A Survey on Ear Recognition

Spyridon Antakis

University of Twente
Department of Mathematics & Computer Science
Enschede, The Netherlands, P.O. Box 217, 7500 AE
Email: {s.antakis}@student.tue.nl

February 23, 2009

Abstract

In the last decades, a lot of research effort has been made in the field of biometric recognition. More precisely, there are many experiments which are conducted in upcoming biometrics such as ears. Ear is one of the new comers in biometric recognition techniques and until today it has not been decided whether the ear can be a good candidate for a human biometric. This paper, presents an in depth survey in the area of ear biometric, it summarizes the most important recognition methods for ears and it discusses their performance. Furthermore, it makes a comparative assessment with respect to other biometric modalities. Finally, the paper concludes if the ear recognition methods are at this point mature enough, to be used in real life applications.

1 Introduction

Biometrics is the science of *identifying* or *verifying* the identity of a person based on physiological or behavioral characteristics. Nowadays, there are different ways to define or verify the identity of someone. A person can use *possessions* like cards, badges and keys, *knowledge* like userid, passwords, and PINs and also *biometrics* such as fingerprint, ear, face and iris. Biometrics, offer much higher accuracy than the traditional methods. Possession can be lost, forgot or replicated easily and knowledge can be forgotten. Both possession and knowledge, can be stolen or shared with other people. In biometrics, these drawbacks do exist, but in a much smaller scale.[1] An ideal biometric must be *universal*, *unique*, *permanent* and *collectable*. However, in practice a biometric with characteristics that satisfy all these requirements maybe is not suitable for real life applications. In biometric systems, there are additional requirements that must be considered such as *performance*, *acceptability* and *circumvention*. Moreover, if the biometric system is used for authentication purposes, then there are several viewpoints that must be taken into account before the implementation starts, such as the *convenience*,*accuracy*, *availability* and *cost* of the system. [2, 3, 4]

Ear seems to have some advantages over the other recognition technologies. Its structure is not changing during the lifetime of an adult and is unaffected by facial expressions. It is located on the side of the head which makes detection easier because the immediate background is predictable, unlike that of the face. Data collection is convenient in comparison to other biometric characteristics and during the measurement procedure will not cause anxiety. Furthermore, ear is missing some unwanted properties that can lead to an easy replication (i.e. gummy fingerprints). Finally, it is detectable and easily captured from a long distance and its appearance it is not altered by make-up, spectacles, beards or glasses, although, it is often occluded by hair and earrings.

This paper, in section 2 addresses the most important ear recognition methods and the results reported on their performance, from contucted experiments. In section 3, a comparative assessment with respect to other biometric characteristics is presented and an evaluation about the discussed ear methods is made. In section 4, the paper concludes whether the ear recognition methods are mature enough at this point, to be used in every day applications.

2 Ear Recognition Methods

In general, the goal in an *identification method* is to be able to verify successfully, if the biometric extracted from a subject, sufficiently matches the previous acquired biometric for that subject. Since the subject and environment change over time, a certain tolerance in the matching criterion must be permitted. This tolerance can be defined in terms of the *False Reject Rate (FRR)* and the *False Acceptance Rate (FAR)* exhibited by the system. A system is usually designed to be tunable to minimize either the FAR or the FRR depending the security level required. If the system that we create aims on *recognition*, then the problem is really harder than that of identification, since the system must determine whether the subject's identity can be verified against any previously enrolled subject. Until now, human ear shape is not commonly used in applications. Through the last decades, different scientific methods for ear identification have been developed. In this section, the most important and popular *ear recognition techniques* are described.

2.1 Iannarelli's System

The first attempt of building a classification system for ear shapes was made by Alfred Iannarelli.[5] He has made two large scale ear identification studies. The first study, compared 10,000 ears drawn from a randomly selected sample and the second study examined identical twins and triplets. In both studies, all examined ears were found to be unique, though identical multiple birth siblings had similar, but not identical ear structures. Iannarelli's system was based on twelve measurements taken on ear photographs.

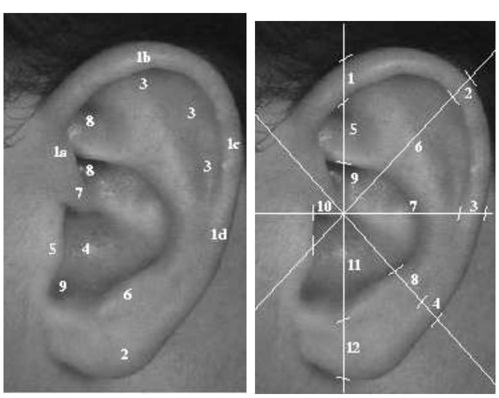

Figure 1: *"Iannarelli's System" - (a) Anatomy, (b) Measurements : 1.Helix rim, 2.Lobule, 3.Anti-Helix, 4.Concha, 5.Tragus, 6.Antitragus, 7. Crus of Helix, 8.Traingular Fossa, 9.Incisure Intertragica.*

The measurements were taken on specially aligned and normalized photograph of the right ear. While developing the negative, the image was projected to fit on a predefined easel. The distance between each of the numbered areas was measured and then an integer distance value, was assigned. (Fig. 1) However, later, it was discovered that this method is not suitable for machine vision because of the difficulty of localizing the anatomical points. If the first point is not defined accurately, none of the measurements is useful. Iannarelli himself, had also recognized this weakness.[6, 7, 8]

2.2 Voronoi Diagrams

M.Burge and W.Burger introduced a graph matching method for passive ear identification that was based on *Voronoi Diagrams* (App. A.3). This method, avoids the Iannarelli's System problem of localizing anatomical points and the weakness of basing all subsequent feature measurements on a single point. In practice, each subject's ear is modeled as an adjacency graph. Then, using *curve segments*, which is an essential technique to create graphics that appear smooth at fixed resolutions, the Voronoi diagram is built.[8]

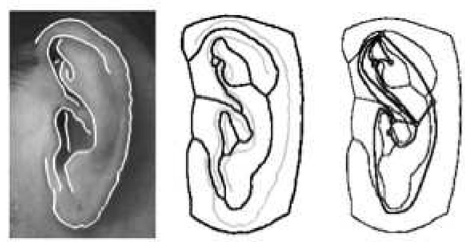

Figure 2: *"Neighborhood Graphs based on Voronoi Diagrams*

The algorithm is based on the following steps:

- *Step 1 - Acquisition:* A 300x500 grayscale image is taken of the subject's head using a camera. After that, the location of the ear is found.
- *Step 2 - Localization:* The ear is located by using a simple deformable contours on a Gaussian pyramid representation of the image gradient.[9]
- *Step 3 - Edge extraction:* Edges are computed using the Canny operator (App. A.4) and thresholding with hysteresis using upper and lower thresholds of 46 and 20 (Fig. 2b).
- *Step 4 - Curve extraction:* Edge relaxation is used to form larger curve segments, and to remove the remaining small curve segments. Performing an identification at this stage, was proved unreliable due to differences in lighting and positioning, so in order to

achieve invariance under affine transformations a Voronoi neighborhood graph of the curves is constructed.

- *Step 5 - Graph model:* A generalized Voronoi diagram of the curves is built and a neighborhood graph is extracted (Fig. 2c).

Using the above steps the result is a high FRR, due to variations in the graph models and underlying differences in the spatial relations of the extracted curves. M.Burge and W.Burger, in order to improve the FRR, proposed a devised novel match algorithm which took into account the erroneous curve segments that could occur in the ear image because of lighting, shadowing and occlusion. Nevertheless, the problem of occlusion by hair was identified as a major obstacle for this method and the possible solution of using *thermograms* was proposed.[8]

2.3 Compression Networks

In 1999, B. Moreno experimented with two different approaches.[1] The first approach, was based in the detection and analysis of *facial features* such as eye distance and chin's angle. The second approach, avoided using feature extraction and processed faces as general images with the appropriate tools (i.e. neural networks). One of the three classifiers that they used, was a *compression network classifier* which was based on the second approach.

In the technique that uses a compression network classifier for ear identification, there are two stages. At the first stage, a network called *compression network* is trained auto-associatively on the original ear image, in order to extract its statistically salient properties of the image data or macro-features. This vector, which is an intermediate codified representation of the original image, is the *compression vector*. It constitutes the input to a single perceptron that performs the identification task. In the identification task, each of the outputs corresponds to one of the individuals to be identified. Compression networks are trained as autoassociative memories, they allow the coding of neural patterns in a small dimensional subspace by extracting salient features. It has been proved, that a compression network with h hidden units can span the space of the h first eigenvectors of the covariance matrix corresponding to the input image. (App. A.2.1 & A.2.2)

Moreno et. al. conducted several experiments using this method, achieving a recognition rate of 93% on a data set of 168 images (6 images x 28 individuals), where individuals were invited to change expressions and face orientation. Also, they observed that in general, the combination of classifiers does not increase the identification rate, since the classifiers are not independent.

2.4 Force Field Transformations

In 2002, D.J.Hurley et. al. developed a new method, called *force field transformations*. This method uses an invertible linear transformation that transforms the entire image into a force field. To succeed this, it is supposed that each pixel exerts an isotropic force on all the other pixels which is proportional to the pixel's intensity. There is a potential energy surface associated with this force field, which in the case of an ear can be likened to a small mountain with a few peaks joined by ridges. The peaks are called *potential energy wells* and the ridges joining them, are called *potential energy channels*. The directional property of the force field is exploited to automatically locate these potential wells and channels, which then form the basis of the ear's signature. [11, 12]

2.4.1 Pixel's Force Exertion

In order to calculate the force exerted by a pixel, the ear image is transformed. Ear image, is considered to consist of an array of N *Gaussian attractors*, which act as the source of a force field. Every pixel, is assumed that it generates a spherically symmetrical force field (Fig. 3). From that force field, the force $F_i(r)$ exerted by a remote pixel, with position vector r_i and pixel intensity $P(r_i)$, on a pixel at the location with position vector r, can be calculated by the following equation:

$$F_i(r) = P(r_i)\frac{r_i - r}{|r_i - r|^3}$$

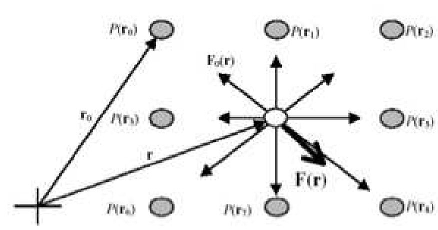

Figure 3: *"Pixel's Force Exertion"*

The total force acting on a pixel at a given position is the vector sum of all the forces due to the other pixels in the image and is given by:

$$F(r) = \sum_i F_i(r) = \sum_i P(r_i)\frac{r_i - r}{|r_i - r|^3}$$

So, the force field for the entire image can be calculated by just applying the previous equation at every pixel position in the image. Also, closely associated with the force field generated by each pixel there is a spherically symmetrical scalar potential energy field $E_i(r)$ (Fig. 4a). This potential energy is imparted to, by the energy field of a remote pixel, with position vector ri and pixel intensity $P(r_i)$, and it is given by the equation:

$$E_i(r) = \frac{P(r_i)}{|r_i - r|}$$

Again, in order to find the total potential energy at a particular pixel location on the image, the scalar sum is taken over the values of the overlapping potential energy functions (Fig. 4b) of all the image pixels at that precise location. Thus, the total potential energy can be calculated by the following equation:

$$E(r) = \sum_i E_i(r) = \sum_i \frac{P(r_i)}{|r_i - r|}$$

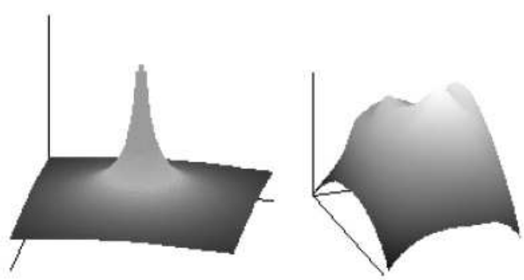

Figure 4: *"(a) Pixel's Potential Energy Function. (b) Pixel's Potential Energy Surface.*

Furthermore, a very important relation between the *vector force field* and the *scalar potential energy fields* is given by the equation:

$$F(r) = -grad(E(r) = - \bigtriangledown E(r)$$

2.4.2 The Invertible Linear Transformation

In order to show that the force field transformation is a *linear transformation*, it is sufficient to show that the force field transformation has a corresponding matrix representation, since linear transformations between finite-dimensional vector spaces are precisely those transformations that have matrix representations. So, for example for a trivial 2x2 pixel image, the matrix equation would be: $Ap = F$, where p = P_i and $d_{ij} = \frac{r_i - r}{|r_i - r|^3}$.

$$\begin{pmatrix} 0 & d_{01} & d_{02} & d_{03} \\ d_{10} & 0 & d_{12} & d_{13} \\ d_{20} & d_{21} & 0 & d_{23} \\ d_{30} & d_{31} & d_{32} & 0 \end{pmatrix} \begin{pmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{pmatrix} = \begin{pmatrix} F_0 \\ F_1 \\ F_2 \\ F_3 \end{pmatrix}$$

It is also possible, to construct a matrix representation corresponding to the potential energy transformation. It has been found, that the force field matrix is singular if the number of image pixels is odd, but that it is invertible if the number is even. The potential energy matrix would be invertible in either case. That means, that the original image is in principle recoverable from the potential energy surface, therefore all the information contained in the image is preserved in the transformation.

2.4.3 Appliance into ear images

In general, it is not possible to see a force field directly because it consists of vectors. As a result of that, it is converted to a scalar field by taking the magnitude of each vector. Then, a number of test pixels must be selected and arranged in an ellipse shaped array, around the ear sector. After that, the pixels are iterated to generate a set of field lines (Fig 5a & 5b).

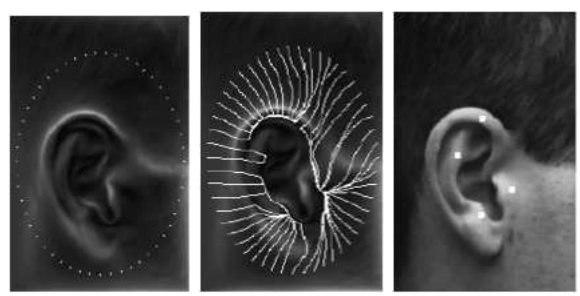

Figure 5: *"(a) Ellipse shaped array". (b) Field Lines. (c) Ear Signature.*

These field lines flow into potential channels and continue until they terminate in potential wells. The extraction of those potential wells will give a number of points that constitute the signature of the ear in the image (Fig: 5c).

Hurley et. al. achieved a recognition rate of 99.2% on a data set of 252 images (4 images x 63 individuals) and they demonstrated that this method has a very good noise tolerance and a remarkable invariance to initialization, scale and rotation.[11, 12]

2.5 Principal Component Analysis

Principal Components Analysis (PCA) has been one of the most popular techniques to ear and face recognition and is also known as "*eigenfaces*". In the past, a comparison between face and ear recognition has been made, using PCA. [13, 14]. This technique is a way of identifying patterns in data, and expressing the data in such a way that the similarities and differences are highlighted. PCA becomes a powerful tool for analyzing data, especially over data of high dimensions, where the graphical representation is not available, The main advantage of this method is that you are able to reduce the number of dimensions without loosing important information. This subsection, describes the basic steps of the method.

Principal Component Analysis basic steps:

- *Step 1:* The image is cropped and expressed in a vector.
- *Step 2:* The mean is subtracted (average across each dimension) from each of the data dimensions.
- *Step 3:* The covariance matrix is calculated. (App. A.1.3)
- *Step 4:* The eigenvectors and eigenvalues of the covariance matrix are calculated. (App. A.2.2 & A.2.1)
- *Step 5:* The eigenvector with the highest eigenvalue is the principle component of the data set. In order to create a final data set that has fewer dimensions, the components with the less significance (low eigenvalues) are ignored.
- *Step 6:* The final data set is created using the *feature vector*, which is just the chosen eigenvectors from Step 5.

In general, this method is easy to use and easy to implement. It produces excellent results, if the provided images are accurately registered and closely cropped to exclude any extraneous information. Nevertheless, PCA suffers from very poor invariance.

Victor et. al. used PCA to both face and ear recognition and conclude that the face demonstrated better performance than ear. [13] However, K. Chang et. al. conducted again an experiment and demonstrated that no significant difference was observed between face and ear biometrics, when using PCA technique. They suggested that the reason for the discrepancy, maybe is a result of the data set's low control over earrings, hair, and lighting. Furthermore, they conducted experiments using both face and ear biometric characteristics and they reported a recognition rate of 90.9% using this multimodal approach. [14]

2.6 3D Ear Shape Recognition

All the methods described until now, were using features from the ear's appearance in 2D intensity images. However, the last years a small number of researchers introduced different approaches, in which a 3D ear shape is used.[16][17] Three-dimensional data seem to offer more flexibility to problems that appear in two-dimensional data, such as pose and illumination. The most recent experiments that were conducted using this method, showed competitive performance for real biometric applications.

P.Yan and K.Bowyer, created a fully automated system for ear biometrics using 3D shapes. In this method, a scanner was used to capture ear's depth and color information. Then, a 3D shape recognition matching, based on Iterative Closest Point (ICP) (App. A.5), was used and both 2D and 3D data contributed in an automatic ear extraction. During the extraction, hair and earrings were separated from the ear's image and a curvature estimation was used to detect the ear pit.

The 4 steps to detect *ear pit* are:

- *Step 1:* Prepossessing and locating the nose. (results to a sector that includes the ear)
- *Step 2:* Using skin detection and isolating the ear region and the face from hair and clothes.
- *Step 3:* A surface curvature estimation is performed, in order to detect the pit regions depicted in black in the image.
- *Step 4:* A surface segmentation and classification is performed. Using a systematic voting method, a selection of the most likely pit region that corresponds to the ear pit is made.

The detected ear pit was then used to initialise an active contour algorithm to find the ear outlines. Both 2D color and 3D depth were used for the active contour algorithm. Experiments, showed that using only color or depth information is not powerful enough, since there are cases in which there is no clear colour or depth change around the ear contour. So, in order to improve the robustness of the algorithm a combination of both is needed. [7, 15, 17]

Yan and Bowyer, achieved a recognition rate of 97.8% on a time-lapsed data set of 1.386 images over 415 subjects for an identification scenario and a *Equal Error Rate (EER)* of 1.2% for a verification scenario. They also compared an ICP approach on a point-cloud representation of the 3D data, with a PCA-based approach on a range image representation of the 3D data and they found that there is a better performance using an ICP approach. The problem with using a range image representation of the 3D data is that landmark points must be selected ahead of time, to use for normalizing the pose and creating the range image. Therefore, errors or noise in this process can lead to recognition errors in the PCA or other algorithms that use the range image. [15]

G. Passalis et. al. developed also a novel 3D ear recognition method and performed an identification experiment where each of the 506 probe data sets of the database are compared with each of the 525 gallery data sets. They achieved a recognition rate of 94.4% and they observed that by decreasing the computational cost from 30 seconds to 15, they had a performance penalty approximately 1%, a tradeoff that could be desirable in a real application. [16]

2.7 Acoustic Ear Recognition

A. Akkermans et. al. introduced a new method that can be used for ear recognition. [18] This method used the acoustic properties of the pinna (the outer flap of the ear and the ear canal). They showed that the acoustic properties can be measured relatively easy with an inexpensive sensor and feature vectors can be derived with little effort. In their experiments, they used three different devices: (i). Headphone with microphones. (ii) Earphone pieces with microphones and (iii) A mobile telephone with extra microphone.

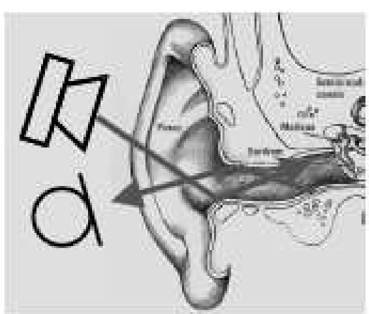

Figure 6: *"Determining the Acoustic Transfer Function"*.

In this method, the ear by virtue of its special shape behaves like a filter so that a sound signal played into the ear is returned in a modified form (Fig. 6). An ear signature is generated by probing the ear with a sound signal which is reflected and picked up by a small device. Then, the shape of the pinna and the ear canal determine the acoustic transfer function. This acoustic transfer function forms the basis of the acoustic ear signature. An obvious commercial use is that a small microphone might be incorporated into the earpiece of a mobile phone to receive the reflected sound signal and the existing loudspeaker could be used to generate the test signal.

Akkermans et. al. achieved *Equal Error Rates (EER)* in the order of 1.5% - 7%, depending on the application device that was used to do the measurement. They observed also, that headphones and earphones gave roughly the same performance resulting in an ERR of respectively 7% and 6%. Furthermore, they demonstrated that the worst-case EERs over all possible splits, were 8% for the headphones, 8.4% for the earphones and 15% for the mobile phone.

As a second experiment, they applied Fisher Linear Discriminant Analysis (LDA) (App. A.6) to the three ear databases, in order to select the most discriminating components amongst the subjects. They reported, that the results were significant better, as the new worst-case EERs they achieved, were 1.4% for the headphones, 1.9% for the earphones and 7.2% for the mobile phone. Finally, Akkermans et. al. investigated how the applied frequency range used in the excitation signal would influence the classification performance. After they conducted some experiments in different frequency ranges, that used also the LDA-technique, they concluded that a wider frequency range would give better classification results. [18]

3 Comparing Biometrics

In general, if we try to answer the question, "Which is the best biometric?", we are going to realize that there is no trivial answer. There is not one biometric modality that is proper for all biometric implementations. Many factors are involved and each must be considered separately, depending on the usage and the environment. Key factors for selecting a biometric technology include evaluating the environment, population size and demographics, security risks, task (identification or verification), ergonomics, interoperability with existing systems and user considerations. The careful evaluation of the key factors, plays a crucial role in the success of the selected technology.

Moreover, it is very important to note that biometric modalities are in varying stages of maturity. For example, fingerprint recognition has been used for over a century, while ear recognition is some decades old. It should be clear to the reader, that maturity may not be related to the best technology, but can always be an indicator of which technologies have more implementation experience. The effectiveness of a biometric characteristic is dependent on how and where it is used.

3.1 Ear & Other Biometrics

Having a closer look to each biometric modality, we are going to observe that every biometric has its own *strengths* and *weaknesses*, that should be taken into account before proceeding to an application.

Fingerprint, is easy to use with some training and the size of the capture system can be really small. On the other hand, ear may need more training, but the capture system can be a simple camera, that efficiently minimizes the cost. In comparison to ear, fingerprint is more effective in the large-scale systems and provides a better handling for the amounts of existing data. However, the latest methods on the ear recognition are really promising, as they have reported much better results on handling large datasets. Fingerprint is unique to each finger of each individual, and the ridge arrangement remains permanent during one's lifetime. Similarly, the ear has also a unique and permanent structure for an adult. An individual's age and occupation may cause some sensors difficulty in capturing a complete and accurate fingerprint image, a problem that is not occurring in ear images. It is important to mention, that ear is more difficult to be replicated than fingerprint (e.g. gummy fingerprints) and that sometimes fingerprint has negative public perceptions (e.g. it is related to criminal implications).

Face and *ear* are closely related, since they are both located in the human head and some of the methods used for their recognition are exactly the same. An extensive research has been performed comparing the performance of these two biometrics and the results reported no significant difference (section 2.5). Comparing ear with face, we observe a lot of similarities, but also some important differences. Both characteristics do not require any contact in order to be captured and the sensors that can be used, are commonly available (e.g. cameras). Moreover, almost the same drawbacks do exist concerning obstruction, as face view can be obstructed by hair, glasses, hats, make-up, scarves, etc. and ear view can be obstructed by hair and earrings. Nevertheless, ear is unaffected by facial expressions and it is not as sensitive as face, to changes such as lighting, expression and pose. Ear is located on the side of the head and the prediction of the background is easier than that of the face. The changes occurring in the face over time and the poor-quality of video images can lead to inaccurate results. However, face seems to handle better the large datasets in comparison to ear and it is a characteristic that can be easily verified by a human.

Iris, is another biometric characteristic that does not require any contact to be captured. It is a protected internal organ and thus compared to ear is less prone to injuries. Furthermore, iris is highly stable over lifetime and the uniqueness of eyes, even between the left and right eye of the same person, makes iris scanning more powerful for identification purposes than ear. Nevertheless, the problem of obstruction appears also to iris, as there is a difficulty on the capturing procedure for some individuals and this procedure can be easily obscured by eyelashes, eyelids, lenses, and reflections from the cornea. Compared to ear, acquisition of an iris image requires more training and attentiveness and a smaller capture distance is needed. Moreover, iris is impossible to be verified by a human and sometimes the capturing procedure causes anxiety to the public.

DNA is unique and permanent to each individual. However, some of the DNA properties, that do not appear to any other biometric, including ear, make it unpopular. The privacy concerns about the additional information of the individual that could be obtained (e.g. diseases), are becoming a serious obstacle to the DNA usage. In comparison to any other recognition system, the DNA recognition system involves real-time authentication capabilities and techniques with high computational resources, which are difficult to be automatized since they require some chemical processes. In comparison to ear, DNA seems more vulnerable to frauds, because it may be impossible to create a DNA copy, but it is always possible to steal a piece of DNA from an individual and use this information for fraudulent purposes.

Hand geometry, is a characteristic that is very easy for users and requires only to place your hand on the device. The amount of data required to uniquely identify a user in a system is very small and comparable to ear's. Hand geometry, is believed to be a highly stable pattern over the adult lifespan and its use requires training. However, it is intended only for verification purposes. In comparison to ear, the hand geometry recognition system needs a larger and more expensive hardware. Another obstacle is that hands are more vulnerable to injuries than ear and that can lead to lack of accuracy in the recognition system.

Vein patterns are unique to each individual and invariant to time, even in the case of identical twins. In comparison to ear, vein patterns can be used to high-security applications and they are more difficult to be copied, because they lie under the skin surface. Nevertheless, the main disadvantage of this technology, is that the infrared sensors are really expensive and large, especially compared to a common camera and therefore are not proper for mass deployment.

Retina, is unique to each individual and it is very difficult to be duplicated. It is a highly accurate biometric for identification and authentication and compared to the ear, it can be used wherever high security is the major concern. However, the retina recognition system is not as user friendly as the ear recognition system. A low-intensity light source through an optical coupler scans the unique patterns of the layer of blood vessels, while the user looks into a receptacle and focuses on a given point. Unfortunately, that makes retina technology very intrusive to the user.

Voice/Speaker enjoys the public acceptance and no contact is required in order to be captured. Voice recognition uses the acoustic features of speech, that have been found to differ between individuals. These acoustic patterns reflect both anatomy (e.g., size and shape of the throat and mouth) and learned behavioral patterns (e.g., voice pitch, speaking style). The sensors used for capturing, are commonly available sensors such as telephones and microphones. In general, voice compared to ear, is not sufficiently distinctive for identification over large databases and it changes through aging. Furthermore, it is difficult to control sensor and channel variances, something that has a significant impact to the capabilities of the voice recognition system.

3.2 Overview of the experiments

The results from the presented experiments in the field of ear recognition, are really promising. Most of the applied methods discussed, reported a *Recognition Rate (RR)* over 90%. (Fig. 7)

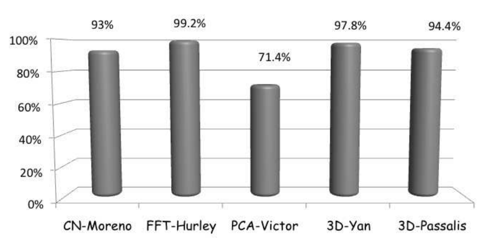

Figure 7: *"Recognition rates reported"*

More precisely, *compression networks* succeed a RR of 93% on a data set of 168 images and *force field transformation* achieved an excellent RR of 99,2% when was used on a data set of 252 images. Experiments that used *principal component analysis* reported a RR of 71.4%, they showed that the result is comparable with the 70.6% of face recognition and that a RR of 90.6% is possible, when both biometrics are used. Furthermore, both of the experiments that presented and used *3D-Shape* recognition method, achieved first-class rates. Yan's experiment gave a RR of 97.8% using a data set of 1.386 images over 415 subjects and Passalis's experiment denoted a RR of 94.4% on a data set of 1031 images over 525 subjects.

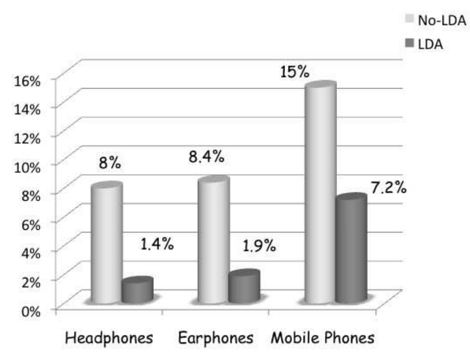

Figure 8: *"Acoustic Recognition - Worst reported EERs for the devices"*

A different approach, that used the ear acoustic was presented by Akkermans. The conducted experiment, stated a significantly lower *Equal Error Rate (EER)* when *Linear Discriminant Analysis (LDA)* was applied and also showed that headphones and earphones have almost the same performance, however mobile phones give a higher EER. In figure 8, the worst-case of EER is demonstrated for each of the used devices, with and without applying LDA.

3.3 Evaluation of ear methods

Ear methods at this point seem to be mature enough. Iannareli's manual system, showed the way and as it was expected modern and more efficient automatic recognition methods, followed.

Voronoi diagrams succeeded to avoid the problem of localizing anatomical points and the weakness of basing all feature measurements on a single point. However, this method did not solve the problem of occlusion by hair and earrings. The use of *thermograms* was proposed, but no further investigation was made and the technique remained unused. [8]

Neural networks and *Principal Components Analysis*, were already known to the research community for their use in other biometrics, such as face. However, these methods had never been applied for ear recognition, in the past. Compression network classifier (neural networks) demonstrated promising identification results, however when it was combined with facial features classifiers, it did not increase the identification rate, as the classifiers were completely independent. Principal Component Analysis was already a popular approach, used in face recognition. It produces very good results, but unfortunately the problem of poor invariance remains a major issue for this method. Two different studies were made, using PCA. The second one, showed that the performance for ear is comparable to facial biometrics and that the different results, that were obtained from the first one, were caused by differences in the quality of the images.

The *force field transformations* and the *3D-shape*, introduced completely different ways that can be used for automatic ear recognition. Force field transformation achieved remarkable results and reported features like robustness, reliability, invariance and excellent noise tolerance. Nevertheless, there are still obstacles concerning the computation load and the probable extension of the method to higher-dimensions. 3D shape recognition showed that the three-dimensions can handle better the problems of obstruction (hair,earrings) and they can give excellent results. The total performance of the technique, seemed ideal even for real life applications, however, the conducted experiments were made in a small range of data and thus more investigation is needed.

Acoustic recognition, demonstrated a totally different technique that is based on acoustic patterns and it takes advantage of the ear's acoustic features. The sensors used (low cost, public acceptance) and the promising results, may be an advantage for the adoption of this method. However, more tests and experiments are needed, in order to conclude about the efficiency of this technique in systems with larger datasets. Finally, it is probable, in the future, that acoustic recognition could be used also as a multimodal biometric and be related to *voice/speaker* biometric.

4 Conclusions

Although ear biometrics is a relatively new topic, researchers have already come up with various approaches for its use as an automated recognition system. Some of the techniques have already been used in the field of human recognition, while some other presented a whole new perspective. The evaluation of the upcoming methods, such as *force field transformation*, *3D-shape recognition* and *acoustic recognition*, offered a competitive performance and a promising efficiency. However, it is true that ear at the moment, does not have any commercial use. Ear has weaknesses, but the performed research is expected to improve the existing methods and discover new ones. All recent conducted experiments showed that ear biometric is already capable to be used in real life applications, nevertheless, the adoption of a new biometric is something that takes a lot of time. Thus, it is more probable, that ear will be introduced first as a supplementary biometric technique (e.g. face + ear) and then evolve through practice.

References

[1] *Moreno B., Sanchez A. and Velez J.F.,* "On the use of outer ear images for personal identification insecurity applications", Security Technology, International Carnahan Conference,(1999).

[2] *N.K. Ratha, A. Senior and R.M. Bolle,* "Automated Biometrics", Advances in Pattern Recognition - Second International Conference Rio de Janeiro, Brazil,(2001).

[3] *A. Jain, L. Hong and S. Pankati,* "Biometric Identification", Communications of the ACM,(2000).

[4] *Hanna-Kaisa Lammi,* "Ear Biometrics", Lappeenranta University of Technology,(2004).

[5] *Alphonse Bertillon,* "Signaletic Instructions including the theory and practice of anthropometrical identification", Gauthier-Villars, Paris,(1890).

[6] *M. Burge and W. Burger,* "Ear Biometrics - In A. Jain, R. Bolle and S. Pankanti, editors: "Biometrics: Personal Identification in a Networked Society", Kluwer Academic,(1998).

[7] *D.J. Hurley, B. Arbab-Zavar and M.S. Nixon,* "The Ear as a Biometric", University of Southampton,(2007).

[8] *M. Burge and W. Burger,* "Ear Biometrics in Computer Vision", IEEE ICRP,(2000).

[9] *K. Lai and R. Chin,* "Deformable Contours: Modeling and extraction", PAMI,(1995).

[10] *B. Moreno, A. Sanchez and J.F. Velez,* "On the Use of Outer Ear Images for Personal Identification in Security Applications", IEEE,(1999).

[11] *D.J. Hurley, M.S. Nixon and J.N.B. Carter,* "Automatic Ear Recognition by Force Field Transformations", IEE Colloquium on Biometrics,(2000).

[12] *D.J. Hurley, M.S. Nixon and J.N.B. Carter,* "Force Field Energy Functionals for Image Feature Extraction", Image and Vision Computing Journal, vol. 20, no. 5-6, 311-318, (2002).

[13] *B. Victor, K. Bowyer and S. Sarkar,* "An Evaluation of Face and Ear Biometrics", In Proceedings of International Conference on Pattern Recognition, pages 429-432, (2002).

[14] *K. Chang, K. Bowyer, S. Sarkar and B. Victor* "Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, No 9, (2003).

[15] *P. Yan and K. Bowyer* "Biometric Recognition Using 3D Ear Shape", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, No 9, (2007).

[16] *G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici and T. Papaioannou* "Towards Fast 3D Ear Recognition for Real-Life Biometric Applications", AVSS IEEE Conference, (2007).

[17] *P. Yan and K. Bowyer* "Empirical Evaluation of Advanced Ear Biometrics", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2005).

[18] *A.H.M. Akkermans, T.A.M. Kevenaar and D.W.E. Schobben,* "Acoustic Ear Recognition for Person Identification", Philips Research, Proc AutoID, (2005).

Appendices

A Mathematical Background

A.1 Statistics

Statistics are based around the idea that you have a big set of data, and you want to analyze, that set in terms of the relationships between the individual points in that data set. In the following subsections, some of the needed measures for this paper, are presented.

A.1.1 Mean of Sample & Standard Deviation

The Standard Deviation is *the average distance from the mean of the data set to a point.*

Let X, be the following data set,

X = [1 2 4 6 12 15 25 45 68 67 65 98] ,

Suppose also that X_i refers to the number in the i position and n to the number of elements in the data set X, then

The *mean of sample* is given by : $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$.

The mean doesnt tell us a lot about the data except for a sort of middle point. For example, the data sets [0 8 12 20] and [8 9 11 12], have exactly the same mean (10), but are obviously quite different. It is the *spread* or else the *Standard Deviation* of the data that is different.

The *standard deviation* is given by : $s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}}$.

The standard deviation for a set with the same data, like [10, 10 , 10, 10], is 0, but the mean is equal to 10.

A.1.2 Variance

Variance is another measure of the spread of data in a data set. In fact, it is almost identical to the standard deviation.

The *variance* is given by: $s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}$.

It is simply the standard deviation squared. Standard deviation is the most common measure, but variance is also used.

A.1.3 Covariance

Standard deviation and variance only operate on 1 dimension, so that you could only calculate the standard deviation for each dimension of the data set independently of the other dimensions. However, it is useful to have a similar measure to find out how much the dimensions vary from the mean with respect to each other. Covariance is such a measure. Covariance is always measured between 2 dimensions. If you calculate the covariance between one dimension and itself, you get the variance. So, if you had a 3-dimensional data set (x,y,z), then you could measure the covariance between the x and y dimensions, the x and z dimensions, and the y and z dimensions. Measuring the covariance between x and x, or y and y and z and z would give you the variance of the x,y and z dimensions respectively.

The *covariance* is given by: $cov(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$.

The exact value of covariance is not as important as its sign. If the value is *positive*, then that indicates that both dimensions increase together. If the value is negative, then as one dimension increases, the other decreases. And in the last case, if the covariance is zero, it indicates that the two dimensions are independent of each other. Also, cov(X,Y) equal to cov(Y,X).

A.2 Matrix Algebra

This section serves to provide a background for the matrix algebra required for the some of the methods described in section 2 of this paper. Nevertheless, a basic knowledge of matrices is assumed.

A.2.1 Eigenvectors & Eigenvalues

An *eigenvector* is a vector that is scaled by a linear transformation, but not moved. Think of an eigenvector as an arrow whose direction is not changed. It may stretch, or shrink, as space is transformed, but it continues to point in the same direction. The scaling factor of an eigenvector is called its *eigenvalue*. An eigenvalue only makes sense in the context of an eigenvector.

Let A be an n x n matrix. The number λ is an *eigenvalue* of A, if there exists a non-zero vector v such that, $Av = \lambda v$. In this case vector v is called eigenvector of A corresponding to λ. In order to compute the eigenvector and eigenvalues, we can rewrite the condition $Av = \lambda v$ as $(A - \lambda I)v = 0$, where I is the n x n identity matrix. In order for a non-zero vector v to satisfy this equation, $(A - \lambda I)$ must not be invertible. That is the determinant of $(A - \lambda I)$ must equal 0. We call p(λ) = det$(A - \lambda I)$ the characteristic polynomial of A. The eigenvalues of A are simply the roots of the characteristic polynomial of A.

A.2.2 Covariance Matrix

When we have a n-dimensial data set you can calculate $\frac{n!}{(n-2)*2}$. different covariances values.

A useful way to get all the possible covariance values, between all the different dimensions, is to calculate them all and put them in a matrix. Then, the matrix is called *covariance matrix*.
The definition for the covariance matrix for a set of data with n dimensions is :

$$C^{nxn} = (c_{i,j}, c_{i,j}, = cov(Dim_i, Dim_j))$$

where C^{nxn} is a matrix with n rows and n columns, and Dim_x is xth dimension.

For example, a 3-dimensional covariance matrix would be :

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

Notice, that down the main diagonal the covariance value is between one of the dimensions and itself and that the matrix is symmetrical about the main diagonal.

A.3 Voronoi Diagrams

In short, a *voronoi diagram* records information about the distances between sets of points in any dimensional space. For path planning, voronoi tends to be used in two dimensional space, where sets of points all lie within a plane. There are many approaches to constructing Voronoi diagram, but some methods are more efficient in terms of time than others.

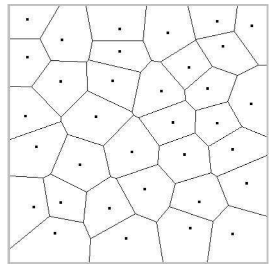

Figure 9: *"Voronoi Diagrams*

As seen from the Fig. 9 above, a plane is divided into cells so that each cell contains exactly one site. For every point in the cell, the Euclidean distance of the point to the site within the cell, must be smaller than the distance of that point to any other site in the plane. If this rule is followed across the entire plane, then the boundaries of the cells, known as *Voronoi edges*, will represent points equidistance from the nearest 2 sites. The point where multiple boundaries meet, is called a *voronoi vertex* and is equidistance from its 3 nearest sites.

For a further study the interested reader can refer to: *Franz Aurenhammer, Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure. ACM Computing Surveys, 23(3), 345-405, (1991).*

A.4 Canny Operator

The Canny operator was designed to be an optimal edge detector (according to particular criteria). It takes as input a gray scale image, and produces as output an image showing the positions of tracked intensity discontinuities. It works in a multi-stage process. First of all the image is smoothed by Gaussian convolution (is used to 'blur' images and remove detail and noise). Then a simple 2-D first derivative operator is applied to the smoothed image to highlight regions of the image with high first spatial derivatives. Edges give rise to ridges in the gradient magnitude image. The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output, a process known as non-maximal suppression. The tracking process exhibits hysteresis controlled by two thresholds: T1 and T2, with T1 > T2. Tracking can only begin at a point on a ridge higher than T1. Tracking then continues in both directions out from that point until the height of the ridge falls below T2. This hysteresis helps to ensure that noisy edges are not broken up into multiple edge fragments.

A.5 Iterative Closest Point

Iterative Closest Point (ICP) is an algorithm employed to match point-sets. This matching is used to reconstruct 3D surfaces from different scans and achieve optimal path planning. The algorithm is very simple and accurate and is widely used for 3D shape matching, however it is computationally expensive. It iteratively estimates the transformation (translation, rotation) between two raw scans. It takes as an input the two raw scans, an initial estimation of the transformation and the criteria for stopping the iteration and it outputs a refined transformation.

Briefly, the ICP's algorithm steps are :

- *Step 1:* Associate points by the nearest neighbor criteria.
- *Step 2:* Estimate the parameters using a mean square cost function.
- *Step 3:* Transform the points using the estimated parameters.
- *Step 4:* Re-associate the points and so on (iterate).

For a further study the interested reader can refer to: *P. J. Besl, N. D. McKay, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 239-256, (1992).*

A.6 Fisher Linear Discriminant Analysis (LDA)

Fisher Linear Discriminant Analysis (LDA) is a method used in statistics and machine learning to find the linear combination of features which best separate two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification. LDA is closely related to *Principal Component Analysis (PCA)* and *Factor Analysis* in that both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities.

For a further study the interested reader can refer to: *Mika, S. et al. Fisher Discriminant Analysis with Kernels. IEEE Conference on Neural Networks for Signal Processing IX, (1999).*