

UBEAR: A Dataset of Ear Images Captured On-the-move in Uncontrolled Conditions

Rui Raposo

University of Beira Interior
Covilhã, Portugal
m3642@ubi.pt

Edmundo Hoyle

University of Beira Interior
Covilhã, Portugal
edhoyle@ubi.pt

Adolfo Peixinho

University of Beira Interior
Covilhã, Portugal
m4067@ubi.pt

Hugo Proença

University of Beira Interior
IT-Instituto de Telecomunicações
Covilhã, Portugal
hugomcp@di.ubi.pt

Abstract—In order to broad the applicability of biometric systems, the data acquisition constraints required for reliable recognition are receiving increasing attention. For some of the traits (e.g., face and iris) significant research efforts were already made toward the development of systems able to operate in completely unconstrained conditions. For other traits (e.g., the ear) no similar efforts are known. The main purpose of this paper is to announce the availability of a new data set of ear images, which main distinguishing feature is that its images were acquired from on-the-move subjects, under varying lighting conditions and without demanding to subjects any particular care regarding ear occlusions and poses. The data set is freely available to the research community and should constitute a valuable tool in assessing the possibility of performing reliable ear biometric recognition in such d challenging conditions.

I. INTRODUCTION

Due to increasing concerns about safety and security in the modern societies, the use of biometric systems has been encouraged by both governmental and private entities to replace or improve effectiveness of the traditional human recognition systems. Several traits were already acknowledged as possessing the key features of a biometric trait: universality (ability to be collected in as much subjects as possible), collectibility (easiness in performing data acquisition), distinctiveness (high variability between different subjects) and stability (low variability over a single subject in human lifetime). For these traits (e.g., fingerprint, iris, face, retina, palm vein...) several recognition systems were already deployed and operate with remarkable success.

Among other traits that are still in embryonal development stages, the human ear is presently accepted as a promising biometric trait: two studies conducted by Iannarelli [1] provide substantial evidence of the distinctiveness of the ear biometric trait. The first study compared over 10 000 ears drawn from a randomly selected sample in California and the second examined fraternal and identical twins, in which the appearance of most physiological features appears to be similar. These studies support the hypothesis that each ear contains unique physiological features: all examined ears were found to be unique though identical twins were found to have similar ear structures. When compared to other biometric traits, the ear has several major advantages:

- its structure does not change over lifetime, from the birth into mature age;

- its surface is relatively small, allowing systems to cope with reduced spatial resolution images;
- it has a uniform color distribution;
- its appearance does not change according to different facial expressions.

Ears have played a significant role in forensic sciences for many years. In 1949, Iannarelli created his anthropometric identification technique based upon ear biometrics, based in twelve measurements illustrated in figure 1. The identification process relies in these twelve measures plus gender and ethnicity information. In order to support the development of automated ear recognition methods, several data sets were constructed and made publicly available, all of these containing images of relatively good quality acquired in high constrained conditions and environments.

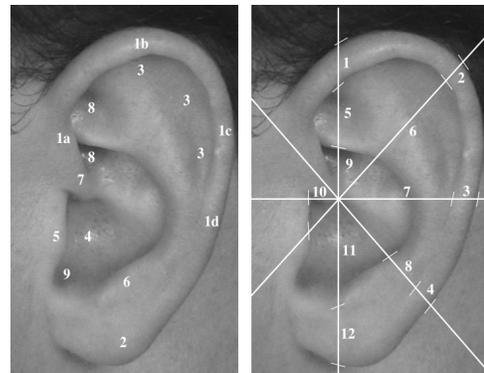


Fig. 1. Iannarelli system [1].

The main purpose of this paper is to announce the availability and to describe a new free available data set (UBEAR) of ear images, which major discriminating features were that data was collected from on-the-move subjects, under dynamic lighting conditions and without requiring to subjects and special care regarding the ears occlusions or poses. This facts turn the UBEAR into a preferable tool in evaluating the robustness of the currently developed ear recognition methods and in the research for new methods able to operate in uncontrolled conditions, toward the applicability of ear recognition systems in real-world scenarios.

The remaining of this paper is organized as follows: section II discusses the related works, section III overviews the existent data sets of ear images. A detailed description of the UBEAR data set, its imaging framework and statistical significance can be found in section IV. Section V describes our experiments and, finally, section VI concludes the paper.

II. RELATED WORK

It's possible to rearrange the proposed methods used in ear recognition, in 3 main categories.

A. 2D Images

- Burge and Burger [2] modeled each subject's ear as an adjacency graph built from the Voronoi diagram of its curve segments. They introduced a novel graph matching based algorithm which takes into account the erroneous curve segments likely to occur, and proposed the use of thermogram images to overcome the problem of hair occlusion.
- Moreno *et al.* [3] investigated the performance of various neural classifiers and combination techniques in ear recognition. The ear image is analyzed by 3 neural classifiers using outer ear feature points, ear morphology and macro features extracted by a compression network.
- Z. Mu *et al.* [4] used two feature vectors to ear recognition. The first vector is composed by features of the outer ear region, the second contains structural features of the inner region. A back propagation neural network is used as classifier.
- B. Arbab-Zavar and M. S. Nixon [5] used a log-Gabor filter to create a template of the ear, which was previously represented in polar coordinates, followed by an occlusion test. Previously, B. Arbab-Zavar *et al.* [6] used the Scale Invariant Feature Transform (SIFT), to detect ear features. They compared the performance of SIFT versus the Principal Components Analysis (PCA) method in the occlusion test.
- A. F. Abate *et al.* [7] proposed the use of a rotation invariant descriptor, the Generic Fourier Descriptor (GFD) [8], to extract meaningful data from ear images. The GFD is applied in the polar representation of the ears images.
- D. Hurley *et al.* [9] treated the ear image as an array of mutually attracting particles that act as source of a Gaussian force field. The original image is described by a set of potential channels and positions of potential wells.
- L. Yaun and Z. Mu [10] used an automatic ear normalization method based on improved Active Shape Model (ASM). The ear images are rotated so that all ears have the same rotation angle. Full-space Linear Discriminant Analysis (FLDA) is applied for ear recognition. Also, they found the acceptable head rotation range between 10 and 20 degrees, to right and left rotations respectively.

B. 3D Images

- Yan and Bowyer [11] used a Minolta VIVID 910¹ range scanner to capture both depth and color information. They developed a fully automatic ear biometric system using Iterative Closest Point (ICP) based in 3D shape matching for recognition, and used both 2D appearance and 3D depth data for automatic segmentation of the iris, also separating it from hair and earrings. They reported a decrease in matching performance where both ear images to be matched have more than 15 degrees of difference. Previously, they tested different approaches [12] and concluded that ICP based matching achieves the best performance.
- Chen and Bhanu [13] proposed to fuse range and color images to detect ears, and both global and local features for the extraction of the meaningful information. Also, they used the ear helix/anti-helix and a LSP (Local Surface Patch) representation for estimating the initial rotation and translation between a gallery/probe pair and then the modified ICP algorithm to compensate the distortion.

C. Acoustic

- Akkermans *et al.* [14] exploited the acoustic properties of the ear for recognition purposes. It turns out that 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. This acoustic transfer function forms the basis of the acoustic ear signature.

III. RELATED DATASETS

Three different datasets are most widely used in the evaluation of ear recognition proposals: the UND² (University Notre Dame), the XM2VTS³ (the extended M2VTS Database) and the USTB⁴ (University of Science and Technology Beijing). Other datasets were developed or referred in academic research, but are not publicly available and not constitute the scope of this section. In the following we briefly describe the main features of its dataset, together with a set of example images shown in figure 3.

A. UND

The University of Notre Dame supplies four collections of ear datasets:

- CollectionE: 464 visible-light profile ear images from 114 human subjects;
- CollectionF: 942 3D ,plus corresponding 2D profile ear images from 302 human subjects;
- CollectionG: 738 3D, plus corresponding 2D profile ear images from 235 human subjects;
- CollectionJ2: 1800 3D, plus corresponding 2D profile ear images from 415 human subjects.

¹<http://www.konicaminolta.com/instruments/products/3d/non-contact/vivid910/index.html>

²[http://www.nd.edu/\\$\sim\\$scvrl/CVRL/Data/_Sets.html](http://www.nd.edu/\simscvrl/CVRL/Data/_Sets.html)

³<http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>

⁴<http://www.ustb.edu.cn/resb/>

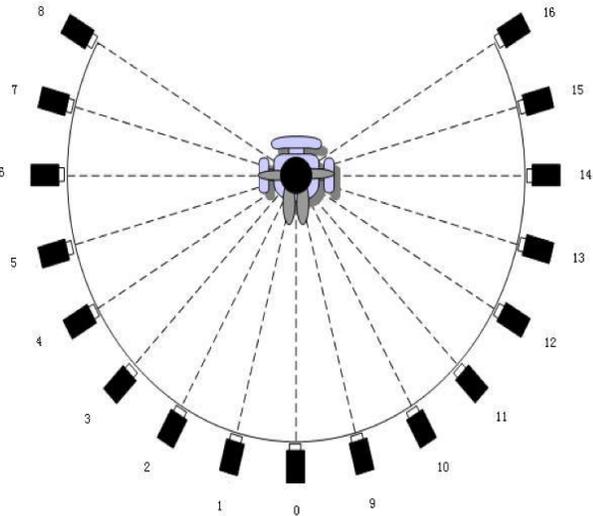


Fig. 2. The platform of the camera system from USTB⁴ dataset.

B. XM2VTS

The University of Surrey supplies several collections of image, audio and video datasets. For ear recognition purposes, the focus should be putted in the head image dataset, described as : 1 left and 1 right profile image (720x576) per person and session, for a total of 2 360 images.

C. USTB

The University of Science and Technology of Beijing (USTB) supplies four datasets of ears, with multi-pose and angles data faces:

- Dataset I - 60 subjects, 3 images each from the right ear, with some of the ears experiencing some shearing and rotation;
- Dataset II - 77 subjects, 4 images per subject. The distance between subject and camera is about 2 meters with variations in terms of illumination and angles. 2 images for different lighting setups and the remaining for pose variations, with rotations of -30 degrees and +30 degrees. Each image is 24-bit true color with 300x400 pixels.
- Dataset III - this dataset is divided into 2 sub sets: the first contains 79 subjects with right and left rotation. The second includes 24 subjects, and each one has 6 images with different ranges of occlusion. The images are 768x576 pixels, 24-bit true color.
- Dataset IV - The capture process consists of 17 CCD cameras which are distributed in a circle with radius being 1 meter and the subject is placed in the center, as illustrated in figure 2. The volunteers (500 in total) were required to look eye level, look upwards, look downwards, look right and look left for the photograph.

IV. UBEAR DATASET

As above stated, our fundamental purpose was to unconstraint the image acquisition scenario, so that images ap-

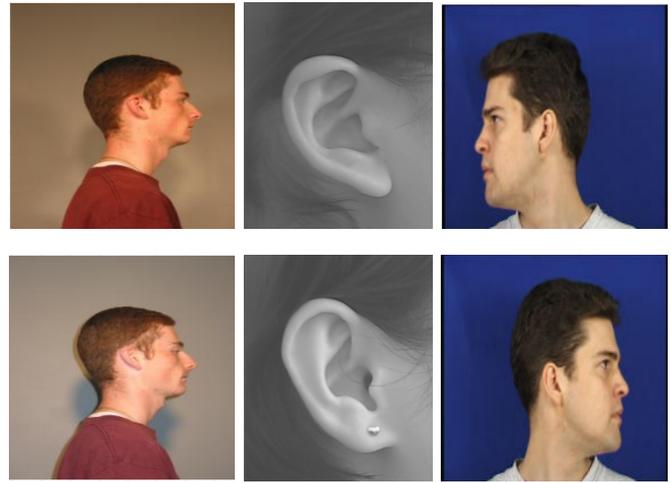


Fig. 3. UND dataset in first column, the second column show USTB dataset and the last column XM2VTS dataset.

pear to be captured in *real-world* conditions, i.e., with subjects on-the-move and without requiring them any particular care about occlusions of the ears and poses. Also, the lighting conditions were highly varying between different sessions so that the typical imagers from each session have notorious differences. This peculiar setup were devised to simulate the covert acquisition of biometric data.

A. Imaging Framework

The setup of the video and imaging framework is given in table I. Samples of the collected images and its corresponding binary ear masks (manually made) are illustrated in figure 4. The video capture sequence starts with all subjects facing front and 3 meters apart from the camera sideways, it's also required of the individual to move his head upwards, downwards, outwards, towards. After these, subjects should step ahead and backwards from the initial position. For each subject both ears were captured in two different sessions, giving a total of four videos per subject. Each video sequence was manually analyzed and 17 frames were selected according to the following criteria:

- 5 frames when the subject was stepping ahead and backwards,
- 3 frames of the subject's head moving upwards,
- 3 frames of the subject's head moving downwards,
- 3 frames of the subject's head moving outwards,
- 3 frames of the subject's head moving towards.

B. Data Variability and Statistics

As illustrated in figure 5, there are three major varying factors in the UBEAR images: 1) lighting variations, either by natural or artificial light; 2) multiple head poses either with yaw and pitch rotations and 3) occlusions due to hair and earrings. Apart from these, images of the UBEAR data set have significantly heterogenous levels of image quality, which can be useful to evaluate the algorithms robustness to changes

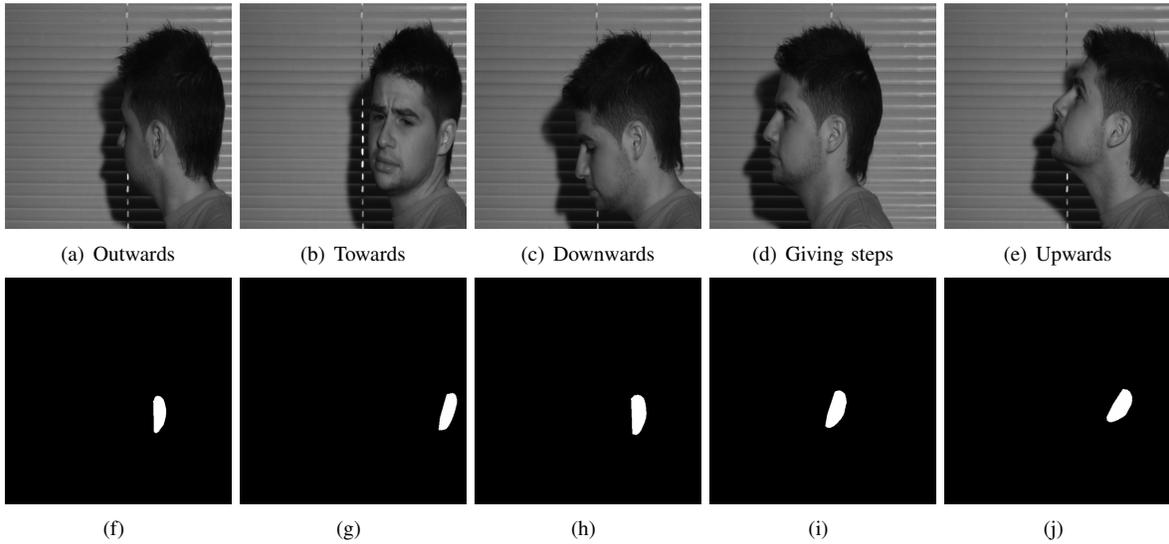


Fig. 4. Images from the UBEAR dataset and this corresponding binary ear mask.

TABLE I
VIDEO AND IMAGE FRAMEWORK

Video Acquisition Framework and Set-Up	
Camera	Stingray F-504B 2/3"
Focal length	35mm
Color Representation	gray scale
Video Resolution	1280x960 pixels
Frames per second	15
Videos Codec	Avi uncompressed
Details of the selected frames	
Image Resolution	1280x960 pixels
Color Representation	gray scale
Image Codec	tiff
Volunteers	
Totals = Subjects 126;	Gender = Male: 44.62%
Ears 252; Images 4430	Female: 55.38%
Age = [0, 20] 43.85%	
[21, 25] 47.69%	
[26, 30] 3.85%	
[31, 35] 2.31%	
[36, 99] 2.30%	

in data quality and to assess the actual deterioration in the corresponding error rates.

Figure 6 gives five histograms that describe some of the major features of the ears regions contained in our dataset. The area histogram (figure 6a) gives the values that correspond to the sum of all pixels inside the ear region, in order to give an idea of the ears size. The eccentricity histogram (figure 6b) describes the proportion between the major axis of the ears. For almost round ears, as in figure 7e, it will return a value near to 0, while for more suchlike elliptical ears (figure 7d), the value will approach 1. The length histogram (figure 6c), exhibits the length of the major axis of the visible part of the ears, as illustrated in figure 7. The orientation histogram (figure 6d) gives the typical orientation of the major axis of the ears, that can be high varying, as illustrated in figures 7(b) and (c). Finally, the perimeter histogram (figure 6e), corresponds to the sum of all pixels in the ear boundaries, which can be



Fig. 5. Comparison between a good quality image and several types of non-ideal images of the UBEAR dataset.

useful for recognition purposes based in boundary descriptors methods.

C. Statistical Significance of the UBEAR dataset

In this section we address the problem of whether the experiments performed on the UBEAR dataset produce statistically significant results. Let α be the confidence interval. Let P be the error rate of a classifier and \hat{P} be the estimated error rate over a finite number of test patterns. At an α -confidence level, we want that the true error rate not exceeds the estimated error rate by an amount larger than $\varepsilon(N, \alpha)$. Guyon et al. [15] fixed $\varepsilon(N, \alpha) = \beta P$ to be a given fraction of P . Assuming that recognition errors are Bernoulli trials, authors concluded that the number of required trials N to achieve $(1 - \alpha)$ confidence in the error rate estimate is $N = -\ln(\alpha)/(\beta^2 P)$. A typical value for α is 0.05 and a typical value for β is 0.2. Based on these values, Guyon et al. [15] recommended the simpler form

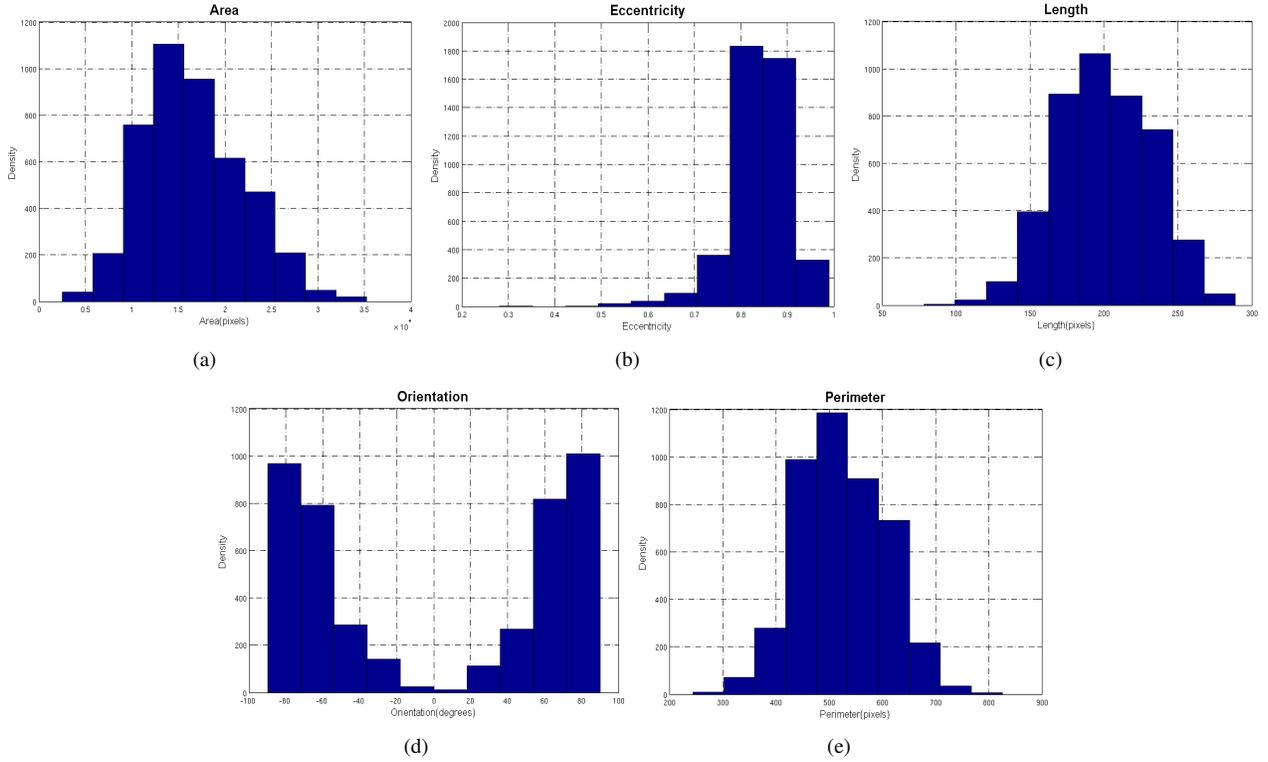


Fig. 6. Statistics of the UBEAR dataset images.

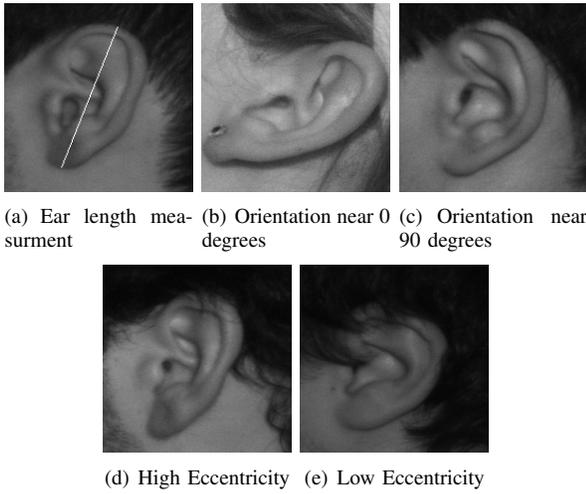


Fig. 7. Some cases that illustrate the histogram extreme cases.

$N \approx \frac{100}{P}$. We had a varying number of subjects that offered as volunteers for the first, second and for both imaging sessions. However, assuming that each iris image can be used to generate a biometric template, that the remaining images from the same eye can be used to analyze intra-class variability and the remaining images from different eyes can be used to analyze inter-class variability, it is possible to obtain a bound for the error that is possible to be tested with statistical significance. The 4 429 images of the UBEAR dataset enable respectively

41 278 and 9 764 528 intra-class and inter-class comparisons. This guarantees statistical significance in experiments with an empirical error rate \hat{P} down to $1,02 \times 10^{-5}\%$. However, we stress that this is a lower bound value that would be increased if we do not assume the independence between images and error correlations are taken into account.

V. EXPERIMENTS

The main purpose of this section is to describe the deteriorations in performance of typical ear recognition methods due to factors that degrade the quality of the acquired data. Based in our earliest experiments, we observed that current methods are particularly sensitive to rotations of the ears, which became the focus of our further experiments. To avoid that other factors bias our results, we selected a sub set of good quality images: those that are sharp, without significant occlusions. Then, data was divided into 5 subsets: 624 images without significant ear rotations (subset 1), 358 images with ears rotated upwards (subset 2), 332 images with ears rotated downwards (subset 3), 309 images with ears rotated outwards (subset 4) and 343 images with ears rotated towards (subset 5). Figure 4 illustrates the four types of rotations that were the main criterium for the division of the data sets.

- Aligned: we compare all images of subset 1,
- Aligned-Upwards: Each image of subset 1 is compared with all images from subset 2,
- Aligned-Downwards: Each image of subset 1 is compared with all images from subset 3,

- Aligned-Outwards : Each image of subset 1 is compared with all images from subset 4,
- Aligned-Towards: Each image of subset 1 is compared with all images from subset 5.

A. SIFT

The Scale Invariant Feature Transform (SIFT) is one of the most popular descriptors for image point matching [16]. The SIFT is known to be invariant to image point scale and rotation and robust to affine distortion, changes in 3D viewpoint, addition of noise and changes in illumination. Its application domain has been extended to human identification and the results are quite promising [17]. Here, keypoints are represented by vectors indicating scale, orientation and location. The keypoints location is refined by fitting it to nearby data and one or more orientations can be assigned using local image gradient directions for each keypoint [17]. The feature descriptor is computed by accumulating the orientation histograms on the 4x4 subregions. Each histogram has 8 bins, thus the SIFT feature descriptor has 128 elements. Finally, the feature vector is normalized to reduce the effects of illumination change [17]. The ratio between the distance of the closest neighbor and the second-closest neighbor, is used in the search for corresponding matching points [17] for recognition strategies. In this paper, we used the D. Lowe’s implementation of the SIFT operator⁵.

B. Results

In order to avoid that segmentation errors carry some bias to the obtained results, we manually segmented all the images used in this experiment, producing a binary segmentation mask that distinguishes between the noise-free regions of the ear and all the remaining types of data in the image, as it is illustrated in figure 4. Thus, the SIFT method was applied exclusively to the regions that comprise the un-occluded ears. As performance measures, we elected the well known receiver operating characteristic curves (ROC), the area under curve (AUC) and the equal error rate (EER) and the decidability [18] index, given by $\frac{|\mu_I - \mu_E|}{\sqrt{0.5 \sigma_I^2 + 0.5 \sigma_E^2}}$, where μ_I and μ_E denote the means of the intra-class and inter-class observations and σ_I and σ_E the corresponding standard deviations. The ROC curve is a graphical plot of the sensitivity, or true positive rate, vs. false positive rate. The AUC can be perceived as a measure based on pairwise comparisons between classifications of two classes. With a perfect ranking, all positives examples are ranked higher than the negatives ones and the area equal to 1. Any deviation from this ranking decreases the AUC. The EER of a verification system, when the operating threshold for the accept/reject decision is adjusted so that the probability of false acceptance and false rejection becomes equal.

The obtained ROC curves are illustrated in figure 9, and the decidability, EER and AUC results are given in table II. The test 1 obtained the best performance, as shown by the ROC curve, with higher decidability, a lower EER and

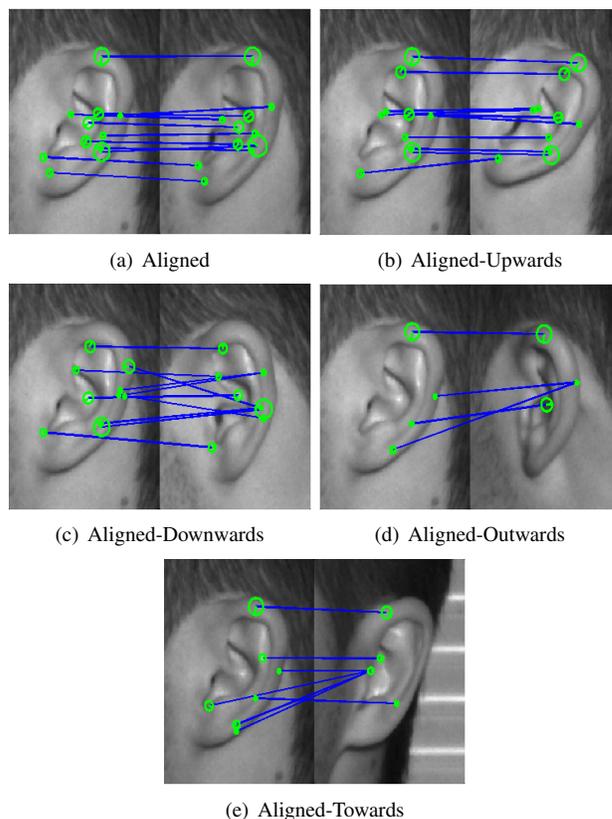


Fig. 8. These images show the matching keypoints according to the used SIFT encoding matching strategies.

higher AUC. Obviously, this was a priori expected, as all the matched images do not have significant differences in pose and SIFT maximally correlates them, as can be seen in figure 8(a). Oppositely, the worst results was obtained in tests 4 and 5, which from our viewpoint can be explained by the fact that significant toward and outward rotations alter the perception of the ear shape, essentially because the ear is far from planar and such rotations lead in some cases to occlusions of portions of the ears. The other types of rotations (upwards and downwards) didn’t significantly deteriorate the results, which is in agreement to our perception of the iris structure according to these rotations.

Figure 8 illustrates the insights of the obtained results, giving the comparisons between the key points that were typically matched in each test. It can be confirmed that maximal correlation was obtained when both ears were aligned with the camera (figure a). Upward and backward rotations didn’t significantly changed the number of matched key points (figures b and c), in opposition with towards and outwards rotations (figures d and e). From these experiments, we concluded that significant research efforts are required to compensate for different poses of the subjects, which will be a crucial step for the development of ear recognition methods able to operate under less constrained conditions. Hopefully, the UBEAR dataset will contribute for such achievement.

⁵<http://www.cs.ubc.ca/lowe/keypoints/>

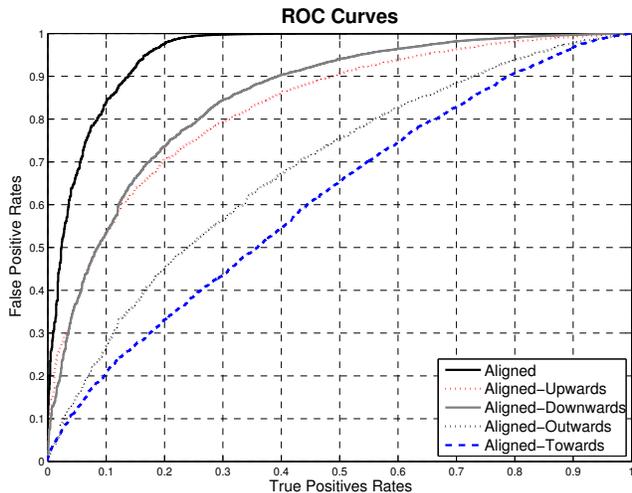


Fig. 9. ROC curves of evaluated tests.

TABLE II
RECOGNITION RATES OF EACH TEST.

	DEC	EER(%)	AUC
Aligned	2.64	12.20	0.95
Aligned-Upwards	1.32	24.99	0.83
Aligned-Downwards	1.43	22.86	0.85
Aligned-Outwards	0.68	36.24	0.69
Aligned-Towards	0.44	42.36	0.62

VI. CONCLUSIONS AND DATA SET AVAILABILITY

In this paper we presented a new dataset of ear images for biometric purposes, which major discriminating point is that it simulates the acquisition of data in real-world scenarios, under varying lighting conditions on moving subjects and without requiring them any particular care regarding ear occlusions and poses. Our experiments show that — as it will be expected — the performance of the most well known ear recognition methods significantly decreases according to the quality of the acquired data. Hence, we hope that the UBEAR dataset constitute as a valuable tool for the research of ear recognition systems more robust to degraded data, namely due to the dynamics lighting recognition, subject movements and perspective. Finally, the value given to the UBEAR dataset should have be directly correspondent to the number of persons that use it in their experiments. Thus, we decided to make UBEAR public and freely available through the UBEAR datasets web site: <http://www.ubear.di.ubi.pt>.

ACKNOWLEDGMENTS

We acknowledge all volunteers that agreed to participate in the imaging sessions of the UBEAR data set. Also, the financial support given by "FCT-Fundação para a Ciência e Tecnologia" and "FEDER" in the scope of the PTDC/EIA/103945/2008 re-

search project "NECOVID: Negative Covert Biometric Recognition" is acknowledged too.

REFERENCES

- [1] A. V. Iannarelli, *Ear Identification (Forensic Identification Series)*, 1st ed. Fremont, California: Paramount Publishing Company, 1989.
- [2] M. Burge and W. Burger, "Ear biometrics in computer vision," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 2, 2000, pp. 822–826 vol.2.
- [3] B. Moreno, A. Sanchez, and J. Velez, "On the use of outer ear images for personal identification in security applications," in *Security Technology, 1999. Proceedings. IEEE 33rd Annual 1999 International Carnahan Conference on*, 1999, pp. 469–476.
- [4] Z. Mu, L. Yuan, Z. Xu, D. Xi, and S. Qi, "Shape and structural feature based ear recognition," in *Advances in Biometric Person Authentication, 5th Chinese Conference on Biometric Recognition, SINOBIOETRICS 2004, Guangzhou, China, December 13-14, 2004, Proceedings*, vol. 3338. Springer, 2004, pp. 663–670.
- [5] B. Arbab-Zavar and M. Nixon, "Robust log-gabor filter for ear biometrics," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 8-11 2008, pp. 1–4.
- [6] B. Arbab-Zavar, M. Nixon, and D. Hurley, "On model-based analysis of ear biometrics," in *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, 27-29 2007, pp. 1–5.
- [7] A. Abate, M. Nappi, D. Riccio, and S. Ricciardi, "Ear recognition by means of a rotation invariant descriptor," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4, 2006, pp. 437–440.
- [8] D. Zhang and G. Lu, "Shape-based image retrieval using generic fourier descriptor," *Signal Processing: Image Communication*, vol. 17, no. 10, pp. 825–848, 2002.
- [9] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field feature extraction for ear biometrics," *Computer Vision and Image Understanding: CVIU*, vol. 98, no. 3, pp. 491–512, Jun. 2005.
- [10] L. Yuan and Z. C. Mu, "Ear recognition based on 2D images," in *Biometrics: Theory, Applications, and Systems*, 2007, pp. 1–5.
- [11] P. Yan and K. W. Bowyer, "Biometric recognition using 3D ear shape," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1297–1308, Aug. 2007.
- [12] P. Yan and K. Bowyer, "Ear biometrics using 2D and 3D images," in *Advanced 3D Imaging for Safety and Security*, 2005, pp. III: 121–121.
- [13] H. Chen and B. Bhanu, "Human ear recognition in 3d," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 718–737, Apr. 2007.
- [14] A. H. M. Akkermans, T. A. M. Kevenaar, and D. W. E. Schobben, "Acoustic ear recognition," in *Biometric Authentication*, 2006, pp. 697–705.
- [15] R. S. I. Guyon, J. Makhoul, and V. Vapnik, "What size test set gives good error rate estimates?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 52–64, Feb. 1998.
- [16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [17] K. Mikołajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
- [18] J. Daugman and G. Williams, "A proposed standard for biometric decidability," in *Proceedings of the CardTech/SecureTech Conference*, pp. 223–234, 1996.