Proceedings of 2017 4th International Conference on
Control, Decision and Information Technologies (CoDIT'17) / April 5-7, 2017, Barcelona, Spain

# Ear Recognition Using Local Color Texture Descriptors From One Sample Image Per Person

**Amir Benzaoui [a,b] and Abdelhani Boukrouche [b]**
[a] Department of Electrical Engineering, Faculty of Science and Applied Sciences
University of Akli Muhand Oulhadj - Bouira (Algeria)
[b] Laboratory of Inverse Problems, Modeling, Information and Systems (PIMIS)
Department of Electronics and Telecommunications, Faculty of Sciences and Technology
University of May 08th 1945 - Guelma (Algeria)
amirbenzaoui@gmail.com hani.boukrouche@gmail.com

***Abstract*— Morphological shape of the human ear presents a rich and stable information embedded on the curved 3D surface, which has invited lot attention from the forensic and engineer scientists in order to differentiate and recognize people. However, recognizing identity from morphological shape of the human ear using one sample image per person in training-set, with insufficient and incomplete training data, dealing with strong person-specificity can be very challenging. To address such problem, we propose a simple yet effective approach which uses and exploits local color texture descriptors in order to achieve faster and more accurate results. Support Vector Machine (SVM) is used as a classifier. We experiment with USTB-1 database consisting of several RGB ear benchmarks of different natures taken under varying conditions and imaging qualities. The experiments show excellent results beyond the state-of-the-art.**

***Keywords—Biometrics; Ear Recognition; Local Color Texture Descriptors; Local Binary Patterns (LBP).***

## I. INTRODUCTION

Biometrics is the science which makes it possible to recognize the identity of a person on the basis of her physiological, chemical, or behavioral characteristics, such as: face, iris, fingerprint, odor, DNA, gait, or electronic signature [1-2]. Research in the field of facial recognition is motivated not only by the fundamental challenges of this problem but also by the important numbers of practical applications where the human identification is necessary. Face recognition, as one of the pilot biometric technologies, has become increasingly significant due to the fast progress in technologies, such as: internet, smart phones, and digital cameras; as well as the augmentation in the requirement of security. Face recognition has several advantages compared to other modalities: it is natural, non-intrusive, and requires less cooperation. However, the conditions of acquisition, such as: position of the face to the camera, lighting, resolution of the camera, partial occlusion, as well as natural aging, facial expressions, disguises, and spoofing attacks cause several changes in the facial appearance; in real world applications, those challenges affect negatively on the performances of recognition [3-4]. In the other hand, the human ear is considered as a new viable class of biometrics with some additional advantages compared to the face. Indeed, the ear is rich in terms of characteristics; it has a stable structure which does not changes significantly during aging and its form does not vary by facial expressions, it can be also captured from a distance and without any cooperation of the user, but sometimes it can be hidden by hair, headphones, scarf, or loops.

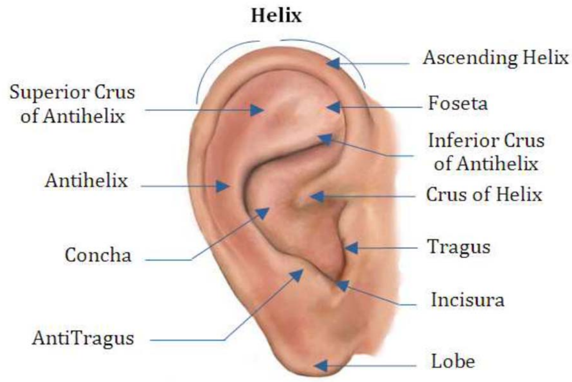


Fig. 1. Anatomy of the Human Ear [7].

The use of the human ear in forensic and biometric applications has become a quite interesting way in the last years. It is considered as a new class of biometrics which is not yet used in real context or in commercial applications. The external shape of the human ear is characterized by a rich structure which provides sufficient and important information to differentiate and recognize people; we can visualize 10 features and 37 sub-features from 2D ear imaging. The terminology of the human ear is presented in Fig. 1; this terminology is made up of standard features. It includes an outer rim (Helix) and ridges (Antihelix) parallel to the helix, the concha (hollow part of the ear), the lobe and the tragus (small prominence of cartilage) [5-7]. The human ear has several advantages compared to other modalities: it has a rich structure, smaller object (small resolution), stable over time, modality accepted by people, the acquisition of the ear imaging can be affected without participation of the subject and can be captured from distance, and not affected by changes in age, facial expression, position, and rotation.

An ear recognition system can be divided into three main steps: ear normalization, feature extraction, and classification. In the normalization step, the ear image must be normalized to standard size and direction according to the long axis of outer ear contour (Fig.2). The long axis was defined as the line crossing through the two points which have the longest distance on the ear contour [7]. After normalization, the long axes of different ear images were normalized to the same length and same direction. The next steps are to represent the ear by appropriate features and design effective classifier. Most researches on ear biometrics are focused on feature extraction and classification. In real applications, however, the big challenge of ear recognition is how to find an efficient descriptor to represent and to model the human ear in a real context where the ear can be affected by illumination variation, pose variation, noise, or occlusion.

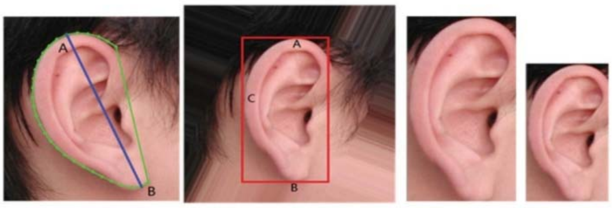

Fig. 2. Example of ear normalization to standard size and direction [7]. **(a)** long axis detection, **(b)** rotation, **(d)** cropping, and **(e)** resize.

The objective of our research work is simply to find a suitable and effective approach to represent and to model the human ear in a real context, in other words, we search to improve the performances of recognition in the unconstrained conditions by only using one sample image per person in training-set. The rest of the paper is organized as follows: in the next section, we describe some related work. In Section 3, our proposed approach is presented. In Section 4, we present our experimental results by applying the proposed approach on USTB-1 database. Finally, a conclusion related to this work is given in Section 5.

## II. Related Work

The ability to identify people by using the shape of the outer ear was discovered for the first time by the French criminologist *Alphonse Bertillon* (1890) [8] and confirmed by the US police officer *Alfred Iannarelli* (1949) [9] who proposed the first automatic system of ear identification, based on only seven characteristics. In our day, the majority of biometric systems based on the ear modality consist in extracting a set of features and comparing these features with the templates stored in the database. Based on this principle, previous research into the problem of ear recognition can be divided into three classes: local approaches, statistical (holistic or global) approaches, and hybrid approaches.

Algorithms based on local correspondences seek to find a specific landmark points in the image, and then extract the features around these landmarks. Local features are more robust than holistic approaches in uncontrolled conditions. Despite the capacity of these algorithms to tackle the problem of variations in pose and size, they have difficulty in finding a precise and exact method to locate the landmark points. The *Force Field Transform* (*Hurley et al.* (2002) [10] and *Abdel-Mottaleb and Zhou* (2006) [11]), The *Scale Invariant Feature Transform* (SIFT) (*Bustard and Nixon* (2010) [12-13]), the *Speeded Up Robust Features* (SURF) (*Prakash and Gupta* (2013) [14]) are considered as the most well-known methods of this approach.

Statistical approaches are based on pixel information; all the pixels of the image are treated as a single vector. In fact, the total number of pixels represents the size of the vector. Most methods in this approach use another representation space (sub-space) in order to reduce the number of pixels and to eliminate redundancies. Principal Component Analysis (PCA) (*Victor et al.* (2002) [15], *Chang et al.* (2003) [16], and *Alaraj et al.* (2010) [17]), Independent Component Analysis (ICA) (*Zhang and Mu* (2008) [18]), Sparse Representation (*Naseem et al.* (2008) [19]) are the most popular subspace methods used for the problem of ear recognition.

Hybrid approaches make it possible to associate the advantages of the methods based on local landmarks and statistical transformations by combining the detection of holistic features with the extraction of the local characteristics of its appearance. This increases the stability of the recognition performances under different changes in pose, size, and illumination. *Lu et al.* (2006) [20], *Yuan and Mu* (2007) [21], *Jeges and Mate* (2007) [22], as well as *Zhang et al.* (2014) [23] are the most well-known hybrid ear recognition methods.

## III. Proposed Approach

In our previous work on ear recognition [24-25], we tested and compared three recent gray local texture descriptors, namely: Local Binary Pattern (LBP) [26], Local Phase Quantization (LPQ) [27], and Binarized Statistical Image Features (BSIF) [28]. These methods can easily derive an effective feature model which combines the global form of the analyzed object and the local texture of its appearance in a single feature vector; it imitates the capacities of the human being in the recognition of objects or faces, by its ability to codify low and high frequency components which contribute to local and global information, respectively. With this type of descriptor, the entire image is scanned pixel by pixel, providing local information, and the co-occurrences of the texture descriptor are accumulated in a discrete histogram, providing global information. In addition, these approaches codify and collect in a histogram the co-occurrence of the micro-features. They are characterized by a very high discriminative power, simplicity of calculation, and invariance to any monotonic changes in gray level.

### 3.1. Local Texture Descriptors

The **Local Binary Patterns (LBP)** is a texture analysis operator defined as a gray-scale invariant texture measure and derived from a general definition of texture in a local neighborhood. It is a powerful mean of texture description and among its properties in real-world applications are its discriminative power, computational simplicity, and tolerance against monotonic gray-scale changes. In the original LBP operator, the local patterns are extracted by thresholding the 3×3 neighborhood of the eight neighbors of each pixel, from the original image, with the central value. All the neighbors are assigned the value 1 if they are greater than or equal to the current element and 0 otherwise, which represents a binary code of the central element. This binary code is converted into a decimal value by multiplying it with the given corresponding weights and is summed to obtain the LBP code for the central value. The histogram of these $2^8 = 256$ different labels can then be used as a texture descriptor for further analysis. Recently, the LBP operator has been extended to use neighborhoods of different sizes in order to capture the large-scale structures which can be considered as the dominant patterns in the image.

The **Local Phase Quantization (LPQ)** was proposed in order to tackle the relative sensitivity of LBP to blur, based on quantizing the Fourier transform phase in local neighborhoods. The phase can be shown to be a blur invariant property under certain commonly fulfilled conditions. In texture analysis, histograms of LPQ labels computed within local regions are used as a texture descriptor similarly to the LBP methodology. The LPQ descriptor has recently received wide interest in blur-invariant texture recognition. LPQ is insensitive to image blurring, and it is has proven to be very efficient descriptor in pattern recognition from blurred as well as from sharp images.

The **Binarized Statistical Images Features (BSIF)** was recently proposed for texture classification. Inspired by LBP and LPQ, the idea behind BSIF is to automatically learn a fixed set of filters from a small set of natural images, instead of using hand-crafted filters such as LBP and LPQ. BSIF applies learning, instead of manual tuning, to obtain statistically meaningful representation of the images, which enables efficient information encoding using simple element-wise quantization. Learning provides also an easy and flexible way to adjust the descriptor length and to adapt applications with unusual image characteristics. To characterize the texture properties within each image sub-region, the histograms of pixels' BSIF code values are then used. The value of each element (i.e. bit) in the BSIF binary code string is computed by binarizing the response of a linear filter with a threshold at zero. Each bit is associated with a different filter and the desired length of the bit string determines the number of filters used. The set of filters is learnt from a training set of natural image patches by maximizing the statistical independence of the filter responses.

### 3.2. Color Spaces & Color Texture Analysis

Inspired by the aforementioned observations, we propose, in this work, a new recognition method based on local color texture analysis. For example, the color Local Binary Patterns (LBP) descriptor proposed by Young et al. (2010) [29] can be used to extract the joint color-texture information from the ear images. In this descriptor, the LBP histograms are extracted from the individual image bands. Subsequently, these histograms are concatenated to form the final descriptor. To gain insight into which color space is more discriminative to distinguish between ear images, we considered, tested, and compared three color spaces, namely RGB, HSV, and YCbCr.

The RGB (Red, Green, and Blue) is the most used color space in machine and artificial vision. However, its application in image analysis is quite limited due to the high dependence between its three color components (red, green and blue) and the limited separation of the luminance and chrominance information.

In our work, we also tested two other color spaces to explore the color texture information RGB, namely: the HSV and the YCbCr. Both of these color spaces are based on the separation of the luminance and the chrominance information. In the HSV color space, the hue and the saturation dimensions define the chrominance of the image while the value dimension corresponds to the luminance. The YCbCr space separates the RGB components into luminance (Y), chrominance blue (Cb) and chrominance red (Cr).

### 3.3. Application of Local Color Texture Analysis on Ear Biometric Recognition

For a given ear image, we propose the following representation. Firstly, we carry out a *pretreatment* on the original image by applying the median filter and normalization of the histograms of each color component. The objective of the pretreatment is to prepare the representation of the original image in order to facilitate the task of the following modules and to improve the identification performances. Next, the texture descriptor (LBP, 1DLBP, LPQ, or BSIF) is *applied* on each component (ex., for RGB color space, the descriptor is applied separately of the R, G, and B components); consequently, a histogram is extracted from each color component. Finally, the histograms extracted from each color component are *concatenated* in a global histogram of characteristics which represents the ear image.

In the training-set, the vector of characteristics is stored in the database. In the testing-set, the SVM classifier is used to compare this feature vector with the biometric templates stored in the database. The general diagram of the proposed biometric systems is presented as shown in following figure.

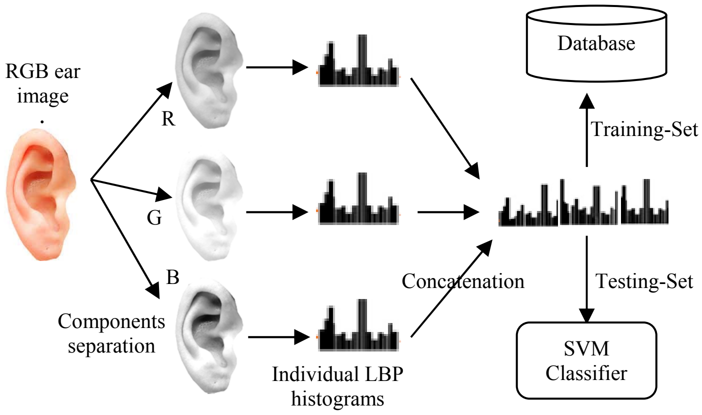


Fig. 3. General diagram of the proposed system.

## IV. EXPERIMENTAL ANALYSIS

To evaluate the performances of the proposed system, we carried out a series of experiments using the local texture descriptors detailed in the previous section, namely: LBP, LPQ, and BSIF, as well as the discussed clolor spaces, namely: RGB, HSV, and YCbCr, in order to extract discriminating features from a set of ear images from the USTB-1 [30] database. The database and the protocol used to evaluate its performance are discussed in the following sub-section.

### 4.1. Dataset & Settings

The first version of the USTB database [30] consists of ear images collected by UST Beijing University. The images of this database were taken under different conditions of lighting and rotation. It contains 185 images of 60 subjects, with at least three images of each subject. This database is delivered with automatically normalized and cropped ear images of 80×150 pixels in size.

The experimental protocol uses only one ear image per person in the training-set and the remaining ear images of the same person are used in the testing-set. As the majority of the subjects in the databases have three images, we carried out three permutations and reported the average rank-1 recognition rate.

### 4.2. Experiments

In the first experiment, we tested the local texture descriptors detailed in the previous section on the gray level space using a single sample image per person in the training-set. The results are presented in Table.1. In fact, each descriptor is characterized by its own parameters: the LBP has two parameters ($P,R$) which represent the sampling points ($P$) on a circle of radius ($R$), the LPQ is defined by the radius of the filter ($P$), and the BSIF depends on the filter size $l$ and the length $n$ of the bit string. In our previous work [24-25], the performances of each descriptor with all possible instances of parameters were measured, tested, and compared in order to find the best texture descriptor and the parameters that give the best recognition performances. A detailed discussion of these tests and experiments can be found in [24-25].

Table 1. Best rank-1 recognition rates of different local texture descriptors applied on the gray level using a single ear image per person in the training-set.

| | **LBP** | **LPQ** | **BSIF** |
|---|---|---|---|
| **USTB-1** | 81.07 | 81.60 | **<u>96</u>** |

It is clear from this table that BSIF descriptor outperforms all other descriptors (LBP and LPQ) in terms of recognition rates. The best results of the LBP descriptor were achieved with $(P,R) = (8,3)$, the best results of the LPQ descriptor were achieved with $P = 03$, while the best results of the BSIF descriptor were achieved with $l = 15 \times 15$ and $n = 12$ (for more details on the parameters of each descriptor, see [24-25]).

Starting from the best configuration of the previous experiment (BSIF descriptor with $= 15 \times 15$ and $n = 12$), we test the performances of recognition by the application of the BSIF descriptor on the components of each color space and compare the performances of these color spaces against the gray level space. These results are presented in Table.2.

Table 2. Best rank-1 recognition rates of different color spaces and BSIF descriptor applied on the color components using a single ear image per person in the training-set.

| | Gray | RGB | HSV | YCbYc |
|---|---|---|---|---|
| **USTB-1 With BSIF descriptor** | 96 | **96.53** | 93.33 | 88.8 |

It can be seen from Table.2 that the results of the gray level exceed the results of the HSV and YCbCr color spaces, while the RGB color space presents more improvements and very interesting results in terms of recognition in comparison to all color spaces and the results achieved in our previous work. It should be noted that most existing approaches in the literature employ two ear (sample) images per person in the training-set, but in real world applications (e.g., ID cards or passports), only one image per person is used in the training phase i.e. only one model of the individual to be identified is registered in the database and available for the task of identification. Our results demonstrate that our approach is reliable and can be implemented in real world applications.

## V. Conclusion

In our investigation, we have successfully implemented a feature extraction approach for automated 2D ear description and recognition using only one sample person in training-set; this problem is very challenging in real world-application. We introduced the use of the local color texture descriptors, which are inspired from the statistics of natural images and produce binary codes. A series of experimental evaluations on the USTB-1 database shows that this implemented approach of feature extraction, based on color RGB BSIF descriptor, given very significant improvements at the recognition rates.

## References

[1] J.A. Unar, W.C. Seng, and A. Abbasi, “A Review of biometric technology along with trends and prospects”. Pattern Recognition (Elsevier). Vol.47(8), pp.2673-2688 (2014).

[2] J. Ashboum, “Biometrics in the new world: The cloud, mobile technology, and pervasive identity”. Springer Publishing Company (2014).

[3] I. Chingovska, N. Erdogmus, A. Anjos, and S. Marcel, “Face recognition systems under spoofing attacks”. Springer International Publishing (2016).

[4] J. Anil and L.P. Suresh, “Literature survey on face and facial expression recognition”. In proceedings of the International IEEE Conference on Circuits, Power and Computing Technologies (ICCPCT), 2016.

[5] A.D. Dinkar and S.S. Sambyal, “Person identification in ethnic Indian goans using ear biometrics and neural networks”. Forensic Science International (Elsevier). Vol.223(1-3), 373.e1-e13, (2012).

[6] H. Chen and B. Bhanu, “Human ear recognition in 3D”. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.29(4), 718-737, (2007).

[7] A. Pflug and C. Busch, “Ear biometrics: a survey of detection, feature extraction and recognition”. IET Biometrics. Vol.1(2), 114-129, (2012).

[8] A. Bertillon, “La photographie Judiciaire, avec un appendice sur la classification anthropométrique”. Technical Report, Gauthier-Villars, Paris, 1980.

[9] A. Iannarelli, “Ear Identification”. Forensic Identification Series, Paramount Publishing Company, Fremont, California, 1989.

[10] D.J. Hurley, M.S. Nixon, and J.N. Carter, “Force Field Energy Functional for Image Feature Extraction”. Image and Vision Computing (Elsevier). Vol.20, No.5-6, 311-317, 2002.

[11] M. Abdel-Mottaleb and J. Zhou, “Human Ear Recognition from Face Profile Images. In Proceedings of the International Conference on Biometrics (ICB). Lecture Notes in Computer Science (Springer). Vol.3822, 786-792, 2006.

[12] D.G. Lowe: “Object Recognition from Local Scale-Invariant Features”. In Proceedings of the International IEEE Conference on Computer Vision (ICCV). Vol.02, 1150-1157, 1999.

[13] J.D. Bustard and M.S. Nixon: “Toward Unconstrained Ear Recognition from Two-Dimensional Images”. IEEE Transactions on Systems, Man, and Cybernetics (Part A: Systems and Humans). Special Issue on Recent Advances in Biometrics. Vol.40(3), 486-494, 2010.

[14] S. Prakash and P. Gupta, “An Efficient Ear Recognition Technique Invariant to Illumination and Pose”. Telecommunication Systems (Springer). Vol.52(3), 1435-1448, 2013.

[15] B. Victor, K. Bowyer, and S. Sarkar: “An Evaluation of Face and Ear Biometrics”. In Proceedings of the 16th International IEEE Conference on Pattern Recognition (ICPR). Vol.01, 429-432, 2002.

[16] K. Chang, K.W. Browyer, S. Sarkar, and B. Vector: “Comparison and Combination of Ear and Face Images in Appearance-based Biometrics”. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). Vol.25(9), 1160-1165, 2003.

[17] M. Alaraj, J. Hou, and T. Fukami: “A Neural Network Based Human Identification Framework using Ear Images”. In Proceedings of the IEEE TENCON Conference. 1595-1600, 2010.

[18] H. Zhang and Z. Mu: “Compound Structure Classifier System for Ear Recognition”. In Proceedings of the IEEE International Conference on Automation and Logistics (ICAL). 2306-2309,2008.

[19] I. Naseem, R. Togneri, and M. Bennamoun: “Sparse Representation for Ear Biometrics”. In Proceedings of the 4th International Symposium on Visual Computing (ISVC). Lecture Notes in Computer Science (Springer). Vol.5359, 336-345, 2008.

[20] L. Lu, Z. Xiaoxun, Z. Youdong, and J. Yunde, "Ear Recognition based on Statistical Shape Model". In Proceedings of the 1st International IEEE Conference on Innovative Computing, Information, and Control (ICICIC). Vol.03, 353-356, 2006.

[21] L. Yuan and Z.C. Mu, "Ear Recognition based on 2D Images". In Proceedings of the 1st International IEEE Conference on Biometrics: Theory, Applications, and Systems (BTAS). 1-5, 2007.

[22] E. Jeges and L. Mate: "Model-Based Human Ear Localization and Feature Extraction". International Journal of Intelligent Computing in Medical Science & Image Processing (Taylor & Francis). Vol.01(2), 101-112, 2007.

[23] B. Zhan, Z. Mu, H. Zeng, and S. Luo: "Robust Ear Recognition Via Nonnegative Sparse Representation of Gabor Orientation Information". The Scientific World Journal (Hindawi). 2014.

[24] A. Benzaoui, A. Hadid, and, A. Boukrouche, "Ear Biometric Recognition using Local Texture Descriptors". Journal of Electronic Imaging (JEI-SPIE). Vol.23(5), 053008, 2014.

[25] A. Benzaoui, N. Hezil, and, A. Boukrouche, "Identity recognition based on the externel shape of the human ear". In Proceedings of the International IEEE Conference on Applied Research in Computer Science and Engineering (ICAR). 2015.

[26] A. Benzaoui and A. Boukrouche, "Face recognition using 1DLBP texture analysis". In Proceedings of the 5th International IARIA Conference on Computational Technologies and Applications (Future Computing). 14-19, 2013.

[27] V. Ojansivu and J. Heikkil: "Blur Insensitive Texture Classification using Local Phase Quantization". In Proceedings of the 3rd International Conference on Image and Signal (ICSIP). Lecture Notes in Computer Science (Springer). Pp.236-243, 2008.

[28] J. Kannala and E. Rahtu: "BSIF: Binarized Statistical Image Features". In Proceedings of the International IEEE Conference on Pattern Recognition (ICPR). Pp.1363-1366, 2012.

[29] J.C. Young, K.N. Plataniotis, and Y.M. Ro, "Using color local binary pattern features for face recognition," in Proceedings of the IEEE International Conference on Image Processing (ICIP). 4541–4544, 2010.

[30] Z. Mu: "USTB Ear Image Database" Beijing, China, 2009, http://www1.ustb.edu.cn/resb/en/index.htm.

[31] A. Benzaoui, I. Adjabi, and, A. Boukrouche, "Person Identification based on Ear Morphology". In Proceedings of the International IEEE Conference on Advances Aspects of Software Engineering (ICAASE). 2016.