# Influence of Alignment on Ear Recognition: Case Study on AWE Dataset

**Metod Ribič**[1] **Žiga Emeršič**[1] **Vitomir Štruc**[2] **Peter Peer**[1]

[1]*Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia*
[2]*Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia*

*E-mail(s): mr3020@student.uni-lj.si, ziga.emersic@fri.uni-lj.si, vitomir.struc@fe.uni-lj.si, peter.peer@fri.uni-lj.si*

## Abstract

*Ear as a biometric modality presents a viable source for automatic human recognition. In recent years local description methods have been gaining on popularity due to their invariance to illumination and occlusion. However, these methods require that images are well aligned and preprocessed as good as possible. This causes one of the greatest challenges of ear recognition: sensitivity to pose variations. Recently, we presented Annotated Web Ears dataset that opens new challenges in ear recognition. In this paper we test the influence of alignment on recognition performance and prove that even with the alignment the database is still very challenging, even-though the recognition rate is improved due to alignment. We also prove that more sophisticated alignment methods are needed to address the AWE dataset efficiently.*

## 1 Introduction

Ever growing need for human recognition over last decades led to new techniques on various biometric modalities. Ear as a biometric modality presents a promising case: it is noninvasive, has a high degree of permanence, distinctiveness and universality [1]. However, one of the most difficult, still-open issues of ear recognition is pose variation: pictures of ears can be captured from various positions as shown in Fig. 5. The first step is therefore to normalize and segment input data as good as possible. Data normalization builds a foundation for the rest of the recognition process. If data is difficult to work with and no normalization is performed, even the best performing state-of-the-art descriptors fail. In this paper we provide a proof of concept using alignment based on normalization technique that uses RANSAC [2] for projective transformation estimation and a calculation of average ears for ear template and segmentation mask calculation.

In Section 2 we discuss existing ear alignment methods. The alignment method used in our experiments is described in Section 3. Section 4 describes experiments. Section 5 presents results. In Section 6 we conclude the paper.

## 2 Existing ear alignment techniques

Ear pose normalization is an open problem. Compared to face alignment it presents a greater challenge – ears do not have a symmetry axis to support landmark positions [3] like faces do. Many authors for accurate ear alignment rely on 3-dimensional information [4].

In [5] authors detect ear outer shape using Canny edge detector and then rotate ear according to the longest distance, which represent the dominant ear axis.

In [6] the authors proposed a combination of active contours techniques and ovoid model for ear fitting to normalize ear features. They achieved encouraging results, using different pitch angles and different distances of ears to camera.

In [7] author extracts outer ear rim out of binarized image and then based on the longest distance aligns ear to a normalized position.

Another promising method CPR (Cascaded Pose Regression) was applied on ear alignment in [3], where outer ear rim is fitted with ellipse. The ellipse, together with ear is then transformed into the normal position.

Other, more general methods could be used as well, such as congealing [8], or using optical flow estimation for smaller corrections in ear pose variations.

The above methods work well, however, only when ear is rotated in its main plane – i.e. when face is moved in the pitch direction (see Fig. 4).

## 3 Alignment method

To acquire the aligned subset of AWE dataset [9] (`http://awe.fri.uni-lj.si`), we used the procedure described below and visualized in Fig. 1.

The first step is SIFT (Scale Invariant Feature Transform) [10] keypoint detection and SIFT keypoint description. Based on the calculated data, RANSAC [2] estimates planar transformation of each image to the average ear image. The transformation is then applied to the image, together with ear mask. To make further processing easier all images are resized with suitable factor to match 100 pixels in height and width corresponding to height. All images are converted to grayscale as well, because color data does not provide any viable information for Local Phase Quantization (LPQ) [11] descriptor that we are using in the recognition evaluation. At the end of this procedure histogram equalization was performed as well.

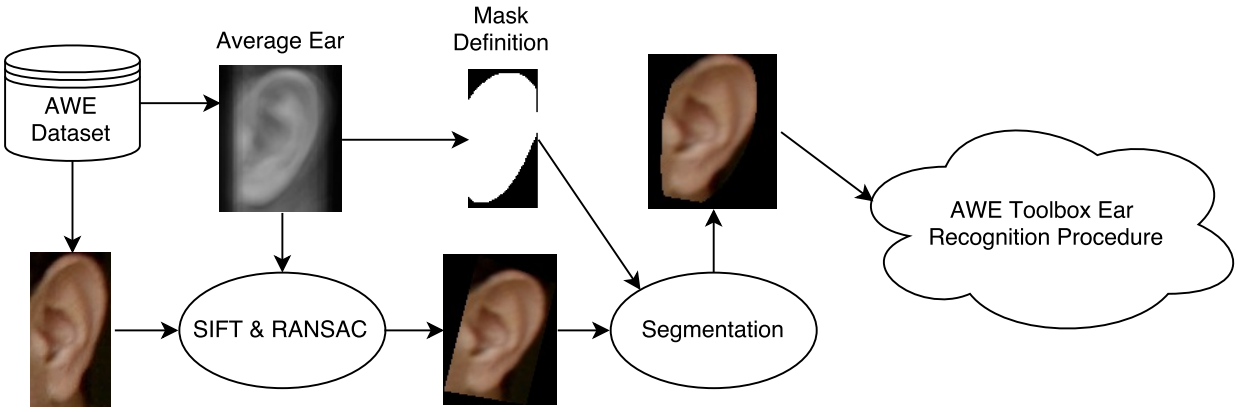


Figure 1: Flowchart of the alignment precedure.

### 3.1 SIFT

For feature extraction we used SIFT detector algorithm provided in vlFeat library, however, due to small images parameters need to be set appropriately:

- The peak selection threshold: $0$.
- The non-edge selection threshold: $1000$.
- The minimum l2-norm of the descriptors before normalization: $2$ (descriptors below the threshold are set to zero).
- The number of levels per octave of the DoG scale space: $200$.
- The descriptor magnification factor: $5$. The scale of the keypoint is multiplied by this factor to obtain the width (in pixels) of the spatial bins.

### 3.2 RANSAC

After feature extraction we estimate the homography from extracted pairs of points and corresponding descriptors. All descriptors and pair of points are passed to RANSAC, which decides, depending on outliers distance, whether a pair of points is viable or not. Iterations for estimating transformation is by default set to $5,000$ iterations but then increased to $10,000$ and $25,000$ iterations if needed. If after that the alignment score is still bellow $40\%$, the image is rejected. The alignment score is defined as $m/a$, where $m$ is number of matched SIFT keypoint pairs that RANSAC managed to set in a plane and $a$ is number of all pairs. These thresholds were set experimentally.

### 3.3 Average ear

All images in our alignment procedure are aligned to an average ear. The average ear is acquired by summing and averaging all pixels of all images, which are annotated as perfectly aligned – meaning that roll, pitch and yaw axis are close to $0°$ and they are not occluded or contain accessories. Because all ears in AWE are annotated, we had the information if ear contains any accessories, if ear is overlapped and most importantly if ear is already in normal position, and if not, on which axis. Left average ear was summed using 26 and right was summed using

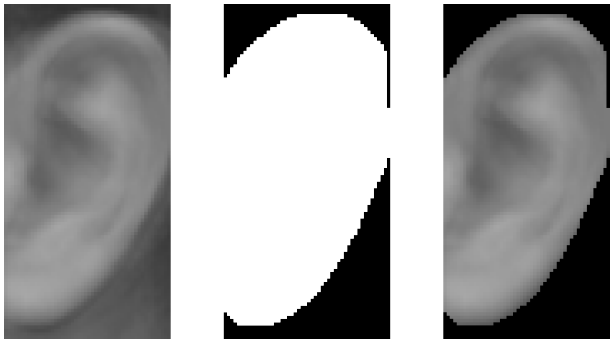

Figure 2: First two images show average ear and mask, respectively, the third image shows the resulting image after mask has been applied.

29 images that met conditions listed above. To be able to perform summing pixel by pixel all images must be same sized, therefore images are padded with black on sides to fit the largest picture. Padding ensure all ears are center and are the same size, but at the same time vertical lines on final image appears and because of this the average ear is cropped to most inner vertical line. The final average ear is shown in the first image in Fig. 2.

### 3.4 Segmentation

For further ear recognition procedure all aligned images are cropped to fit the perfect ear and this is done using binary mask gained from average ear. Since every ear is aligned to average ear, binary mask provide perfect crop and with that we ensure recognition is performed only on area showing the ear. Binary mask of average ear is shown in the second image in Fig. 2. Fig. 3 shows results of three stages: original image, aligned image and masked image.

## 4 Experiments

### 4.1 Data

For experimental evaluation we use AWE dataset, where images are annotated with pose variation angles. Experiments are performed only on images containing left ears – $520$ images. On these images two separate subsets are defined, according to the severity of pose variations. Pose

Figure 3: Images showing three stages alignment step before features are extracted.

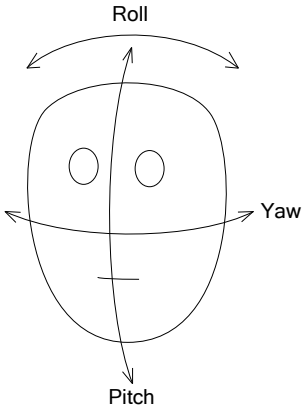


Figure 4: Diagram showing pitch, roll and yaw.

variations are shown in Fig. 4. Sample images, showing properties of the two subsets, are shown in Fig. 5.

In the first subset only images with mild yaw and roll angles (angles of up to 10°) are considered, with arbitrary pitch angle (there are 358 such images). In the second subset all images, regardless of angles are used (all 520 images). However, in both subsets images that RANSAC is not able to transform, according to our rules (as described in Section 3.2), are additionally removed. Furthermore, images where only 1 image per person remain are removed as well. The final subsets sizes are 105 and 163 images for the first and the second subset, respectively. Each aligned image has original counterpart, which serves as a base dataset for the experimental evaluation.

### 4.2 Evaluations

To evaluate recognition performance we use AWE Toolbox [9] that is freely available at `http://awe.fri.uni-lj.si`. For both subsets we perform two evaluations: on non-aligned images and on aligned images. To evaluate recognition performance we use Local Phase Quantization (LPQ) [11] for feature extraction on image data.

## 5 Results and discussion

The results on the subset I show improvement of aligned ear images over non-aligned ear images with Rank-1 recognition rate of 72.8% vs. 66.3%. Equal Error Rate (used in verification setups) reduced from 27.65% to 24.45%,

Figure 5: Four images in the first line have significant pitch angles and negligible roll and yaw and present the easier dataset (subset I) for ear alignment. Four images in the second line have all angles significant and present the more difficult dataset (subset II) for ear alignment.

Table 1: Results of evaluation on subset I, showing performance improvement using aligned images.

| [%] | Non-aligned | Aligned |
|---|---|---|
| Rank-1 | $66.3 \pm 11.0$ | $72.8 \pm 16.7$ |
| EER | $27.7 \pm 13.1$ | $24.5 \pm 10.9$ |

Table 2: Results of evaluation on subset II, showing performance drop when aligning difficult images.

| [%] | Non-aligned | Aligned |
|---|---|---|
| Rank-1 | $65.6 \pm 7.0$ | $57.1 \pm 6.0$ |
| EER | $26.3 \pm 4.8$ | $30.0 \pm 7.3$ |

proving the usability of alignment method on images that contain pose variation of up to 10° (Table 1).

However, the results on subset II, which contains images of higher roll and yaw angles (more than 10°), do not show improvement. Here Rank-1 recognition rate falls from 65.6% to 57.1%, while undesirable increase of EER from 26.3% to 30.0% is also evident (Table 2). This shows that normalization of images taken under severe pose variations regarding roll and yaw (with angles larger than 10°) cannot be addressed properly.

The results show that RANSAC planar transformation estimation succeeds on images, where mild roll and yaw angles are present with only pitch angles being severe and fails when roll or yaw angle are severe as well. Furthermore, to the best of our knowledge, this problem has not been successfully addressed in literature yet, using single 2-dimensional images only. This shows that proper ear pose normalization is still an open problem.

## 6 Conclusion

We have used RANSAC-based method for ear pose aligment. Experiments showed that ear alignment to average ear, together with masking ear area, improve recognition when head pitch variations are present in ear images. However, aligning ear images that contain severe pose variation in roll and yaw still remains an open problem. We plan to further improve ear alignments and to evaluate other existing ear normalization techniques such as CPR, optical-flow based and build on top of them.

## References


[1] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1160–1165, 2003.

[2] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[3] A. Pflug and C. Busch, *Secure IT Systems: 19th Nordic Conference, NordSec 2014, Tromsø, Norway, October 15-17, 2014, Proceedings*. Cham: Springer International Publishing, 2014, ch. Segmentation and Normalization of Human Ears Using Cascaded Pose Regression, pp. 261–272. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11599-3_16

[4] A. Pflug and C. Busch, "Ear biometrics: a survey of detection, feature extraction and recognition methods," *Biometrics, IET*, vol. 1, no. 2, pp. 114–129, June 2012.

[5] A. P. Yazdanpanah and K. Faez, *Emerging Intelligent Computing Technology and Applications: 5th International Conference on Intelligent Computing, ICIC 2009, Ulsan, South Korea, September 16-19, 2009. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, ch. Normalizing Human Ear in Proportion to Size and Rotation, pp. 37–45. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-04070-2_5

[6] E. Gonzalez, L. Alvarez, and L. Mazorra, "Normalization and feature extraction on ear images," in *Security Technology (ICCST), 2012 IEEE International Carnahan Conference on*, Oct 2012, pp. 97–104.

[7] W. Shu-zhong, "An improved normalization method for ear feature extraction," *Shandong College of Information Technology, Weifang, China*, 2013.

[8] G. B. Huang, V. Jain, and E. Learned-Miller, "Unsupervised joint alignment of complex images," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[9] Ž. Emeršič and P. Peer, "Toolbox for ear biometric recognition evaluation," in *EUROCON 2015-International Conference on Computer as a Tool (EUROCON), IEEE*. IEEE, 2015, pp. 1–6.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[11] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Image and signal processing*. Springer, 2008, pp. 236–243.