2014

# Face recognition with the RGB-D sensor

Cesare Ciaccio

# FACE RECOGNITION WITH THE RGB-D SENSOR

**Cesare Ciaccio**

**Thesis submitted**
**to the College of Engineering and Mineral Resources**
**at West Virginia University**

**in partial fulfillment of the requirements for the degree of**

**Master of Science**
**in**
**Electrical Engineering**

**Guodong Guo, Ph.D., Chair**
**Xin Li, Ph.D.**
**Marvin Cheng, Ph.D.**

**Department of Computer Science and Electrical Engineering**

**Morgantown, West Virginia**
**2014**

**Keywords: RGB-D face recognition, Kinect face recognition, 3D face recognition, pose correction**

# Abstract

FACE RECOGNITION WITH THE RGB-D SENSOR

Cesare Ciaccio

Face recognition in unconstrained environments is still a challenge, because of the many variations of the facial appearance due to changes in head pose, lighting conditions, facial expression, age, etc. This work addresses the problem of face recognition in the presence of 2D facial appearance variations caused by 3D head rotations. It explores the advantages of the recently developed consumer-level RGB-D cameras (e.g. Kinect). These cameras provide color and depth images at the same rate. They are affordable and easy to use, but the depth images are noisy and in low resolution, unlike laser scanned depth images. The proposed approach to face recognition is able to deal with large head pose variations using RGB-D face images. The method uses the depth information to correct the pose of the face. It does not need to learn a generic face model or make complex 3D-2D registrations. It is simple and fast, yet able to deal with large pose variations and perform pose-invariant face recognition. Experiments on a public database show that the presented approach is effective and efficient under significant pose changes. Also, the idea is used to develop a face recognition software that is able to achieve real-time face recognition in the presence of large yaw rotations using the Kinect sensor. It is shown in real-time how this method improves recognition accuracy and confidence level. This study demonstrates that RGB-D sensors are a promising tool that can lead to the development of robust pose-invariant face recognition systems under large pose variations.

*Alla mia famiglia.*

# Acknowledgements

Before starting the presentation of my work, I would like to take the time to thank a few people that accompanied and supported me during my studies at the West Virginia University. It wouldn't have been possible for me to get to the day of my graduation without them.

First of all, I owe my deepest gratitude to my advisor, Dr. Guo, for his experienced guidance and support. He gave me the opportunity to work on exciting projects and develop valuable skills. He constantly encouraged me to look at my work from different perspectives and open my mind. He made this thesis possible.

I wish to thank my lab mates for their constant encouragement and always valuable suggestions. In particular Lingyun, who I had the pleasure to work with several times and who participated in this project.

My sincere gratitude goes also to Dr. Cheng and Dr. Li for serving on my committee and taking the time to carefully review my work and provide me with precious feedback.

I also owe my gratitude to my fiancé Anna for her personal support and great patience.

I'm thankful to my parents and my brother for always being there for me, supporting my decisions and dealing with my absence.

Last but not least, I would like to thank all my friends in Morgantown that became my extended family during this experience.

# Contents

# List of Figures

# List of Tables

# Chapter 1. Introduction

This chapter introduces the problem addressed in this work, gives a brief overview of previous approaches and explains the motivations behind this project. Most of the faces shown in this work are publicly available as part of the CurtinFace dataset [2] (figures 11, 12, 16, 17, 19, 20, 21, 22, 23, 36) or the Eurecom dataset [3] (figures 1, 2, 6, 14). The remaining images are publicly available on the Internet.

## 1.1 Problem definition

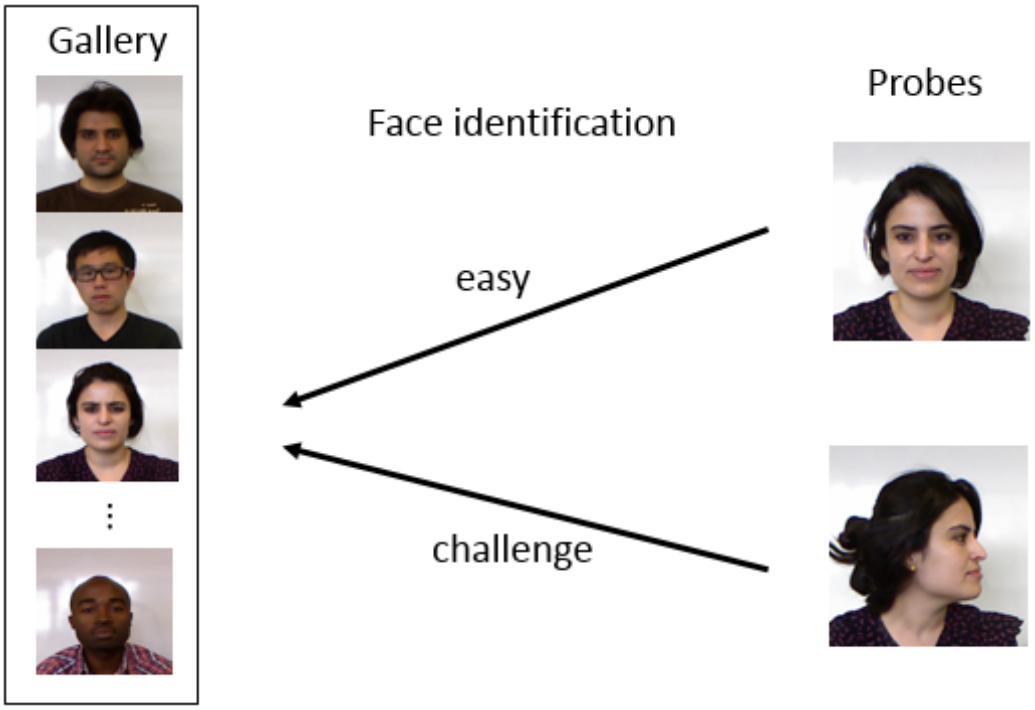


*Figure 1. Problem statement – When both gallery image and probe are frontal, state-of–the-art algorithms can easily recognize faces. However, when the pose of the probe is not frontal, the identification becomes more challenging.*

Biometrics techniques have many practical applications in security-related fields like access control, identity management, video surveillance, law enforcement, homeland security, etc [4, 5, 6, 7]. They can also be useful to

improve human-machine interaction [8, 9, 10]. As one of these techniques, face recognition is an active field of research that received a great deal of attention in the last few decades [11, 12, 13]. The research community has made significant progress since the very first face recognition attempts about 50 years ago [14]. The first scientific publication dates back to 1971 [15]. Several face recognition algorithms have been developed since then, and great performances have been achieved in experiments with controlled environments. However, face recognition in practical applications is still a challenge, because of the variations of the facial appearance that can be caused by head pose, lighting conditions, facial expression, aging, etc [16, 17, 18, 19, 20, 21]. Among all the possible variations, probably the biggest challenge is represented by changes of the facial appearance in the 2-dimensional space caused by 3-dimensional head rotations. To address this problem, several methods have been proposed and investigated [22-32]. Face recognition with significant pose variations is the focus of this study. From now on, the assumption is that all the images under analysis were taken under the same uniform light conditions (no light variations), neutral face (no facial expressions) and within a relatively short temporal distance between them (no aging). Figure 1 gives a visual representation of the problem addressed in this work. The general structure of a face recognition system is fairly simple. There is always a set of face images stored in memory usually referred as the “gallery”. The gallery contains face-identity pairs, and it represents the “knowledge” of the system. When

the system is given a test image (or probe) to identify, it compares the probe to each image in the gallery, looking for the most similar face (nearest neighbor classifier). The identity of the most similar face is returned as the identity of the probe. If the probe and the gallery face images are both frontal, state-of-the-art algorithms are able to give almost perfect results. However, when the probe face is not frontal, the problem becomes more challenging, and the difficulty is proportional to the size of the rotation angle.

## 1.2 RGB-D based face recognition

Researchers in the field of face recognition have developed many methods to deal with head pose variations. These methods can be grouped in two big categories: 2D-based methods and 3D-based methods [33, 34].

2D-based methods use only 2D images of the face. The common denominator of these methods is the attempt to learn the relationship between 2D face images in front view and 2D face images in side view, and then apply that knowledge to recognize side view test faces while the gallery contains only front view images [35-47]. Figure 2 shows the general idea. The set of faces on the left are used to learn the relationship between the 2D frontal images and the 2D side view images. The mapping that is learned is then used to compare side view test faces to a front view gallery and infer the identity of the probes. Figure 3 gives a representation of the method presented in [47]. The algorithm projects the features extracted from the faces to a common latent space where the distance between the two

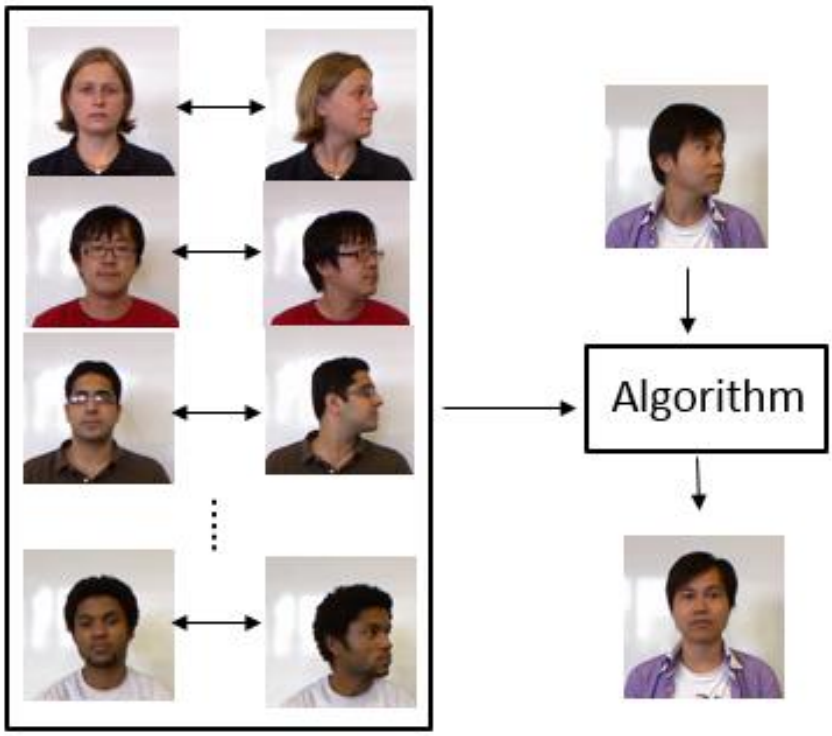


*Figure 2. 2D-based approach – This type of method tries to learn a relationship between the frontal face and the side face. Then, that knowledge is used to recognize side view test faces while the gallery contains only front view images.*

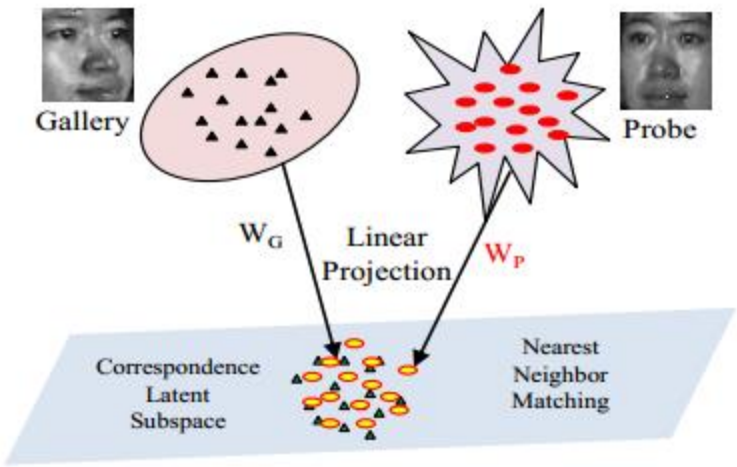


*Figure 3. 2D-based example – This image is shown in [47]. The method consists in projecting the feature vectors in a different subspace where the two versions of the face (side and frontal) are "close". After the images are projected in the new space, the nearest neighbor classifier is used.*

instances of the same face (front and side) is small. After the projection, the nearest neighbor classifier is used to infer the identity. These types of methods are best used with small rotation angles (up to 30 or 40 degrees), but have difficulty handling larger rotations.

3D-based methods address the problem of face recognition using the 3D structure of the face [48]. Some of these methods try to infer the 3D structure of the face from 2D face images [49-54]. Most of the 3D-based methods use 3D information provided by depth images of the faces acquired with laser scanners (Figure 4). These devices use light rays to create a 3D map of the face. A depth map is an image whose pixel values represent distances from the camera. A depth map gives an approximation of the 3D structure of the objects. Some of these approaches require the 3D depth map of both the gallery faces and the test faces [55-62]. Other approaches use the depth map of the gallery faces while the probe is just a 2D image [63-68]. These approaches consist in learning a 3D face model and then adapt the 3D model to a given 2D face image. For example, the classic 3D face recognition method based on the 3D morphable model [69] learns a generic 3D face model from a number of subjects. This 3D model is used to estimate the 3D shape of the face from a single 2D face image by using a 3D to 2D registration process. The 3D shape inferred (described by two parameters) can be used to compare two faces images (Figure 5). However, the 3D to 2D registration is not simple. It is a time-consuming process because there are several parameters to adjust. This process also

needs a good initial estimation of the head pose in order to achieve acceptable results. These types of algorithms are complex and computationally expensive. Also, acquiring the depth images with a laser scanner is a long process (it takes several minutes per image) and the laser scanners are expensive and not easy to manage.

The approach presented in this work explores the advantages of the recently developed consumer-level RGB-D cameras (e.g. Kinect – see Figure 6). RGB-D cameras could play a key role in solving real-world problems in the near future. They caught the attention of the research community after becoming more affordable and more commercially mainstream. Many publications appeared recently, and are spread over a variety of research fields like computer vision, robotics, human-computer interaction and others [70-78]. There are several papers that address the problem of face detection/tracking [79-86]. For an introduction about basics and underlying principles of RGB-D perception and more information about applications, see [87]. More recently, researchers have tried to assess the impact that these devices could have in the field of biometrics, in particular face recognition [2, 88, 89, 90]. The purpose of this study is to show that RGB-D sensors could have a major impact in the field of face recognition. RGB-D cameras provide color and depth images at the same rate. They are affordable and easy to use, but the depth images are noisy and low resolution, unlike laser scanned depth images (Figure 7). The proposed approach is able to handle large head pose variations using RGB-D face

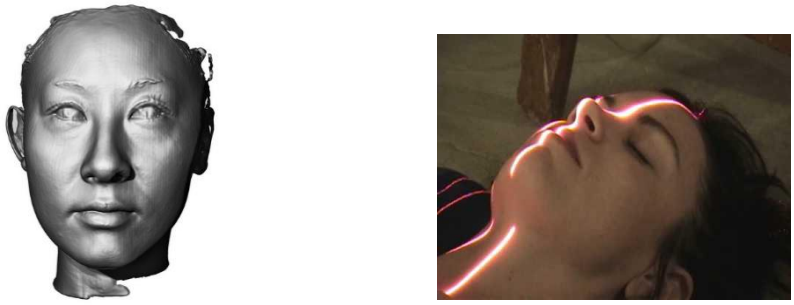

*Figure 4. Laser scanned face images – The image on the left[1] shows an example of a high definition 3D face image obtained with a laser scanner. The image on the right[2] shows the acquisition process.*

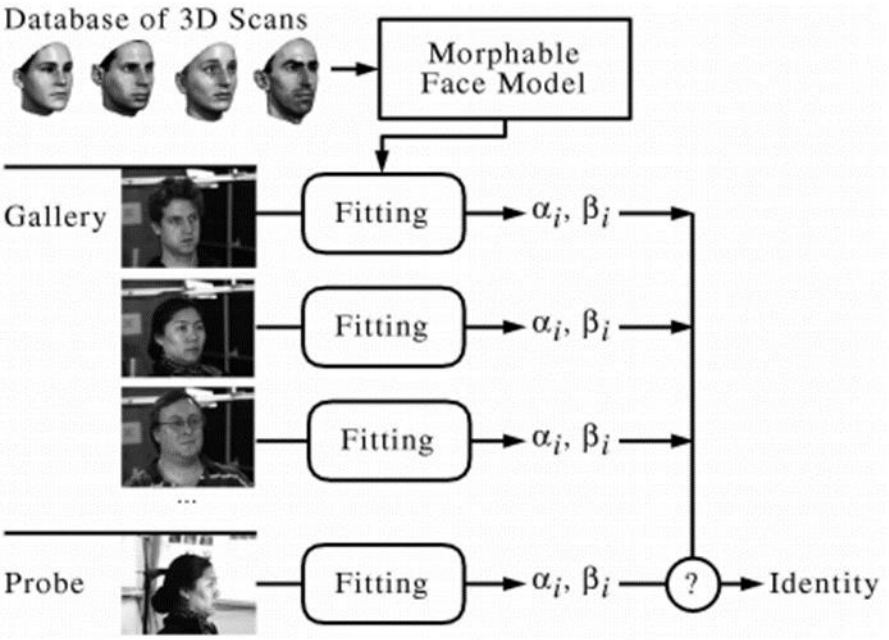


*Figure 5. 3D-based example – This figure is shown in [65]. It represents the face recognition method based on the well-known 3D morphable model. The process consists in learning a generic 3D face model from a number of subjects and then using this 3D model to estimate the 3D shape of a face from a single 2D image. The 3D shape inferred (in form of parameters) is used to compare two faces images.*

[1] www.clonesculptures.com/3dscanning
[2] www.youtube.com/watch?v=4XZfR1at-AQ

images. The method uses the depth information to correct the pose of the face. It does not need to learn a generic face model or make a complex 3D-2D registration. It is simple and fast, yet able to deal with large pose variations and perform pose-invariant face recognition. Experiments on a public database confirm that the approach is effective and efficient. Also, the underlying idea is used to develop a face recognition system that is able to achieve real-time face recognition in the presence of large yaw rotations using the Kinect sensor. It is possible to show, in real-time, how the proposed method improves recognition accuracy and confidence level. A new scheme for pose correction and a new face representation based on integration of the covariance descriptor and the popular local binary patterns are also introduced.


*Figure 6. Kinect RGB-D image – The Kinect sensor shown above[3] is able to capture RGB and depth images at the same rate and resolution.*

*Figure 7[4]. Laser vs. RGB-D – Laser scanners can create accurate depth images (left), but the acquisition process is quite slow. RGB-D cameras can acquire depth images at 30fps, but the depth images are noisy (right).*

[3] www.xbox.com/en-US/kinect

[4] www.kickstarter.com/projects/45699157/fuel3d-a-handheld-3d-scanner-for-less-than-1000

## 1.3 Previous works

To the best of the author's knowledge, there are only 4 previous works that address the problem of face recognition using RGB-D sensors [2, 88, 89, 90]. These approaches are different from the one presented in this study. For example, in [88] only the depth images are used for face recognition. The method uses the point cloud for face identification. The color images are not used and the problem of pose variation is not addressed. Figure 8 shows the registration process of the point clouds proposed in [88]. Both color and depth images are used for identification in [89]. This method computes an entropy measure for both RGB and depth images and combines them to improve face recognition. But, the faces are all in front view and there is no pose variation. Figure 9 gives a visual representation of the process. The work in [90] is quite different. A sequence of RGB-D images (video) is used to improve face identification over time. Figure 10 gives a representation of the method. Only one of these previous works ([2]) considers pose variations, but the goal and experiment setup are very different. The gallery consists of 18 images per subject (Figure 11). For each subject, there are images with different poses, different expressions and different light conditions. The test images present two of these variations at the same time (Figure 12). The goal is to identify a face when there are two different variations at the same time. Also, the method in [2] requires a complex registration of each face to a 3D face model obtained from a laser scanner.

The approach presented in this document is different from the ones previously described. The major focus of this study is exploring the advantage of using RGB-D cameras for pose-invariant face recognition. The method proposed addresses the problem of pose-invariant face recognition using both color and depth images. Also, the gallery contains only one frontal RGB-D image for each subject and the probes consist of a single side view RGB-D image per subject.

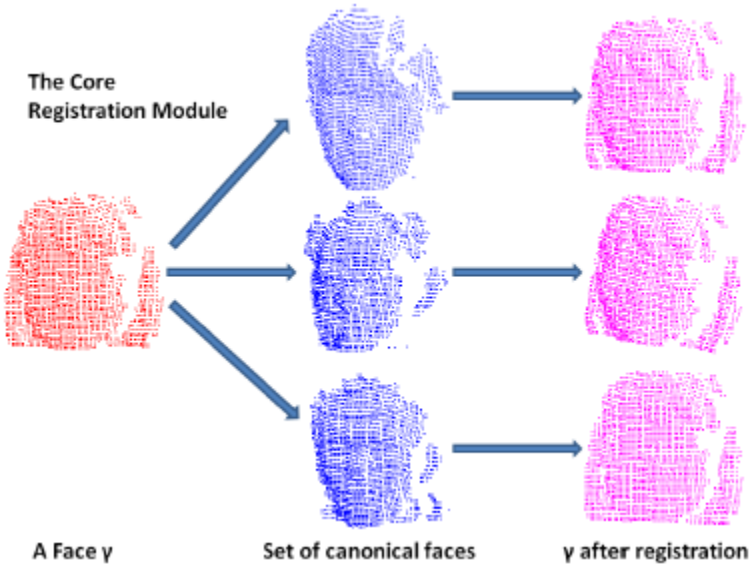


*Figure 8. Point cloud registration – This figure is shown in [88]. The method represented uses only the depth information for face recognition. It doesn't not address the problem of pose variation and requires a registration of 3D point clouds.*

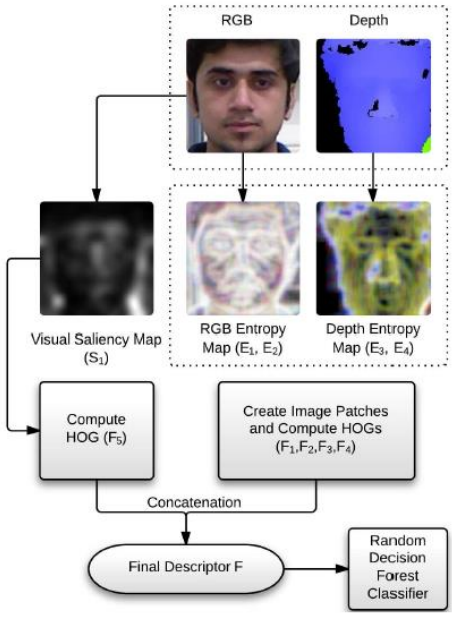


*Figure 9. Entropy-based method – This figure is shown in [89]. This method combines entropy measures of the color and depth images in order to improve face recognition. It also does not address the problem of pose variation.*

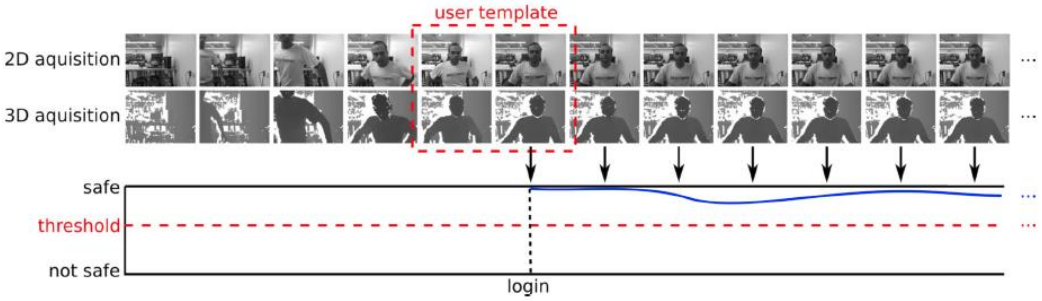


*Figure 10. Video-based face recognition – This figure is shown in [90]. The method presented in [90] is completely different from the one proposed in this work because a sequence of face images in a time interval is used to improve the face recognition.*

*Figure 11. Gallery in [2] – The gallery contains 18 images for each subject. There are images with different pose, illumination and expression. In this study the gallery contains only one neutral front view image per subject.*

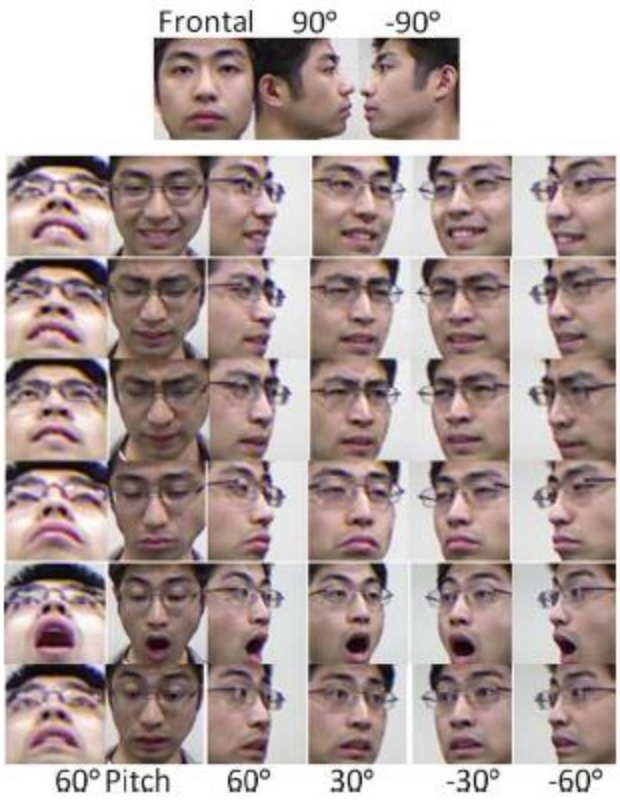


*Figure 12. Test images in [2] – The test images contain two variations at the same time. In this figure you can see pose and expression. The goal in [14] is recognizing a face that shows pose AND expression variations, while the images in the gallery present ONLY pose OR expression variation.*

# Chapter 2. Face representation

Comparing two face images requires extracting features from those two images. These features are a numerical representation of the face. The representation used in this study is a combination of two different descriptors: one is the popular Local Binary Patterns (LBP) and the other is the covariance descriptor. This work investigates whether these two types of features can be complementary for face recognition and if a better representation can be obtained by integrating the two.

## 2.1 Uniform Local Binary Patterns

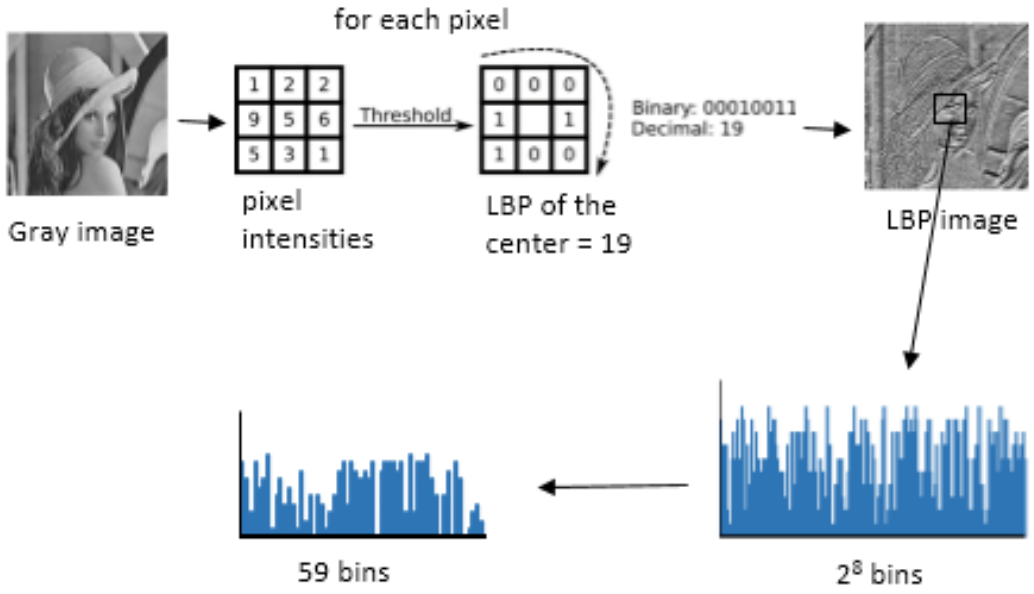


*Figure 13[5]. Uniform LBP – Each pixel is represented by 8 binary digits. The digits are calculated using the center pixel as threshold. Each image patch is represented by a histogram with 59 bins: 58 bins for the "uniform" patterns (at most 2 transitions from 0 to 1 or 1 to 0) and 1 bin for the "non-uniform" patterns.*

Originally designed for texture description [91, 92], LBP [93] features are a well-known and popular method for face recognition [94, 95, 96, 97]. They

[5] The face in the figure comes from http://en.wikipedia.org/wiki/Lenna

have been used extensively and give state-of-the-art results in many face recognition applications. Several variants of this descriptor have been proposed and investigated [98-104]. In this study, the Uniform LBP is used. Uniform LBP [105] is a variant of the original operator that reduces the length of the feature vector. Certain local binary patterns called “uniform” patterns are more descriptive than the “non-uniform” ones. The term “uniform” refers to the uniform appearance of the local binary patterns. Within uniform patterns, there are at most two bitwise transitions (e.g. 1000000 is uniform, 01101111 is not uniform). These uniform patterns represent the majority of the 3x3 texture patterns. Also, the most frequent uniform binary patterns correspond to features such as edges and corners; this makes them the best matching patterns. Figure 13 shows the uniform LBP calculation process on a gray scale image. The LBP operator returns an 8-digit binary number for each pixel in the image. To create the binary descriptor for each image pixel, the operator compares the 3 x 3 neighboring pixels to the value of the center. If the value of the neighboring pixel is greater than the value of the center, a “1” digit is assigned to that descriptor; otherwise, a “0” digit is assigned. Once all neighboring pixels are evaluated, an 8-digit binary code is created and assigned to the center pixel. When the LBP image is created, patches of this image can be described by a 2^8 bins histogram. If a pixel is represented by 8 digits, there are 256 possible patterns but only 58 are uniform. To describe a patch, a 59 bins histogram is used: 58 bins for each one of the uniform patterns and 1 for

the non-uniform patterns. To compare histograms, the popular chi-squared distance is used. It is defined as:

$$d(H_1, H_2) = \frac{1}{2}\sum_{i=1}^{n}\frac{[H_1(i) - H_2\,(i)]^2}{H_1(i) + H_2(i)}$$

## 2.2 Covariance descriptor

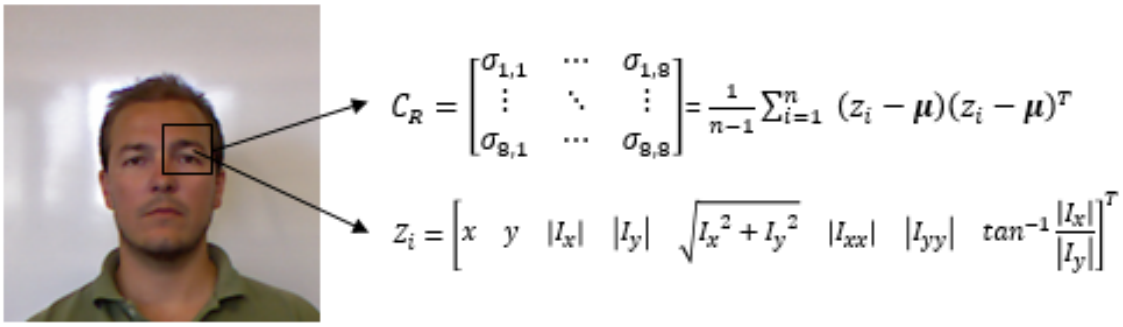


*Figure 14. Covariance descriptor – Each pixel is described by multiple values and the descriptor is considered a realization of a random vector. The descriptor of a patch is the sample covariance matrix of the pixel descriptors contained in the patch.*

The formal definition of the covariance descriptor was first given in [106]. The covariance descriptor has been successfully used for pedestrian detection, object tracking, action recognition [107, 108] and facial recognition [109, 110, 111, 112]. The usage of covariance matrices as a region descriptor provides some advantages. One of them is that the representation proposes a natural way of fusing multiple features. Each pixel is described by multiple values and the descriptor is considered a realization of a random vector. The pixel location, intensity derivatives, and the edge orientation are computed as features for each pixel. This work

uses the same settings as in [106] to form the feature vector $z_i$ for each pixel (x, y),

$$z_i = \left[ x \quad y \quad |I_x| \quad |I_y| \quad \sqrt{{I_x}^2 + {I_y}^2} \quad |I_{xx}| \quad |I_{yy}| \quad tan^{-1}\frac{|I_x|}{|I_y|} \right]^T$$

Each region R is represented by a covariance matrix. If *d* is the dimensionality of the feature vectors $z_i$, the region R is represented by the *d*×*d* sample covariance matrix of the feature vectors:

$$C_R = \frac{1}{n-1}\sum_{i=1}^{n} (z_i - \boldsymbol{\mu})(z_i - \boldsymbol{\mu})^T$$

where μ is the mean vector and *n* is the number of points in region R. Covariance matrices are symmetric and positive semi-definite, hence they reside in the Riemannian manifold [113]. The distances between two covariance matrices can be calculated in the Riemannian manifold. The distance between points is given by the length of the geodesic (the minimum length curve connecting two points on the manifold). It is defined as ([106]):

$$d(C_1, C_2) = \sqrt{\sum_{j=1}^{k} ln^2 \lambda_j(C_1, C_2)}$$

where $\{\lambda_j(C_1, C_2)\}_{j=1\ldots k}$ are the generalized eigenvalues.

To compute the covariance descriptor and the distance between covariance matrices, the code available at [114] is used.

## 2.3 Integration scheme

The covariance descriptor is very different from the classic LBP features. These two features cannot be concatenated directly because the covariance descriptor is a matrix while the LBP descriptor is a histogram. A probability-based integration scheme that combines the covariance descriptor with the LBP features is proposed. This method transforms the distances measured into probabilities, then multiplies them. Let $d_1$ be the distance between a probe and a gallery face in the covariance space. Let $d_2$ be the distance in the LBP space. The new descriptor is defined as:

$$P(d_1, d_2) = A \cdot e^{\left(\frac{d_1}{\lambda_1}\right)} \cdot e^{\left(\frac{d_2}{\lambda_2}\right)}$$

where $\lambda_1$ and $\lambda_2$ are empirically adjusted parameters used to balance the inputs from the two different features. In the experiments, the values used are $\lambda_1$ = 1 and $\lambda_2$= 10. *A* is a constant to maintain a probability-based measure. The proposed integration scheme transforms the two distances into a probability.

# Chapter 3. Pose-invariant face recognition

This chapter starts with an overview of the approach and then describes the pre-processing method, the gallery creation process, the face alignment procedure and the patch selection. The following content is based on [1].

## 3.1 Method overview

The goals are developing a face recognition algorithm that can handle large head pose changes and identifying the advantages of using the RGB-D face images. In the literature, the user's head pose is often defined by three angles: yaw, pitch, and roll (see Figure 15). This scheme is borrowed from aeronautics, where the same three angles define the orientation of the aircraft. The coordinate system of reference is usually the Cartesian system with the origin being in the camera. In this work, the focus is on facial appearance changes caused by head pose variations only in the yaw direction. This means that the gallery contains one RGB-D face image in front view for each enrolled subject (that is yaw, pitch and roll equal to 0), while the probe faces have pitch equal to 0, roll equal to 0, and yaw in the range of -90 to 90 degrees. As discussed in the introduction, this is a challenging problem for traditional 2D-based face recognition. Traditionally, a pose correction approach would try to correct the pose of the probe by creating a front view face based on the side view and then match the newly generated face against the front view gallery images [115, 116, 117, 118]. The method described in this study uses the opposite approach. It uses the

RGB-D front view face images to create side view (yaw at -90 to 90 degrees) RGB face images (Figure 16 and 20); the test images are then matched against all the generated face images and the best match is considered for identification.

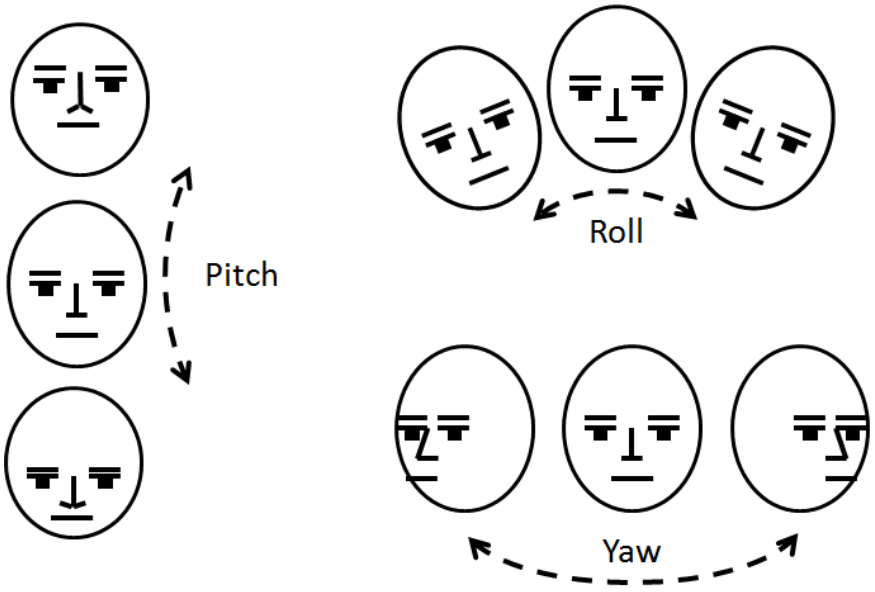


*Figure 15[6]. Head angles definition*

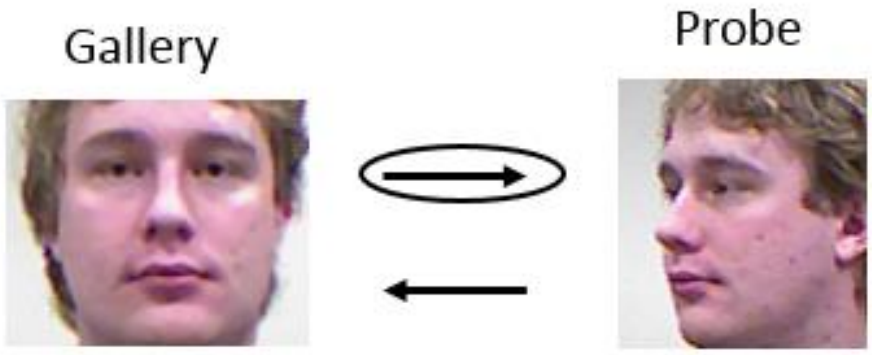


*Figure 16. Rotation direction – Unlike traditional approaches, this method rotates the gallery face to the side and then match the probe against the newly generated image.*

[6] msdn.microsoft.com/en-us/library/jj130970.aspx

## 3.2 Face images pre-processing

The first steps of the process are face detection and landmarks localization on the RGB-D images. There is an assumption that the RGB and depth images are aligned, because the RGB-D sensors have this capability. Face detection is performed on the color image only, using a recently developed technique [119]. The landmark points of interest in this work are the eye corners, the mouth corners and the nose tip, which can also be detected with the method presented in [119]. However, it is useful and easy to refine the nose tip position estimation using the depth image; it is the closest point to the RGB-D camera. The accuracy of the location of the nose tip is very important because the nose tip is used to estimate the position of the head center. The head center is crucial in the process of side view face generation. As stated before, the depth images provided by RGB-D cameras are noisy. Therefore, the depth images are filtered with a median filter and then a Gaussian filter. This filtering is important to improve the landmarks detection and the quality of the generated images. Also, the depth information is used to perform background subtraction on the RGB face images. Figure 17 illustrates the pre-processing steps.

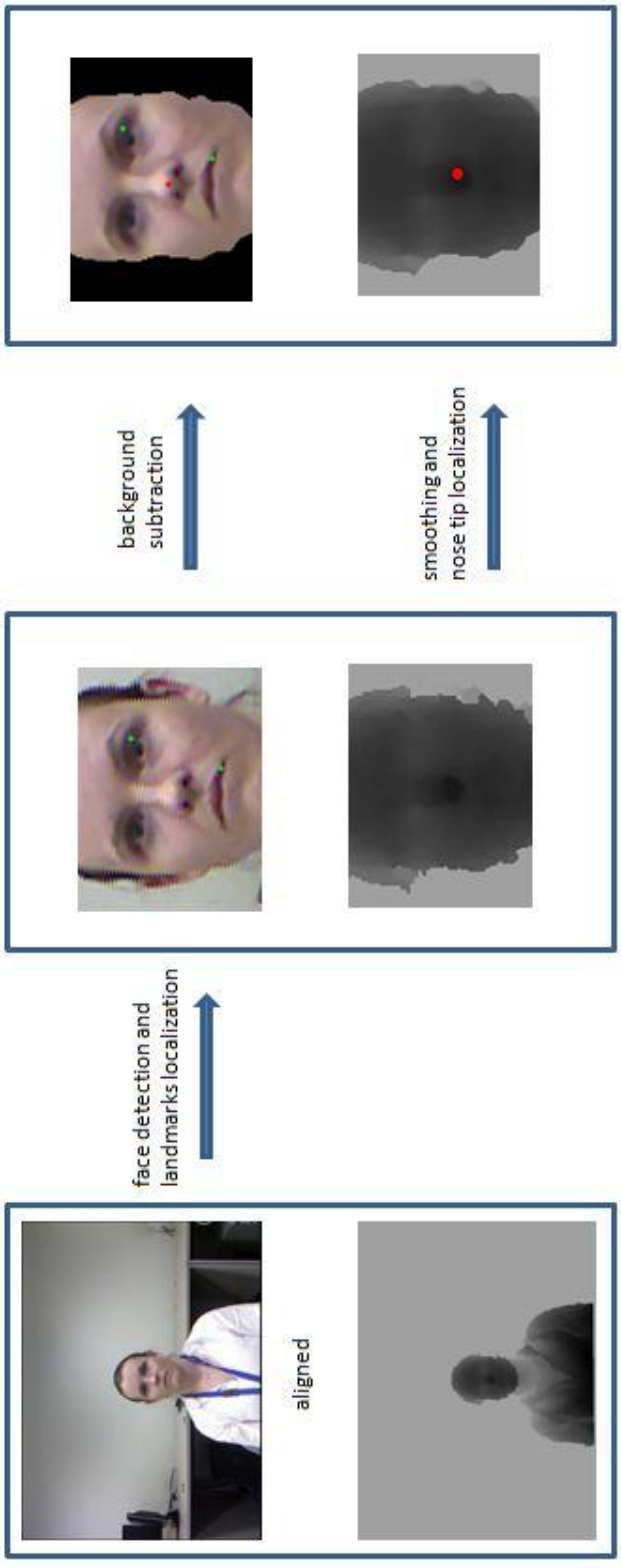


*Figure 17. Pre-processing of RGB-D face images – Face detection and landmarks localization are performed on the RGB image. The depth image is first smoothed and then used for nose tip position refinement and background subtraction.*

## 3.3 Face synthesis

From each RGB-D face image in the gallery, a set of face images with different pose angles is automatically generated by applying a 3D rotation to the RGB-D point cloud. In order to do this, the first step is computing the axis of rotation. The axis chosen is the vertical axis that passes by the center of the head. The center of the head is estimated as:

$$(x_0, y_0, z_0) = (n_t(x), n_t(y), n_t(z) + \delta)$$

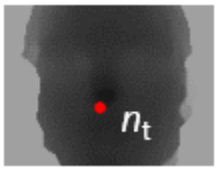


*Figure 18. Nose tip*

where $n_t$ is the nose tip position (Figure 18), and $\delta$ is the distance between the vertical axis passing by the nose tip and the vertical axis passing by the center of the head (axis of rotation). In the experiments, the value used is $\delta = 50\ mm$ and the results show that this setting works well for all the faces in the database. Each head is rotated around the Y axis of the coordinate system with the origin in the center of the head. This simple but effective approach to head rotation is illustrated in Figure 19, and it is presented in Algorithm 1 in detail. In the algorithm, *T* is the transformation given by:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = T \begin{pmatrix} x \\ y \\ z \\ C \\ \theta \end{pmatrix} = \begin{bmatrix} \cos\theta & 0 & \sin\theta & x_0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & z_0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x - x_0 \\ y \\ z - z_0 \\ 1 \end{bmatrix}$$

T rotates a 3D point (x, y, z) on the original face surface to a new location (x', y', z') with a yaw angle θ. *C*=(x0, y0, z0) is the center of the head. In theory the rotation can be any combination of the yaw, pitch, and roll angles. In this study the focus is on the yaw rotation. A sample sequence of generated face images is shown in Figure 20.

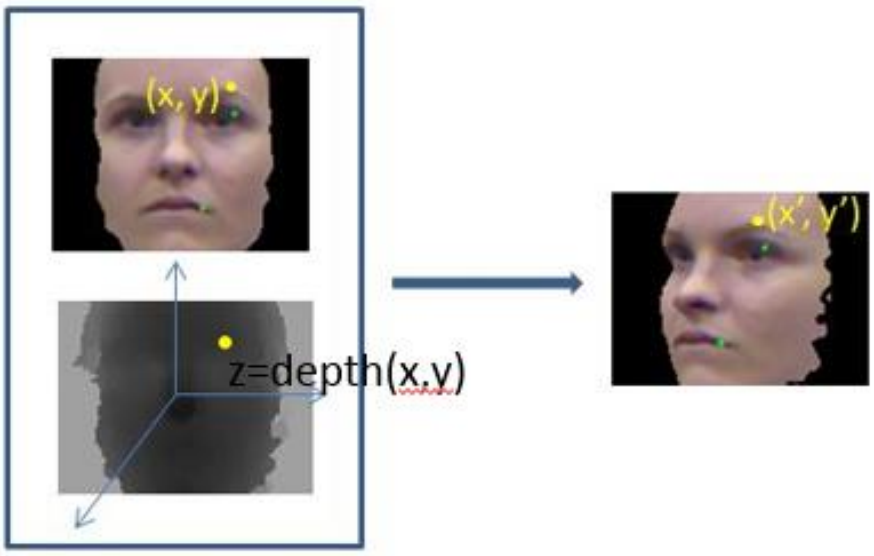


*Figure 19. Face synthesis – On the left, the RGB-D point cloud is shown. The depth image represents the 3D shape of the face; the RGB image gives the color information of each 3D point. Each 3D point carries with it the color information when it is rotated. Because of numerical approximation, the generated side view image has holes, so interpolation is needed.*

**Algorithm 1.** Synthesis of side-view face images

```
for each front-view RGB-D gallery image do
  estimate center of the head C = (x0, y0, z0)
  for θ = 5 to θ = 90 do
    create a blank RGB image for the target side-view face
    for each foreground pixel (x, y) in the front-view image do
      (x', y') = T(x, y, Depth(x, y), C, θ)
      side_view_RGB(x', y') = front_view_RGB(x, y)
    end
    do interpolation on the created image
  end
end
```

*Table 1. Algorithm*

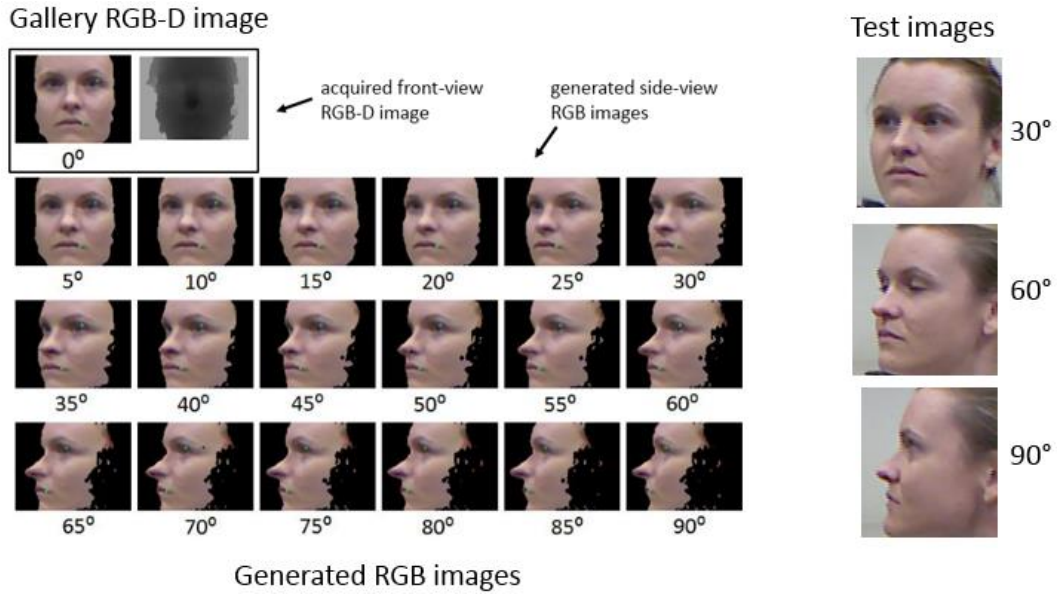


*Figure 20. Sample sequence of generated gallery images – The test images are compared against all the gallery images generated with the method outlined in Table 1.*

## 3.4 Face alignment

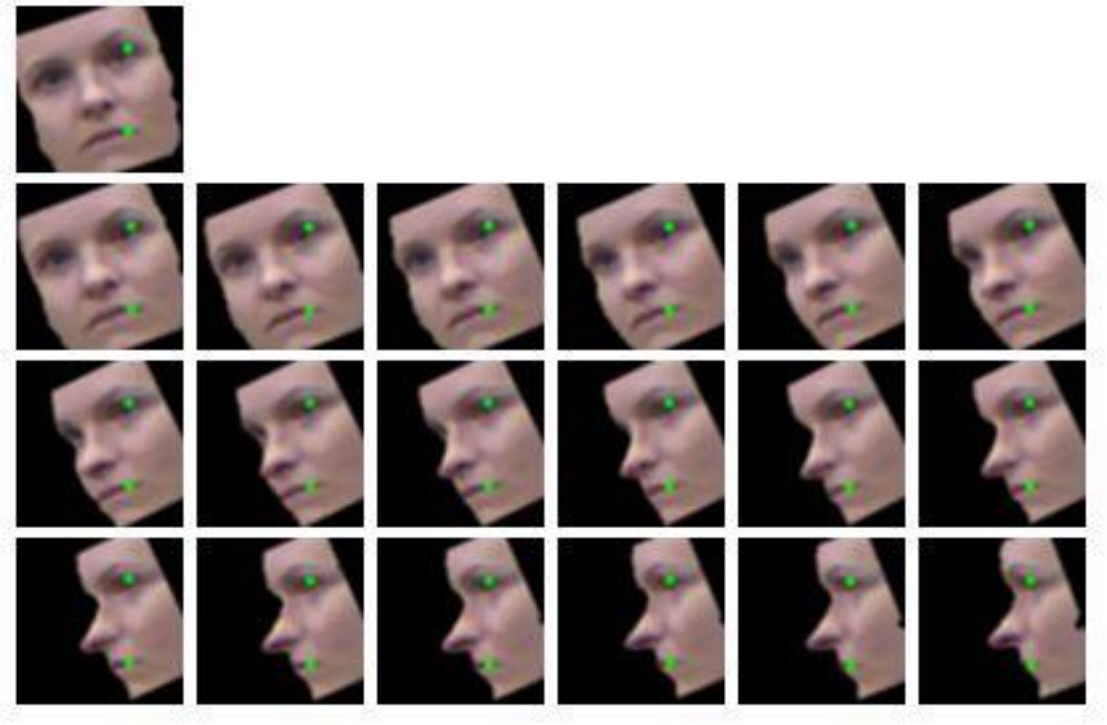

*Figure 21. Face alignment – Sample gallery faces aligned and resized according to eye corner and mouth corner positions.*

The face representation used (LBP and covariance features) requires face alignment. Before comparing two images, they need to be scaled to the same size and aligned according to specific key points. In this work, the eye corners and mouth corners are used to align the face images. Since there are times that the eye and mouth corners are not detected precisely (especially in the horizontal direction when the head rotations are large), the alignment step uses the vertical distance between the eyes and mouth. All the faces are aligned to have the same vertical distance between the eyes and mouth. Traditionally, the two eyes are used for face alignment, which is

not possible in this method because with head rotations close to 90 degrees one eye cannot be seen due to self-occlusion. After the alignment, each face is cropped and resized to 60x60 pixels to then perform face identification.

## 3.5 Patch selection

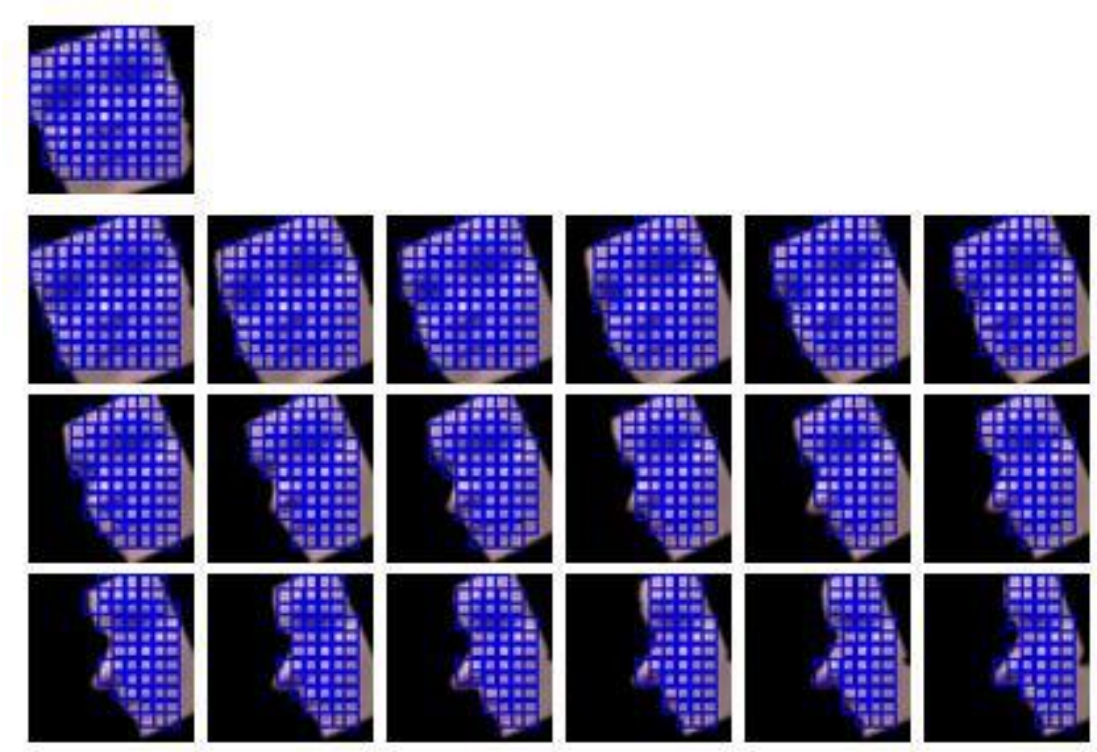

*Figure 22. Patch selection – 10x10 overlapping patches are selected from the face. Corresponding patches from different images are pulled and compared.*

The matching step requires selecting overlapping patches from the face images. Corresponding patches from different images are pulled and features from within those two patches are compared. The distance between two face images (either $d_1$ or $d_2$ referred to in Chapter 2, Section

2.3) is the average of the distances of corresponding patches. The patch size used is 10x10. The patches overlap, and they are selected by shifting every 5 pixels. The number of patches that can be selected in each face image may be different because of different pose angles. Therefore, when comparing two face images, only the patches that contain face details in those images are considered (see Figure 22).

# Chapter 4. Experiments

This chapter presents the experiments conducted to evaluate the proposed approach.

## 4.1 Dataset

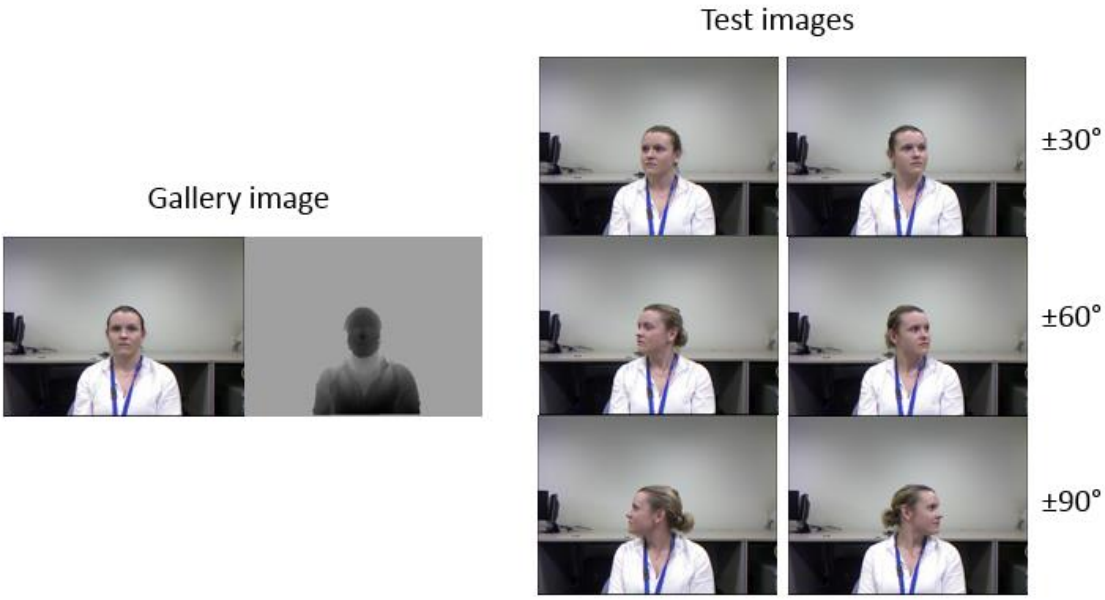


*Figure 23. Sample subject from the CurtinFaces dataset.*

Since using RGB-D for face recognition represents a recent development in this field, there are no standard benchmark databases for RGB-D based face recognition experiments. For the experiments in this work the CurtinFaces database [2] is used which is publicly available. This dataset contains RGB-D face images of 52 subjects. For each subject there are several images acquired under different pose, illumination, facial expression, and disguise. In this study, only a subset of this database is used because only pose-invariant face recognition is addressed. Only the front view face images with neutral expressions and uniform illumination are

used as the content of the gallery. The faces with neutral expressions and uniform illumination but with yaw angles that are different from 0 degrees are used as test images. The dataset contains 6 different yaw angles (±30, ±60, ±90). Figure 23 shows an example.

## 4.2. Results

The experimental results are shown in Tables 2, 3, and 4. Different experiments were conducted in order to get to the final method.

### 4.2.1 Rotation direction

As said in the previous chapter, the goal is to bring gallery images and test images to the same pose before performing face identification. In the first experiment, two different pose correction schemes are evaluated. One rotates the test face images to the front view (others->front), and then compares it to the gallery face images also in front view; the second scheme does the opposite, i.e. each gallery face image is rotated in order to obtain the same pose of the test image (front->others). For this experiment, LBP features only are used. As shown in Table 2, the second scheme is significantly better than the first, especially for large pose variations (60 and 90 degrees).

There are two possible explanations for this result. First, the front view of a face image carries more information about the identity, compared to other views with large angles. Rotating the front view face image simply results in a loss of information. On the other hand, when rotating the other views to

the front, missing information needs to be estimated. Also, estimating the center of the head in non-frontal poses is more difficult. When rotating the frontal face, the nose tip can be used to get a good estimate of the center of the head. Also, the nose tip in a front view can be estimated more accurately using the depth map. On a profile face, the nose tip detection is less accurate; also, there is no easy way to get a reasonable estimate of the head center. In the following experiments, the second scheme (front -> others) is adopted.

| | 30° | 60° | 90° |
|---|---|---|---|
| **LBP (others -> front)** | 75.0% | 59.3% | 35.4% |
| **LBP (front -> others)** | 76.9% | 71.1% | 57.6% |

*Table 2. Rotation direction – Rotating the frontal gallery images to the pose of the probes gives better results than rotating the probes to front view.*

### 4.2.2 Rotation density

When rotating the front view to the side view, it is necessary to know the exact pose angle of the side view. The exact estimation of the head pose from 2D images is a challenging problem. In the second experiment, it is shown that this problem can be mitigated by creating multiple face images with different pose angles, and then compare the probe face to all those synthesized gallery face images. The one with the highest match score will

be used as the identification result. Table 3 shows the results for this experiment. Two schemes are compared. One rotates the gallery images to the estimated pose of the probes (only 30, 60, 90 degrees are present in the dataset). The other scheme generates 18 different poses varying the yaw angles from 5 to 90 degrees in 5 degrees increments, and then it compares the probe with each of the generated face images. Table 3 shows that the latter approach gives better results.

| | **30°** | **60°** | **90°** |
|---|---|---|---|
| **LBP (limited poses)** | 76.9% | 71.1% | 57.6% |
| **LBP (dense poses)** | 92.3% | 73.0% | 65.3% |

*Table 3. Rotation density – Generating images at multiple pose angles gives better results than estimating the probe pose.*

### 4.2.3 Features comparison

In the last experiment, the LBP and covariance features are compared, and the new face representation that integrates both descriptors is evaluated. The results presented in Table 4 demonstrate the effectiveness of the covariance descriptor in the context of face recognition. It can be observed that the covariance descriptor gives better results than LBP for pose angles of 60 and 90 degrees. Most importantly, the experiment shows that the new

representation that integrates the two features gives a higher accuracy in all poses. This demonstrates that the two features can be complementary to each other, and the combination of the two can improve the face recognition accuracy significantly.

| | 30° | 60° | 90° |
|---|---|---|---|
| **LBP** | 92.3% | 73.0% | 65.3% |
| **COV** | 92.3% | 76.9% | 69.2% |
| **LBP+COV** | **94.2%** | **84.6%** | **75.0%** |

*Table 4. Features comparison – The covariance features give better results than LBP. The integration of both features gives an even higher accuracy.*

# Chapter 5. Face recognition with Kinect

This chapter describes the face recognition software developed as part of this project. This software uses the Kinect sensor and its recently developed face tracking capabilities.

## 5.1 Kinect face tracker

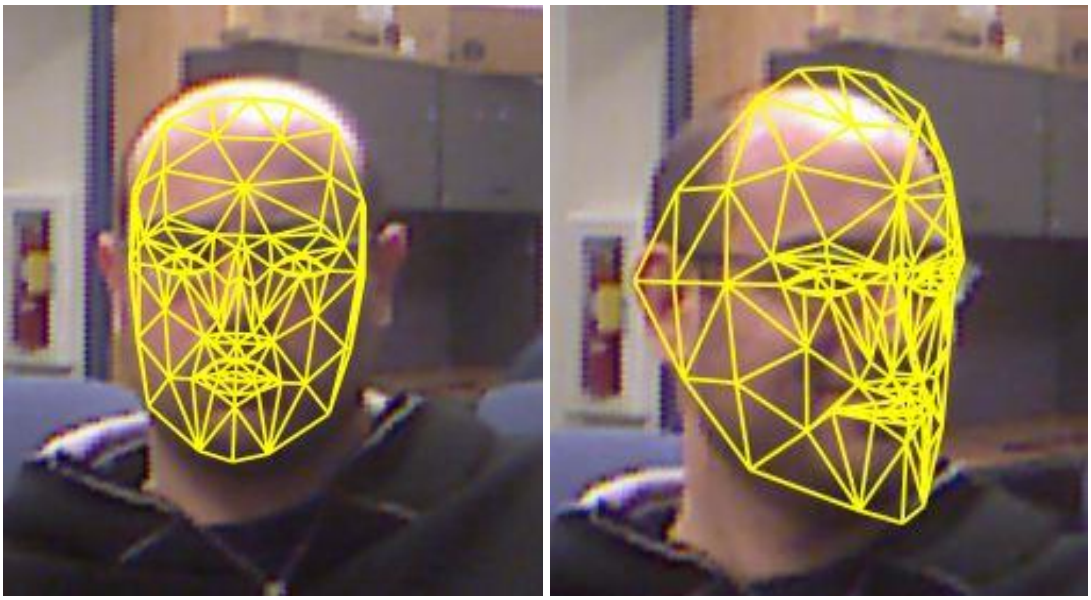

*Figure 24. Kinect face tracking – The Kinect is able to track human faces and follow head moves in real-time. It also fits a 3D triangular mesh to the depth map of the face and detects key points (eyes, nose, mouth, etc.) of the face using the color image.*

The Kinect RGB-D sensor has face tracking capabilities and is able to follow users' head moves in real-time. It is also able to fit a 3D triangular mesh to the depth map of the face (Figure 24) and detect key points (eyes, nose, mouth) of the face using the color image. The Kinect face tracking engine

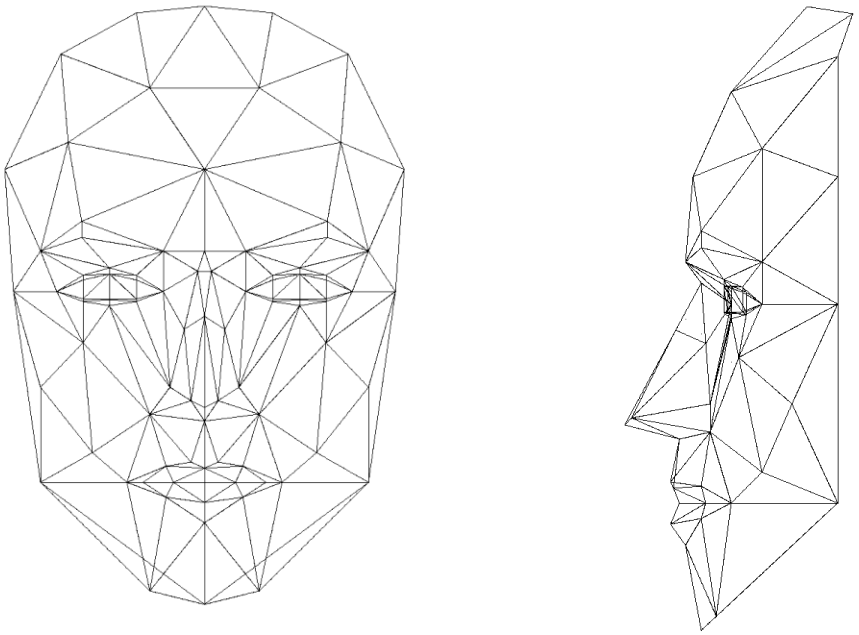

Figure 25[7]. Candide-3 model - This model specifies a set of 3D points (about 200) and triangles (the points are vertices of the triangles).

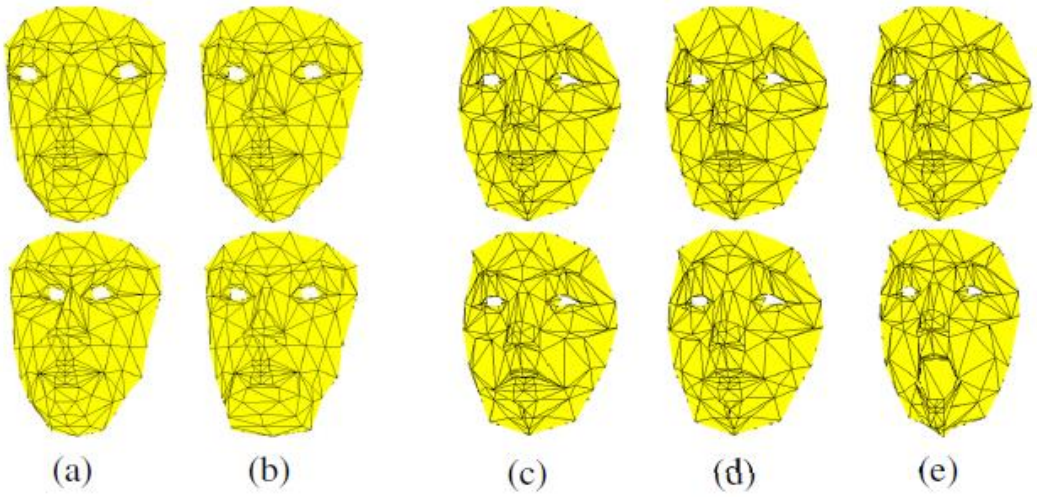


*Figure 26. SUs and AUs examples - This image is shown in [79]. The SUs adapt the face model to the shape of a particular face (a, b). The AUs adapt the model to the movements of the face (c, d, e).*

[7] www.icg.isy.liu.se/candide/

analyzes the input from the Kinect camera, uses the color and depth images to compute the head pose and facial expressions, and makes that information available to an application in real-time. This enables users to create applications that can track human faces at a 30fps rate. The method used by the Kinect to perform real-time face tracking is similar to the one presented in [79]. The 3D model of the face used is based on the well-known Candide-3 model [120] (Figure 25). This model specifies a set of 3D points (about 200) and triangles (the points are vertices of the triangles). The Candide-3 model is linearly deformable. The deformations are expressed in terms of parameters called Shape Units and Animation Units. The SUs define the unique characteristics of each user's facial shape in their neutral position. (Figure 26, a and b). These parameters are specific for each face. SUs specify the displacement of each vertex from the standard positions in order to make the model fit a particular face. The AUs are deltas from the neutral shape that can be used to adapt the mesh to the movements of the face in the presence of face expressions (Figure 26, c, d and e). Each AU is expressed as a numeric weight varying between -1 and +1. The X, Y, and Z position of each vertex is based on a right-handed coordinate system with the origin located at the camera's optical center (the sensor), Z pointed towards the user and Y pointed up (Figure 27). The user's head pose is captured by the yaw, pitch and roll angles. The units used are meters for translations and degrees for angles.

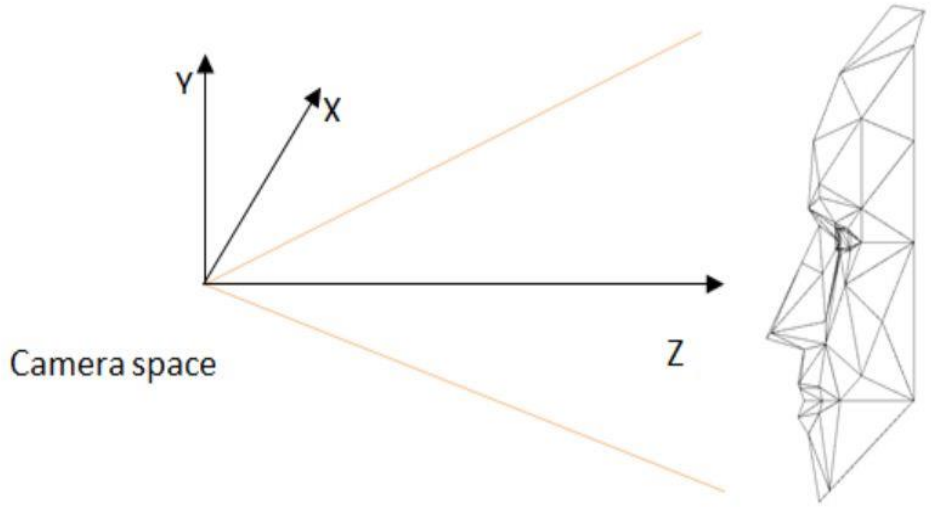


*Figure 27[8]. Kinect coordinate system – The X, Y, and Z position of each vertex is based on a right-handed coordinate system with the origin located at the camera's optical center (the sensor), Z pointed towards the user and Y pointed up.*

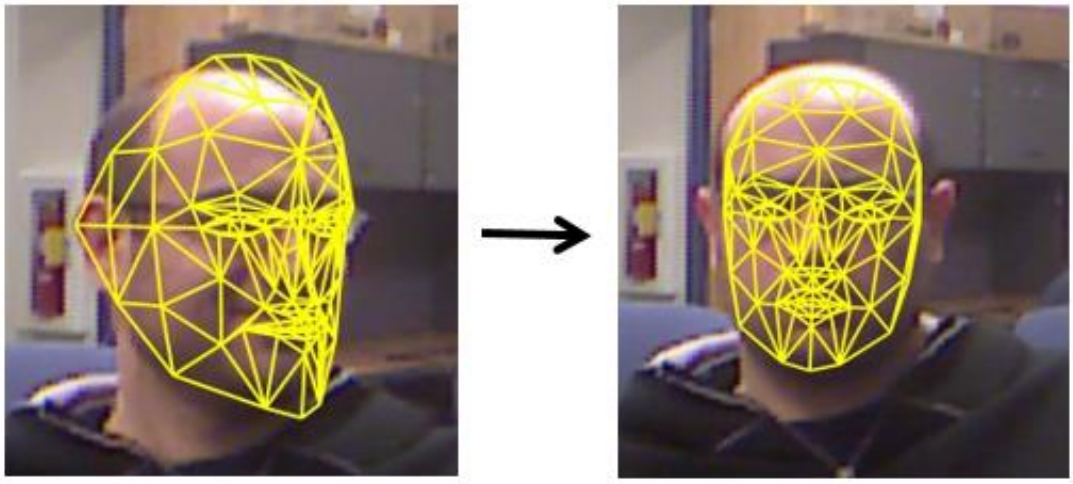

*Figure 28. Pose correction problem - Given the image on the left, the goal is to get an estimation of the image on the right using the face structure (in the form of the 3D mesh) and the color information.*

[8] msdn.microsoft.com/en-us/library/jj130970.aspx

## 5.2 Pose correction

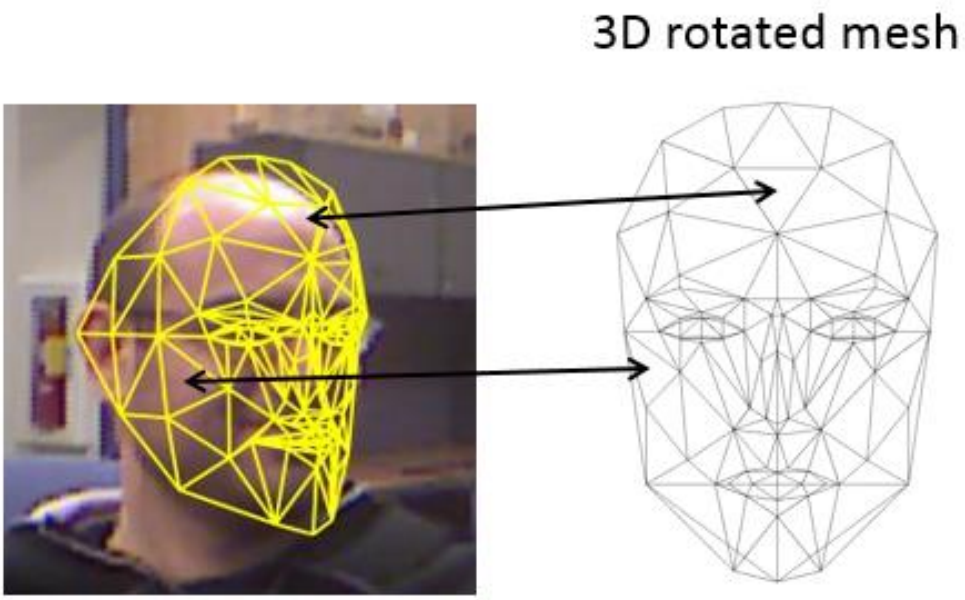


*Figure 29. Pose correction method - The side view mesh is rotated to front view. Between the side view mesh and the frontal mesh there is a triangle-to-triangle correspondence. To obtain the texture information in the blank mesh, the 2D projections of the 3D triangles are warped using the barycentric coordinates technique.*

The face recognition system described in the next section uses the Kinect sensor and its recently developed face tracking capabilities. Inspired by the approach described in the previous chapters, this system uses the 3D mask provided by the Kinect to improve the performance of the face recognition algorithm by doing pose-correction on the user face.

Figure 28 gives a representation of the pose correction problem. Suppose the target to identify is the face on the left. That face is not frontal, and therefore a traditional face recognition algorithm will fail, because it compares the side view face to front view gallery faces. Given the image on the left, the goal is getting an estimation of the image on the right using the

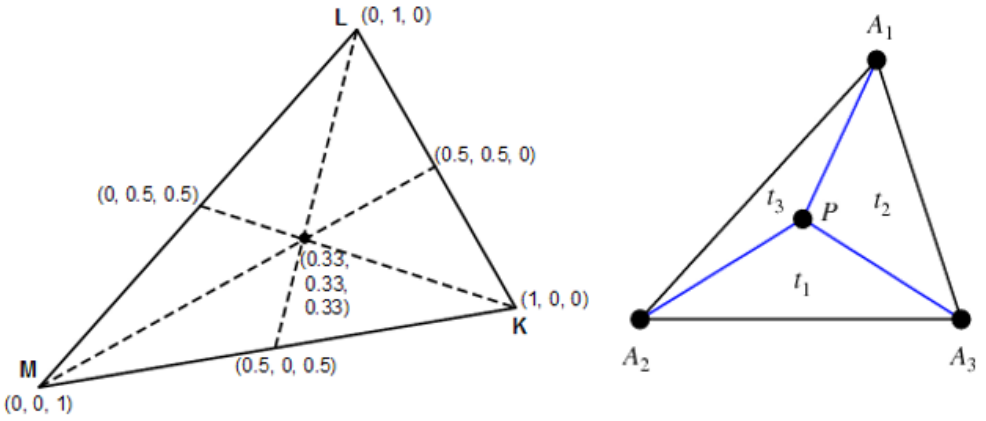


*Figure 30[9] Barycentric coordinates – The position of a point is defined w.r.t. the triangle vertices They can be calculated as ratios of areas (e.g. t1=Area(A2,A3,P)/Area(A1,A2,A3) where (t1, t2, t3) are the barycentric coordinates of P in the triangle on the right).*

face structure (in the form of the 3D triangular mesh) and the texture information. The pose-corrected version of the face ideally looks similar to the image on the right, and this newly generated image is used to identify the face. The proposed pose correction method is simple but effective. The first step consists in creating a blank 3D mesh with the same parameters (SUs and AUs) of the mesh in the side view and then setting yaw, pitch and roll of the blank mesh to 0 (Figure 29). This is equivalent to rotating the 3D mesh of the side view face to front view. Once the frontal mesh is created, there is a triangle-to-triangle correspondence between the two 3D meshes. However, the 2D projections of the 3D triangles are different because of the rotated positions in the 3D space. To estimate the texture information in the blank mesh, the 2D projection of each triangle within the side view mesh is

[9] mathworld.wolfram.com/ArealCoordinates.html
www.codeproject.com/Articles/625787/Pick-Selection-with-OpenGL-and-OpenCL

warped to the shape of the corresponding 2D triangle within the blank mesh using the barycentric coordinates technique. The barycentric coordinates define the position of a point w.r.t. the triangle vertices. They can be calculated as ratios of areas (see Figure 30). A backward warping is implemented with the following:

- compute the barycentric coordinates of a pixel in the blank mesh;
- retrieve the RGB value of the corresponding pixel in the side view image;
- assign the RGB value retrieved to the pixel in the blank mesh.

Figure 31 shows the result of this pose correction process. Looking at the newly generated face image, only the right side of the face has been reconstructed. The left side of the face is generated by mirroring the right side (not shown in Figure 31).

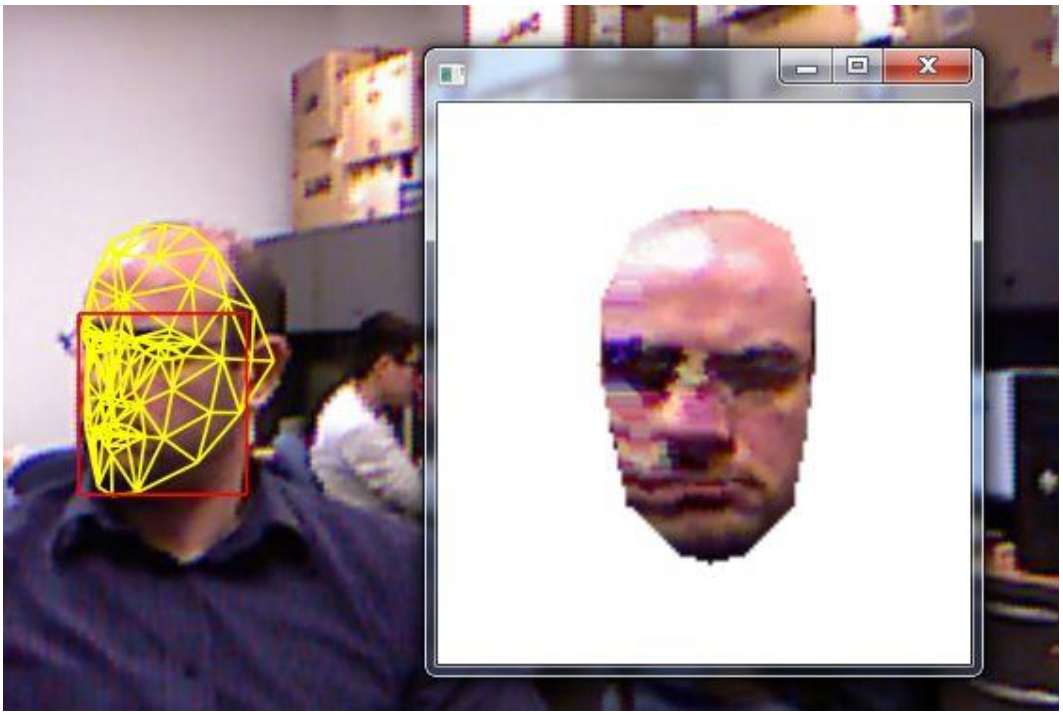

*Figure 31. Face warping - The left side of the generated image will be replaced with a mirrored version of the right side before performing face recognition.*

## 5.3 Face recognition demo


*Figure 32. Face recognition software interface – The system is working without pose correction (the check box is not checked, see red ellipse on the right).*

This section presents the face recognition software that was developed as part of this project. Figure 32 shows the graphical user interface. The system first performs face detection using the Kinect capabilities. The detected face is the portion shown in the white square. The same face is shown in the top left image. The image in the middle left is the pose-corrected face while the image at the bottom left is the best match found in the gallery. Subjects need to be enrolled in the gallery to be identified. On the right there is the button to enroll subjects. There is also the radio button to select the type of features, and a check box to activate the pose correction. In Figure 32, the system is working without pose correction (the check box is not checked, see red ellipse on the right). The system shows

the identity under the white box in the original image. The percentage beside the identity is the confidence level of the decision. In this case the face recognition succeeds because the face is frontal. The system also shows the estimated distance of the face from the camera and the estimated yaw, pitch and roll angles. Figure 34 shows a situation where the system is still working without pose correction and recognition fails because of the large pose variations. Figure 35 shows the same situation after the pose correction box is checked. The system now uses the pose-corrected face for recognition and is able to pull the correct identity even though it shows a low confidence level.

### 5.2.3 Confidence level

As mentioned above, this face recognition software shows a confidence level next to the presumed identity (Figure 32, 34, 35). This paragraph explains how this confidence level is computed. In the software, to compare histograms of LBP features, the chi-squared distance is used. This distance is defined as:

$$\chi^2(H_1, H_2) = \frac{1}{2}\sum_i \frac{[H_1(i) - H_2(i)]^2}{H_1(i) + H_2(i)}$$

The chi-squared distance between two histograms with n bins is approximately distributed as a chi-square distribution with n-1 degrees of freedom. Given the distance between two histograms, the significance level

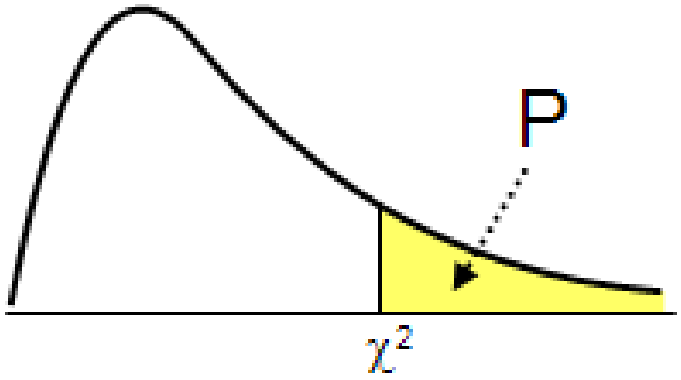


*Figure 33[10]. Confidence level - The $\chi^2$ distance between two histograms with n bins is approximately distributed as a $\chi^2$ distribution with n-1 degrees of freedom. The significance level associated with the distance (yellow area) is considered as the confidence level of the decision.*

associated with the distance is considered as the confidence level of the decision. Figure 33 shows a typical chi-squared distribution function. The yellow area is the significance level associated with that distance and represents the probability that the two histograms belong to the same face. In the case under analysis there are 59 bins, so a chi-squared distribution with 58 degrees of freedom is considered.

[10] www.medcalc.org/manual/chi-square-table.php


*Figure 34. Face recognition failure - The recognition fails because of the pose variation.*

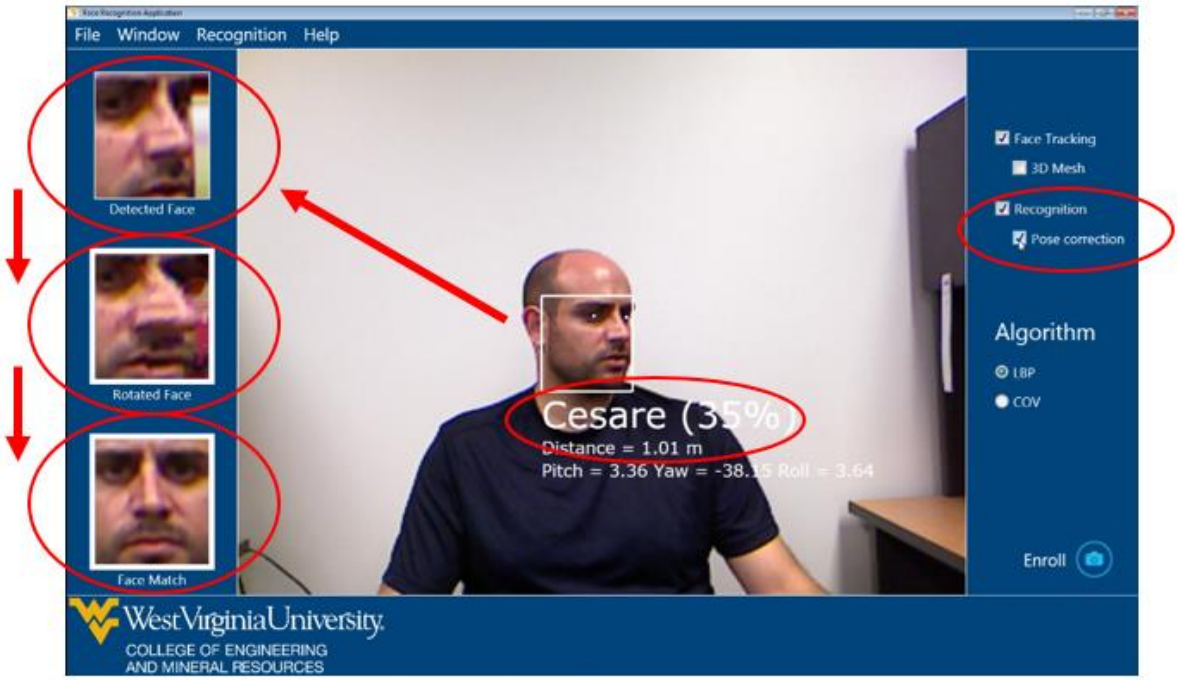


*Figure 35. Face recognition success - Face recognition succeeds thanks to pose correction (see red ellipse on the right).*

# Chapter 6. Future works

In the future, the presented method will be evaluated on other RGB-D face databases as they become available. Also, some improvements and extensions have been considered. For example, the method could be extended to deal with arbitrary 3D head pose variations. Sometimes the probe face may also present pitch and roll rotations (Figure 36). A method able to deal only with yaw rotation will likely fail. To improve performances, the method needs to be able to perform pose correction in the presence of an arbitrary value of the yaw, pitch and roll angles. In addition, the method could be extended to simultaneously perform face recognition and head pose estimation. The face recognition algorithm could be used to drive the

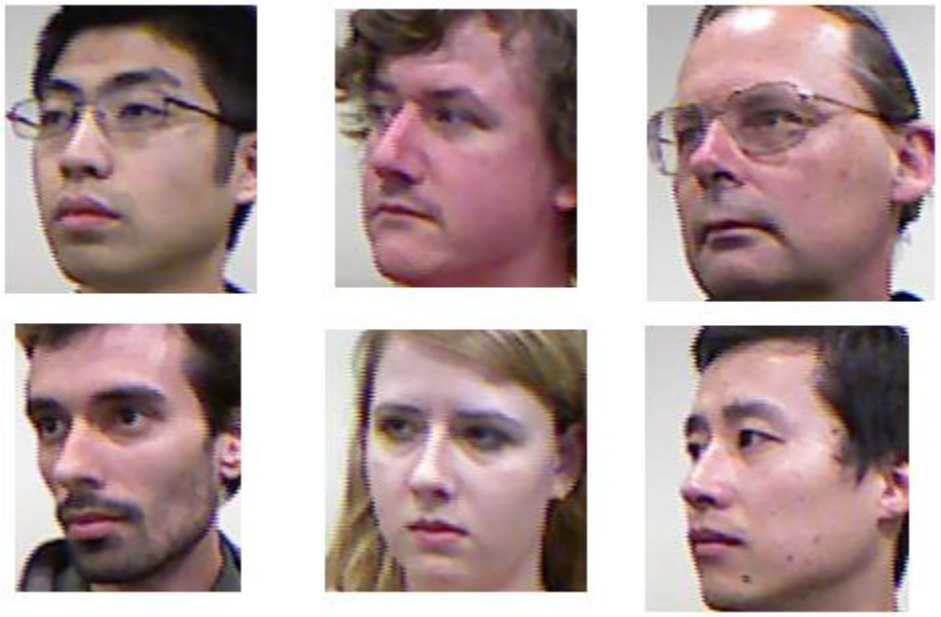

*Figure 36. Arbitrary pose - Sometimes probe faces present also pitch and roll rotations. A method able to deal only with yaw rotation will likely fail. To improve the performance the method needs to perform pose correction in presence of an arbitrary values of the yaw, pitch and roll angles.*

pose estimation algorithm, and the face recognition results could be improved with a more accurate pose estimation. A minimization technique like gradient descent could be adopted, where the cost is represented by the distance between faces. Ultimately, the RGB-D sensor could be very helpful for face expression correction and light normalization. Figure 37 shows a preliminary result of expression correction with Kinect.

*Figure 37*. Expression correction with Kinect

# Chapter 7. Conclusion

In this study the problem of face recognition in the presence of 2D facial appearance variations caused by 3D head rotations was addressed. The advantages of the recently developed consumer-level RGB-D cameras (e.g. Kinect) were explored. The proposed face recognition approach is able to deal with large head pose variations using RGB-D face images. The method uses the depth information to correct the pose of the face with a simple smoothing and 3D rotation of the point cloud; it is simple and fast, yet able to deal with large pose variations and perform pose-invariant face recognition. Experiments on a public database show that the approach is effective and efficient under significant pose changes. The method doesn't need to learn any face models. Also, a new face representation that combines the covariance descriptor with the LBP features was presented, and the experiments showed that the new representation is effective in the context of face recognition. Finally, a face recognition software able to achieve real-time and accurate face recognition in the presence of large yaw rotations using the Kinect sensor was presented. The results of this study demonstrate that RGB-D sensors are a promising tool that can lead to the development of robust pose-invariant face recognition systems under large pose variations.

# Bibliography


[1] Ciaccio C., Lingyun Wen, Guodong Guo, "*Face recognition robust to head pose changes based on the RGB-D sensor",* in Biometrics: Theory, Applications and Systems (BTAS), pages 1-6, 2013.

[2] B. Y. Li, A. S. Mian, W. Liu, and A. Krishna. "*Using kinect for face recognition under varying poses, expressions, illumination and disguise",* in IEEE Workshop on Applications of Computer Vision, pages 186–192, 2013.

[3] http://rgb-d.eurecom.fr/

[4] A. Jain, A. Ross, S. Prabhakar. "*An introduction to biometric recognition",* in IEEE Transaction on Circuit and System for Video Technology, 2003.

[5] Gagandeep Kaur, Gurpreet Singh and Vineet Kumar. "*A Review on Biometric Recognition".* International Journal of Bio-Science and Bio-Technology. Vol.6, No.4, pp.69-76, 2014.

[6] S. Liu and M. Silverman. "*A Practical Guide to Biometric Security Technology".* IEEE IT Pro, vol. 3, no. 1, pp. 27-32, 2001.

[7] A. Jain, A. Ross, S. Pankanti. "*Biometrics: A tool for information security".* IEEE Trans. Inform. Forensics Security 1 (2), 125-143, 2006.

[8] H. Gamboa, A. Fred. "*A behavioral biometric system based on human-computer interaction"*. In SPIE 5404 - Biometric Technology for Human Identification, pp. 381-392, August 2004.

[9] R. Jiang, A. Sadka and D. Crookes. "*Multimodal biometric human recognition for perceptual human-computer interaction".* IEEE Trans. Syst. Man Cybern. C, Appl. Rev., vol. 40, no. 6, pp.676 -681, 2010.

[10] H. Gamboa, A. L. N. Fred, and A. K. Jain. "*Web biometrics: user verification via web interaction"*, in Proc. Biometrics Symp., Baltimore, MD, pp. 1-6, 2007.

[11] W. Zhao, R. Chellappa, P.J. Phillips and A. Rosenfeld. "*Face Recognition: A Literature Survey".* ACM Computing Surveys, vol. 35, no. 4, pp. 399-458, Dec. 2003.

[12] V. Vijayakumari. "*Face Recognition Techniques: A Survey".* World Journal of Computer Application and Technology 1(2), 41-50, 2013.

[13] Gupta, Gaurav. "*A Literature Review of Face Recognition".* Journal of Bioinformatics and Intelligent Control, Volume 3, Number 1, pp. 1-7(7), March 2014.

[14] http://www.biometrics.gov/Documents/FaceRec.pdf

[15] A. J. Goldstein, L. D. Harmon, and A. B. Lesk. "*Identification of Human Faces".* Proc. IEEE, Vol. 59, No. 5, 748-760, May 1971.

[16] D. J. Beymer. "*Face recognition under varying pose".* In Proc. of IEEE Conf. on CVPR, pp. 556-761, Seattle, Washington, June, 1994.

[17] F. J. Huang, H.-J. Zhang, T. Chen, and Z. Zhou. "*Pose invariant face recognition",* In IEEE Proc. Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pages 245-250, 2000.

[18] Zhang Xiaozheng, Gao Yongsheng. "*Face recognition across pose: A review",* Pattern Recognition, Volume.42, Issue.11, pp.2876, 2009.

[19] X. Zou, J. Kittler and K. Messer. "*Illumination invariant face recognition: A survey",* IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems, pp.1 -8 2007.

[20] SaxenaDevi Hidangmayum, Meitei Thounaojam Dalton, Laishram Romesh. "*An approach to Illumination and Expression Invariant Multiple Classifier Face Recognition",* International Journal of Computer Applications, vol. 91, issue 15, pp. 34-37, 2014.

[21] J.S Nayak, M. Indiramma. "*Efficient face recognition with compensation for aging variations".* Fourth International Conference on Advanced Computing (ICoAC), pages 1–5, 2012.

[22] T. Kanade and A. Yamada. "*Multi-Subregion-Based Probabilistic Approach toward Pose-Invariant Face Recognition".* Proc. IEEE Int'l Symp. Computational Intelligence in Robotics and Automation, pp. 954-959, 2003.

[23] Byeong Doo Ahn, Hanseok Ko. "*Pose-invariant face recognition using cylindrical model and stereo camera"*. Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), pages 34–38, 2004.

[24] X. Liu and T. Chen. "*Pose-robust face recognition using geometry assisted probabilistic modeling".* Proc. Int. Conf. Comput. Vis. and Pattern Recog., vol. 1, pp.502 -509, 2005.

[25] X. Liu and T. Chen. "*Pose-robust face recognition using geometry assisted probabilistic modeling".* In CVPR, June 2005.

[26] Chai, X., Shan, S., Chen, X., and Wen, G. "*Locally Linear regression for pose-invariant face recognition"*, IEEE Transactions on Image Processing, 16 (7), p. 1716-1725, 2007.

[27] R. Singh, M. Vatsa, A. Ross, and A. Noore, "*A Mosaicing Scheme for Pose-Invariant Face Recognition",* IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 37, pp. 1212-1225, 2007.

[28] Khaleghian, S.; Rabiee, H.R.; Rohban, M.H. "*Face recognition across large pose variations via Boosted Tied Factor Analysis",* IEEE Workshop on Applications of Computer Vision (WACV), page(s): 190-195, 2011.

[29] Arashloo, S.R.; Kittler, J. "*Energy Normalization for Pose-Invariant Face Recognition Based on MRF Model Image Matching",* Pattern Analysis and Machine Intelligence, IEEE Transactions on, On page(s): 1274 - 1280 Volume: 33, Issue: 6, June 2011

[30] Arashloo, Shervin Rahimzadeh; Kittler, Josef; Christmas, William J. *"Pose-invariant face recognition by matching on multi-resolution MRFs linked by supercoupling transform"*, Computer Vision and Image Understanding, Volume.115, Issue.7, pp.1073, 2011.

[31] Shahdi, S.O.; Abu-Bakar, S.A.R. *"Varying pose face recognition using combination of discrete cosine & wavelet transforms"*, 4th International Conference on Intelligent and Advanced Systems (ICIAS), page(s): 642 - 647 Volume: 2, 12-14 June 2012.

[32] Arashloo, S.R.; Kittler, J. *"Efficient processing of MRFs for unconstrained-pose face recognition"*, Biometrics: Theory, Applications and Systems (BTAS), page(s): 1-8, 2013.

[33] Shyam, R, Singh, Y.N. *"A taxonomy of 2D and 3D face recognition methods"*, International Conference on Signal Processing and Integrated Networks (SPIN), pages: 749-754, Feb. 2014.

[34] Andrea F. Abate, Michele Nappi, Daniel Riccio, Gabriele Sabatino, *"2D and 3D face recognition: A survey"*, Pattern Recognition Letters, Volume 28, Issue 14, Pages 1885–1906, 15 October 2007.

[35] C.W. Chen, J.S. Huang, *"Human face profile recognition from a single front view"*, Int. J. Pattern Recognit. Artif. Intell. , Volume 06, Issue 04, 571 - 593, October 1992.

[36] D. Beymer, T. Poggio, *"Face Recognition From One Example View"*, Proc. 5th Int. Conf. Computer Vision, Cambridge, pp. 500-507, 1995.

[37] Y. Gao, M. K. H. Leung, W. Wang, and S. C. Hui, *"Fast face identification under varying pose from a single 2-D model view"*, Vision, Image and Signal Processing, IEE Proceedings-, vol. 148, pp. 248-253, 2001.

[38] S. Ting, B. C. Lovell, and C. Shaokang, *"Face Recognition Robust to Head Pose from One Sample Image"*, 18th International Conference on Pattern Recognition, pp. 515-518, 2006.

[39] S. Lucey and T. Chen. *"Learning patch dependencies for improved pose mismatched face verification"*, in CVPR, June 2006.

[40] C. D. Castillo and D. W. Jacobs, *"Using Stereo Matching for 2-D Face Recognition Across Pose"*, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[41] Gonzalez-Jimenez, D.; Alba-Castro, J.L. *"Toward Pose-Invariant 2-D Face Recognition Through Point Distribution Models and Facial Symmetry"*, IEEE Transactions on Information Forensics and Security, page(s): 413–429, Volume: 2, Issue: 3, Sept. 2007.

[42] A. B. Ashraf, S. Lucey, and T. Chen. *"Learning patch correspondences for improved viewpoint invariant face recognition"*, in IEEE Conf. on Computer Vision and Pattern Recognition, pages 1–8, 2008.

[43] L. Annan, S. Shiguang, C. Xilin, and G. Wen, *"Maximizing intraindividual correlations for face recognition across pose differences"*, in Computer Vision and Pattern Recognition (CVPR) , pp. 605-611, 2009.

[44] D. Shan and R. Ward, *"Component-wise pose normalization for pose invariant face recognition"*, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 873-876, 2009.

[45] Castillo, C.D.; Jacobs, D.W. *"Using Stereo Matching with General Epipolar Geometry for 2D Face Recognition across Pose"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, page(s): 2298–2304, Volume: 31, Issue: 12, Dec. 2009.

[46] Xiao Zeng; Hua Huang, *"Super-Resolution Method for Multiview Face Recognition From a Single Image Per Person Using Nonlinear Mappings on Coherent Features"*, Signal Processing Letters, IEEE, page(s): 195 - 198 Volume: 19, Issue: 4, April 2012.

[47] A. Sharma, M. A. Haj, J. Choi, L. S. Davis, and D. W. Jacobs. *"Robust pose invariant face recognition using coupled latent space discriminant analysis"*. Comput. Vis. Image Underst., 116(11):1095-1110, Nov. 2012.

[48] Chenghua Xu; Yunhong Wang; Tieniu Tan; Long Quan *"Depth vs. intensity: which is more important for face recognition?"*. Proceedings of the 17th International Conference on Pattern Recognition, page(s): 342 – 345, Volume: 1, Aug. 2004.

[49] Y.X.Hu, D.L.Jiang, S.C Yan, L. Zhang, H.J. Zhang. *"Automatic 3D Reconstruction for Face Recognition"*. Proceedings of IEEE Automatic Face and Gesture Recognition, pp. 843-848, 2004.

[50] Irfanoglu, M.O.; Gokberk, B.; Akarun, L. *"3D shape-based face recognition using automatically registered facial surfaces"*. Proceedings of the 17th International Conference on Pattern Recognition, page(s): 183 – 186, Volume: 4, Aug. 2004.

[51] X.J. Chai, S.G. Shan, L.Y.Qing, X.L. Chen, W.Gao. *"Pose and Illumination Invariant Face Recognition Based on 3D Face Reconstruction".* Journal of Software, vol.17(3), pp.525-534, 2006.

[52] Widanagamaachchi, W.N.; Dharmaratne, A.T. *"3D Face Reconstruction from 2D Images"*, Digital Image Computing: Techniques and Applications (DICTA), page(s): 365 - 371, 2008.

[53] Martin D. Levine, Yingfeng (Chris) Yu, *"State-of-the-art of 3D facial reconstruction methods for face recognition based on a single 2D training image per person"*, Pattern Recognition Letters, v.30 n.10, p.908-913, July, 2009.

[54] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. V. Rohith. *"Fully automatic pose invariant face recognition via 3D pose normalization".* Proc. ICCV, pages 937-944, 2011.

[55] G.G. Gordon, *"Face Recognition based on Depth Maps and Surface Curvature"*, Proc. SPIE Geometric Methods in Computer Vision, Vol.1570, 1991.

[56] C.S. Chua, F. Han, and Y.K. Ho, *"3D Human Face Recognition Using Point Signiture"*, Proc. FG, pp.233-239, 2000.

[57] C. Hesher, A. Srivastava, and G.Erlebacher, *"A Novel Technique for Face Recognition Using Range Imaging"*, Inter. Multiconference in Computer Science, 2002.

[58] Y. Lee, K. Park, J. Shim, and T. Yi, *"3D Face Recognition Using Statistical Multiple Features for the Local Depth Information"*, Proc. ICVI, 2003.

[59] Xiaoguang Lu; Colbry, D.; Jain, A.K. *"Three-dimensional model based face recognition"*. Proceedings of the 17th International Conference on Pattern Recognition, page(s): 362 – 366, Volume: 1, Aug. 2004.

[60] Cook, J.; Chandran, V.; Sridharan, S.; Fookes, C. *"Face recognition from 3D data using Iterative Closest Point algorithm and Gaussian mixture models"*, 2nd International Symposium on 3D Data Processing, Visualization and Transmission, page(s): 502 – 509, 2004.

[61] Le Zou; Cheng, S.; Zixiang Xiong; Mi Lu; Castleman, K.R. *"3-D Face Recognition Based on Warped Example Faces"*, IEEE Transactions on Information Forensics and Security, page(s): 513 – 528, Volume: 2, Issue: 3, Sept. 2007.

[62] Mohammadzade, H.; Hatzinakos, D. *"Iterative Closest Normal Point for 3D Face Recognition"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, page(s): 381 – 397, Volume: 35, Issue: 2, Feb. 2013.

[63] Blanz, V., Romdhani, S., Vetter, T. *“Face identification across different poses and illuminations with a 3D morphable model”*, Proc. IEEE 5th Int. Conf. Automatic Face and Gesture Recognition, p. 100–105, 2002.

[64] S. Romdhani, V. Blanz, and T. Vetter, *"Face Identification by Fitting a 3D Morphable Model using Linear Shape and Texture Error Functions"*, Proc. ECCV, pp.3-19, 2002.

[65] V. Blanz and T. Vetter, *"Face recognition based on fitting a 3D morphable model"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 1063-1074, 2003.

[66] B. Weyrauch , B. Heisele , J. Huang , V. Blanz, *“Component-Based Face Recognition with 3D Morphable Models”*, Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 5, p.85, June 27-July 02, 2004.

[67] Q. Chen, J. Yao and W. Cham, *"3D model-based pose invariant face recognition from multiple views"*, IET Comput. Vis., vol. 1, no. 1, pp.25 -34, 2007.

[68] L. Wang, L. Ding, X. Ding and C. Fang, *"Improved 3-D assisted pose-invariant face recognition"*, Proc. ICASSP, pp.889 -892, 2009.

[69] Volker Blanz and Thomas Vetter, *"A morphable model for the synthesis of 3D faces"*. In: Proceedings of ACM SIGGRAPH 99 Conference, pp.187-194, 1999.

[70] S. Izadi, R. Newcombe, D. Kim, O. Hilliges, D. Molyneaux, S. Hodges, P. Kohli, A. Davison, and A. Fitzgibbon, *"Kinectfusion: Real-time dynamic 3d surface reconstruction and interaction"*. ACM SIGGRAPH, 2011.

[71] L. Spinello and K. Arras, *"People detection in rgb-d data"*. International Conference on Intelligent Robots and Systems, 2011.

[72] Cruz, L.; Lucio, D.; Velho, L. *"Kinect and RGBD Images: Challenges and Applications"*, Graphics, Patterns and Images Tutorials (SIBGRAPI-T), page(s): 36 – 49, 2012.

[73] Jungong Han; Ling Shao; Dong Xu; Shotton, J. *"Enhanced Computer Vision With Microsoft Kinect Sensor: A Review"*, IEEE Transactions on Cybernetics, page(s): 1318 – 1334, Volume: 43, Issue: 5, Oct. 2013.

[74] Shuran Song, Jianxiong Xiao, *"Tracking Revisited Using RGBD Camera: Unified Benchmark and Baselines"*, IEEE International Conference on Computer Vision (ICCV), page(s): 233 – 240, 2013.

[75] Du Yong Kim; Ba-Tuong Vo; Ba-Ngu Vo, *"Data fusion in 3D vision using a RGB-D data via switching observation model and its application to people tracking"*, International Conference on Control, Automation and Information Sciences (ICCAIS), page(s): 91 – 96, 2013.

[76] Anasosalu, P.K.; Thomas, D.; Sugimoto, A. *"Compact and Accurate 3-D Face Modeling Using an RGB-D Camera: Let's Open the Door to 3-D Video Conference"*, IEEE International Conference on Computer Vision Workshops (ICCVW), page(s): 67 – 74, 2013.

[77] R. Hg, P. Jasek, C. Rofidal, K. Nasrollahi, T. Moeslund, and G. Tranchet. *"An rgb-d database using microsoft's kinect for windows for face detection"*. In Eighth Int'l Conf. on Signal Image Technology and Internet Based Systems, pages 42–46, 2012.

[78] T. Huynh, R. Min, and J.-L. Dugelay. *"An efficient lbp-based descriptor for facial depth images applied to gender recognition using rgb-d face data"*. In ACCV Workshop on Computer Vision with Local Binary Pattern Variants, 2012.

[79] Qin Cai, David Gallup, Cha Zhang, Zhengyou Zhang. *"3D Deformable Face Tracking with a Commodity Depth Camera"*. Proc. of European Conference on Computer Vision (ECCV), 2010.

[80] R. Mattheij, E. Postma, Y. van den Hurk, and P. Spronck. *"Depth-based detection using haar-like features"*. In Proc. of the BNAIC 2012 conference, Maastricht University, The Netherlands, pages 162–169, 2012.

[81] Fei Yang, Junzhou Huang, Xiang Yu, Xinyi Cui, Metaxas. D. *"Robust face tracking with a consumer depth camera"*. International Conference on Image Processing, pages 561-564, 2012.

[82] Gabriele Fanelli, Matthias Dantone, Juergen Gall, Andrea Fossati, Luc Gool, *"Random Forests for Real Time 3D Face Analysis"*. International Journal of Computer Vision, v.101 n.3, p.437-458, 2013.

[83] Songnan Li, King Ngi Ngan, Lu Sheng, *"A Head Pose Tracking System Using RGB-D Camera"*. Proceedings of the 9th international conference on Computer Vision Systems, pages 153-162, 2013.

[84] Rekik, A., Ben-Hamadou, A., and Mahdi, W. *"3d face pose tracking using low quality depth cameras"*. In VISAPP, pages 223–228, 2013.

[85] Loris Nannia, Alessandra Luminib, Fabio Dominioa, Pietro Zanuttigha, *"Effective and precise face detection based on color and depth data"*. Applied Computing and Informatics, Volume 10, Issues 1–2, Pages 1–13, 2014.

[86] Hai Xuan Pham, Vladimir Pavlovic, *"Hybrid On-Line 3D Face and Facial Actions Tracking in RGBD Video Sequences"*. International Conference on Pattern Recognition, 2014.

[87] http://files.is.tue.mpg.de/jgall/tutorials/visionRGBD13.html

[88] R. Min, J. Choi, G. Medioni, and J.-L. Dugelay, *“Real-time 3d face identification from a depth camera”*. In Int’l Conf. on Pattern Recognition, pages 1739–1742, 2012.

[89] Gaurav Goswami, Samarth Bharadwaj, Mayank Vatsa, and Richa Singh, *“On RGB-D Face Recognition using Kinect”*. In Biometrics: Theory, Applications and Systems (BTAS), pages 1-6, 2013.

[90] Mauricio Pamplona Segundo, Sudeep Sarkar, Dmitry Goldgof, Luciano Silva, Olga Bellon, *“Continuous 3D Face Authentication using RGB-D Cameras”*. Computer Vision and Pattern Recognition Workshops (CVPRW), pages 64-69, 2013.

[91] T. Ojala, M. Pietikäinen, and D. Harwood, *"Performance evaluation of texture measures with classification based on Kullback discrimination of distributions"*, Proceedings of the 12th International Conference on Pattern Recognition (ICPR), vol. 1, pp. 582 – 585, 1994.

[92] T. Ojala, M. Pietikäinen, and D. Harwood, *"A Comparative Study of Texture Measures with Classification Based on Feature Distributions"*, Pattern Recognition, vol. 29, pp. 51-59, 1996.

[93] Sheryl Brahnam, Lakhmi C. Jain, Loris Nanni, Alessandra Lumini, *“Local Binary Patterns: New Variants and Applications”*. Studies in Computational Intelligence, Volume 506, pages 191-217, 2014.

[94] T. Ahonen, A. Hadid, and M. Pietikainen, *“Face recognition with local binary patterns”*, in Eur. Conf. on Comput. Vision, pages 469–481, 2004.

[95] T. Ahonen, A. Hadid and M. Pietikainen, *"Face description with local binary patterns: Application to face recognition"*, IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 12, pp.2037 -2041, 2006.

[96] G. Heusch, Y. Rodriguez and S. Marcel, *"Local binary patterns as an image preprocessing for face authentication"*, Proc. Int. Conf. Autom. Face Gesture Recog, pp.9 -14, 2006.

[97] Di Huang, Caifeng Shan, Ardabilian, M., Yunhong Wang, Liming Chen, *“Local Binary Patterns and Its Application to Facial Image Analysis: A Survey”*. IEEE Transactions on Systems, Man, and Cybernetics, Biometrics Compendium, pages 765–781, 2011.

[98] W. Zhang, S. Shan, W. Gao, X. Chen and H. Zhang, *"Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition"*, Proc. IEEE Int',l Conf. Computer Vision, pp. I: 786-791, 2005.

[99] J. Zhao, H. Wang, H. Ren and S.-C. Kee, *"LBP discriminant analysis for face verification"*, presented at the IEEE Workshop Face Recog. Grand Challenge Exp., 2005.

[100] Y. Rodriguez and S. Marcel, *"Face Authentication Using Adapted Local Binary Pattern Histograms"*, Proc. Ninth European Conf. Computer Vision, pp. IV: 321-332, 2006.

[101] C. Chan, J. Kittler and K. Messer, *"Multi-scale local binary pattern histograms for face recognition"*, Proc. Int. Conf. Biometrics, pp.809 -818 2007.

[102] S. Liao and A. C. S. Chung, *"Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude"*, Proc. Asian Conf. Comput. Vis., pp.672 -679, 2007.

[103] S. Liao and S. Z. Li, *"Learning multi-scale block local binary patterns for face recognition"*, Proc. Int. Conf. Biometrics, pp.828 -837, 2007.

[104] Z. Lei, S. Liao, R. He, M. Pietikainen and S. Z. Li, *"Gabor volume based local binary pattern for face representation and recognition"*, Proc. IEEE Int. Conf. Autom. Face Gesture Recog., pp.1 -6, 2008.

[105] Ojala T., Pietikainen M., Maenpaa T. *"Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns"*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, July 2002.

[106] O. Tuzel, F. Porikli, and P. Meer. *"Region Covariance: A Fast Descriptor for Detection and Classification"*, in ECCV, vol. 3952, pp. 589-600, 2006.

[107] O. Tuzel, F. Porikli, and P. Meer. *"Pedestrian detection via classification on Riemannian manifolds"*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(10):1713–1727, 2008.

[108] M. T. Harandi, C. Sanderson, A. Wiliem, and B. C. Lovell, *"Kernel analysis over riemannian manifolds for visual recognition of actions, pedestrians and textures"*. IEEE Workshop on Applications of Computer Vision (WACV), pages 433–439, 2012.

[109] Y. Pang, Y. Yuan, and X. Li, *"Gabor-Based Region Covariance Matrices for Face Recognition"*, TCSVT, vol. 18, pp. 989-993, 2008.

[110] S. Guo and Q. Ruan, *"Facial expression recognition using local binary covariance matrices"*. In Wireless, Mobile & Multimedia Networks, page 237-242, 2011.

[111] Qin H.F., Qin L., Xue L., Yu C.B. *"Gabor-based weighted region covariance matrix for face recognition"*. Electronics Letters, volume 48, issue 16, pages 992 - 993, 2012.

[112] J. Krizaj, V. Struc, S. Dobrisek, *“Combining 3D Face Representations using Region Covariance Descriptors and Statistical Models”*, Face Gesture Recognition Workshops: 3D Face Biometrics, Shanghai, China, 2013.

[113] W. Forstner and B. Moonen, *“A metric for covariance matrices”*. Quo vadis geodesia, pages 113-128, 1999.

[114] http://mmprec.iais.fraunhofer.de/asbach

[115] V. Blanz, P. Grother, P. J. Phillips, and T. Vetter, *“Face recognition based on frontal views generated from non-frontal images”*. In CVPR, June 2005.

[116] H.-S. Lee and D. Kim, *“Generating frontal view face image for pose invariant face recognition”*. Pattern Recognition Letters, 27(7):747-754, 2006.

[117] Z. Xiaozheng, G. Yongsheng, and L. Maylor, *"Recognizing Rotated Faces From Frontal and Side Views: An Approach Toward Effective Use of Mugshot Databases"*, IEEE Transactions on Information Forensics and Security, vol. 3, pp. 684-697, 2008.

[118] Li Yuelong, Feng Jufu, *"Frontal face synthesizing according to multiple non-frontal inputs and its application in face recognition"*, Neurocomputing, 2012.

[119] X. Zhu and D. Ramanan, *“Face detection, pose estimation, and landmark localization in the wild”*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2879–2886, 2012

[120] J. Ahlberg, *“CANDIDE-3 -- an updated parameterized face”*, Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001.