

# Human Ear Localization: A Template-Based Approach

Alaa Halawani

Department of Applied Physics and Electronics (TFE), Umeå University, Umeå, Sweden  
Email: alaa.halawani@umu.se

Haibo Li

School of Computer Science & Communication, Royal Institute of Technology (KTH), Stockholm, Sweden  
Email: haiboli@kth.se

**Abstract**—We propose a simple and yet effective technique for shape-based ear localization. The idea is based on using a predefined binary ear template that is matched to ear contours in a given edge image. To cope with changes in ear shapes and sizes, the template is allowed to deform. Deformation is achieved by dividing the template into segments. The dynamic programming search algorithm is used to accomplish the matching process, achieving very robust localization results in various cluttered and noisy setups.

**Index Terms**—ear detection, localization, dynamic programming, shape template, segments, edge noise, partial occlusion

## I. INTRODUCTION

The use of human ear as a biometric has gained attention recently [1], [2]. Having a unique pattern that does not change with age or facial expressions, the ear seems to serve as a reliable biometric for human verification and identification. Beside the use as a biometric trait, the ear can be used as a facial feature to achieve better head tracking or pose estimation [3]. In order to fully automate the ear-based identification/verification techniques, the systems should be able to correctly detect the ear position in the acquired images. We call this here “ear localization”.

The ear localization problem comes with many challenges that should be dealt with. First of all the ear usually constitutes a very small portion of the captured image, meaning that the rest of the image should be treated as noise or unnecessary clutter. Occlusion caused by hair or earrings is another serious challenge. Moreover, variations of the ear shape from one user to another makes the situation worse. Practical systems should also take imaging conditions into considerations. Change in lighting conditions, for example, should be tolerated.

In literature, several ear localization approaches have been investigated. Active contours were used in [4] to fit the ear in the image. In [5], the authors used deformable contours [6] to localize the ear. Hurley *et al.* [7] accomplished the task by applying the so-called “Force

field feature extraction”. However, all these contributions require an initialization step that is crucial to the success of the detection. Without initializing the search around the ear in the image, the localization will fail. To avoid manual initialization, [1] tries to estimate the ear pit location as a starting point for active contour fitting algorithm. However, estimating the ear pit location involves several steps including preprocessing, skin detection, curvature estimation, and surface segmentation and classification. Moreover, in order to be able to accomplish this task, the nose tip must be reliably detected, the matter that adds more complexity to the system. Finally, if the ear pit is invisible due to some occlusion or any other reason, the system will definitely fail. Other researchers relied on AdaBoost classifiers [8], [9] in a Viola-Jones-like fashion. The method works well and can tolerate occlusion, but processing speed seems to be slow. In [8], it is reported that processing time is about 26 seconds per image.

In this paper, we present a very robust shape-based ear localization technique that can deal with the above-mentioned challenges. We embrace a template-based localization paradigm where single predefined binary template of the ear is used to accomplish the task. The template is divided into a set of segments that are fed to a dynamic programming search algorithm to detect the ear in the target images. The algorithm works robustly under extensive edge image noise, and can tolerate partial ear occlusion.

Please keep in mind that our goal in this contribution is not to identify people using their ear biometrics, but rather to localize the ear so that it can be used in applications like biometric recognition or verification.

The paper is organized as follows: Section II introduces the concept of ear localization, followed by Section III that shows how to achieve the localization using dynamic programming. Results are summarized in Section IV, and concluding remarks are given in Section V.

## II. EAR LOCALIZATION

Shape-Based detection can in general be achieved by either of two ways: Statistical modeling or deterministic

modeling. Statistical modeling, like active shape model (ASM) [10], usually requires training of the model using a training set of collected shapes. Deterministic modeling, on the other hand, relies on object features like edges. Since they do not require any training, deterministic methods are usually faster and easier to implement.

Our methodology is classified as deterministic, in which we model the ear as a set of contours in an edge image. Suppose that a binary template of the ear contour is available. The task then is to match the available template to the ear contour in the edge image.

The concept thus is based on storing a *single* contour template of an ear and then using it to match the contours present in the edges that are extracted from the query image. We search in edge information extracted from the acquired image using the very standard Canny edge detector. To achieve robust ear localization we use a dynamic programming (DP) search algorithm in order to insure a globally optimal solution. The procedure is abstracted in Fig. 1.

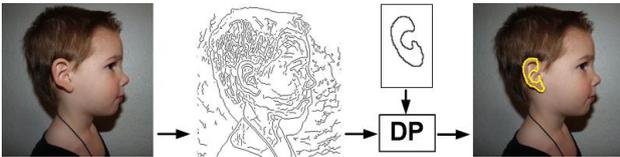


Figure 1. Using template matching with dynamic programming (DP) to localize the ear in a query image.

The scheme possesses several advantages that are able to overcome the previously-mentioned challenges. First of all, it is robust to illumination changes, since it relies only on edge information, which is insensitive to lighting changes. Secondly, it is also robust to the presence of clutter in the background. Even when the edge image is noisy, the detection results are highly reliable, thanks to the powerful DP search we are using. It is worth noting that we do not put any constraints on the background used. The algorithm is assumed to work for any given background content.

Moreover, the allowed template deformations enable tolerating scale and orientation changes to certain limit (can reach  $\pm 50\%$ ). This flexibility makes it possible to use a *single* ear template to accomplish localization for *all* system users. I.e., there is no need to store a separate template for each user. The technical details are briefly described in the following section.

### III. TEMPLATE MATCHING VIA DP

DP is an algorithm that is used to ensure a globally optimal solution of a problem, as it always returns the highest score match. It has proven to be successful in many contributions [11]-[13]. Hence, we decided to adapt the DP algorithm to the task of ear localization. Namely, we use DP to search for the best fit of an ear template in a given image. Following is a brief clarification of the concept.

As can be seen in Fig. 1, the inputs to the DP module are the binary template and the binary edge image. A *single* template is acquired offline and stored to be used for online ear detection for *all* users. Two main issues are

to be considered here: Template deformation and template matching.

#### A. Deformable Templates

In order for the template to be usable, it should adapt to different ear shapes and tolerate certain range of transformations. This cannot be achieved if it is treated as a rigid entity. Instead, it is divided into shorter segments that are allowed to move within a certain range during the matching process. Deformation is introduced as a result of the controlled segment movements.

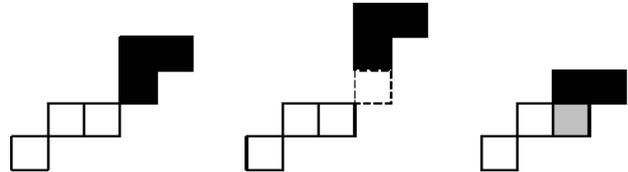


Figure 2. (left) Original arrangement of two segments (3 pixels each) in the template. (middle) Deformation caused by introducing a 1-pixel gap between the segments. (right) Deformation caused by introducing a 1-pixel overlap between the segments.

The example in Fig. 2 illustrates the concept. It is assumed that the template is divided into several 3-pixel segments. The left column of the figure shows the original spatial arrangements of two segments in the template. By allowing the segments to move one pixel, a set of deformations can be achieved by introducing a gap (middle column) or an overlap (right column) between the segments.

Segments can be shifted to any position in the image to be searched, provided that the relative displacement between two consecutive segments is not greater than one pixel. I.e., if two consecutive segments are shifted by  $o = (o_x, o_y)$  and  $p = (p_x, p_y)$  respectively, then relative displacement is governed by:

$$|o - p| = \max(|o_x - p_x|, |o_y - p_y|) \leq 1 \quad (1)$$

The degree of template flexibility is then governed by the segment length. If each segment contains 3 pixels, then an overall shrinkage or enlargement of around  $1/3 = 33\%$  can be introduced.

A mixture of gaps and overlaps will result in a set of possible deformations. Fig. 3 shows a random selection of possible deformations of a template (upper left) with segment length of 2 pixels (up to 50% flexibility).



Figure 3. A random selection of possible deformations of the original template (upper left). Each segment is two pixels in length.

#### B. Template Matching

The Viterbi DP algorithm is used for the matching process. If the process is viewed as a trellis, each column corresponds to a segment, and nodes in that column

correspond to possible shift values of that segment in the image. Arcs connecting nodes in two consecutive columns are governed by (1).

We search for the path through the trellis that maximizes the accumulated score,  $R$ . For segment  $i$  shifted by  $p$ ,  $R$  is given by:

$$R(p,i) = \max_{p' \in \sigma(p)} \{R(p',i-1)\} + V(p,i), \quad (2)$$

$$\sigma(p) = \{p' \in \Omega : |p' - p| \leq 1\}.$$

where  $\Omega$  is the set of all possible shifts.  $V(p,i)$  is the local reward given to node  $(p,i)$  and it is equal the number of edge pixels segment  $i$  will cover if placed at position  $p$ .

When  $R$  has been calculated for the last segment the algorithm backtracks the optimal path starting from the node with the highest  $R$  value in the last column. This path is returned as the best search result, showing the positions to which the template segments should be shifted in order to cover the target ear in the image.

#### IV. RESULTS

##### A. Setup

To run our experiments, we constructed a dataset of 212 face profile images downloaded from the internet. When collecting the images, we were careful to diversify setups and difficulty levels. The database included male, female, young, and old face profiles with many noise and clutter instances including hair, earrings, and tattoos. Within-range transformations (scaling and rotation) were also included.

To assess the detection quality, ground-truth ear bounding boxes were defined and saved for each image. Fig. 4 shows some samples of the collected images.



Figure 4. Samples of the collected dataset.



Figure 6. Sample results for ear localization. The used edge image is depicted to the right of each result to show how noisy the search space is.

Processing was carried out using a PC with a core 2 Duo 2.93 GHz Processor with 8 GB of memory.

A single ear template was acquired offline and used for localizing ears in images. Edge images were extracted using the standard Canny edge detector.

For matching with DP, the segment length was set to 2 pixels. This achieves a flexibility of  $\pm 50\%$  of the original template size.

##### B. Ear Localization

Ground-truth bounding boxes are used as a comparison basis for the ear localization. The receiver operating characteristic curve (ROC curve) is chosen as a quantitative evaluation measure. The ROC curve depicts the correct detection rate versus the average false positives per image (FPPI). The highest detection rate with the lowest FPPI is desirable.

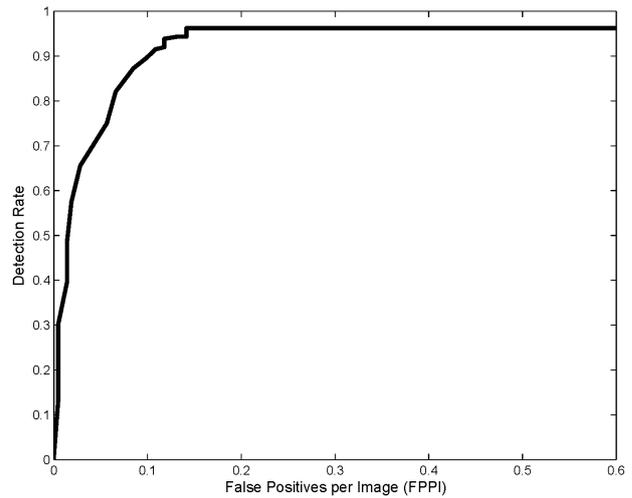


Figure 5. Performance of the ear detection system in terms of ROC curve.

For our system, the ROC curve is shown in Fig. 5. From the figure it can be noticed that our system achieves this goal. The system already reaches a detection rate of approximately 65% with an FPPI that is close to zero (0.02). The best performance is reached with an FPPI of 0.14 only. At that point the detection rate is 96.2%. An FPPI of 0.14 is considered very low in the literature, and this proves the robustness of the technique.



Figure 7. Additional sample results for ear localization. Shown results demonstrate the robustness of the algorithm to different situations and transformations.

Qualitative localization results are shown in Fig. 6 and Fig. 7. The purpose of results shown in Fig. 6 is to stress the difficulty of the search space. In most (if not all) of the database images, the extracted edge images are very noisy constituting a big challenge to any search algorithm. Thanks to the robust DP search algorithm and the flexibility of the template representation, our system was able to accurately localize the ear in the images despite the intensive noise present. This can be clearly seen in the edge images depicted to the right of the detection results shown in Fig. 6.

Fig. 7 shows extra localization results. The results shown in this figure are meant to demonstrate the robustness of the algorithm to different situations and transformations. First of all, one can notice how the detection tolerates different ear shapes for different subjects. This is possible due to the segment representation of the used ear template.

Moreover, the system can tolerate changes in ear size (up to  $\pm 50\%$  of original template size as discussed earlier), as can be clearly noticed in the figure.

Within-range rotation is also robustly handled by our algorithm. This can be seen in most of the shown results but is more eminent in the first two instances (from the left) of the upper row of the Fig. 7.

Finally, the algorithm can very well handle cases of partial ear occlusion, caused by hair and/or earrings. Hair occlusion can be observed in the last two instances of the upper row of Fig. 7, especially in the last instance where occlusion is more eminent. The first three instance of the lower row show results with earring occlusion. In all of the shown cases the algorithm was able to skirt around occlusion and to localize the ear robustly.

Counter to many existing localization algorithms [8], which can only detect the ear up to a bounding box, our system is able to define ear contours, making the localization results more accurate. This can be clearly seen in all results shown in Fig. 6 and Fig. 7.

## V. CONCLUSION

We presented a technique for robust shape-based ear localization in edge images. A dynamic programming search, with a binary ear template, is carried out to match the corresponding ear contours in the image. A segment based representation of the template is adopted in order to allow for necessary template deformations. The method was tested on a dataset of 212 images containing face

profiles, each showing an ear instance that can be in a noisy surrounding, partially occluded, or moderately scaled and/or rotated. The results demonstrated the robustness of the technique in localizing ear instances in such situations. This was both quantitatively and qualitatively illustrated.

## REFERENCES

- [1] P. Yan and K. W. Bowyer, "Biometric recognition using 3D ear shape," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1297-1308, 2007.
- [2] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*, Springer-Verlag New York, Inc., 2008.
- [3] T. Vatahska, M. Bennewitz, and S. Behnke, "Feature-Based head pose estimation from images," in *Proc. 7th IEEE RAS International Conference on Humanoid Robots*, 2007, pp. 330-335.
- [4] L. Alvarez, E. González, and L. Mazorra, "Fitting ear contour using an ovoid model," in *Proc. 39th Annual 2005 International Carnahan Conference on Security Technology*, 2005, pp. 145-148.
- [5] M. Burge and W. Burger, "Ear biometrics in computer vision," in *Proc. 15th International Conference on Pattern Recognition (ICPR)*, 2000, pp. 822-826.
- [6] K. F. Lai and R. T. Chin, "Deformable contours: Modeling and extraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 11, pp. 1084-1090, 1995.
- [7] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force field feature extraction for ear biometrics," *Computer Vision and Image Understanding*, vol. 98, no. 3, pp. 491-512, 2005.
- [8] S. M. Islam, M. Bennamoun, and R. Davies, "Fast and fully automatic ear detection using cascaded AdaBoost," in *Proc. IEEE Workshop on Applications of Computer Vision*, 2008, pp. 1-6.
- [9] M. Castrillón-Santana, J. Lorenzo-Navarro, and D. Hernández-Sosa, "A study on ear detection and its applications to face detection," in *Advances in Artificial Intelligence*, Springer, 2011, pp. 313-322.
- [10] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, 1995.
- [11] P. F. Felzenszwalb and R. Zabih, "Dynamic programming and graph algorithms in computer vision," *PAMI*, vol. 33, no. 4, pp. 721-740, Apr. 2011.
- [12] A. Shashua and S. Ullman, "Structural saliency: The detection of globally salient structures using a locally connected network," in *Proc. ICCV*, 1988.
- [13] R. Nohre, "Deformed template matching by the Viterbi algorithm," *PRL*, vol. 17, no. 14, pp. 1423-1428, Dec. 1996.



**Alaa Halawani** was born in Hebron, Palestine in 1974. He received his B.Sc. degree in Computer Systems Engineering from the Palestine Polytechnic University, Hebron, Palestine, in 1996. In 1997 he was awarded a "German Academic Exchange Service" (DAAD) scholarship to continue his education in Jordan. He received his M.Sc. in Electrical Engineering - Computer Major - from Jordan University of Science and Technology, Irbid, Jordan, in January 2000. In 2001 he was awarded a second DAAD scholarship to attain his PhD degree in Germany, where he received his PhD (Dr.-Ing.) from the Institute of Pattern Recognition and Image Processing, University of Freiburg, Germany in 2006. From February 2000 to October 2001 he worked as a lecturer at the Electrical and Computer Engineering Department at the Palestine Polytechnic University. After PhD, he worked as an assistant professor at the same department between 2007 and 2009. In February 2009, Dr. Halawani joined the Department of Applied Physics and Electronics (TFE) at Umeå University in Sweden as a Postdoc. As of August 2011, he is an Assistant Professor at the same department. His research interests include Artificial Intelligence, Content-Based Image Retrieval, and Object Detection and Recognition.



**Haibo Li** received the Technical Doctor degree in Information Theory from Linköping University, Sweden, in 1993. His doctoral thesis dealt with advanced image analysis and synthesis techniques for low bitrate video. Dr. Li got the “Nordic Best PhD Thesis Award” in 1994. In 1997 Dr. Li was awarded the title of “Docent in Image Coding”. From 1990 to 1993 he was a teaching assistant of digital video at Linköping University. After graduation Dr. Li

joined the technical faculty of Linköping University first as an Assistant Professor and then promoted to an Associate Professor in 1998. During his period at Linköping University he developed advanced image and video compression algorithms, including extremely low bitrate video

compression, 3D video transmission, and tele-operation and tele-presence. He joined the Department of Applied Physics and Electronics (TFE) at Umeå University in 1999 as a full professor and was directing the Digital Media Lab, Umeå Center for Interaction Technology (UCIT), Umeå University, and worked on advanced Human, Thing and Information interaction techniques. As of 2012 he is a Full Professor at the School of Computer Science & Communication, Royal Institute of Technology (KTH), Stockholm, Sweden. Prof. Li has been chairing sections at relevant international conferences and was actively involving in MPEG activities in low bitrate video compression. He has contributed to several EU projects, like VIDAS, SCALAR, INTERFACE and MUCHI. He has published more than 100 technical papers including chapters in books and holds six international patents as the first inventor in multimedia area.