

2D-3D POSE INVARIANT FACE RECOGNITION SYSTEM FOR MULTIMEDIA APPLICATIONS

Authors:

Antonio Rama¹, Francesc Tarrés¹

Jürgen Rurainsky²

{tonirama, tarres}@gps.tsc.upc.edu

rurainsky@hhi.fraunhofer.de

¹Department of Signal Theory and Communications
Universitat Politècnica de Catalunya (UPC)

²Image Processing Department
Fraunhofer Institute for Telecommunications
Heinrich-Hertz-Institut (HHI)

Automatic Face recognition of people is a challenging problem which has received much attention during the recent years due to its potential multimedia applications in different fields such as 3D videoconference, security applications or video indexing. However, there is no technique that provides a robust solution to all situations and different applications, yet. Face recognition includes a set of challenges like expression variations, occlusions of facial parts, similar identities, resolution of the acquired images, aging of the subjects and many others. Among all these challenges, most of the face recognition techniques have evolved in order to overcome two main problems: illumination and pose variation. Either of these influences can cause serious performance degradation in a 2D face recognition system. Some of the new face recognition strategies tend to overcome both research topics from a 3D perspective. The 3D data points corresponding to the surface of the face may be acquired using different alternatives: A multi camera system (stereoscopy), structured light, range cameras or 3D laser and scanner devices. The main advantage of using 3D data is that geometry information does not depend on pose and illumination and therefore the representation of the object does not change with these parameters, making the whole system more robust. Howev-

er, the main drawback of the majority of 3D face recognition approaches is that they need all the elements of the system to be well calibrated and synchronized to acquire accurate 3D data (texture and depth maps). Moreover, most of them also require the cooperation or collaboration of the subject during a certain period of time. All these requirements can be available during the training stage of many applications. When enrolling a new person in the database, it could be performed off-line, with the help of human interaction and with the cooperation of the subject to be enrolled. On the contrary, the previous conditions are not always available during the test stage. The recognition will be in most of the cases in a semi-controlled or uncontrolled scenario, where the only input of the system will probably consist of a 2D intensity image acquired from a single camera. This leads to a new paradigm where 2D-3D mixed face recognition approaches are used. The idea behind this kind of approaches is that these take profit of the 3D data during the training stage but then they can use either 3D data (when available) or 2D data during the recognition stage. Belonging to this category, some of 2D statistical approaches like Eigenfaces or Fisherfaces have been extended to fit in this new paradigm leading to the Partial Principal Component Analysis (P²CA) approach. This algorithm intends to cope with big pose variations ($\pm 90^\circ$) by using 180° cylindrical texture maps for training the system but then only images acquired from a single, normal camera are used for the recognition. These training images provide pose information from different views (2.5D data). Nevertheless they can also be extended to a complete 3D multimodal system where depth and texture information is used. This chapter is structured as follows: First, a brief overview of the state-of-the-art in face recognition is introduced. The most relevant methods are grouped by multimedia scenarios and concrete applications. Afterwards, novel 2D-3D mixed face recognition approaches will be introduced.

2D-3D POSE INVARIANT FACE RECOGNITION SYSTEM FOR MULTIMEDIA APPLICATIONS.....	1
1. Introduction: Face Recognition in Multimedia Applications	4
1.1 Access Control Application Scenario	4
1.2 Intelligent Human Computer Interface (IHCI)	5
1.3 Virtual Collaboration	6
1.4 Video Indexing and Retrieval	7
1.5 Surveillance Application.....	8
2. Mixed 2D-3D Face Recognition Schemes	10
2.1 Overview of Partial Principal Component Analysis (P ² CA) Approach.....	10
2.2 Computation of the face space using P ² CA	12
3. Improvements in P²CA Approach	14
3.1. P ² CA in the frequency domain	14
3.2. Automatic Creation of the Texture Maps and a Local Alignment method	15
3.2.1 Introduction	15
3.2.2 Image stitching	16
3.2.3 Image Alignment.....	19
4. EXPERIMENTAL RESULTS	21
4.1. Dataset and Experiment Description.....	21
4.2. Experimental Results	21
5. CONCLUSIONS.....	23
5.1. Performance of the system in the different application scenarios	23
5.2. Future Work.....	23
ACKNOWLEDGEMENT	24
REFERENCES	25

1. Introduction: Face Recognition in Multimedia Applications

One of the reasons face recognition has attracted so much research attention and development over the past 30 years is its great potential in numerous multimedia commercial applications. Zhao and Chellappa [Zhao06] grouped face recognition technology into five different categories of multimedia applications and described their advantages and disadvantages. Another example of the evolution of face processing research is the continuously proposal of different benchmarks and performance evaluation initiatives [FRVT02, FRGC05], indicating that the field is far from maturity. In this subsection, face processing approaches are briefly described or mentioned depending on how suitable they are for the addressed multimedia application scenario. A total of five different scenarios have been proposed: Access Control Points, Intelligent Human Computer Interfaces, Virtual Collaboration, Video Indexing and Retrieval and Video Surveillance.

1.1 Access Control Application Scenario

Nowadays, there are multiple access points where the verification or identification of the person is required. Most of them use a card that should be passed through a magnetic card reader. The main problem is that the card is not associated with a person. If the card is lost or stolen, anyone that gets it, will access the restricted area without any further problem. Thus, it is becoming more common the use of biometric technology together with the card to improve the security of the system. Access control by face recognition has several advantages in comparison with other biometrics systems: First, there are no requirements for expensive or specialized equipment since the system consists on a simple video camera and a personal computer. Secondly, the system is passive and the complete cooperation of the user is not needed. Any user walking or staying in front of the camera is processed by the recognition system. Generally speaking, control access applications are not only entering a physical place but also a virtual one. For instance, imagine a communication system or any other multimedia application where the person is recognized and some settings are loaded depending on the privileges of this person.

The main characteristics of this kind of scenario are that conditions are controlled and the subject to be recognized may cooperate in acquiring an almost frontal image without any kind of occlusions and with a neutral facial expression. Furthermore, the illumination should not vary significantly since the access points are in most of the cases indoors localizations. Hence, this scenario implies a low-medium challenge for the current face detection and recognition technologies. Skin detectors [McKenna98, Albiol01, Jones02, Vezhnevets03, Chenaoua06, Aureli07, Pierrard07], template-based face detectors [Craw92, Yuille92, Sinha94, Kwon94, Samal95, Miao99] or the well known state-of-the-art Adaboost face de-

tector [Viola01, Lienhart02] may achieve a very good performance in this kind of applications. For the recognition stage, 2D statistical approaches like Principal Component Analysis (PCA) or Eigenfaces [Turk91], Fisherfaces [Belhumeur97], Independent Component Analysis [Barlett02], Kernel PCA [Kim02], or 2D PCA [Yang04] may be enough for solving the problem with a high performance.

1.2 Intelligent Human Computer Interface (IHCI)

In Intelligent HCI applications a key goal is usually to give machines some perceptual abilities that allow them to interact naturally with people: To recognize people and remember their preferences and peculiarities, to know what they are looking at, and to interpret their gestures and facial expression changes. This could be very useful to facilitate typical tasks with machines especially for illiterate, blind or dumb people. An example of these tasks could be getting money from a cash machine. The IHCI may be able to recognize and interpret the movements of the lips or gestures of the hands as different commands. This can be extended to other environments such as smart rooms where a gesture may indicate to switch on the light projector or to pass the slide of a presentation.

The a priori requirements can vary from controlled conditions (office desk) to semi-controlled conditions (outdoor cash machine). Thus, the methods proposed have to be more robust towards some challenges such as small (mid-) pose and illumination variations, or even they have to consider the possibility of some occlusions like wearing sun glasses or any other clothing accessories. Nevertheless, since the goal is the interaction with the machine it is also supposed that the user will collaborate with the system; thus, an almost frontal image of the face with an acceptable resolution may be acquired.

For this reason, the previous commented face detection and recognition methods proposed in the “access control scenario” may also be used. However, since some small pose and illumination variations should be taken into account, face detection approaches based on neural networks [Propp92, Rowley98] and variations of the Adaboost method [Lienhart02] may produce better results. The recognition stage requires also pose robust algorithms; thus, variations of the previous statistical approaches like Self-Eigenfaces [Torres00], Eigenfeatures [Pentland94, Emidio00] can be applied. Additionally, 2D-3D mixed schemes like the Partial PCA or the Partial LDA, that will be described in detail in the next section, show their robustness towards pose variations and can be more efficient for this kind of IHCI applications. Finally, Active Appearance Models (AAM) [Cootes01, Batur05] and Elastic Graph Matching (EGM) [Lades93, Wiskott99] are also more suitable in this IHCI multimedia scenario since there are supposed to be very robust for even big pose variations.

1.3 Virtual Collaboration

Intelligent Human Computer Interface is only a small, but important, part from bigger applications for a virtual collaboration scenario. Virtual collaboration encloses different areas or possibilities like smart rooms, e-learning, or entertainment. In these cases, face detection and recognition should be applied under semi-controlled conditions. Here there are a lot of possibilities or requirements since this kind of applications could go from a conventional 2D video conference with only one single camera available, to a complete 3D tele-immersive system with multiple cameras acquiring different images simultaneously. Nevertheless, almost all virtual collaboration scenarios are indoors so that the illumination conditions should be relative well controlled to avoid problems of shadows. Furthermore, occlusions may also not be one of the biggest challenges in this kind of applications since the person is supposed to collaborate with the environment. However, pose and facial expression changes may occur continuously.

Thus, one main multimedia application of the Virtual Collaboration scenario represents a meeting between people in different places. An example is represented in Figure 1. A team of four individuals from a university (Terminal 1) were teamed up with a fifth individual, who is at a remote office (Terminal 2), through the Virtual Collaboration Desk for a discussion on producing a promotional video for the university.



Figure 1 Sample screenshots from a Virtual Collaboration recordings at terminal T1 (left) and terminal T2 (right)

One important point is to know which person is talking for producing, for instance, automatic reports or summaries of the meeting. Thus, face recognition may be fused with other information such as audio to detect who is talking at each moment. As an example, the audio of the sequence can be recorded not only using a conventional microphone but also a microphone array. This is useful for audio source localization which can be combined with face detection in order to identify which person is talking. Another possibility could be to fuse speaker recognition with face recognition. A visual example of output of such automatic e-meeting reports can be illustrated in Figure 2.

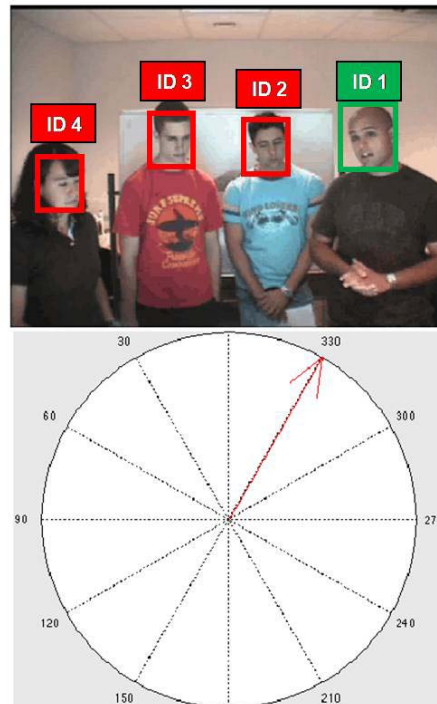


Figure 2 E-meeting multimedia application for automatic minutes extraction

Thus, in this kind of multimedia applications face detection methods that may cope with different view angles should be developed. Examples are neural network approaches or techniques that use deformable templates. In the case of the identification stage, 3D face recognition methods [Cartoux89, Lee90, Beymer95, Beumier01, Georghiades01, Blanz03, Bowyer04, Papatheodorou04, Tsalakani-dou04, Bronstein05, Chang05, Lu05, Feng06, Lu06, Onofrio06, Samani06, Antini06, Mahoo07, Yueming07, Sen07, Kakadiaris07, Faltemier08] can improve face recognition results in virtual collaboration applications due to the possibility of extracting 3D data in a more or less accurate way since multiple cameras can be integrated and the view area of the application may be spatially limited.

1.4 Video Indexing and Retrieval

Nowadays, the amount of created digital data each year exceeds the 250 Exabytes, and more or less the 90% of this corresponds to audiovisual data. The storage of this data is a quite small problem compared with the management and the intelligent reuse of this data. Therefore, video indexing and retrieval has become a main issue in the video processing field. It has been demonstrated that faces are

an important clue for searching visual data (images or videos) in large databases. Thus, automatic labeling of persons in visual databases is an important application scenario for face processing. Regarding personal multimedia applications, nowadays people use digital cameras and usually they shot hundreds of pictures during the holidays. All these pictures are stored in folders and sometimes the quantity of them is so huge that these are never again viewed. This occurs basically due to the time needed for searching to some specific picture in those folders. Thus, face recognition may be a very useful tool for making intelligent searches like “I want to seek pictures where person A and person B appear”. This automatic photo album indexing is one of the main issues in face processing and particularly in face recognition. In this case, a 100% of recognition accuracy is not needed since it can be improved by the user by means of a verification process.

Particularly, video indexing and retrieval of faces depends a lot on the features of the data that wants to be catalogued. For this reason, faces can vary from controlled conditions to uncontrolled conditions depending on the nature of the video. Nevertheless, it is not necessary that the recognition rate is above the 100% since errors can be solved manually in most of the cases. Additionally, the detection and recognition stages should not be performed in real time since all the video indexing process may be done off-line. Consequently, the methods proposed in this application area have to be more robust towards more challenges such as pose and illumination variations, or some occlusions independently of the computational cost.

Face detection methods robust to illumination variations and occlusions such as techniques using illumination robust features like Local Binary Patterns (LBP) [Heusch06, Ahonen06] or Gabor filters can be developed; or algorithms based on topology verification like the component-based face detector proposed by Goldmann et al. [Goldmann07]. In the case of face recognition for video indexing applications methods based on multiple still images [Arandjelovic05] or video-based approaches [Li01, Zhou03] would be a good selection.

1.5 Surveillance Application

Video surveillance is one of the most common application scenarios in face processing. One or various cameras placed in some special positions in order to detect and recognize suspicious or dangerous persons without the cooperation of those. This kind of scenarios represents a big challenge in face processing since the conditions may range from semi-controlled to uncontrolled. Faces might be hidden either intentionally (with sunglasses or a mask), or unintentionally (crowded places, scarves). Furthermore, the scenario permits the presence of multiple frontal and non-frontal faces in a crowded background, with some inclination due to the angle of the camera, with changes in illumination, and another important challenge: The low resolution of the images (faces). Nevertheless, in some small or medium places the cameras could be installed in strategic places where

almost frontal faces can be acquired with relatively controlled conditions. An example of this kind of VS applications is illustrated in next Figure.



Figure 3 Sample screenshots from the Video Surveillance recordings using (a) Standard Definition camera in the frontal desktop and (b) High Definition camera pointing the main entrance

If a more complex scenario is defined, such as VS for huge places like airports or train stations, then the results produced by face recognition systems may not be very reliable. Only recent infrared-based face recognition approaches [Cutler96, Wilder96, Socolinsky04, Jinwoo06] can represent an alternative to cope with all the challenges mentioned for this application scenario.

The most important multimedia application scenarios for face processing have been presented at this point of the chapter together with some references to the more representative methods of the literature that work better in such scenarios or challenges. Now an alternative method for face recognition is described. Concretely, this method is an extension of conventional statistical approaches that tries to add a new concept: Training the system with 3D data, but then performing the test stage using 3D or 2D data depending on the main features of the scenario, i.e. if it is possible or not to acquire the 3D data during the test stage.

The rest of the chapter is structured as follows: Section 2 describes the novel concept of 2D-3D face recognition schemes. Section 3 explains some improvements in terms of accuracy and computational cost together with a novel method for the automatic creation and alignment of the 3D training using a multi-camera system. Finally, section 4 presents some results and also how the method may perform in the scenarios explained in Section 1. Finally, conclusions and future work are briefly commented.

2. Mixed 2D-3D Face Recognition Schemes

Recently some of the new face recognition strategies tend to overcome the different challenges from a 3D perspective. The 3D data points corresponding to the surface of the face may be acquired using different alternatives: a multi camera system (stereoscopy), structured light, range cameras or 3D laser and scanner devices. The main advantage of using 3D data is that depth information does not depend on pose and illumination and therefore the representation of the object do not change with these parameters, making the whole system more robust. However, the main drawback of the majority of 3D face recognition approaches is that they need all the elements of the system to be well calibrated and synchronized to acquire accurate 3D data (texture and depth maps). Moreover, most of them also require the cooperation or collaboration of the subject making them not useful for uncontrolled or semi-controlled scenarios where the only input of the algorithms will be a 2D intensity image acquired from a single camera. For this reason, the main objective of this work is to intend to ask the following question: “*It is possible to develop a face recognition framework that takes advantage of 3D data during the training stage of the system, but then, use either 2D or 3D data in the test stage depending on the possibilities of the scenario?*” Thus, this work is focused on the development of a 2D-3D face recognition framework. This framework would provide the recognition system a great flexibility so that it could be adapted to the application scenario, i.e. the recognition system will use only the information available in each situation.

2.1 Overview of Partial Principal Component Analysis (P²CA) Approach

The objective of P²CA is to implement a mixed 2D-2.5D method (2D-3D when using texture and depth maps [Onofrio06]), where either 2D (pictures or video frames) or 2.5D data (180° texture images in cylindrical coordinates) can be used in the recognition stage. However, the method requires a cylindrical representation of the 2.5D face data for the training stage. Like in the majority of face recognition methods, in P²CA the dimensionality of the face images is reduced through the projection into a set of M optimal vectors which composed the so called *feature space* or *face space*. The vectors representing the i^{th} individual are obtained as:

$$\mathbf{r}_k^i = \mathbf{A}_i^T \cdot \mathbf{v}_k \quad k = 1, \dots, M \quad (1)$$

,where \mathbf{A}_i^T is the transposed of the $H \times W$ image representing individual i , and \mathbf{v}_k are the M optimal projection vectors that maximize the energy of the projected vectors \mathbf{r}_k averaged over the entire database. These vectors could be interpreted as unique signatures that identify each person. The projection described in equation

(1) is depicted in Figure 4. The (training) texture map of the subject i is represented by the M vectors \mathbf{r}_k^i . Each vector \mathbf{r}_k^i has W elements, where W is the width of the matrix A_i .

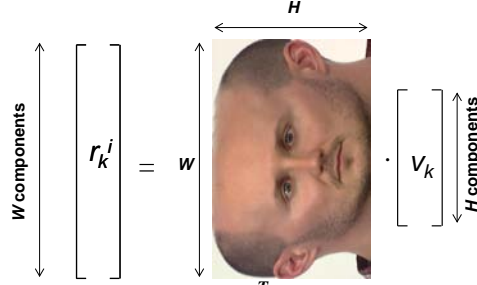


Figure 4 Description of a texture map (A_i^T) by means of projection vectors using P^2CA (training stage)

The main advantage of this representation scheme is that it can also be used when only partial information of the individual is available. Consider, for instance, the situation depicted in Figure 5, where it is supposed that only one 2D picture of the individual is available. In this case, the M vectors \mathbf{r}_k representing the 2D picture, have a reduced dimension W' . However, it is expected that these W' components will be highly correlated with a section of W' components in the complete vectors \mathbf{r}_k^i computed during the training stage.

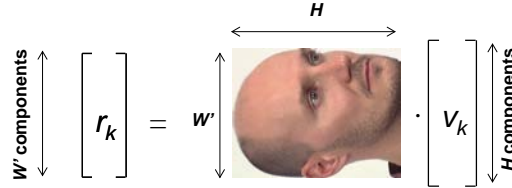


Figure 5 Projection of a "partial" 2D image through the vector set \mathbf{v}_k (recognition stage)

Therefore, the measure proposed below has been used to identify the partial available information (W' components) through the vectors \mathbf{r}_k^i [Rama05]:

$$\min_{(i,j)} \left\{ \sum_{k=1}^M \sum_{l=1}^{W'} (r_k(l) - r_k^i(l+j))^2 \right\}$$

$$i = 1, \dots, L; \quad j = 0, \dots, W - W' \quad (2)$$

, with L being the total number of gallery images (subjects) of the database.

In other words, the training texture maps coded 180° angle information of the person in the database. So during the test stage, any pose view image acquired from a single camera can be used to perform the recognition. It is expected that this 2D partial information after the projection in the face space is highly correlated with some part of the higher dimensionality vector of the texture map as shown in Figure 6.

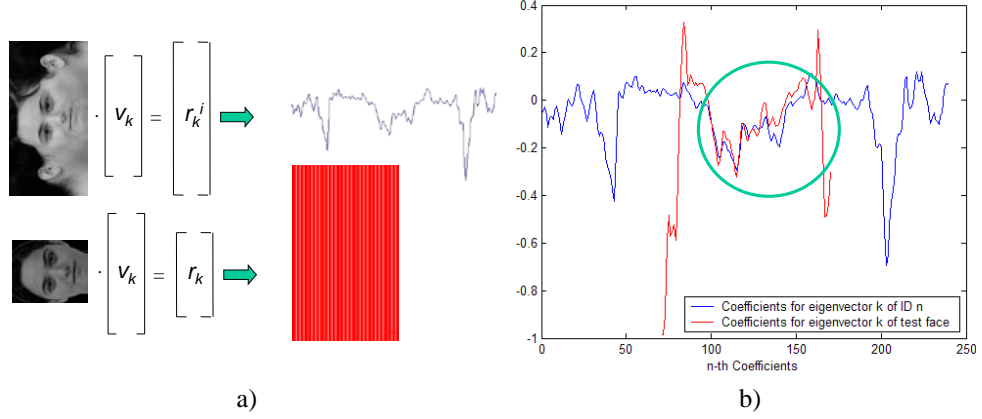


Figure 6 a) Projection of the training texture map and a test image using only one eigenvector (face space is one dimensional). b) High correlation area of the test feature vector and the training feature vector

The Square Differences (SD) similarity measure of equation (2) has been proposed due to its relatively low computational cost. However, other measures have been used with better results for pattern matching like Normalized Cross Correlation (NCC) but at the expense of a higher computational cost:

$$\max_{(i,j)} \left\{ \frac{\sum_{k=1}^M \sum_{l=1}^{W'} (r_k(l) \cdot r_k^i(l+j))}{\sqrt{\sum_{k=1}^M \sum_{l=1}^{W'} r_k(l)^2 \cdot \sum_{k=1}^M \sum_{l=1}^{W'} r_k^i(l+j)^2}} \right\}$$

$i = 1, \dots, L; \quad j = 0, \dots, W - W'$ (3)

In section 3, the computation of P^2CA in the frequency domain is formulated allowing us the introduction of the NCC for identification efficiently. The other limitation of the method is that the training image (Figure 4) should be well aligned when computing the face space [Rama05, Rama06]. If this is not the case, the correlation between test and the gallery images will be reduced after projecting them into the face space leading to a reduction of the face recognition rate. This is especially important for the relevant facial features since these are the key points used for the creation of the training texture maps. A process for a more accurate alignment of these texture maps is presented.

2.2 Computation of the face space using P^2CA

The set of vectors which maximize the projection of Eq. (1) may be obtained as the solution to the following optimization problem: Find v_k for $k=1, \dots, M$ such that

$\xi = \sum_k \sum_n (r_k^n)^T \cdot r_k^n$ is maximum, where r_k^n is defined as the projection of image n through the vector v_k and n accounts for the number of images in the training set. The function to be maximized may be expressed as:

$$\xi = \sum_k \sum_n (A_n^T v_k)^T \cdot (A_n^T v_k) = \sum_k v_k^T \left(\sum_n A_n \cdot A_n^T \right) \cdot v_k \quad (4)$$

,which states that the optimum projection vectors may be obtained as the eigenvectors associated to the M largest eigenvalues of the mxm positive definite matrix C_s

$$C_s = \sum_n A_n \cdot A_n^T \quad (5)$$

This vector set will be used for feature extraction and recognition from partial information:

$$\{\mathbf{v}_1, \dots, \mathbf{v}_M\}$$

The procedure for feature extraction from an intensity image A consists in projecting the *transposed* image through every eigenvector:

$$\mathbf{r}_k = A^T \cdot \mathbf{v}_k \quad k = 1, \dots, M$$

Therefore, a total of M feature vectors are available, with n (width) components each, for the image. The image has been compressed to a total of nxM scalars with M always being smaller than m .

When a complete image sample A (mxn) is available, the recognition stage is straightforward. First, the projected vectors of the sample image are computed using the previous equation and then, the best match is found as the individual i whose representing vectors minimize the Euclidean distance:

$$\min_i \left\{ \xi_k = \sum_{k=1}^M \sum_{l=1}^n (r_k(l) - r_k^i(l))^2 \right\} \quad i = 1, \dots, L \quad (6)$$

,where L represents the number of individuals in the database.

The most outstanding point of this procedure is that the image projected in the n -dimensional space does not need to have dimension mxn during the recognition stage so that partial information can be used. It is possible to use a reduced pxn ($p < m$) image which is projected to a smaller subspace.

If only partial information is used, a classification method is needed to compare the partial projection with the data in the whole space. In this case, it is not possible to use nearest neighbour classifier like in conventional PCA and correlation of partial difference methods like the criteria defined in (2) or (3) have to be applied.

The procedure is quite different from conventional PCA. Certainly, in PCA a scalar number is obtained when the vector image is projected to one eigenvector, whereas in P²CA, an n -dimensional vector (\mathbf{r}_k) is obtained, when the image (in matrix form) is projected to an eigenvector. It can seem that the P²CA approach demands more computational cost because it uses vectors instead of numbers to

represent the projections. However, the number of eigenvectors $\{\mathbf{v}_k\}$ needed in P²CA for an accurate representation is much lower than in PCA.

It should be mentioned that the mathematical theory behind this approach is similar to one recent method which has extended the conventional PCA method [Turk91] from 1D to 2D; this technique was called 2DPCA [Yang04].

3. Improvements in P²CA Approach

3.1. P²CA in the frequency domain

In [Lewis95] the author presented an efficient way of computing NCC (equation (3)) in the frequency domain. Based on this work a new formulation of P²CA is presented. The numerator of the NCC in equation (3) can be expressed again as:

$$c(i, j) = \sum_{k=1}^M \sum_{l=-\frac{(W'-1)}{2}}^{\frac{(W'-1)}{2}} (r_k(l) \cdot r_k^i(l+j)) \quad i=1, \dots, L; \quad j = -\frac{W-W'}{2}, \dots, \frac{W-W'}{2} \quad (7)$$

For convenience we have accepted a change in the vector index l , choosing the zero coordinate in the center. In this case, we have to correlate vectors of L components with vectors of W' components in $1/2(W-W')$ lags. This condition can be implemented very efficiently in the frequency domain. So taking the Discrete Fourier Transform (DFT) of the inner sum of the previous equation:

$$S(u) = R_k(u) \cdot R_k^{i*}(u) \quad (8)$$

The correlation between r_k and r_k^i has a total of $W+W'-1$ lags from which only $W-W'-1$ samples are interesting for the computation. Thus, we have to avoid that spectral overlapping occurs when transforming these $W-W'-1$ central. For this reason it is necessary to compute the W -points-DFT of r_k^i and the W -points-DFT of r_k^{2D} which consists on the zero padded version of r_k . Now taking equation (1), $R_k(u)$ can be expressed as follows:

$$R_k(u) = DFT[r_k^i] = DFT[\mathbf{A}_i^T \cdot \mathbf{v}_k] = DFT2D[\mathbf{A}_i^T] \cdot \mathbf{v}_k' \quad (9)$$

,where \mathbf{v}_k' are the eigenvectors that minimize the energy projection of a given training set after applying two dimensional DFT. The demonstration of this statement have already been proved for the 1D case in [Savvides04] but the same parallelism can be followed if the images (\mathbf{A}_i) are treated like matrices.

The test stage of the P²CA in frequency domain will be summarized as follows:

- Given an image A_{test} , normalize this image (same as P²CA in spatial domain [Rama05]).
- Extend A_{test} with columns of zeros until we have A_{test}^{zp} with W columns (same width as the training texture maps).
- Compute the DFT-2D of A_{test}^{zp} and project the result into the face space \mathbf{v}_k obtained during the training stage.
- Obtain the product between the frequency domain test coefficients and the frequency domain weights of each identity (equation (8)).
- Compute the IDFT of (8) for the M different coefficients vectors where only the $W-W'-1$ central samples should be considered because these are not affected by aliasing.
- Identify the identity of the database that gets a max value of the sum of the M IDFT vectors.

3.2. Automatic Creation of the Texture Maps and a Local Alignment method

3.2.1 Introduction

The goal is it to create a cylindrical projection of a face, like the one shown in Figure 7 with a desired view angle range of $\pm 90^\circ$ using multi view images. Such image can be used as texture map for 3D models or pose invariant face recognition tasks. The creation of texture maps from one or multiple views are described in many publications. The common idea is to project the captured images onto a cylinder using a more or less detailed surface representation or to find registration data, which related one captured view to another. Blending rules at the overlapping areas define the quality of the synthetic image. The approach implemented here is based on a simple approximation of a human head in contrast to a detailed surface representation used for similar results. Therefore a detailed reconstruction of the face is not required. Registration techniques require high frequency parts, which maybe not available for the complete face or hair. Marker points are not very likely, because of the more complicated capture and removal process. On the other hand, ghost edges are very likely, if the cylindrical approximation is not placed at the right position or the system calibration is not adequate. Another important issue in the creation of this kind of images is the correct aligned of the images for the posterior face space creation.



Figure 7 Texture map created from an image set of nine images captured with a multi-view camera setup

3.2.2 Image stitching

Image stitching algorithms blend images in a seamless manner, taking care to deal with potential problems such as blurring or ghosting caused by parallax and scene movement as well as varying image exposures. Our goal is the creation of cylindrical face projections from several different views. Thus, the idea is to use a cylinder as approximation of a human head. The combination of several views captured with a system as shown in Figure 8 to a synthetic image as shown in Figure 7 requires information about the surface of the captured object. Such information can range from an exact 3D model over some points in 3D space to an approximation of the basic shape. If such surface information is not available, they can be extracted from the captured view images with a wide variety of methods. Triangulation only requires corresponding point pairs in two views. Surface point reconstruction for a couple of corresponding point pairs allows the definition of an approximation of the desired surface up to a detailed representation. If the captured images show a unique silhouette, a method known as shape from silhouette can be used to reconstruct a volume model. This method is limited to convex surfaces. Other methods use registration information, created from gradients or marker positions.

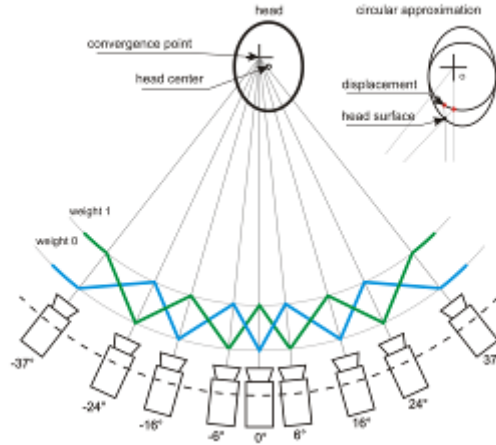


Figure 8 Capture system for training data acquisition consists of nine cameras at different horizontal angles. (left) face surface point displacement due to the circular approximation. (middle) Blending rule for the combining of the input images

A cylindrical approximation is adapted to the captured head and positioned in the middle of the assumed head axis. Head dimensions are defined in the book of Farkas [Farkas95] with average values of 151mm for the width of a head and 197mm for the depth of a head. Similar numbers can also be found in the report of Young [Young93]. Considering a head with hair and two ears, a circular assumption is one possible approximation to the ellipse shaped head therefore suitable for the given problem. Besides the head size (radius for the cylinder), which can be taken from the mentioned statistical publications, there are other remaining unknown positions. Both offsets (x - and z -axis) have to be estimated.

Therefore the following equation describes the problem.

$$r^2 = (x - o_x)^2 + (z - o_z)^2 \quad (10)$$

This non-linear equation can be solved by Levenberg-Marquadt approach as described in the book of Scales [Scales85]. We have used the 3D reconstructed locations of both eye pupil centers as well as left and right eye corner of both eyes as fitting locations. The radius was fixed to a slightly bigger value than given for the head depth by the statistical publications, because of the requirements for the resulting texture map ($\pm 90^\circ$). The determined cylinder offset parameter for the x - and z -axis are used to place the cylinder at the face surface with the highest impact for the alignment, like the eyes and mouth. Therefore, the edges of these features are aligned and the displacement error for other features are not visible, e.g. nose tip. The fitted cylinder and a face surface model are shown in Figure 9. Therefore, some face surface points are reconstructed using the calibration data and perspective projection. Due to the fact, that several views provide 2D locations of the same feature point, a multi view approach for the reconstruction is used for more

reliable 3D feature point locations. The handling of outliers, which can be a result of the point correspondences or calibration data, is crucial at this step. The average of permuted reconstructed 3D point locations using two views as well as a closed-form solution for all views has the drawback of moving the solution to one or more outliers. We have used a combination of the permutation and closed-form solution, which takes advantage of the back projection error in all considered views and is therefore an iterative solution. The consideration of the minimum back projection error as measurement for the 3D point localization accuracy leads to reliable surface points and identifies possible problems with 2D point locations. The evaluation result can not only be used for more exact 3D point locations by excluding outliers, but also to define weights according to the quality of back projection and therefore to define the influence of each view to the to be created texture map.

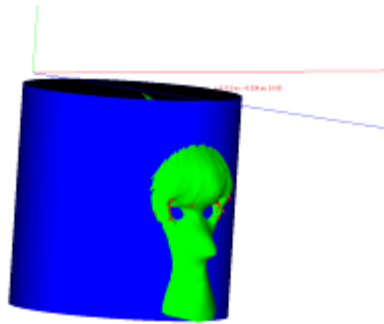


Figure 9 Interpolated face surface model with a cylinder fitted at both eyes.

The composition is mainly described by the projection of each view onto the object surface and in our case a cylindrical approximation. The perspective warping of images onto a cylindrical surface is described in the technical report of Szeliski [Szeliski06]. Afterwards a projection will take place, in order to create the texture map. Sampling the cylinder surface with the desired resolution of the virtual view is used for this approach. This method allows a set of DOFs, like the horizontal and vertical resolution as well as defining a specific region of interest. A linear blending rule, like the one shown in Figure 8 is used to incorporate adjacent views. For each horizontal position along the circumference left and right views are selected according to the camera rotation around the y-axis. The angle differences between the view vector and the selected left and right views are converted to weights. This method is constrained on the assumption, that the views are now rotating around the head and cylinder center and not longer around the convergence point of all views. The 3D reconstructed face features show a very small offset along the x axis and a strong offset along the z axis, so that middle view still refers to the most frontal face view. Using a linear interpolation between adjacent views leads to a weight of 0.5 in the middle of these both views.

The result is a texture map shown in Figure 7, where nine views are incorporated to one cylindrical projected face image showing $\pm 90^\circ$. At the top and the bottom of the created image some ghost parts are visible. These wrong stitched parts are caused by the cylindrical assumption and the difference to the real object surface. A more detailed 3D object would decrease such miss alignments of the regions below the chin.

3.2.3 Image Alignment

The importance of local aligned face features for recognition is described by Tsapatsoulis *et al.* [Tsapatsoulis98]. There, the local alignment is described as re-sizing in contrast to a complete affine transformation. In most face recognition systems, the eye centers are used for the alignment of the images, whereas other face features have usually not received attention although it has been demonstrated that they are as important as eyes centers for performing recognition [Kouzani99]. In this work, a two step alignment approach is proposed. First, a global alignment is applied to the training texture maps, which is based on 2D affine transformation. The parameters for the transformation are determined by using a limited number of manually selected face feature locations. The average feature point location is used as reference and all transformations are calculated with respect to these data.

In order to achieve a better alignment result, the regions of the selected face features are additionally aligned locally. The generic triangle mesh of Figure 10 is adapted and placed at the global transformed face feature locations and the associated texture information for each triangle is extracted. The face feature locations are transformed to the desired position and thus affine warping is performed locally for each triangle.

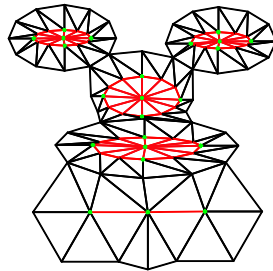


Figure 10 Adapted triangle mesh for the local alignment of face features including eyes, nose, mouth and chin.

Using this approach leads to a better alignment of the local features as depicted in Figure 11. Figure 11a) shows the point clouds for the selected feature points on the texture maps (only frontal features). Figure 11b) represents the same points after applying the 2D affine global transformation to each image. And finally, Figure 11c) illustrates the consistency of all the facial feature points after the local affine transformation. Since alignment of the feature points is done jointly for each

feature, the feature points shown in Figure 11c) do not match in one center. A more visual comparison between just global or global plus local aligned texture maps is given in Figure 12.

Figure 11 Facial Feature localizations of the facial images a) before any transformation, b) after global transformation, and c) after local transformation using the wired mesh

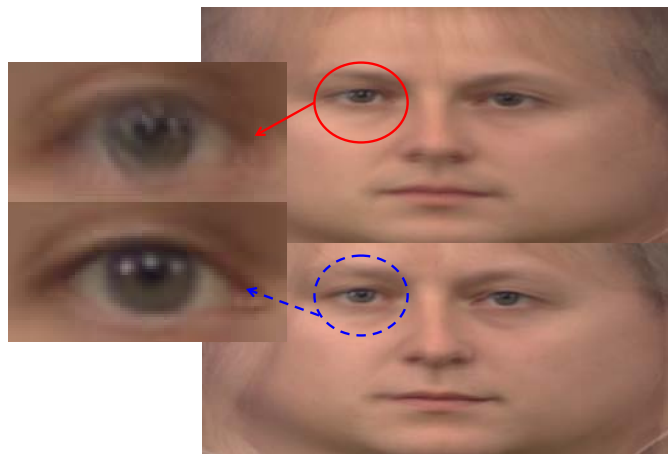


Figure 12 Average image. Top: after global alignment. Bottom: after global plus local alignment.

4. EXPERIMENTAL RESULTS

4.1. Dataset and Experiment Description

Two different databases are used for the experimental results. The first one is the HHI face database which is composed of 10 different subjects. The HHI database contains one 180° texture map for each subject that have been aligned using only global alignment; and another texture map using the global and local alignment process described in Section 3.2. These texture maps are used as the training and gallery sets. The test set for recognition is composed of a total of 9 views with different head poses (0° , $\pm 6^\circ$, $\pm 16^\circ$, $\pm 25^\circ$ and $\pm 37^\circ$) for each subject (90 test images) acquired in a second session.

The second database is the UPC database [UPC-FaceDatabase]. This database includes a test set of 30 persons with 9 pictures per person which correspond to different pose views (0° , $\pm 30^\circ$, $\pm 45^\circ$, $\pm 60^\circ$ and $\pm 90^\circ$). Furthermore, a total of 30 different 180° texture maps have been created by morphing only five views [Rama05], these texture maps are aligned and used as the training and gallery ensemble.

The first database is used for testing the improvement of the recognition rate when using the alignment images with the proposed approach of section 3.2, whereas the second one is used for testing the computational time when performing the P^2CA approach in the frequency domain. Finally, both databases are used for comparing the results using the SD and the NCC measures. A final experiment is focused on a comparative between the Partial Principal Component Analysis technique and the conventional *Eigenfaces* approach using the UPC-Face Database.

4.2. Experimental Results

Table 1 summarizes the recognition accuracy for the two correlation methods proposed in equations (2) and (3). From the results it is clear that using NCC improves the recognition rate results since this measure is more robust towards slight changes in illumination. For the UPC database the improvement is more visible since 10 out of the 30 identities have been enrolled on the database in a different session with slightly different illumination conditions.

Table 1. Recognition Accuracy

Dataset	SD	NCC
HHI	94.44%	96.66%
UPC	81.85%	89.63%

Table 2 presents the results for face recognition when using the alignment process described above. Although only 10 different persons are used in the experiments, results show that the locally aligned images present a slight improvement in the recognition rate. The improvement of this rate has been obtained for the 0° and $\pm 6^\circ$ views since these enclose all the face features used for the alignment.

Table 2. Face Recognition Results using NCC

Dataset	FR (global alignment)	FR (global + local alignment)
HHI	93.33%	96.66%

The computational time is analyzed when using the P²CA approach in the spatial and in the frequency domain. Simulations have been run in MATLAB using a 2.0 GHz μ P with 1GB of RAM. Table 3 shows the computational time for P²CA in Spatial and in Frequency Domain depending on the total number of eigenvectors used for computing the face space. This time comprises the matching of one image to the 30 enrolled persons. Results illustrate the importance of performing P²CA in FD for higher dimensions since the reduction in computational time is between 70 and 150.

Table 3. Computational Time for NCC measure

dim	P²CA in SD	P²CA in FD	factor
1	0.14 sec	0.011 sec	12
20	0.96 sec	0.042 sec	22
60	8.15 sec	0.114 sec	71
122	32.9 sec	0.222 sec	150

In the final experiment, we verify the robustness of P²CA in front of the conventional 2D strategies. Thus, Eigenfaces (PCA), and 2DPCA have been implemented. For the training of the conventional 2D strategies, 5 different face views for each subject have been used as training and gallery data (the same 5 images used for the creation of the texture maps as explained in [Rama05]).

Table 4. Recognition Accuracy of the different algorithms

Method \ Exp.	Neutral illumination	3 illuminations
PCA / Eigenfaces	72.22%	60.45%
2DPCA	75%	61.24%
P²CA	89.63%%	72.9%

The results presented in Table show that the novel 2D-3D mixed scheme (P²CA) outperforms its respective two dimensional approaches (PCA and 2DPCA) when varying pose.

5. CONCLUSIONS

5.1. Performance of the system in the different application scenarios

Although the experiments have been carried out on a small database, the idea is to present a framework for the extension of any statistical approach such as LDA, ICA, Kernel-PCA to this novel 2D-3D mixed framework. Thus, the performance of the Partial PCA may decrease for some scenarios such as Video Surveillance since PCA is sensitive to illumination changes. However, if there were enough training samples, a Partial LDA or even a Partial ICA could be implemented to make the face recognition system more robust towards these illumination changes.

Partial PCA is supposed to get good results for Access Control, HCI or VC scenarios. In those scenarios, the main challenge is pose variation and as already demonstrated in the previous section P²CA is developed considering this challenge. In these 3 scenarios the illumination may not vary considerably and the resolution of the images is medium high. For the Video Surveillance it may depend a lot on the type of room or place of the scenario. In a bank-hall like the one illustrated in Figure 3, P²CA may perform acceptable since the face can be acquired in semi-controlled conditions. Nevertheless, the accuracy may decrease considerably in VS for airports or train stations where the resolution of the face is smaller, the inclination of the cameras or, for all of them, illumination and occlusions. In the case of video and image indexing and retrieval, the illumination conditions may vary a lot. Nevertheless, the main problem is to acquire for all the persons of the photo album or video database different views for the creation of the 3D training data. In this case, manual interaction for the creation of the texture and depth maps may be needed. This could be one of the main drawbacks of the 2D-3D mixed face recognition schemes for the image and video indexing scenario.

5.2. Future Work

Face processing is a very hot topic for several multimedia applications like video surveillance, virtual collaboration, HCI, video indexing or control access points. In each multimedia application there are different challenges that should be overcome, but pose is one common problem in all of them. Recently, a new trend of 3D face recognition approaches showed an increase in the recognition rate under the presence of big pose and illumination variations if 3D data is available. Nevertheless, cost of the set-up, acquisition time and cooperation of the subjects are still some of the requirements for obtaining accurate 3D data that may not be available during the recognition stage. Thus, we have presented here a possible al-

ternative following a mixed 2D-3D face recognition philosophy, i.e. the system is trained with 3D data but it can use either 2D or 3D data in the test stage. However, this philosophy may be extended also to other face recognition statistical approaches like LDA or ICA which have shown a higher robustness in the presence of illumination variations. Additionally, we have presented an automatic approach for the creation of aligned virtual view images using nine different views. These aligned virtual view images are used as training data for the P²CA technique. The virtual view image is created by using a cylindrical approximation for the real object surface.

Furthermore, two improvements for the P²CA approach have been proposed: First, a local alignment method of the training images. Second, a reformulation of the complete approach in the frequency domain is proposed. Both improvements lead to an increase of the recognition accuracy and a reduction of the computational time making it suitable for face recognition multimedia scenarios such as security control access point, Human Computer Interface or Virtual collaboration applications where the processing should be on real time.

ACKNOWLEDGEMENT

The work presented was developed within VISNET II, a European Network Of Excellence funded under the EC IST FP6 programme.

REFERENCES

- [Ahonen06] Ahonen, T.; Hadid, A.; Pietikainen, M.; "Face Description with Local Binary Patterns: Application to Face Recognition" in IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 28, Issue 12, Dec. 2006 Page(s):2037 – 2041
- [Albiol01] A. Albiol, L. Torres, Ed. Delp. Optimum Color Spaces for Skin Detection. IEEE International Conference on Image Processing, Thessaloniki, Greece, October 7-10, 2001
- [Antini06] G. Antini, S. Berretti, A. Del Bimbo, P. Pala, "3D Face Identification Based On Arrangement Of Salient Wrinkles", in IEEE International Conference & Multimedia Expo, Toronto, Canada, July 9th-12th 2006
- [Arandjelovic05] O. Arandjelovic, G. Shakhnarovich, J. Fischer, R. Cipolla, and T. Darrell, "Face recognition with Image Sets using Manifold Density Divergence" Proc. IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, San Diego, USA, 2005
- [Aureli07] Aureli Soria-Frisch, Rodrigo Verschae and Aitor Olano, Fuzzy Fusion for Skin Detection Fuzzy Sets and Systems, Vol. 158, Issue 3, pp. 325-336, 2007
- [Barlett02] M. S. Barlett, J.R. Movellan, T.J.Sejnowski, "Face Recognition by Independent Component Analysis". IEEE Trans. On Neural Networks, vol 13, no. 6, Nov 2002
- [Batur05] A.U. Batur, M.H. Hayes, "Adaptive active appearance models" in IEEE Transactions on Image Processing, Volume 14, Issue 11, Nov. 2005 Page(s):1707 – 1721
- [Belhumeur97] R.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection" IEEE Trans. Pattern Analysis and Machine Intelligence, vol 19, No. 7, July 1997
- [Beumier01] Charles Beumier and Marc Acheroy. Face verification from 3D and grey level clues. Pattern Recognition Letters, 22:1321- 1329, 2001
- [Beymer95] D. Beymer and T. Poggio, "Face Recognition from One model View", in Proc. Fifth Int'l Conf. Computer Vision, 1995
- [Blanz03] V. Blanz, and T. Vetter, "Face Recognition based on fitting 3D morphable model", in IEEE Trans. Pattern Analysis and Machine Intelligence, 25(9):1063-1074, 2003
- [Bowyer04] K. Bowyer, K. Chang, and P. Flynn, "A Survey of Approaches to 3D and Multi-Modal 3D+2D Face Recognition," IEEE Intl.Conf. on Pattern Recognition, 2004.
- [Bronstein05] A. M. Bronstein, M. M. Bronstein, R. Kimmel, "Three-dimensional face recognition ,, in International Journal of Computer Vision Vol.64/1, pp. 5-30, August 2005
- [Cartoux89] J. Y. Cartoux, J. T. Lapreste, and M. Richetin. "Face authentication or recognition by profile extraction from range images." In Workshop on Interpretation of 3D Scenes, pages 194- 199, 1989

- [Chang05] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An Evaluation of Multimodal 2D+3D Face Biometrics", in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.27, pp 619-624, April 2005
- [Chenaoua06] K. Chenaoua, A. Bouridane; "Skin Detection using a Markov Random Field and a New Color Space", in *IEEE International Conference on Image Processing*, 2006, 8-11 Oct. 2006 Page(s):2673 - 2676
- [Cootes01] T. Cootes, G. Edwards, and C. Taylor. "Active Appearance Models", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:681-685, 2001
- [Craw92] I. Craw, D. Tock, and A. Bennett, "Finding Face Features," *Proc. Second European Conf. Computer Vision*, pp. 92-96, 1992.
- [Cutler96] R. Cutler "Face Recognition Using Infrared Images and Eigenfaces", Technical Report. UMI Order Number: CSC 989., University of Maryland at College Park, 1996
- [Emidio00] T. Emidio, R. Schmidt, R. Marcondes, "A Framework in Face Recognition from Video Sequences using GWN and Eigenfeature Selection" *I WAICV - Workshop on Artificial Intelligence and Computer Vision*. São Paulo, Brazil. 2000
- [Faltemier08] T. C. Faltemier; K. W. Bowyer, P.J. Flynn; "A Region Ensemble for 3-D Face Recognition" *IEEE Transactions on Information Forensics and Security*, Volume 3, Issue 1, March 2008 Page(s):62 – 7
- [Feng06] S. Feng, H. Krim, I. Gu, M. Viberg, "3D Face Recognition Using Affine Integral Invariants" in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, May 14th- 19th, 2006
- [FRGC05] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," *IEEE Conf. Computer Vision and Pattern Recognition(CVPR)*, 2005
- [FRVT02] P. J. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and J. Bone, "Face Recognition Vendor Test 2002: Evaluation Report". NISTIR 6965, National Institute of Standards and Technology, 2003. <http://www.frvt.org>
- [Georghiades01] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition" in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643-- 660, 2001
- [Goldmann07] L. Goldmann, U. J. Mnich, and T. Sikora. "Components and their topology for robust face detection in the presence of partial occlusions." in *IEEE Transactions on Information Forensics and Security*, 2, Sep 2007
- [Heusch06] Heusch, G.; Rodriguez, Y.; Marcel, S.; "Local binary patterns as an image preprocessing for face authentication" *7th International Conference on Automatic Face and Gesture Recognition*, 2006. FGR 2006. 10-12 April 2006 Page(s):6 pp
- [Jinwoo06] Jinwoo Kang; Borkar, A.; Yeung, A.; Nong, N.; Smith, M.; Hayes, M.; "Short Wavelength Infrared Face Recognition for Personalization" *IEEE International Conference on Image Processing*, 2006 8-11 Oct. 2006 Page(s):2757 – 2760
- [Jones02] M. J. Jones and J. M. Rehg, *Statistical Color Models with Application to Skin Detection*, *Int. Journal of Computer Vision*, vol. 46(1), pp. 81-96, 2002.

- [Kakadiaris07] Kakadiaris, I.A.; Passalis, G.; Toderici, G.; Murtuza, M.N.; Yunliang Lu; Karampatziakis, N.; Theoharis, T.; “Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach “ in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 29, Issue 4, April 2007 Page(s):640 – 649
- [Kim02] K.I.Kim, K. Jung, H.J. Kim, “Face Recognition Using Kernel Principal Component Analysis”, *IEEE Signal Processing Letters*, Vol.9, No. 2, February 2002
- [Kouzani99] A. Kouzani, and K. Sammut, “Quadtree principal component analysis and its application to facial expression classification” in *IEEE Int. Conf. on Systems, Man, and Cybernetics*, October 1999
- [Kwon94] Y. H. Kwon and N.da Vitoria Lobo, “Face detection using templates”, in *Proc. of IAPR International Conference on Pattern Recognition*, pp. 764-767, 1994
- [Lades93] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, W. Konen. “Distortion Invariant Object Recognition in the Dynamic Link Architecture” *IEEE Transactions on Computers* archive Volume 42 , Issue 3 Pages: 300 – 311. 1993
- [Lee90] John Chun Lee and E. Miliou. Matching range images of human faces. In *Proc. IEEE International Conference on Computer Vision*, pages 722- 726, 1990
- [Lewis95] J.P. Lewis, “Fast Normalized Cross-Correlation,” *Vision Interface*, 1995
- [Li01], Y. Li, S. Gong, and H. Liddell, “Constructing facial identity surfaces in a nonlinear discriminating space”, in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition* 2001
- [Lu05] X. Lu, and A.K. Jain, “Integrating Range and Texture Information for 3D Face Recognition”, in *Proc. IEEE WACV*, Breckenridge, Colorado 2005
- [Lu06] X. Lu, and A.K. Jain, “Matching 2.5D face scans to 3D models”, in *IEEE Transactions Pattern Analysis and Machine Intelligence*, Jan. 2006 Volume: 28, pp: 31- 43
- [Mahoo07] M.H. Mahoo ; M. Abdel-Mottaleb, “3D Face Recognition Based on 3D Ridge Lines in Range Data” in *IEEE International Conference on Image Processing, 2007. ICIP 2007*, Volume 1, Sept. 16 2007-Oct. 19 2007 Page(s):I - 137 - I – 140
- [McKenna98] S. McKenna, S. Gong, and Y. Raja, “Modelling Facial Colour and Identity with Gaussian Mixtures,” *Pattern Recognition*, vol. 31, no. 12, pp. 1883-1892, 1998.
- [Miao99] J. Miao, B. Yin, K. Wang, L. Shen, and X. Chen “A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template”, *Pattern Recognition*, vol.7, no. 32, pp. 1237-1248, 1999
- [Onofrio06] D. Onofrio, A. Rama, F. Tarres, S. Tubaro, “P2CA: How Much Information is needed”, *IEEE International Conference on Image Processing*, Atlanta, USA, October 2006
- [Papatheodorou04] T. Papatheodorou, D. Rueckert, "Evaluation of Automatic 4D Face Recognition Using Surface and Texture Registration," *fgv*, p. 321, Sixth

- IEEE International Conference on Automatic Face and Gesture Recognition, 2004
- [Pentland94] A. Pentland, B. Moghaddam, and T. Starner. View based and modular eigenspaces for face recognition. In IEEE Conf. on Computer Vision and Pattern Recognition, 1994
- [Pierrard07] J.-S. Pierrard; T. Vetter; "Skin Detail Analysis for Face Recognition" IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07. 17-22 June 2007 Page(s):1 – 8
- [Pretzel07] Pretzel and Lotz. Research project: "Face recognition as a search tool" Technical report, Bundeskriminalamt Wiesbaden, 2007
- [Propp92] M. Propp and A. Samal, "Artificial neural network architecture for human face detection", *Intell. Eng. Systems Artificial Neural Networks* 2, 1992, 535–540
- [Rama05] A.Rama, and F.Tarrés, "P2CA: A new face recognition scheme combining 2D and 3D information", IEEE International Conference on Image Processing, Genova, Italy, September 11-14, 2005
- [Rama06] A. Rama, F. Tarres, "Partial PCA Vs Partial LDA", IEEE International Conference on Multimedia and Expo, Toronto, Canada, July 9-12, 2006
- [Rowley98] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection", *IEEE Trans. Pattern Anal. Mach. Intell.* 20, January 1998, 23–38
- [Samal95] A. Samal and P. A. Iyengar "Human face detection using silhouettes", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 9, no.6, pp. 845-867, 1995
- [Samani06] A. Samani, J. Winkler, M. Niranjana, „Automatic Face Recognition Using Stereo Images”, in IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, May 14th- 19th, 2006
- [Savvides04] M. Savvides, B. V. Kumar, and P. K. Khosla, "Eigenphases vs. Eigenfaces" *Int. Conf. On Pattern Recognition*, Washington DC, 2004
- [Sen07] Sen Wang; Yang Wang; Miao Jin; Gu, X.D.; Samaras, D.; "Conformal Geometry and Its Applications on 3D Shape Matching, Recognition, and Stitching" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 29, Issue 7, July 2007 Page(s):1209 – 1220
- [Sinha94] P. Sinha "Object recognition via image invariants: A case study," *Investigative Ophthalmology and Visual Science*, vol. 35, no. 4, pp. 1735-1740, 1994
- [Socolinsky04] Socolinsky, D.A.; Selinger, A.; "Thermal face recognition in an operational scenario", in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. Volume 2, 27 June-2 July 2004 Page(s):II-1012 - II-1019 Vol.2
- [Szeliski06] Richard Szeliski "Image alignment and stitching: a tutorial" in *Found. Trends. Comput. Graph. Vis.*, Vol. 2, No. 1. (2006), pp. 1-104.
- [Torres00] L.Torres, L.Lorente, J. Vila. "Automatic Face Recognition of Video Sequences Using Self-eigenfaces" *International Symposium on Image/video Communication over Fixed and Mobile Networks*, Rabat, Morocco, April 17-20, 2000

- [Tsalakanidou04] F. Tsalakanidou, S. Malassiotis, and M. Strintzis, “Integration of 2D and 3D Images for enhanced face authentication”. Sixth International Conference on Automated Face and Gesture Recognition, pages 266-271, May 2004
- [Tsapatsoulis 98] N. Tsapatsoulis, N. Doulamis, A. Doulamis, and S.Kollias, “Face extraction from non-uniform background and recognition in compressed domain” *IEEE ICASSP*, Seattle, WA, USA, May 1998
- [Turk91] M. A. Turk, A. P. Pentland, “Face recognition using eigenfaces”, Proceedings of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 586-591, Maui, Hawaii 1991
- [UPC-FaceDatabase] “UPC Face Database” in <http://gps-tsc.upc.es/GTAV>
- [Vezhnevets03] V. Vezhnevets, V. Sazonov, and A. Andreeva. “A survey on pixel-based skin color detection techniques” In Proc. Graphicon, 2003.
- [Viola01] Viola P., Jones M.: Rapid Object Detection using a Boosted Cascade of Simple Features, Computer Vision and Pattern Recognition, 2001
- [Wilder96] Wilder, J., Phillips, P. J., Jiang, C., and Wiener, S. “Comparison of visible and infra-red imagery for face recognition” In Proceedings of the 2nd international Conference on Automatic Face and Gesture Recognition (FG '96) (October 14 - 16, 1996). FG. IEEE Computer Society, Washington, DC, 182, 1996
- [Wiskott99] L. Wiskott, J.M. Fellous, N. Kruger, and C. von der Malsburg, “Face recognition by elastic bunch graph matching” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775—779. Revised version 1999
- [Yang04] J. Yang, D. Zhang, A.F. Frangi, and J.Yang, “Two-Dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition”, in *IEEE Trans. on Pattern Analysis and Machine. Intel.*, Jan. 2004
- [Young93] Young, J.W.: Head and Face Anthropometry of Adult U.S. Civilians. Technical Report ADA268661, Civil Aeromedical Institute, Federal Aviation Administration (1993)
- [Yueming07] Yueming Wang; Gang Pan; Zhaohui Wu; “3D Face Recognition in the Presence of Expression: A Guidance-based Constraint Deformation Approach” *IEEE Conference on Computer Vision and Pattern Recognition*, 2007. CVPR '07. 17-22 June 2007 Page(s):1 – 7
- [Yuille92] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, “Feature extraction from faces using deformable templates,” *International Journal of Computer Vision* vol. 8, no. 2, 99–111, 1992
- [Zhao06] W. Zhao, and R.Chellappa, “Face Processing: Advanced modeling and methods”, Academic Press, 2006
- [Zhou03] S. Zhou, V. Krueger, and R.Chellappa, “Probabilistic recognition of human faces from video” in *Computer Vision and Image Understanding*, 91:214-245, 2003