stereo face recognition using discriminant eigenvectors[15]. We first discuss the face recognition system for the sake of understanding the characteristics of the face database, followed by the image retrieval mechanism in Section 5. The face recognition system consists of the modules of performing a linear discriminant analysis on a set of sampled signatures on the 3D face image. These modules are explained below.

4.1 Fisher/Linear Discriminant Analysis

Recently, practical face recognition systems have been developed based on eigenface representations. Systems using Linear/Fischer discriminant analysis as the classifier have also been very successful. Such classifiers perform LDA training via scatter matrix analysis. For an M class classification, the within- and between-class scatter matrices S_w and S_b respectively, are computed as given by equations 9.1 and 9.2 as follows:

$$S_w = \sum_{i=1}^{M} Pr(\omega_i) C_i \tag{9.1}$$

$$S_b = \sum_{i=1}^{M} Pr(\omega_i)(m_i - m_o)(m_i - m_o)^T \tag{9.2}$$

where $Pr(\omega_i)C_i$ is the prior class probability and usually replaced by $1/M$ in practice with the assumption of equal probability. S_w and S_b show the average scatter C_i of the sample vectors x of different classes ω_i around their respective means mi:

$$C_i = E[(x - m_i)(x - m_i)^T | \omega = \omega_i \tag{9.3}$$

Similarly, S_b represents the scatter of the conditional mean vectors m_i around the overall mean vector m_o. Various measures are available for quantifying the discriminative power, a commonly used one being the ratio of the determinant of the between- and within-class scatter matrices of the projected samples:

$$V_{opt} = \arg\max \left| \frac{V^T S_b V}{V^T S_w V} \right| = [\lambda_1, \lambda_2, ..., \lambda_k] \tag{9.4}$$

Let us denote the optimal projection matrix which minimizes V_{opt} by V; then V can be obtained by solving the generalised eigenvalue problem:

$$S_b \zeta_i = \zeta_i S_w \lambda_i \tag{9.5}$$

The Fisher-face method uses a subspace projection prior to LDA to avoid the possible singularity in S_w. This is the approach followed in this paper. Let a training set of N face images represent M different subjects. The face images in the training set are two dimensional arrays of disparity values, represented as vectors of dimension n. Different instances of a person's face are defined to be in the same class and faces of different subjects to be from different classes.

For the scatter matrices defined in Equations 9.1 and 9.2, the matrix cannot be found directly from Equation 9.4 because in general the matrix S_w is singular. This stems from the fact that the rank of S_w is less than $N - M$, and in general the number of pixels in each image(n) is much larger than the number of images in the learning set N. In [1], the fisherfaces method avoids S_w being singular by projecting the image set onto a lower dimensional space so that the resulting within class scatter is non-singular. This is achieved by using

Principal Component Analysis (PCA) to reduce the dimension of feature space to $N - M$ and then applying the standard linear discriminant on the resulting separation matrix defined in Equation 9.5 to reduce the dimension to M – 1.

$$V = V_{fisher} V_{pca} \tag{9.6}$$

$$V_{pca} = \arg\max |TV^T CV| \tag{9.7}$$

$$V_{fisher} = \arg\max \frac{|V^T V_{pca}^T S_b V_{pca} V|}{|V^T V_{pca}^T S_w V_{pca} V|} \tag{9.8}$$

Equation 9.6 forms the feature vector for cluster analysis. Hence every sample in the set of N face images is projected onto this feature vector corresponding to the columns of V_{fisher} and a set of features is extracted for each sample image in the training set. Alternatively, average of feature vectors may be determined for each class. This provides a generalised feature vector for each class and minimises the number of searches during matching.

4.2 Face Classification in NIVA

Using Euclidean distance in the feature space performs the recognition task. In[1], a weighted mean absolute/square distance with weights obtained based on the reliability of the decision axis was used. We stick to the Euclidean distance measure:

$$\mathcal{D}(\mathcal{T}, \mathcal{E}) = \sum_{v=1}^{k} \frac{(\mathcal{T}_v - \mathcal{E}_v)^2}{\sum_{\mathcal{E} \in S} (\mathcal{T}_v - \mathcal{E}_v)^2} \tag{9.9}$$

where $\mathcal{T}$ and $\mathcal{E}$ are the projections of the test image and example image respectively on vector v . S is the set of image instances. Therefore, for a given face image $\mathcal{T}$, the best match is given by

$$\mathcal{E}' = \arg\min_{\mathcal{E} \in S} \mathcal{D}(\mathcal{T}, \mathcal{E}) \tag{9.10}$$

4.3 Pattern Vectors

The LDA algorithm described above is applied on the set of signatures derived from the disparity images. These signatures are a result of sampling the disparity depth image along the y-axis. This procedure is similar in approach to wavelet decompositions along possible directions[15] and can be repeated for other directions here. The result of such sampling is a set of 25 signatures across the depth image (Figs.2b and 3). Fig.3 gives an image representation of the set of all signatures obtained for a face with varying intensity levels. These depth signatures could be obtained at fiducial points on the face deriving local features and then we can apply eigenspace analysis. Here, we consider equidistant sampling points along the height of the face and use any features obtained for the eignspace analysis.

For each such signature, a set of higher order central moments[14] are obtained as features (Figs. 4 and 5). This results in a feature vector of dimension 40 subjects × 10 views × 25 signatures × 6 moments. These features are used not only as a discriminating feature amongst the individual faces but also used as automatic indices for dynamic partitioning of the database. Such a mechanism enables lower response times of match and/or retireival.

5. NIVA Dynamic Indexing to Database and Recognition

In an earlier paper, we define a ***static*** database indexing mechanism for face recognition [18] that takes the approach of a relational database system for indexing. In that paper, a cluster analysis is performed along each dimension of the feature space and a conditional check is made in the multi-dimensional feature space for classifying an image to a partition. Such a partition typically includes replication of images under different partitions. An index tree structure gives the possible candidate partitions that could be searched for at any time. This is an offline process.

At the end of this process, what the system learns is what are the common features between the various images in the database and those that make them different. If so, how could we use this knowledge to partition the database appropriately. Such knowledge is particularly useful for a vision system that recognizes faces just as in the human process of registering and recognizing faces. When a query is posed, index matching is first performed which prunes the database tree. At the second level, the actual database elements are matched as in a recognition system. The system is interfaced with an RDBMS (relational database management system). This system is currently under development progress.

In this paper, we describe the NIVA indexing mechanism that ***dynamically*** partitions the database. That is, the partitioning is biased by the query and is not a pre-process. Instead it takes place on the fly. Hence each time a new query is posed, the partitioning looks different. The same result as above is obtained but with an index tree that has a better hit rate and smaller size of target sets. The target set is now pruned by using the Fisher's Discriminant Analysis (FDA) and a distance metric as described in Section 4. The result is higher efficiency with real-time performance and can adapt to an incremental learning very easily. (This is an improvement the static database partitioning).

6. NIVA Implementation of Indexing and Recognition

In this section, we describe in detail the indexing mechanism that is crucial for a high hit rate. The key to success of an efficient index tree is through the use of multivariate analysis on the feature set that provides the discriminating ability.

Much work in visual object recognition deal with different views of objects which are analysed in a way that allows access to view-invariant descriptions. Generalisation from one profile viewpoint to another is poor, though generalisation from one three-quarter view to the other is very good. Fortunately, for

face recognition the differences in the 3D shapes of different face objects are not dramatic. This is especially true after the images are aligned and normalised. Using a statistical representation of the 3D heads, PCA was suggested as a tool for solving the parametric SFS problem. The inherent nature of the 3D volumetric data and the model-based approach adopted in NIVA enable sufficient overlap of information in the sample set. The usefulness of higher order moments in vision have been demonstrated in earlier papers [14]. We use second and higher-order moments as a feature set and build a view invariant description across the canonical views. The same set of features is also used to index the database.

6.1 Feature Space

Let M be the set of unique faces in the database and N be the number of samples per face. Also let Z be the number of signatures per image. In NIVA, $M = 40$ and $N = 10$ and $Z = 25$, typically. Let $I = \{I_i \mid i = 1, ..., M.N\}$ be the set of images; each image is described by 25 signatures, so $I = \{I_i \mid i = 1, ..., Z\}$ is the set of signatures of the i-th image. The signatures are characterized by the set of 6 central moments, which are real numbers. We know that the first central moment is zero. Hence $I_{ij} = \{m_{ij}^k \mid k = 2, ..., 6\}$ is the set of the central moment values for the j-th signature of the i-th image from the database. In the discussion that follows index i refers to the image number in the database, index j to the signature's number, and index k to the moment's number.

Let the central moments be represented by the set S given by:

$$S_j^k = \{m_{ij}^k \mid i = 1, ..., M.N, j = 1, ..., N\} \tag{9.11}$$

which defines the k-th moment of the j-th signature of all images.

The sets

$$S^k = \bigcup_{j=1}^{Z} S_j^k \tag{9.12}$$

describe the partitioning of the k-th moments of all signatures of all images, and, finally, the set

$$S = \bigcup_{k=2}^{7} S^k \tag{9.13}$$

contains all the information about all the moments of all signatures for all images in the database. For computational purpose, the set S could be thought of as a 3-dimensional matrix of dimension 400 × 25 × 6.

6.2 Query Processing

The query Q is a single image to be matched with the ones from the database; the information about the query is then contained in 25 signatures, each of which

is characterized by 6 central moments. Then, we have the following:

$$Q_j^k = \{m_{Qj}^k \mid \forall j \in N, \forall k \in Z\} \tag{9.14}$$

$$Q = \bigcup_{k=1}^{Z} Q_j^k \tag{9.15}$$

Equations 9.14- 9.15 take the same definitions as in Equations 9.11- 9.15 except that it is for a specific query.

Let $\overline{S}^k$ and $\overline{Q}^k$ be the means of moments and signatures respectively taken over the samples of individual face images of the database and the query respectively. These contain a generalised representation for every subject in the database. For computational purpose, the set S is a 3-dimensional matrix of dimension 40 × 25 × 6.

The next step is a projection of the central moments of the query onto the feature space. This is achieved by applying the LDA on the feature set. The LDA classifier requires as input, the training set given by Equation 9.12, the query given by Equation 9.15 and a group vector to identify the group to which each sample of the training set belongs. The classifier determines the group into which each sample is classified and by computing the distance metric given in Equations 9.9 and 9.10. The result is a classified subset of the original database, based on evidence accumulation using the distance metric:

$$F_j = \{f_{ij} = \sum_{k=2}^{2} \delta(\overline{S}^k, \overline{Q}^k) \mid \forall j \in N, \forall k \in Z\} \tag{9.16}$$

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \tag{9.17}$$

and f_{ij} is a frequency count of the moments of the j-th signature of the query Q and the image I_j . The sets F_j are aggregated to produce set F:

$$F_j = \{f_{ij} = \sum_{j=1}^{Z} f_{ij} \mid i = 1, ..., M\}, \tag{9.18}$$

where f_i is the accumulated matches between all the moments of the query and i-th image from the database.

NB: The result of such a matching procedure is a vector with integer entries, describing only the frequency of the match.

6.3 Step 1: Image Indexing

Equation 9.18 gives a matching score for all the images in the database as a first cut. What we need to extract is a subset of candidate images close to match with the query. This is achieved by applying a simple threshold on the scores as follows, assuming equal probability of occurrence of $j \in Z$:

Let

$$C = |F|$$
$$C \subseteq I$$

Let

$$\delta = \sum_i \frac{f_i}{C} \tag{9.19}$$

be the threshold applied for selection. Then we have candidate sets for recognition satisfying the threshold condition as follows:

$$F^0 = \{f_i | f_i \geq \delta\} \tag{9.20}$$

The cardinality of this set is expected to be small and is passed on to the FDA module of the NIVA system.

6.4 Step 2: Face Recognition

We now use the set defined in Equation 9.20 as the most likely set for of matches for the query and perform the Fisher's Discriminant Analysis (FDA) for recognition. This module is as described in Section 3.1.

A key feature of the two stage process of indexing and recognition is that, we start with a generalised representation of the models and hence the indices are deduced from these representations. However, for the recognition process, since the number of possible candidates is narrowed down to a small subset we can afford to have individual representations in terms of the various views of the models. We infer from the test results that the generalised representations are not only compact but provide high discriminative ability so as to index into a narrow candidate subset. On the other hand, the view expansions enable to pin-point the specific pose(s) of a query.

7. Testing and Results

In general, vision systems are tested in a controlled environment against the following criteria of validation, generalisation and rejection. Validation tests are conducted to verify if the system recognizes seen instances of an object. Generalises checks if the system has sufficient generalisation ability to recognize unknown instances of known objects. Rejection simply tests if the system is capable of rejecting objects that do not belong to the database, which is a difficult task. In NIVA vision, three main categories of classification were performed based on representations of the database, query and recognition as shown in table below: Table 9.1a indicates that the database at the time of indexing always maintains the compact representation. The query may take either the compact or individual view representations. Note that the compact representations in fact are the generalised representations derived from the canonical views. Likewise the database representation for recognition through FDA may take either of the representations. Whatever is the condition, the recognition performance is tested to be the same.

Table 9.1a. Feature Representations Used for Testing.

Test	Query	Index	Recognition
1	Generalised	Generalised	Generalised
2	Generalised	Generalised	View Based
2	View based	Generalised	View Based

Table 9.1b. Smaple-set sizes for the databases and Query.

Test	NIVA Module	Sample-sets	Representation
Indexing	Query	8,10	Generalised, View Based
Indexing	Database	8,10	Generalised
Recognition	Query	View Based	View Based
Recognition	Database	8,10	Generalised, View Based

Table 9.1b indicates the possible combinations used in NIVA's testing module. Typically, 10 samples/face are used in the training set both for indexing and for recognition. Other combinations, with fewer samples are also tested. Typically 8 samples in any order are chosen. In this case, the queries specifically include the left out samples. This tests the ability of the system to generalise with fewer samples. Every view of the face is used in testing.

Again, whatever might be the sample size, the representation could be one of two combinations, namely, generalisation (compact) or view-dependent (each view is treated as a sample).

7.1 Indexing and Recognition Performance

NIVA's performance is measured at two levels, namely indexing and recognition. The recognition performance is indeed influenced by that of the indexing module. In this respect, certain performance combinations take place. For instance, hits refer to direct recognition of the query as a result of indexing. There is no doubt on the match. Hence there is no need to proceed with the recognition process. Misses refer to the state when the indexing mechanism completely misses out the right match. No further recognition is carried out. Very few cases such as these occur. Other cases are enumerated below.

Resolving Cases.

1 The hit list has two top ranks: This is resolved automatically through FDA.

2 Query is ranked 1 in the indexed list, but other faces also have same ranking: resolved automatically through FDA. Frequency of occurrence is high.

3 Not in the top of the list. Along with one or more indices of same rank: resolved automatically through FDA. Frequency of occurrence is high.

4 Misses in the index list. No automatic resolution as yet. However, if threshold level is brought to lower that than what is fixed, it might be possible.

5 In the indexed hit list, but missed during recognition: Never occurs.

Ranked indexing. Table 9.2 shows ranks along the first column. These mean that the query exists in the indexed list with the rank specified. This is then followed by the recognition process.

Tables 9.2- 9.4 give part of the results carried out. It is to be noted that the recognition performance 9.3 is 100 per cent in all cases. If the indexing mechanism missed out the actual match, then recognition fails. This brings down the overall performance of the NIVA VISION system as indicated by the overall performance in Table 9.4. However, as long as the query is in the index list, whatever might be its ranking, FDA recognizes the query without fail. The performance is in real-time.

Table 9.2. Indexing Performance.

Test → Ranks ↓	Sample-sets for Database and Query 10/10	10/1-8	10/3-10	1-8/1-8
	Test 1	Test2	Test3	Test4
Hits	42.5	37.5	40.0	37.5
1	37.5	37.5	40.0	45.0
2	7.5	15.0	10.0	12.5
3	7.5	5.0	5.0	5.0
5	-	-	2.5	-
Miss	5.0	5.0	2.5	-
Total	95	95	97.5	100

Table 9.3. Recognition Performance on Indexed Subset.

Test →	Sample-sets for Database and Query 10/10	10/1-8	10/3-10	1-8/1-8
	100	100	100	100

Table 9.4. Overall (NIVA) Performance.

Test →	Sample-sets for Database and Query 10/10	10/1-8	10/3-10	1-8/1-8
	95	95	97.5	100

7.2 Conclusion and Future Work

This paper has described a two-stage process of 3D face recognition through indexing and FDA matching suited to AmI contexts. The system known as NIVA Vision has an indexing mechanism that has proved to be very efficient in narrowing down the matching sets, in computational capacity, and response time. The system has been tested on a moderate size database with 400 images in all. The simplicity of the algorithms has contributed to the performance of system. The use of higher order statistics has provided high degree of discrimination between classes. The results are very promising for pursuing further research in stereo-based face recognition.

Future work will focus on optimization and validation type activities including the following:

1 This work forms the basis for testing on an extensive 3D face database [16] that has been constructed with stereo-camera. Future work includes testing the system with video sequence that is part of the above database.

2 The suitability of extending the system to meet the needs of real-time remote authentication [17] application will be explored.

3 The indexing and recognition mechanisms proposed in [18] will be improved integrating the system to a relational database system, to make it suitable for handling very large databases.

4 Validation of this approach relative to the needs of particular AmI contexts, including e-learning and information systems transactions.

AmI contexts place high requirements in terms of performance, response and usability on vision systems. The NIVA system is a useful step towards improving both the effectiveness and usability of face recognition in AmI contexts.

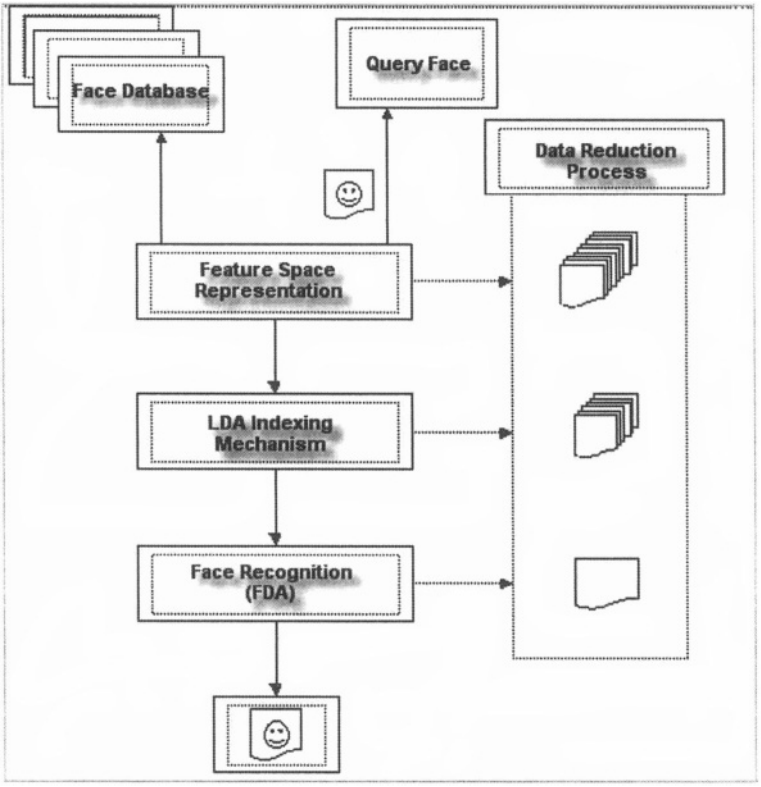

Figure 9.1. NIVA VISION Architecture

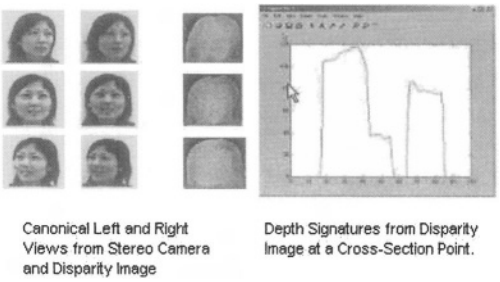

Figure 9.2. Depth Signature from Disparity Image at a Cross-Section Point

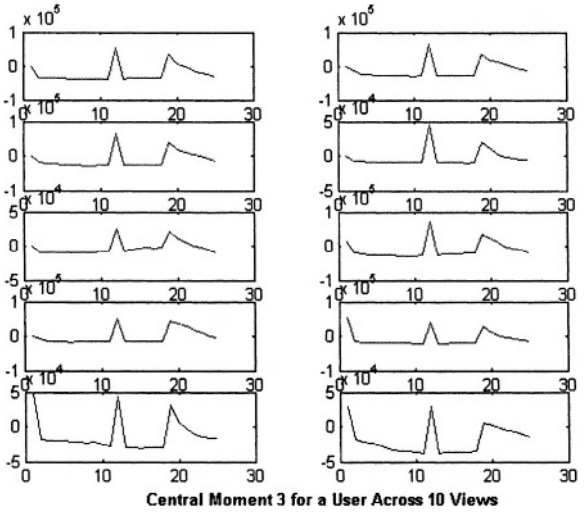

Figure 9.3. Central moments (order=2) on a set of signatures on the disparity image

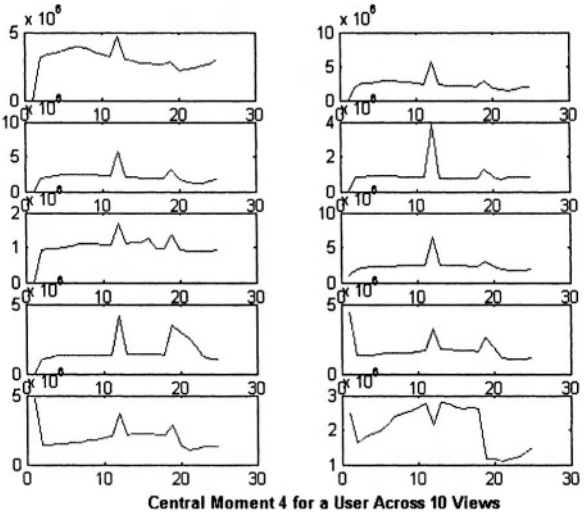

Figure 9.4. Central moments (order=3) for the same user

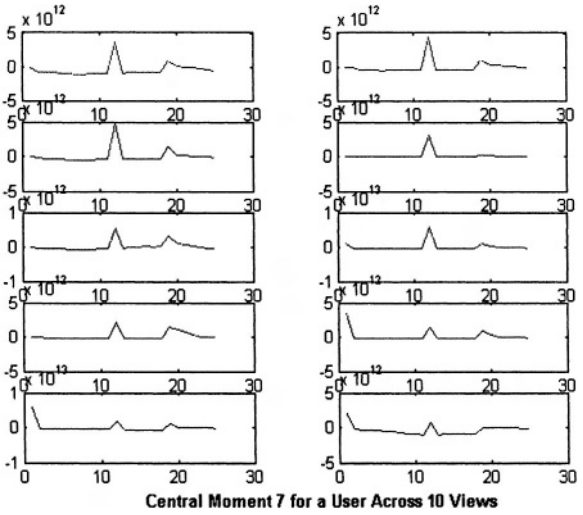

Figure 9.5. Signature Differences for 2 Users (First User)

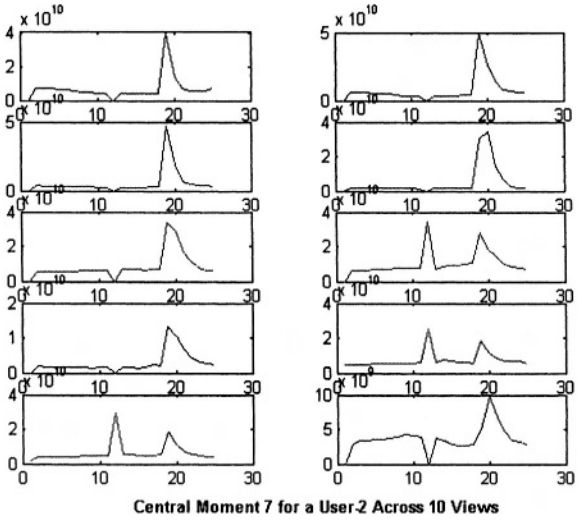

Figure 9.6. Signature differences for 2 users (second User)

References

[1] Belhumeur, P.N., Hespanha, J. P., and David J. Kriegman. (1997). " Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 19 n.7, pp.711-720.

[2] Bach J., Paul S., and Jain R. (1993). "A Visual Information Management System for the Interactive Retrieval of Faces," IEEE Trans. Knowledge and Data Engineering, Vol.5, No.4, p.619-628.

[3] Douglas Decarlo , Dimitris Metaxas. (2000). "Optical Flow Constraints on Deformable Models with Applications to Face Tracking," International Journal of Computer Vision, v.38 n.2, p.99-127.

[4] Gudivada, V. V. Raghavan, and G. S. Seetharaman. (1994). "An approach to interactive retrieval in face image databases based on semantic attributes," In Third symposium on Document Analysis and Information Retrieval, p.319–335.

[5] Konolige K., 1997,"Small Vision Systems: Hardware and Implementation". Eighth International Symposium on Robotics Research. Hayama, Japan. October 1997. http://www.ai.sri.com/ konolige].

[6] Kohonen T. (1990). "The Self -Organising Map," in Proceedings of IEEE:78, Number 9 , p. 1464-1480.

[7] Kirby M., and Sirovich L. (1990). "Application of the Karhunen-Love Procedure for the Characterization of Human Faces," IEEE Trans. PAMI, No.12, Vol. l, p.103-108.

[8] Lee, S.Y., Shan, M.K., and Yang, W. P. (1989). "Similarity Retrieval of Iconic Image Database". Pattern Recognition. Vol.22. No.6. Nov. 1989. pp.675-682.

[9] Martinez A. (1999). "Face Image Retrieval Using HMMs," in Proceedings of IEEE Workshop on Content-Based Access of Images and Video-Libraries.

[10] Mulhem, P., Lim. J. (2002). "Symbolic Photograph Content-Based Retrieval," in Proceedings of the third international conference on Information and knowledge management. p.94-101.

[11] Phillips, P. J., Grother, P. J., Micheals, R. J., Blackburn, D. M., Tabassi, E., and Bone, J. M. (2003). "Face recognition vendor test 2002: Evaluation report." NISTIR 6965. Available online at http://www.frvt.org.

[12] Rabiner L.R.(1989). "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE 77(1):257-285.

[13] Ravela, S. and Manmatha, R. (1997). "Image Retrieval by Appearance". SIGIR'97. In Proc. of the ACM, pp.278-285.

[14] Soodamani, R. and Liu, Z.Q. (1998)."Object Recognition by Fuzzy Modelling and Matching," Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence, vol.1, p. 165-170.

[15] Soodamani, R. and Ronda, V. (2001). "Stereo Face Recognition using Discriminant Eigenvectors." Electrical and Computer Engineering Series. Advances in Signal Processing, Robotics and Communications. WSES Press. p. 164-169.

[16] Soodamani, R. (2001). "Inhouse Project, Stereo Imaging Based Face Recognition - Database Construction," Centre for Signal Processing, Singapore: Nanyang Technological University.

[17] Soodamani, R and Zheng, S. (2002). "Face Recognition Web-based Security," Projects showcased in CoE Technology Week 2002, Singapore: Nanyang Technological University.

[18] Soodamani, R., Vladlena, B. and David, A. (2004). "Image Retrieval and Indexing for Stereo-based 3D Face Database," Technical Report, Middlesex University, 2004.

[19] Wu J.K. (1990). "LEP-Learning based on Experiences and Perspectives," Paris: ICNN-90.

[20] Wu. J. K.,Ang Y. H., Lam, P.C., Moorthy S.K., Narasimhalu A.D.(1993). "Facial Image Retrieval , Identification, and Inference System," Proceedings of the first ACM international conference on Multimedia ACM p. 1-9.

[21] Zhao W., Chellappa R., Phillips P. J. and Rosemfeld A. (2003) "Face Recognition: A Literature Survey". ACM Computing Surveys. Vol. 35. No. 4. December 2003. pp. 399-458.