We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

3D and Thermo-Face Fusion

Štěpán Mráček, Jan Váňa, Radim Dvořák,
Martin Drahanský and Svetlana Yanushkevich

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/3420

1. Introduction

Most biometric-based systems use a combination of various biometrics to improve reliability of decision. These systems are called multi-modal biometric systems. For example, they can include video, infrared, and audio data for identification of appearance (encompassing natural changes such as aging, and intentional ones, such as surgical changes), physiological characteristics (temperature, blood flow rate), and behavioral features (voice and gait) [1].

Biometric technologies, in a narrow sense, are tools and techniques for identification of humans, and in a wide sense, they can be used for detection of alert information, prior to, or together with, the identification. For example, biometric data such as temperature, blood pulse, pressure, and 3D topology of a face (natural or changed topology using various artificial implants, etc.) must be detected first at distance, while the captured face can be further used for identification. Detection of biometric features, which are ignored in identification, is useful in design of Physical Access Security Systems (PASS) [2][3]. In the PASS, the situational awareness data (including biometrics) is used at the first phase, and the available resources for identification of person (including biometrics) are utilized at the second phase.

Conceptually, a new generation of the biometric-based systems shall include a set of biometric-based assistants; each of them deals with uncertainty independently, and maximizes its contribution to a joint decision. In this design concept, the biometric system possesses such properties as modularity, reconfiguration, aggregation, distribution, parallelism, and mobility. Decision-making in such a system is based on the concept of fusion. In a complex system, the fusion is performed at several levels. In particular, the face biometrics is considered to be the three-fold source of information, as shown in Figure 1.

In this chapter, we consider two types of the biometric-based assistants, or modules, within a biometric system:

- A thermal, or infrared range assistant,

- A 3D visual range assistant.

We illustrate concept of fusion at the recognition component, which is a part of more complex decision-making level. Both methods are described in terms of data acquisition, image processing and recognition algorithms. The general facial recognition approach, based on the algorithmic fusion of the two methods, is presented, and its performance is evaluated on both 3D and thermal face databases.
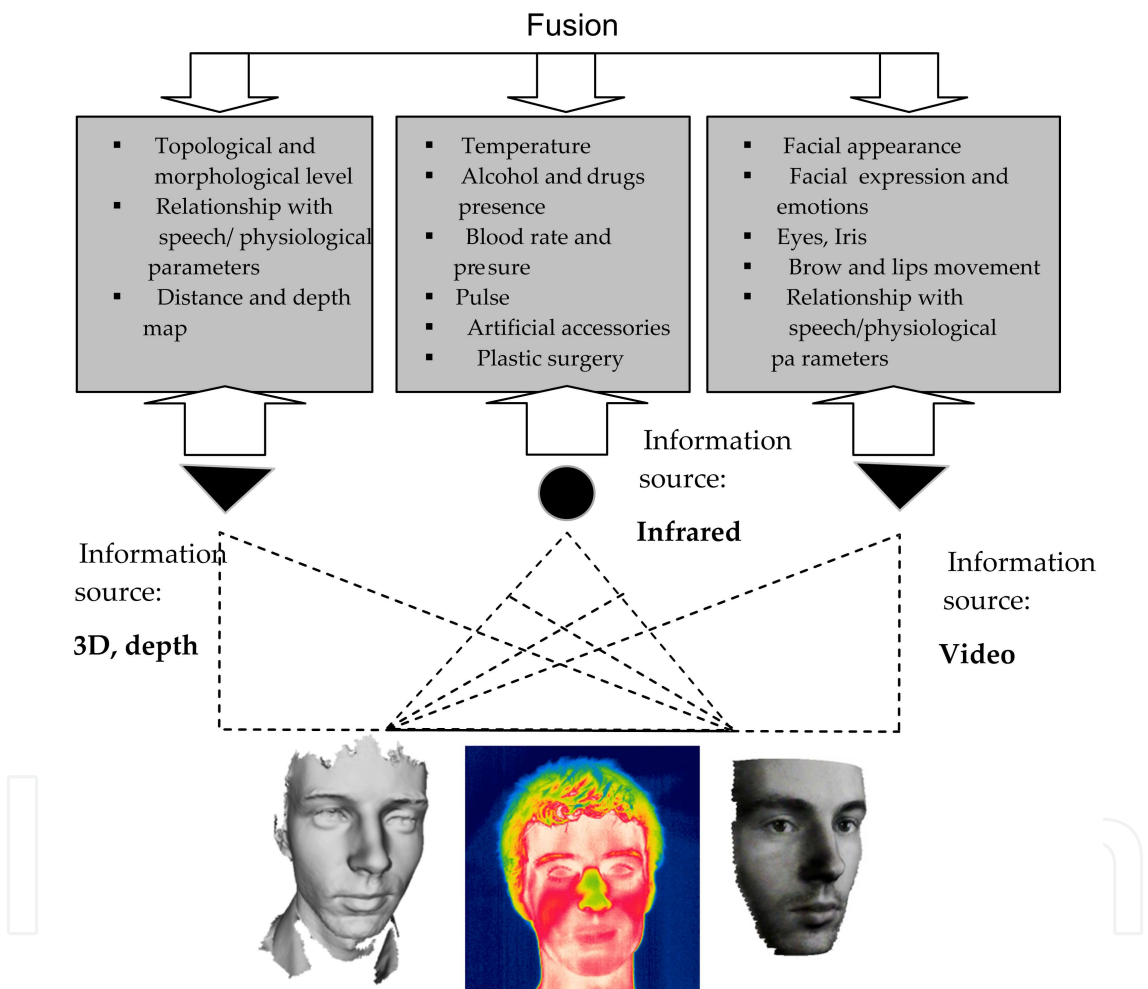
Figure 1. Thee sources of information in facial biometrics: a 3D face model (left), a thermal image (center) and a visual model with added texture (right).

Facial biometric, based on 3D data and infrared images, enchnace the classical face recognition. Adding depth information, as well as the information about the surface temperature, may reveal additional discriminative abilities, and thus improve recognition performance. Furthermore, it is much harder to forge a 3D, or thermal model, of the face.

The following sections provide an overview of how 3D and infrared facial biometrics work, and what is needed in terms of data acquisition and algorithms. The first section deals with the 3D face recognition. Thermal face recognition is described in the second section. Next, a general method for recognition, of both 3D and thermal images, is presented. The fusion on a decision (recognition score) level is investigated. Finally, the performance of the proposed fusion approach is evaluated on several existing databases.

2. Three dimensional face recognition

The three-dimensional (3D) face recognition is a natural extension of the classical two-dimensional approach. Contrary to 2D face recognition, additional possibilities for the recognition are available, due to the added dimension. Another advantage, for example, is a more robust system, in terms of pose variations. An overview of the advantages of the biometric system, based on a 3D face, is shown in Table 1.

Pose variation robustness	Due to the 3D form of the data, the face can be easily rotated into a predefined position.
Lighting condition robustness	Many 3D scanners work in infra-red spectra or emit their own light, for inappropriate lighting conditions do not affect recognition performance.
Out-of-the-box liveness detection	It is much more difficult to spoof fake data on a 3D sensor. While in 2D face recognition, simple systems may be fooled by a photograph or video, it is much more difficult to create an authentic 3D face model.

**Table 1.** Advantages of 3D face recognition.

The wide range of applications of 3D face recognition systems is limited by a high acquisition cost of special scanning devices. Moreover, a 3D face is targeted more at access control systems, rather than surveillance applications, due to the limited optimal distance range between the scanned subject and the sensor.

2.1. Acquisition of 3D data

Most facial 3D scanners use structured light in order to obtain the three-dimensional shape of a face. Structured light scanners project certain light pattern onto the object's surface, which is simultaneously captured by a camera from a different angle. The exact surface is then computed from the projected light pattern distortion, caused by the surface shape. The most commonly structured light pattern in 3D scanning devices consists of many narrow stripes lying side by side. Other methods, either using a different pattern, or one without the structured light, can be also used [4], however, they are not common in biometric systems.

The pattern can be projected using visible or infra-red light spectrum. An advantage to the infrared light is its non-disturbing effect on the user's eyes. On the other hand, it is more

difficult to segment the image and distinguish between the neighboring stripes properly. Therefore, many methods of acquiring the 3D surface use a visible light and color camera. The description of the method, where many color stripes are used, is given in [5]. The authors use the De Bruijn sequence there (see Figure 2), which consists of seven colors, in order to minimize the misclassification between the projected lines and the lines in the image captured by the camera.
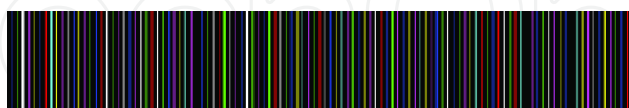
Figure 2. The De Bruijn color sequence [5].

The algorithm for surface reconstruction is composed of several steps. In the first step, two images are taken. In the first one ($IM_{pattern}$), the object is illuminated by structured light, whereas in the second one (IM_{clean}) unstructured light is used. Next, the projected light is extracted from the background by subtracting the two images:

$$IM_{extracted} = IM_{pattern} - IM_{clean} \tag{1}$$

The pattern in $IM_{extracted}$ is matched with the original pattern image. In the last step, the depth information of the points lying on the surface is calculated by the trigonometry principle. In order to calculate the exact depths properly, the precise position of the camera and projector, including their orientation, need to be known. It can be measured, or calculated by the calibration of both devices.

An example of a 3D scanner (commercial solution) is the Minolta Vivid Laser 3D scanner. The light reflected by the object is acquired by the CCD camera. Then, the final model is calculated, using the standard triangulation method. For instance, the scanner was used to collect models from the FRGC database [20].

Figure 3. The examples of an acquired 3D face models, using Artec 3D scanner.

Another example is the Artec 3D scanner [6] which has a flash bulb and camera. The bulb flashes a light pattern onto an object, and the CCD camera records the created image. The

distortion pattern is then transferred to the 3D image, using Artec software. The advantage of the scanner is its ability to merge several models (pattern images) belonging to the same object. When models are taken from different angles, the overall surface model is significantly accurate, and possible gaps in the surface are minimized. On the other hand, the surface of facial hair, or shiny materials, such as glasses, is hard to reconstruct because of a highly distorted light pattern (see Figure 3).

2.2. 3D face preprocessing

A key part of every biometric system is the preprocessing of input data. In the 3D face field, this task involves primarily the alignment of the face into a predefined position. In this section, several possible approaches of the face alignment will be described. In order to fulfill such a task, the important landmarks are located first. Detecting the facial landmarks from three-dimensional data cannot be performed using the same algorithms as in the case of two-dimensional data. It is mainly because two-dimensional landmark detection is based on analyzing color space of the input face picture, which is not usually present in raw three-dimensional data. However, if the texture data is available, the following landmark detection methods, based on the pure 3D model, may be skipped.

The location of the tip the nose is a fundamental part of preprocessing in many three-dimensional facial recognition methods [7][8][9][15]. Segundo et al. [10] proposed an algorithm for nose tip localization, consisting of two stages. First, the y-coordinate is found, then an appropriate x-coordinate is assigned. To find the y-coordinate, two vertical y-projections of the face are computed – the profile and median curves. The profile curve is determined by the maximum depth value in each row, while the median curve is defined by the median depth value of every set of points with the same y-coordinate. A curve that represents the difference between the profile and median curves is created. A maximum of this difference curve along the y-axis is the y-coordinate of the nose. The x-coordinate of the nose tip is located as follows: along the horizontal line, that intersects the y-coordinate of the nose, the density of peak points is calculated; the point with the highest peak density is the final location of the nose tip (see Figure 4).

In order to classify the points on the surface as peaks, curvature analysis is performed. The curvature at that specific point denotes how much the surface diverges from being flat. The sign of the curvature k indicates the direction in which the unit tangent vector rotates as a function of the parameter along the curve. If the unit tangent rotates counterclockwise, then $k > 0$. Otherwise, $k < 0$.

To depart from 2D curve to 3D surface, two principal (always mutually orthogonal) curvatures k_1 and k_2 are calculated at each point. Using these principal curvatures, two important measures are deduced: Gaussian curvature K and mean curvature H [11]:
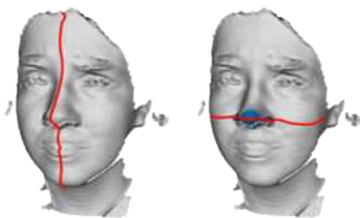
$$K = k_1 k_2 \tag{2}$$

$$H = (k_1 + k_2)/2 \qquad\qquad (3)$$

Classification of the surface points based on signs of Gaussian and mean curvatures is presented in Table 2.

	$K < 0$	$K = 0$	$K > 0$
$H < 0$	saddle ridge	ridge	peak
$H = 0$	minimal	flat	(none)
$H > 0$	saddle valley	valley	Pit

Table 2. Classification of points on 3D surface, based on signs of Gaussian (K) and mean curvatures (H).

Figure 4. The vertical profile curve that is used to determine the y-coordinate of the nose. Once the y-coordinate is located, another horizontal curve is created. Along the new curve, the density of peak points is calculated and exact position of the nose tip is located.

2.3. Overview of methods

2.3.1. Adaptation of 2D face recognition methods

The majority of widespread face recognition methods are holistic projection methods. These methods take the input image consisting of r rows and c columns, and transform it to a column vector. Pixel intensities of the input image directly represent values of individual components in the resulting vector. Rows of the image are concatenated into one single column.

Projection methods

A common attribute of projection methods is the creation of the data distribution model, and a projection matrix, that transforms input vector $v \in \mathbb{R}^{r \cdot c}$ into some lower dimensional space. In this section, the following methods will be described:

- Principal component analysis (PCA)

- Linear discriminant analysis (LDA)

- Independent component analysis (ICA)

Principal component analysis (PCA) was first introduced by Karl Pearson and covers mathematical methods, which reduce the number of dimensions of a given multi-dimensional space. The dimensionality reduction is based on data distribution. The first principal component is the best way to describe the data in a minimum-squared-error sense. Other components describe as much of the remaining variability as possible.

The *eigenface* method is an example of PCA application. It is a holistic face recognition method, which takes grayscale photographs of people, normalized with respect to size and resolution. The images are then interpreted as vectors. The method was introduced by M. Turk and A. Pentland in 1991 [12].

Linear discriminant analysis (LDA), introduced by Ronald Aylmer Fisher, is an example of supervised learning. Class membership (data subject identity) is taken into account during learning. LDA seeks for vectors that provide the best discrimination between classes after the projection.

The *Fisherface* method is a combination of principal component analysis and linear discriminant analysis. PCA is used to compute the face subspace in which the variance is maximized, while LDA takes advantage of intra-class information. The method was introduced by Belhumeur et al. [13].

Another data projection method is *independent component analysis* (ICA). Contrary to PCA, which seeks for dimensions where data varies the most, ICA looks for the transformation of input data that maximizes non-gaussianity. A frequently used algorithm that computes independent components is the FastICA algorithm [14].

Using projection methods for a 3D face

The adaptation of projection methods for a 3D face is usually based on the transformation of input 3D scans into range-images [15]. Each vertex of a 3D model is projected to a plane, where the brightness of pixels corresponds to specific values of z-coordinates in the input scan. An example of an input range image, and its decomposition in PCA subspace, consisting of 5 eigenvectors, is in Figure 5. Projection coefficients form the resulting feature vector, directly.

The face recognition method proposed by Pan et al. [9] maps the face surface into a planar circle. At first, the nose tip is located and a region of interest (ROI) is chosen. The ROI is the sphere centered at the nose tip. After that, the face surface within the ROI is selected and mapped on the planar circle. The error function E that measures the distortion between the original surface and plane is used. The transformation to the planar circle is performed so that E is minimal. Heseltine [15] shows that the application of certain image processing techniques to the range image has a positive impact on recognition performance.

2.3.2. Recognition methods specific to 3D face

So far, the methods that have emerged as an extension of the classical 2D face recognition were mentioned. In this section, an overview of some purely 3D face recognition methods is provided.

9.53   -0.90   6.43   0.43   -4.45
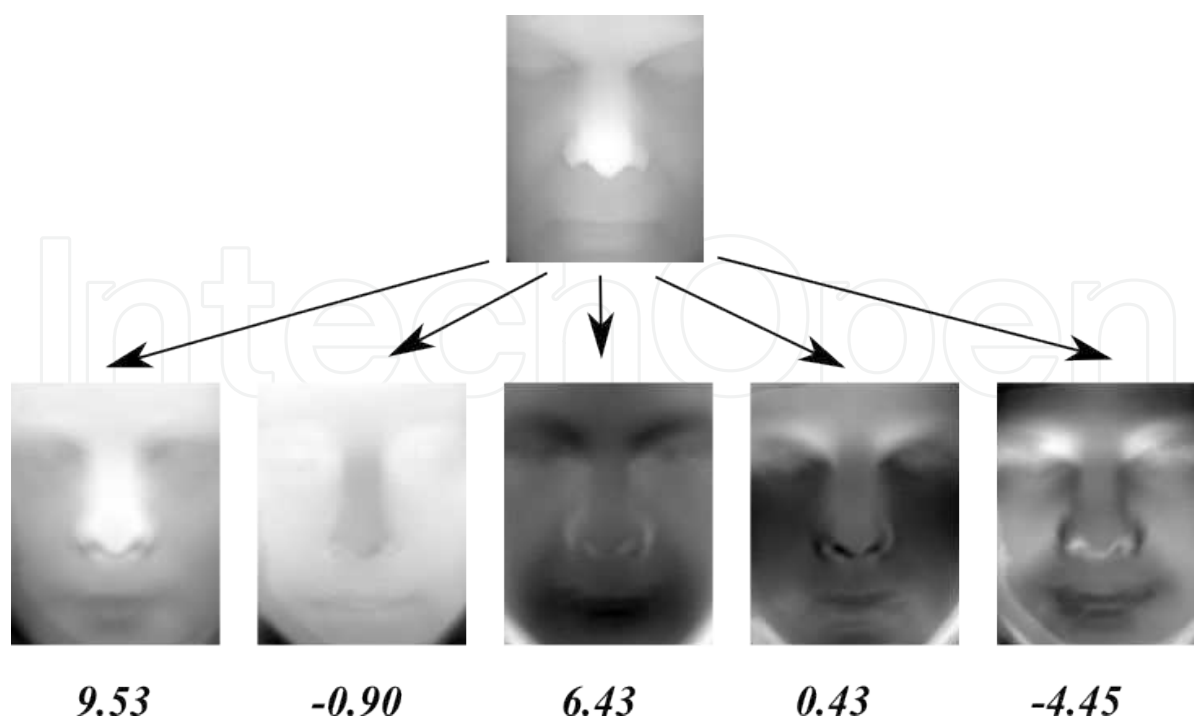
Figure 5. An input range image and its decomposition in PCA subspace consisting of 5 eigenvectors.

### Direct comparison using the hybrid ICP algorithm

Lu et. al [7] proposed a method that compares a face scan to a 3D model stored in a database. The method consists of three stages. At first, landmarks are located. Lu uses the nose tip, the inside of one eye and the outside of the same eye. Localization is based on curvature analysis of the scanned face. These three points, obtained in the previous step, are used for coarse alignment with the 3D model, stored in the database. A rigid transformation of the three pairs of corresponding points is performed in the second step.

A fine registration process, the final step, uses the Iterative Closest Point (ICP) algorithm. The root mean square distance minimized by the ICP algorithm, is used as the comparison score.

Recognition using histogram-based features

The algorithm introduced by Zhou et al. [16] is able to deal with small variations caused by facial expressions, noisy data, and spikes on three-dimensional scans. After the localization of the nose, the face is aligned, such that the nose tip is situated in the origin of coordinates and the surface is converted to a range image. Afterwards, a rectangle area around the nose is selected (*region of interest*, ROI). The rectangle is divided into N equal stripes. Each stripe n contains S_n points. Maximal $Z_{n,max}$ and minimal $Z_{n,min}$ z-coordinates within each stripe are calculated and the z-coordinate space is divided into K equal width bins. With the use of the K bins, a histogram of z-coordinates of points forming the scan, is calculated in each stripe. This yields to a feature vector consisting of $N \cdot K$ components. An example of an input

range image, and a graphical representation of the corresponding feature vector, is shown in Figure 6.
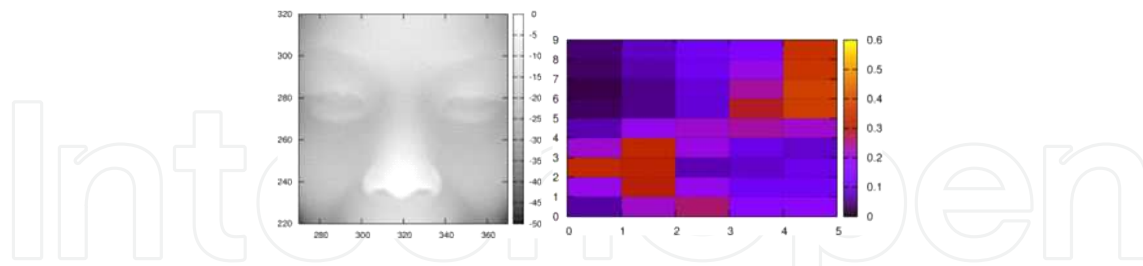
Figure 6. An input range image and its corresponding histogram template, using 9 stripes and 5 bins in each stripe.

Recognition based on facial curves

In recent years, a family of the 3D face recognition methods, which is based on the comparison of facial curves, has emerged. In these methods, the nose tip is located first. After that, a set of closed curves around the nose is created, and the features are extracted.

(a)   (b)
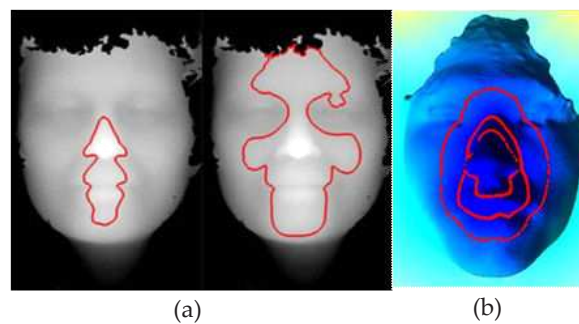
**Figure 7.** Iso-depth (a) and iso-geodesic (b) curves on the face surface [15].

In [18], recognition based on iso-depth and iso-geodetic curves is proposed. The iso-depth curve is extracted from the intersection between the face surface and the parallel plane, perpendicular to the z-axis (see Figure 7(a)). The iso-geodesic curve is a set of all points on the surface that have the same geodesic distance from a given point (see Figure 7(b)). The geodesic distance between two points on the surface is a generalization of the term distance on a curved surface.

There is one very important attribute to a the iso-geodesic curve. Contrary to the iso-depth curves, from a given point, iso-geodesic curves are invariant to translation and rotation. This means that no pose normalization of the face is needed, in order to deploy a face recognition algorithm strictly based on iso-geodesic curves. However, precise localization of the nose-tip is still a crucial part of the recognition pipeline.

There are several shape descriptors used for feature extraction in [18]. A set of 5 simple shape descriptors (convexity, ratio of principal axes, compactness, circular variance, and el-

liptical variance) is provided. Moreover, the Euclidian distance between the curve center and points on the curve is sampled for 120 points on the surface and projected using LDA in order to reduce dimensionality of the feature vector. Three curves are extracted for each face.

The 3D face recognition algorithm proposed in [19] uses iso-geodetic stripes and the surface data are encoded in the form of a graph. The nodes of the graph are the extracted stripes and the directed edges are labeled with *3D Weighted Walkthroughs*. The walkthrough from point $a=(x_a, y_a)$ to $b=(x_b, y_b)$ is illustrated in Figure 8. It is a pair $\langle i, j \rangle$ that describes the sign of mutual positions projected on both axes. For example, if $x_a < x_b \wedge y_a > y_b$ holds, then $\langle i, j \rangle = \langle 1, -1 \rangle$. For more information about the generalization of walkthroughs from points to a set of points and to 3D space, see [19].
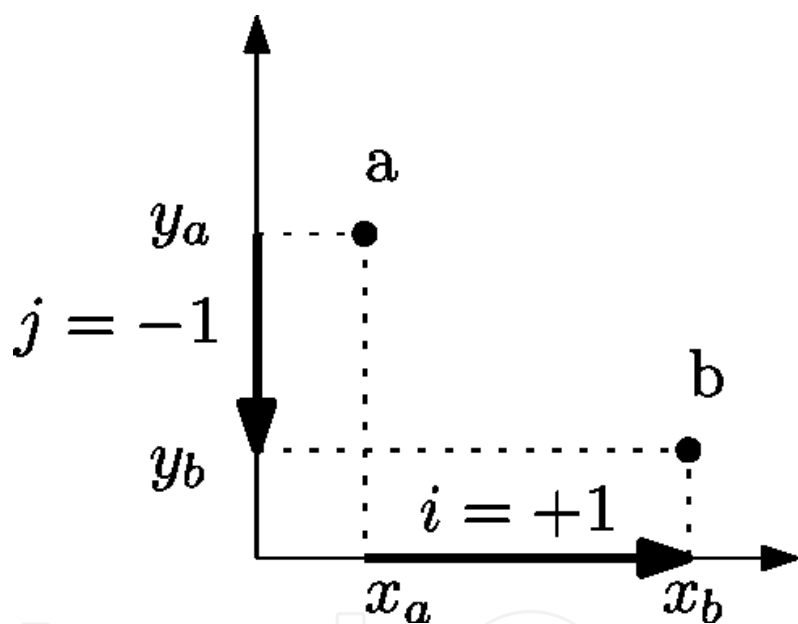
Figure 8. The walktrough $\langle i, j \rangle = \langle 1, -1 \rangle$ from *a* to *b*.

Recognition based on the anatomical features

Detection of important landmarks, described in section 2.2, may be extended to other points and curves on the face. The mutual positions of the detected points, distances between curves, their mutual correlations, and curvatures at specific points may be extracted. These numerical values directly form a feature vector, thus the distance between two faces may be instantly compared using an arbitrary distance function between these two feature vectors.

In [17], 8 facial landmarks and 4 curves were extracted from each input scan (see Figure 9) and form over sixty features. These features may be divided into four categories (see Table 3).

Figure 9. Detected facial landmarks (marked with white circles) and four facial curves: vertical *profile curve*, horizontal *eye curve*, horizontal *nose, or middle, curve* (that intersects the tip of the nose and horizontal curve lying directly between the eye and nose curves).

| Category	Description	Number of features
Basic	Distances between selected landmarks	7
Profile curve	Utilization of several distances between the profile curve, extracted from input scan and the corresponding curve from the mean (average) face	4
Eyes curve	Distances between the eyes curve from an input scan and its corresponding eyes curve from the average face model	4
Nose curve	Distances between the nose curve from an input scan and the corresponding nose curve from the average face model	4
Middle curve	Distances between the middle curve from an input scan and the corresponding middle curve from the average face model	4
1st derivation of curves	Distances between the 1st derivation of facial curves and corresponding curves from the average face model	16
2nd derivation of curves	Distances between the 2nd derivation of facial curves and the corresponding curves from the average face model	16
Curvatures	Horizontal and vertical curvatures on selected facial landmarks	6
		Σ 61

Table 3. Categories of anatomical 3D face features.

A fundamental part of recognition, based on anatomical features, is the selection of feature vector components. This subset selection boosts components, with good discriminative ability, and decreases the influence of features with low discriminative ability. There are several possibilities on how to fulfill this task:

- Linear discriminant analysis. The input feature space consisting of 61 dimensions is linearly projected to a subspace, with fewer dimensions, such that the intra-class variability is reduced, and inter-class variability is maximized.

- Subset selection and weighting. For the selection and weighting based on the discriminative potential, see section 4.2.

2.3.3. State-of-the-art

The developed face recognition system should be compared with other current face recognition systems available on the market. In 2006, the National Institute of Standards and Technology in USA found the Face Recognition Vendor Test (FRVT) [20]. It has been the latest, thus far, in a series of large scale independent evaluations. Previous evaluations in the series were the FERET, FRVT 2000, and FRVT 2002. The primary goal of the FRVT 2006 was to measure progress of prototype systems/algorithms and commercial face recognition systems since FRVT 2002. FRVT 2006 evaluated performance on high resolution still images (5 to 6 mega-pixels) and 3D facial scans.

A comprehensive report of achieved results, and used evaluation methodology is described in [21]. The progress that was achieved during the last years is depicted in Figure 10. Results show achieved false rejection rate, at a false acceptance rate of 0.001, for the best face recognition algorithms. This means that, if we admit that 0.1% of impostors are falsely accepted as genuine persons, only 1% of genuine users are incorrectly rejected. The best 3D face recognition algorithm that has been evaluated in FRVT 2006 was Viisage, from the commercial portion of participating organizations [21].
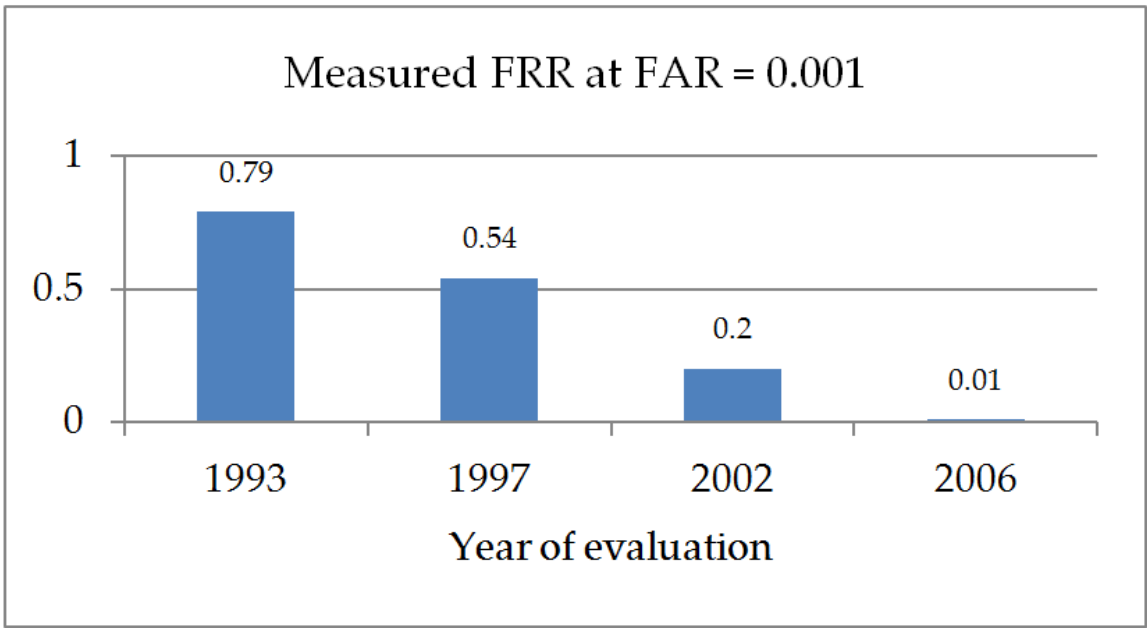
Figure 10. Reduction in error rate for state-of-the-art face recognition algorithms as documented through FERET, FRVT 2002, and FRVT 2006 evaluations.

The upcoming Face Recognition Vendor Test 2012 continues a series of evaluations for face recognition systems. The primary goal of the FRVT 2012 is to measure the advancement in the capabilities of prototype systems and algorithms from commercial and academic communities.

3. Thermal face recognition

Face recognition based on thermal images has minor importance in comparison to visible light spectrum recognition. Nevertheless, in applications such as the liveness detection or the fever scan, thermal face recognition is used as a standalone module, or as a part of a multi-modal biometric system.

Thermal images are remarkably invariant to light conditions. On the other hand, intra-class variability is very high. There are a lot of aspects that contribute to this negative property, such as different head poses, face expressions, changes of hair or facial hair, the environment temperature, current health conditions, and even emotions.

3.1. Thermal-face acquisition

Every object, whose temperature is not absolute zero, emits the so called "thermal radiation". Most of the thermal radiation is emitted in the range of 3 to 14 μm, not visible to the human eye. The radiation itself consists primarily of self-emitted radiation from vibrational and rotational quantum energy level transitions in molecules, and, secondarily, from reflection of radiation from other sources [26]. The intensity and the wavelength of the emitted energy from an object are influenced by its temperature. If the object is colder than 50°C, which is the case of temperatures of a human being, then its radiation lies completely in the IR spectrum.

3.1.1. Temperature measurements

The radiation properties of objects are usually described in relation to a perfect blackbody (the perfect emitter) [27]. The coefficient value lies between 0 and 1, where 0 means none and 1 means perfect emissivity. For instance, the emissivity of human skin is 0.92. The reflected radiations from an object are supposed to be much smaller than the emitted ones, therefore they are neglected during imaging.

Atmosphere is present between an object and a thermal camera, which influences the radiation due to absorption by gases and particles. The amount of attenuation depends heavily on the light wavelength. The atmosphere usually transmits visible light very well, however, fog, clouds, rain, and snow can distort the camera from seeing distant objects. The same principle applies to infrared radiation.

The so-called atmospheric windows (only little attenuation), which lie between 2 and 5 μm (the mid-wave window), and 7.5–13.5 μm (the long-wave window) have to be used for thermo-graphic measurement. Atmospheric attenuation prevents an object's total radiation from

reaching the camera. The correction of the attenuation has to be done in order to get the true temperature, otherwise it will be dropping, with increasing distance.

3.1.2. Thermal detectors

Majority of IR cameras have a microbolometer type detector, mainly because of cost considerations. They respond to radiant energy in a way that causes a change of state in the bulk material (i.e., the bolometer effect) [27]. Generally, microbolometers do not require cooling, which allows for compact camera designs (see Figure 11) to be relatively low cost. Apart from lower sensitivity to radiation, another substantial disadvantage of such cameras is their relatively slow reaction time, with a delay of dozens of milliseconds. Nevertheless, such parameters are sufficient for biometric purposes.

**Figure 11.** FLIR ThermaCAM EX300[28].

For more demanding applications, quantum detectors can be used. They operate based on an intrinsic photoelectric effect [27]. The detectors can be very sensitive to the infrared radiation which is focused on them by cooling to cryogenic temperatures. They also react very quickly to changes in IR levels (i.e., temperatures), having a constant response time in the order of 1μs. However, their cost disqualifies their usage in biometric applications in these days.

3.2. Face and facial landmarks detection

Head detection in a visible spectrum is a very challenging task. There are many aspects making the detection difficult such as a non-homogenous background and various skin colors. Since the detection is necessary in the recognition process, much effort was invested in dealing with this problem. Nowadays, one of the most commonly used methods is based on the Viola-Jones detector, often combined with additional filtering, required for skin color model.

In contrast to the visible spectrum, detection of the skin on thermal images is easier. The skin temperature varies in a certain range. Moreover, skin temperature remarkably differs from the temperature of the environment. That is why techniques based on background and foreground separation are widely used. The first step of skin detection is usually based on a thresholding. A convenient threshold is in most scenarios computed using Otsu algorithm [34]. Binary images usually need more correction consisting of hole removal and contour smoothing [33]. Another approach detects the skin, using Bayesian segmentation [29].

The next step of detection is the localization of important facial landmarks such as the eyes, nose, mouth and brows. The Violla-Jones detector can be used to detect some of them. Friedrich and Yeshurn propose eye brow detection by analysis of local maxima in a vertical histogram (see Figure 12) [33].
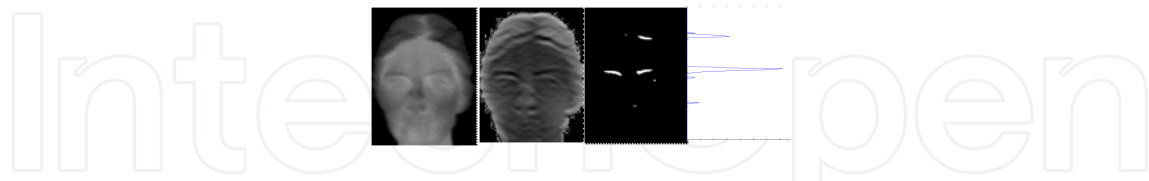
Figure 12. Eye brow detection by [33]: Original image (left), edges enhancement, binarized image and its vertical histogram - sum of intensities in each row (right).

3.3. Head normalization

If a comparison algorithm is performed on raw thermal face images, without any processing, we would get unacceptable results. This is becuase the thermal faces belong to biometrics with high intra-class variability. This makes thermal face recognition one of the most challenging methods, in terms of intra-class variability reducing, while keeping or increasing the inter-class variability of sensory data. The normalization phase of the recognition process tries to deal with all these aspects and decrease the intra-class variability as much as possible, while preserving the inter-class variability.
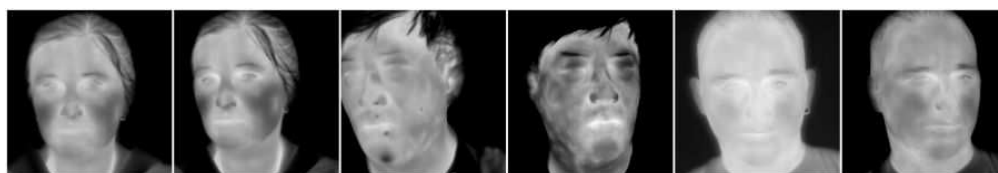
Figure 13. Thermal images of 3 different people. Output of all normalization methods is demonstrated by processing these raw images.

Proposed normalization consists of a pose, intensity and region of stability normalization. All normalization methods are described in the following sections and their output is visualized on the sample thermal images in Figure 13.

3.3.1. Pose normalization

Biometric systems based on face recognition do not strictly demand the head to be positioned in front of the camera. The task of pose (geometric) normalization is to transform the captured face to a default position (front view without any rotation). The fulfillment of this task is one of the biggest challenges for 2D face recognition technologies. It is obvious that a perfect solution cannot be achieved by 2D technology, however, the variance caused by different positions should be minimized as much as possible.

Geometric normalization often needs information about the position of some important points within the human face. These points are usually image coordinates of the eyes, nose and mouth. If they are located correctly, the image can be aligned to a default template.

2D affine transformation

Basic methods of geometric normalization are based on affine transformation, which is usually realized by transformation, using a matrix T. Each point $p = [x, y]$ of original image I is converted to homogenous coordinates $p_h = [x, y, 1]$. All these points are multiplied by the matrix T to get the new coordinates p'.

Methods of geometric normalization vary with the complexity of transformation matrix computation. The *general affine transformation* maps three different facial landmarks of the original image, I, to their expected positions within the default template. Transformation matrix coefficients are computed by solving a set of linear algebraic equations [23].

3D projection

Human heads have an irregular ellipsoid-like 3D shape, therefore, the 2D-warping method works well, when the head is scaled or rotated in the image plane. In the case of any other transformation, the normalized face is deformed (see Figure 15).

The proposed 3D-projection method works with an average 3D model of the human head. A 3D affine transformation, consisting of translation, rotation and scaling, can be applied to each vertex. The transformed model can be perspectively projected to a 2D plane afterwards. This process is well known from 3D computer graphics and visualization.

Model alignment, according to the image, I, is required. The goal is to find a transformation of the model whose orientation after the transformation will reveal each important facial landmark. The texture of the input image, I, is projected onto the aligned model. Then, the model is transformed (rotated and scaled) to its default position and finally, the texture from the model is re-projected onto the resulting image (see Figure 14).
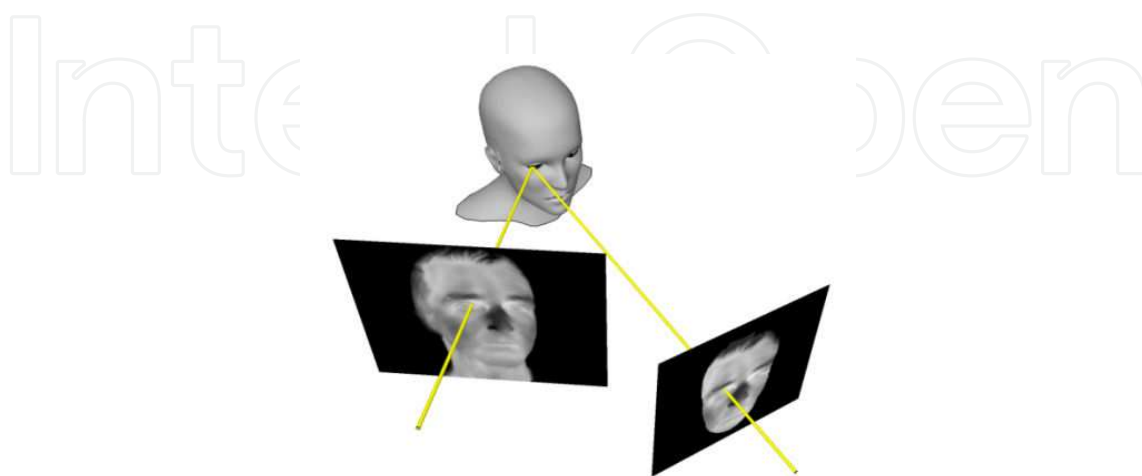

Figure 14. Visualization of the 3D projection method.

This kind of normalization considers the 3D shape of the human face. However, the static (unchangeable) model is the biggest drawback of this method. There are more advanced techniques that will solve this problem by using the *3D Morphable Face Model* [22].
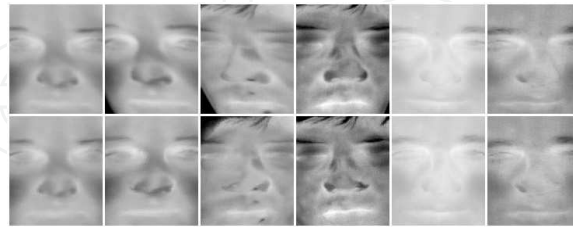
Figure 15. Pose normalization methods overview: 2D affine transformation (first row) and the 3D projection method (second row).
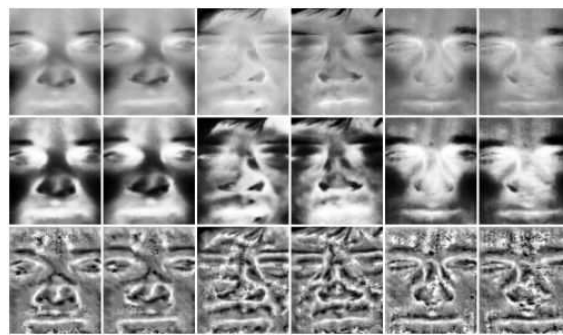

Figure 16. Intensity normalization methods overview: Min-max (first row), Global equalization (send row) and Local equalization (third row).

3.3.2. Intensity normalization

A comparison of thermal images in absolute scale does not usually lead to the best results. The absolute temperature of the human face varies with environmental temperature, physical activity and emotional state of the person. Some testing databases even do not even contain information on how to map pixel intensity to that temperature. Therefore, intensity normalization is necessary. It can be accomplished via global or local histogram equalization of noticeable facial regions (see Figure 16).

3.3.3. Region of stability normalization

The region of stability normalization takes into account the shape of a face, and a variability of temperature emission within different face parts. The output image of previous normalizations has a rectangular shape.
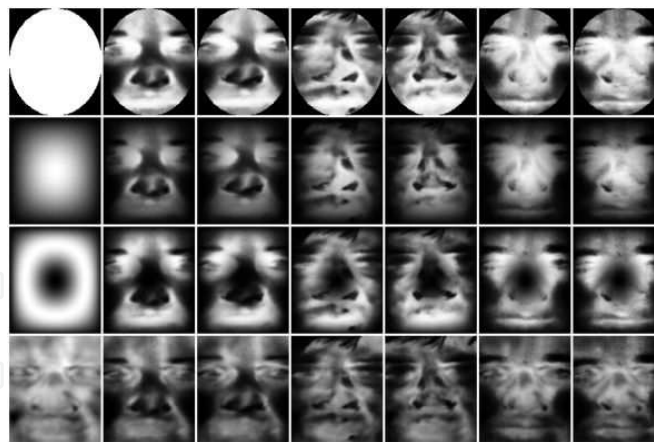
Figure 17. Overview of methods used for region of stability and the normalization. The masks (first column) and normalization responses are displayed for the following methods: Elliptical (top row), Smooth-Elliptical, Weighted-Smooth-Elliptical, Discriminative potential (bottom row).

The main purpose of this normalization is to mark an area where the most important face data are located, in terms of unique characteristics. This normalization is done by multiplying the original image by some mask (see Figure 17).

- Elliptical mask: The human face has an approximately elliptical shape. A Binary elliptical mask is therefore the simplest and most practical solution.

- Smooth-elliptical mask: The weighted mask does not have a step change on the edge between the expected face points and background points.

- Smooth-weighted-elliptical mask: Practical experiments show that the human nose is the most unstable feature, in terms of temperature emissivity. Therefore, the final mask has lower weight within the expected nasal area position.

- Discriminative potential mask: Another possibility to mark regions of stability is by training on part of a face´s database. A mask is obtained by the discriminative potential method described in section 4.2.

3.4. Feature extraction on thermal images

Several comparative studies of thermal face recognition approaches have been developed in recent years. The first recognition algorithms were appearance-based. These methods deal with the normalized image as a vector of numbers. The projected vector turns into a (low dimensionality) subspace, where the separation between impostors and genuines can be efficiently computed with higher accuracy. The commonly used methods are the PCA, LDA and ICA, which are described in more detail in section 4.2.

While appearance-based methods belong to the global-matching methods, there are local-matching methods, which compare only certain parts of an input image to achieve better performance. The LBP (*Local Binary Pattern*) method was primarily developed for texture description and recognition. Nevertheless, it was successfully used in visible and thermal face

recognition [24]. It encodes the neighbors of a pixel, according to relative differences, and calculates a histogram of these codes in small areas. These histograms are then combined with a feature vector. Another comparative study [25] describes other local-matching methods such as the Gabor Jets, the SIFT (*Scale Invariant Feature Transform*), the SURF (*Speeded up Robust Features*) and the WLD (*Weber Linear*) descriptors.

A different approach extracts the vascular network from thermal images, which is supposed to be unique according to each individual. One of the prominent methods [29] extracts thermal minutia points (similar to fingerprints minutia) and compares two vascular networks subsequently (see Figure 18). In another approach, it is proposed to use a feature set for the thermo-face representation: the bifurcation points of the thermal pattern and geographical, and gravitational centers of the thermal face [30].

Figure 18. A thermal minutia point being extracted from a thinned vascular network[29].

4. Common processing parts for normalized, 3D, and thermal images

For both thermal and 3D face recognition, it is very difficult to select one method, out of numerous possibilities, which gives the best performance in each scenario. Choosing the best method is always limited to a certain or specific database (input data). In order to address this problem, multi-algorithmic biometric fusion can be used.

In the following sections, a general multi-algorithmic biometric system will be described. Both 3D face recognition and thermal facial recognition require a normalized image. Since many of these characteristics are similar, we do not distinguish between origins of the normalized image. This section describes generic algorithms of feature extraction, feature projection and comparison, which are being evaluated on normalized 3D, as well as thermal images.

4.1. Feature extraction of the normalized image

The feature extraction part takes the normalized image as input, and produces a feature vector as output. This feature vector is then processed in the feature projection part of the process.

Vectorization is the simplest method of feature extraction. The intensity values of the image $I = w \times h$ are concentrated to a single column vector. Performance of this extraction depends on the normalization method.

Since normalized images are not always convenient for direct vectorization, feature extraction with the use of the **bank of filters** has been presented in several works. The normalized image is convolved with a bank of 2D filters, which are generated, using some kernel function with different parameters (see Figure 19). Each response of the convolution forms a final feature vector.

The **Gabor filter bank** is one of the most popular filter banks [32]. We employed the **Laguerre-Gaussian filter bank** as well due to its good performance in the facial recognition field [31].

Figure 19. The Gabor (first row) and Laguere-Gaussian (second row) filter banks.

4.2. Feature vector projection and further processing

Statistical projection methods linearly transform the input feature vector from an input m-dimensional space into an n-dimensional space, where $n < m$. We utilize the following methods:

- Principal component analysis (PCA, Eigenfaces)

- PCA followed by linear discriminant analysis (LDA of PCA, Fisherfaces)

- PCA followed by independent component analysis (ICA of PCA)

Every projectional method has a common learning parameter, which defines how much variability of the input space is captured by the PCA. This parameter controls the dimensionality of the output projection space. Let k eigenvalues, computed during the PCA calculation, be denoted as $e_1, e_2, \ldots, e_k, \quad (e_1 > e_2 > \ldots > e_k)$. These eigenvalues directly represent the variability in each output dimension. If we want to preserve only 98% of variability, then only the first l eigenvalues and its corresponding eigenvectors are selected, such that their sum forms only 98% of the $\sum_{j=1}^{k} e_j$.

There is an optional step to perform a per-feature z-score normalization after the projection, so that each vector fv is transformed into

$$fv' = \frac{fv - \overline{fv}}{\sigma} , \tag{4}$$

where $\overline{fv}$ is the mean vector and σ is the vector of standard deviations.

Optional processing, after the application of statistical projection methods, is the feature weighting. Suppose that we have a set FV of all pairs of feature vectors, v_j, and their corresponding class (subject) labels, id_j:

$$FV = \{(id_1, fv_1), (id_2, fv_2), \ldots, (id_n, fv_n)\} \tag{5}$$

The individual feature vector components, $v_{j1}, fv_{j2}, \ldots, fv_{jm}$, of the vector fv_j do not have the same discriminative ability. While some component may have positive contribution to the overall recognition performance, the other component may not. We have implemented and evaluated two possible feature evaluation techniques.

The first possible solution is the LDA application. The second option is to make an assumption that the good feature vector component has stable values across different scans of the same subjects, however, the mean value of a specific component across different subject differs to the greatest possible extent. Let the intra-class variability of the feature component i be denoted as $intra_i$, as it expresses the mean of standard deviations of all measured values for the same subjects. The inter-class variability of component i is denoted as $inter_i$, and expresses the standard deviation of means of measured values for the same subject. The resulting discriminative potential, therefore, can be expressed as follows:

$$discriminative potential = inter_i - intra_i \tag{6}$$

4.3. Fusion using binary classifiers

The number of combinations for common feature extraction techniques and optional feature vector processing, yields a large set of possible recognition methods. For example, the Gabor filter bank, consisting of 12 kernels, may be convolved with an input image. The results of the convolution are concatenated into one large column vector, which is then processed with PCA, followed by LDA. Another example is an input image processed by the PCA. The individual features in the resulting feature vector, are multiplied by their corresponding normalized discriminative potential weight.

After the features are extracted, the feature vector is compared with the template from a biometric database, using some arbitrary distance function. If the distance is below a certain threshold, the person, whose features were extracted, is accepted as a genuine user. If we are

using several different recognition algorithms, the simple threshold becomes a binary classification problem. The biometric system has to decide whether the resulting score vector $s = (s_1, s_2, \ldots, s_n)$ belongs to the genuine user or the impostor. An example of a general multimodal-biometric system employing a score-level fusion is in Figure 20, but the same approach may be applied to a multi-algorithmic system, where input is just one sample and more than one feature extraction and comparison method is applied.
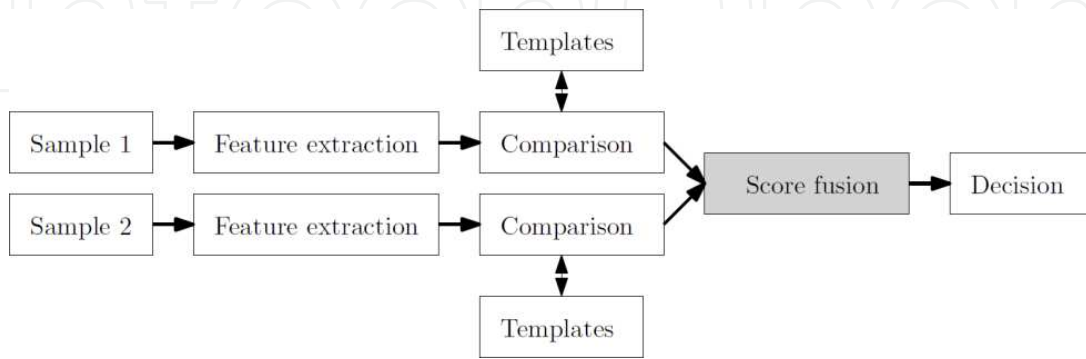
Figure 20. A generic multimodal biometric system using score-level fusion.

In order to compare and fuse scores that come from different methods, normalization, to a certain range, has to be performed. We use the following score normalization: the score values are linearly transformed so that the genuine mean (the score, obtained from comparing the same subjects) is 0, and the impostor mean (score, obtained from comparing different subjects) is 1. Note that individual scores may have negative values. This does not matter in the context of score-level fusion, since these values represent positions within the classification space, rather than distances between two feature vectors.

The theoretical background of general multimodal biometric fusion, especially the link between the correlation and variance of both impostor and genuine distribution between the employed recognition methods, is described in [35]. The advantage provided by score fusion relative to monomodal biometric systems is described in detail in [36].

In our fusion approach, we have implemented a classification using logistic regression [37], support vector machines (SVM), with linear and sigmoidal kernel [37], and linear discriminant analysis (LDA).

4.4. Experimental results

For evaluation of our fusion approach on ethermal images, we used the Equinox [38] and Notre-dame databases [39][40]. Equinox contains 243 scans of 74 subjects, while the Notre Dame database consists of 2,292 scans. Evaluation for 3D face recognition was performed on the "Spring 2004" part on the FRGC database [20], from which we have selected only subjects with more than 5 scans. This provided 1,830 3D scans in total.

The evaluation scenario was as follows. We divided each database into three equal parts. Different data subjects were present in each part. The first portion of the data was used for

training of the projection methods. The second portion was intended for optional calculation of z-score normalization parameters, the feature weighting, and the fusion classification training. The final part was used for evaluation.

To ensure that the particular results of the employed methods are stable and reflect the real performance, the following cross-validation process was selected. The database was randomly divided into three parts, where all parts had an equal number of subjects. This random division, and subsequent evaluation was processed n times, where n depends on the size of the database. The Equinox database was cross-validated 10 times, while the Notredame database and FRGC were cross-validated 3 times. The performance of a particular method was reported as the mean value of the achieved equal error rates (EERs).

4.4.1. Evaluation on thermal images

For thermal face recognition the following techniques were fused:

- Local and global contrast enhancement,

- The Gabor and Laguerre filter banks,

- PCA and ICA projection,

- Weighting based on discriminative potential,

- Comparison using the cosine distance function.

From these techniques, 10 best different recognition methods were selected for final score-level fusion. Logistic regression was used for score fusion. The results are given in Table 4.

4.4.2. Evaluation on 3D face scans

For performance evaluation of the 3D face scans, the following recognition methods were used:

- Recognition, using anatomical features. The discriminative potential weighting was applied on the resulting feature vector, consisting of 61 features. The city-block (Manhattan, L_0) metric was employed.

- Recognition using histogram-based features. We use a division of the face into 10 rows and 6 columns. The individual feature vector components were weighted by their discriminative potential. Cosine metric is used for the distance measurement.

- Cross correlation of shape-index images[7], see Figure 21.

- Shape-index images projected by the PCA, weighting based on discriminative potential, and the cosine distance function.

- Shape-index images projected by the PCA followed by the ICA, weighting based on discriminative potential, and the cosine distance function.
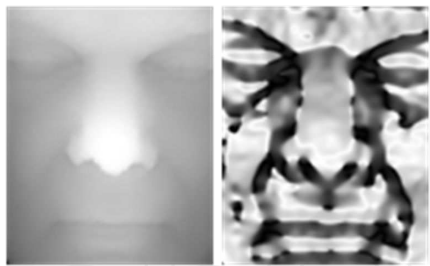
Figure 21. An original range image and shape index visualization.

The result of the algorithms evaluation using FRGC database is given in Table 4.

Database	Best single method name	Best single method EER	Fusion EER	Improvement
Equinox	Global contrast enhancement, no filter bank, ICA, cosine distance	2.28	1.06	53.51%
Notre-dame	Global contrast enhancement, no filter bank, PCA, cosine distance	6.70	5.99	10.60%
FRGC	Shape index, PCA, weighting using discriminative potential, cosine distance	4.06	3.88	4.43%

Table 4. Evaluation of fusion based on logistic regression. For every fusion test, all individual components of the resulting fusion method were evaluated separately. The best component was compared with overall fusion, and improvement was also reported. The numbers represent the achieved EER in %.

5. Conclusion

This chapter addressed a novel approach to biometric based-system design, viewed as design of a distributed network of multiple biometric modules, or assistants. The advances of multi-source biometric data are demonstrated using 3D and infrared facial biometrics. A particular task of such system, namely, identification, using fusion of these biometrics, is demonstrated. It is shown that reliability of the fusion-based decision increases. Specifically, 3D face models carry additional topological information, and, thus, are more robust, compared to 2D models. Thermal data brings additional information. By deploying the best face recognition techniques for both 3D and thermal domains, we showed, through experiments, that fusion increases the overall performace up to 50%.

It should be noted that the important components of the processing pipeline image processing and the correct feature selection greatly influence the decision-making (comparison). Also, choice of another way of fusion may influence the results.

The identification task, combined with advanced discriminative analysis of biometric data, such as temperature and its derivatives (blood flow rate, pressure etc.), constitute the basis

for the higher-level decision-making support, called semantic biometrics [3]. Decision-making in semantic form is the basis for implementation in distributed security systems, PASS of the next generation. In this approach, the properties of linguistic averaging are efficiently utilized for smoothing temporal errors, including errors caused by insufficiency of information, at the local and global levels of biometric systems. The concept of semantics in biometrics is linked to various disciplines; in particular, to dialogue support systems, as well as to there commender systems.

Another extension of the concept of PASS is the Training PASS (T-PASS) that provides a training environment for the users of the system [41]. Such a system makes use of synthetic biometric data [42], automatically generated to "imitate" real data. For example, models can be generated from real acquired data, and can simulate age, accessories, and other attributes of the human face. Generation of synthetic faces, using 3D models that provide the attribute of convincing facial expressions, and thermal models which have a given emotional coloring, is a function of both the PASS (to support identification by analysis through synthesis, for instance, modeling of head rotation to improve recognition of faces acquired from video), and T-PASS (to provide virtual reality modeling for trainees).

Acknowledgement

Author details

Štěpán Mráček[1], Jan Váňa[1], Radim Dvořák[1], Martin Drahanský[1] and Svetlana Yanushkevich[2]

1 Faculty of Information Technology, Brno University of Technology, Czech Republic

2 University of Calgary, Canada

References

[1] Jain A. K., Nandakumar K., Uludag U., LuX. Multimodal biometrics: augmenting face with other cues, In: Face Processing: Advanced Modeling and Methods, Elsevier, 2006.

[2] Yanushkevich S. N., Stoica A., Shmerko V. P. Experience of design and prototyping of a multi-biometric early warning physical access control security system (PASS) and a training system (T-PASS), In: Proc. 32nd Annual IEEE Industrial Electronics Society Conference, Paris, pp. 2347-2352, 2006.

[3] Yanushkevich S. N., Shmerko V. P., Boulanov O. R., StoicaA. Decision making support in biometric-based physical access control systems: design concept, architecture, and applications, In book: N. Bourgoulis and L. Micheli-Tsanakou, K. Platantionis (Eds.), Biometrics: Theory, Methods and Applications, IEEE/Wiley Press, pp. 599-632, 2009.

[4] Peng T."Algorithms and Models for 3-D Shape Measurement Using Digital Fringe Projections", Ph.D. thesis, University of Maryland, 2006.

[5] Fechteler P., Eisert P., Rurainsky J. Fast and High Resolution 3D Face Scanning, In Proceedings of the 14th International Conference on Image Processing (ICIP2007), San Antonio, Texas, USA, 2007.

[6] Artec M™ 3D Scanner. ARTEC GROUP. Artec M™ — Artec 3D Scanners, http://www.artec3d.com/3d_scanners/artec-m (accessed 21 April 2012).

[7] Lu X., Colbry D. & Jain A. Three-Dimensional Model Based Face Recognition, in ICPR 04: Proceedings of the Pattern Recognition, 17th International Conference on Pattern Recognition, pp. 362-366, 2004.

[8] Mahoor, M. H. & Abdel-Mottaleb, M. Face recognition based on 3D ridge images obtained from range data, Pattern Recognition, vol. 42, no. 3, pp. 445-451, 2009.

[9] Pan G., Han S., Wu Z. & Wang Y. 3D Face Recognition using Mapped Depth Images, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, vol. 3, p. 175, 2005.

[10] Segundo M., Queirolo C., Bellon O. & Silva L. Automatic 3D facial segmentation and landmark detection, in ICIAP '07 Proceedings of the 14th International Conference on Image Analysis and Processing, pp. 431-436, 2007.

[11] GrayA. The Gaussian and Mean Curvatures, in Modern Differential Geometry of Curves and Surfaces with Mathematica, pp. 373-380, 1997.

[12] Turk M. & Pentland A. Face recognition using eigen faces, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 591, no. 1, pp. 586-591, 1991.

[13] Belhumeur P., Hespanha J. & Kriegman D. Eigenfacesvs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.

[14] Hyvärinen A. Fast and robust fixed-point algorithms for independent component analysis., IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council, vol. 10, no. 3, pp. 626-34, 1999.

[15] Heseltine T., Pears N. & Austin J. Three-dimensional face recognition using combinations of surface feature map subspace components," Image and Vision Computing, vol. 26, no. 3, pp. 382-396, 2006.

[16] Zhou X., Seibert H. & BuschC. A 3D face recognition algorithm using histogram-based features, Eurographics 2008 Workshop on 3D Object Retrieval, pp. 65-71, 2008.

[17] Mráček Š., Busch C., Dvořák R. &Drahanský M. Inspired by Bertillon - Recognition Based on Anatomical Features from 3D Face Scans, in Proceedings of the 3rd International Workshop on Security and Communication Networks, pp. 53-58, 2011.

[18] Jahanbin S., Choi H., Liu Y. &Bovik A. C. Three Dimensional Face Recognition Using Iso-Geodesic and Iso-Depth Curves, in 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems. 2008.

[19] Berretti S., Del Bimbo A. & Pala P. 3D Face Recognition Using iso-Geodesic Stripes. IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 12, pp. 2162-2177, 2010.

[20] Phillips P. J. et al. Overview of the Face Recognition Grand Challenge, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 947-954, 2005.

[21] Phillips P. J. et al., FRVT 2006 and ICE 2006 large-scale experimental results, IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 5, pp. 831-46, 2010.

[22] Blanz V. Face Recognition based on a 3D Morphable Model, in Proc. of the 7th Int. Conference of Automatic Face and Gesture Recognition, pp. 617–622, 2006.

[23] Raton B. Affine Transformations. In Standard Mathematical Tables and Formulae, D. Zwillinger, pp. 265–266, 1995.

[24] Mendez H., San Martı́n C., Kittler J., Plasencia Y., Garcı́a E. Face recognition with LWIR imagery using local binary patterns, LNCS5558, pp. 327–336, 2009.

[25] Hermosilla G., Ruiz-del-Solar J., Verschae R., Correa M. A comparative study of thermal face recognition methods in unconstrained environments, Pattern Recognition, vol. 45, no. 7, pp. 2445-2459, 2012.

[26] Lloyd J. M. Thermal Imaging Systems, New York: Plenum Press. ISBN 0-306-30848-7, 1975.

[27] FLIR Systems. The Ultimate Infrared Handbook for R&D Professionals: A Resource Guide for Using Infrared in the Research and Development Industry, United Kingdom, 2012.

[28] FLIR Systems, ThermaCAM Reporter - user's manual, Professional Edition, Version 8.1, publ. No. 1558567, 2007.

[29] Buddharaju P., Pavlidis I. Multi-spectral face recognition - fusion of visual imagery with physiological information, in:R. I. Hammoud,B. R. Abidi, M. A. Abidi (Eds.),

Face Biometrics for Personal Identification: Multi-Sensory Multi-Modal Systems, Springer, pp.91–108, 2007.

[30] Cho S. Y., Wang L., Ong W. L. Thermal imprint feature analysis for face recognition, Industrial Electronics, ISIE2009, IEEE International Symposium on, pp.1875–1880, 2009.

[31] Jacovitti, G. & Neri, A. Multiscale image features analysis with circular harmonic wavelets,in Proceedings of SPIE, vol. 2569, pp. 363–374, 1995.

[32] Lee T. Image representation using 2D Gabor wavelets, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 959–971, 1996.

[33] Friedrich G. & Yeshurun Y. Seeing People in the Dark: Face Recognition in Infrared Images, in BMCV '02 Proceedings of the Second International Workshop on Biologically Motivated Computer Vision, pp. 348–359. Springer-Verlag London, UK, 2002.

[34] Otsu N. A Threshold Selection Method from Gray-Level Histograms, IEEE Transactions on Systems, Man and Cybernetics, pages 62-66, 1979.

[35] Poh N. & Bengio S. How Do Correlation and Variance of Base-Experts Affect Fusion in Biometric Authentication Tasks? IEEE Transactions on Signal Processing, vol. 53, no. 11, pp. 4384-4396, 2005.

[36] Puente L., Poza M. J., Ruíz B. & Carrero D. Biometrical Fusion – Input Statistical Distribution, in Advanced Biometric Technologies, InTech, 2011, pp. 87-110, 2011.

[37] Bishop C. Pattern Recognition and Machine Learning, vol. 4, no. 4. Springer, p. 738, 2006.

[38] Equinox, Multimodal face database. http://www.equinoxsensors.com/products/HID.html, (accessed 3 May 2012).

[39] Chen X., Flynn P. & Bowyer K. Visible-light and infrared face recognition, ACM Workshop on Multimodal User Authentication, pp. 48-55, 2003.

[40] Flynn P., Bowyer K. & Phillips P. Assessment of time dependency in face recognition: An initial study, Lecture Notes in Computer Science, vol. 2688, pp. 44-51, 2003.

[41] Yanushkevich S. N., Stoica A. & Shmerko, V. P. Fundamentals of biometric-based training system design, In: S. N. Yanushkevich, P. Wang, S. Srihari, and M. Gavrilova, Eds., M. S. Nixon, Consulting Ed., Image Pattern Recognition: Synthesis and Analysis in Biometrics, World Scientific, 2006.

[42] Yanushkevich S. N., Stoica A., Shmerko V. Synthetic Biometrics, IEEE Computational Intelligence Magazine, Volume 2, Number 2, pp. 60-69, 2007.

[43] Wu S., Lin W., Xie S. Skin heat transfer model of facial thermograms and its application in face recognition. Pattern Recognition 41(8): 2718-2729, 2008.

[44] Lu Y., Yang J., Wu S., Fang Z. Normalization of Infrared Facial Images under Variant Ambient Temperatures, book chapter, Advanced Biometrics Technologies, Intech, Vienna, Austria, ISBN 978-953-307-487-0, 2011.