Ear Detection in 3D Profile Images based on Surface Curvature

Anika Pflug[1], Adrian Winterstein[1], and Christoph Busch[1,2]

[1]Hochschule Darmstadt - CASED, Mornewegstrasse 32, 64293 Darmstadt, Germany
[2]Gjøvik University College, Technologieveien 22, 2815 Gjøvik, Norway

Abstract

Although a number of different ear recognition techniques have been proposed, not much work has been done in the field of ear detection. In this work we present a new ear detection approach for 3D profile images based on surface curvature and semantic analysis of edge-patterns. The algorithm applies edge-based detection techniques, which are known from 2D approaches, to a 3D data model. As an additional result of the ear detection, the outline of the outer helix is found, which may serve as a basis for further feature extraction steps. As our method does not use a reference ear model, the detector does not need any previous training. Furthermore, the approach is robust against rotation and scale. Experiments using the 3D images from UND-J2 collection resulted in a detection rate of 95.65%.

1 Introduction

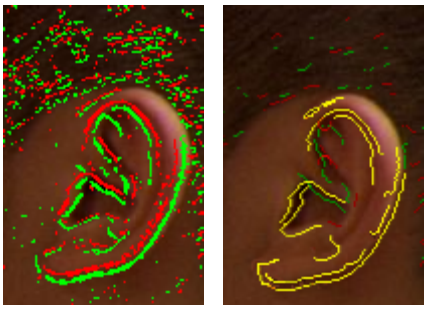
Referring back to the first large-scale study on the suitability of the ear as a biometric characteristic by Iannarelli in 1989 [7], several automated and semi-automated ear recognition systems have been proposed in literature. Since then the ear has been highly valued in forensic image analysis as an individual structure with a high distinctive potential.

The outer ear (also referred to as auricle or pinna) is a richly structured body part, which is composed of cartilage covered by a thin skin layer. Its appearance is not completely random, but rather subject to the somewhat predictable process of cell segmentation. In his work on 'earology', Iannarelli was able to show that the ear is not only unique, but also stable over its entire lifetime. Ear recognition is also more acceptable than face recognition, as people feel more comfortable when a photograph of their ear, as opposed to their face, is taken [5]. In public opinion, ear recognition, unlike fingerprints, is not associated with criminal investigations, and the fact that it does not require any physical contact with the sensor, further contributes to its acceptability.

In the field of 3D ear recognition, a large number of descriptors for ear models have been proposed. Despite this, the problem of ear detection is not addressed by many authors as they are using manually pre-segmented images. Possible solutions to the problem of ear detection have been proposed by Chen an Bhanu [4] and by Yan and Bowyer [8]. Both approaches require the availability of 2D texture images as well as corresponding 3D images. The idea is to reduce complexity by coarsely pre-segmenting the image in 2D space before locating the ear in the 3D model with a helix-template. However, Zhou *et al* showed that the ear can also be localized efficiently without the help of additional 2D texture information [9].

In this work, we introduce a technique for ear detection from a 3D profile image that detects ears from profile images without using color information or making any assumptions about orientation and scale. Our approach is inspired by edge-based 2D ear detection approaches such as [2] or [1] and relies on the fact that edges we see in 2D images are a result of the interaction between extreme curvature values on the object's surface and reflections of ambient light. The algorithm is similar to the first bottom-up ear detection approaches based on depth images by Chen and Bhanu [4]. Our approach, however, exceeds their reported detection rate of 92.4%. In contrast to Chen and Bhanu's method, our work is based on curvature values and uses a number of constraints that provide a general description of the ear.

The upcoming section presents a detailed description of our 3D ear detection approach and will be split into four subsections. Each of these subsections covers one step of the detection algorithm. Subsequently, section 3 presents the results we obtained during the examination of the algo-



(a) (b)

Figure 1. Projected curvature lines after threshold (a) and helix candidates after smoothing, thinning and removing small components (b).

rithm's performance and will also cover the strengths and weaknesses of the approach. Finally, in section 4, the findings of this paper are summarized and suggestions for further improvements are made.

2 Ear Detection Approach

The ear detection algorithm outlined in this paper can be divided into three subsequent steps, namely the curvature calculation and a preprocessing step, the closing of small gaps in the helix contour, and finally the evaluation of helix candidates. In each step, the algorithm reconstructs and combines lines in the image in such a way as to satisfy a number of conditions that determine if the line is part of the ear. In the upcoming subsections, each of these steps will be explained in more detail.

2.1 Mean curvature and binarization

Our detection approach is based on the assumption, that the ear region, with its rich and curved structure, can easily be distinguished from other regions by looking at the local curvature. Hence we need an appropriate measure for quantifying the surface curvature, and we need to assign a curvature value for each point in the 3D model. In our approach, we use the mean curvature H, which is defined as the mean value of the minimum and the maximum principle curvature k_{min} and k_{max} at a point on the surface [3].

$$H = \frac{1}{2}(K_{min} + k_{max}) \tag{1}$$

After calculating a curvature value for each point in the 3D model, all mean curvature values between a minimum value t_1 and a maximum value t_2 are removed from the set of points in the model. These threshold values should be defined according to the actual curvature values that occur in the image, such that enough points are left in the model for later analysis. Generally speaking, it is better to have too many, rather than too few points, left in the model. For UND-J2 we chose $t_1 = -0.5$ and $t_2 = 0.5$. This step removes all points with smooth curvature values, and only leaving points with large curvature $H < t_1$ and $H > t_2$ for further processing and in the point set P. As the ear is a structure with extreme surface curvature, the points representing the ear's outline will be a subset of P.

After applying the threshold all, points in P are projected on a 2D binary image according to their x and y coordinates. The depth information, which was contained in the z value, is declined in this step. Instead the points are divided into two categories, as shown in 1(a). Maximum curvature($H < t_1$) points are depicted in red and minimum curvature ($H > t_2$) are depicted in green. For the rest of this paper, the term curvature always refers to curvature of lines in 2D space, not surface curvature in 3D.

The line-like shapes in the projected 2D image are then smoothed by applying a Gaussian filter and thinned to the width of one pixel. As shown in figure 1(a), the ear contours are prominent larger structures surrounded by smaller ones, which are likely to be noise artifacts. Hence, the last step of the binarization process removes all components that are smaller than a fixed minimum value from the image. For the images in UND-J2, the minimum size for a component was set to 5.

2.2 Reconstructing the helix contour

After having made a coarse selection of the components, likely to be part of the ear contour in the previous processing step, the next step reconstructs contour lines from the remaining components in the image. The goal of this step is to reconstruct components from the image that likely belong to the same contour, but are not connected in the 2D binary image. These missing connections are often a result of occlusion by hair or cluttering by other objects, such as scarves or earrings, as large values for surface curvature cannot be found in occluded regions.

The basis for connecting lines in the image is the calculation of the vector $\vec{a_i}$, which represents the local orientation at an endpoint E_i in the image. The local orientation vector $\vec{z_i}$ is determined by the difference between the x and y coordinates of the endpoint E_i and the x and y coordinates of the nth pixel along the currently examined line (See Figure 2(a)). As the minimum size of a line is 6 pixels (all shorter lines were removed during the preprocessing step), we set the maximum search depth $n = 5$.

$$\vec{z_i} = \begin{pmatrix} x_{E_i} \\ y_{E_i} \end{pmatrix} - \begin{pmatrix} x_n \\ y_n \end{pmatrix} \tag{2}$$

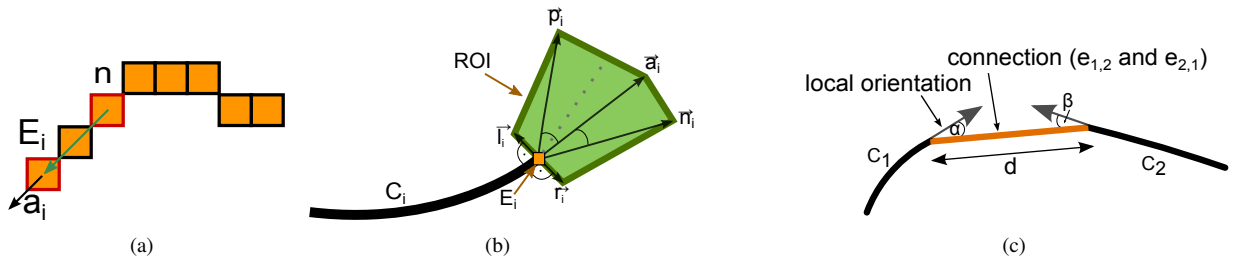


Figure 2. 2(a) illustrates the calculation of the local orientation vector. 2(b) and 2(c) show the definition of the search space used when searching for components to connect with, and an example of a connection between two components with distance d and the two angles α and β, respectively.

Finally, we get the orientation vector $\vec{a_i}$ by normalizing $\vec{z_i}$

$$\vec{a_i} = \frac{\vec{z_i}}{\|\vec{z_i}\|} \tag{3}$$

For narrowing the search space of possible connections, a region of interest (ROI) is defined for each line ending E_i of a component C_i. The ROI is a polygon spanned by the output vector $\vec{a_i}$ of E_i as well as two vectors $\vec{p_i}$ and $\vec{n_i}$ for the angular tolerance and two vectors $\vec{l_i}$ and $\vec{r_i}$ in an orthogonal direction to $\vec{a_i}$. The vector $\vec{p_i}$ is the angular tolerance in the direction of the curvature of C_i, whereas $\vec{n_i}$ is the angular tolerance in the opposite direction. As we want to construct lines that do not change their curvature, the angular tolerance for $\vec{p_i}$ is larger than the tolerance for $\vec{n_i}$. The length of the vectors $\vec{a_i}$, $\vec{p_i}$ and $\vec{n_i}$ is the maximum search distance value d_{max}, which is a predefined fixed value. The length of $\vec{l_i}$ and $\vec{r_i}$ is set to $\frac{d}{2}$ (see Figure 2(b)).

This specific shape of the ROI covers all points near the endpoint E_i while ignoring all points with a large distance. The orthogonal tolerance vectors are necessary because the ROI should also cover lines that are nearby E_i but not directly in front of it. Without the orthogonal offset, either the angular tolerance must be very broad or the maximum distance must be very large, which increases the chances for false connections during the connection step.

For each pair of possible connections, there are two angles α and β, such that $\alpha = e_{1,2} - a_1$ and $\beta = e_{2,1} - a_1$, where $e_i j$ is the vector between the endpoints and d is the distance between the endpoints. Because the connection is only evaluated if the corresponding endpoint is inside the ROI, d is always smaller than d_{max}. In case the two angles α and β are smaller than a given maximum angle, the connection is added to the list of plausible connections. Furthermore, if this list contains more than one possible connection between two endpoints of the components C_1 and C_2, all plausible connections between these points are ranked by using a quality score Q. The likelihood for a connection between C_1 and C_2 increases if α and β and the distance between the endpoints d are small. Moreover, for small values of α and β, a larger value for d may be preferred to a connection with a small d but high values for α and β (see Figure 2(c)).

$$Q = d(\alpha + \beta) \tag{4}$$

A connection between two components C_1 and C_2 is only established if C_2 has an endpoint in the ROI of E_1 (which is an endpoint of C_1) and their score Q is the smallest score of all possible connections in the Region of interest. Moreover, Q must not be larger than a given maximum score Q_{max} and it must not have any intersection points with itself.

After reconnecting components in the image, the ear's outline is now among the largest components. Because of the specific shape of the outer helix, all lines that consist of both positively and negatively curved parts are discarded, as they are unlikely to represent the helix. The algorithm only selects lines that are curved in only one direction. Therefore, the set of possible components representing the ear's outline can once again be reduced by selecting the ten largest components with a single direction of curvature as helix-candidates. In Figure 1(b), the highlighted lines are the ten helix-candidates selected from the set of lines after the reconstruction step. Note that many of the selected lines are already part of the ear. Other long and prominent lines that are frequently selected as helix candidates are hair or hair ties, the outlines of glasses or clothing. In the next step, the helix-candidates are further combined and evaluated to make sure that the only lines kept are those that form an ear.

2.3 Combining Segments and Evaluation of Helix Candidates

In the previous step, each component in the image was examined separately by selecting components due to their size and their curvature. It is, however, unlikely that the ear outline is reflected by only a single component. The algorithm therefore combines helix-candidates and selected additional components, and then estimates the likelihood that

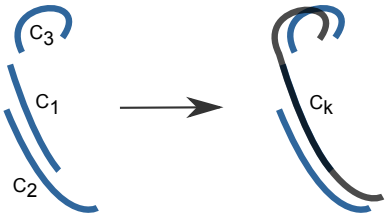


Figure 3. Combination of single shapes to create an ear candidate

the resulting shape is an ear or not. Figure 3 shows an example of the combination of three components C_1, C_2 and C_3 to form a common structure, which reflects the ear contour better than each component alone.

One specific property of the ear is that the helix contour usually consists, with respect to the 2D binary image plane,usually consists of a convex and a concave line, which are parallel in the projected 2D image. Therefore, the first combination step is to search for parallel components of helix-candidates in the image. These parallel components do not necessarily have to be helix-candidates themselves but can be chosen from all reconstructed components in the projected curvature image. The algorithm then combines non-parallel helix-candidates that are close to each other. For each of the possible combinations, a score is calculated with describes the likelihood that the combined shape represents the outline of an ear. If the score drops below a certain threshold, no ear is detected in the image. Otherwise, the combined shape with the largest score is chosen and is marked with a bounding box.

There are two different categories of criteria for evaluating the combined shape. The first category is that of absolute criteria, which are calculated individually for each combined shape. The second category consists of relative criteria, which enable the algorithm to weigh different ear candidates and mark the most likely ear region.

2.3.1 Absolute Criteria

Absolute criteria are used to make a coarse selection of possible combinations and exclude every combined shape that does not fulfill each of them. These criteria are the proportion B_{P_i}, the cumulative curvature B_{K_i}, the ratio of parallel shapes B_{R_i} and the number of corners B_{C_i}. Figure 3 illustrates the absolute criteria with the regards to a sample ear contour. The total score, which estimates the likelihood that the combined shape i is an ear, is denoted as A_i, and is the sum of each of the criteria.

$$A_i = \frac{1}{4}(B_{P_i} + B_{K_i} + B_{R_i} + B_{C_i}) \tag{5}$$

In order to prevent a combined shape from having a high score because it satisfies only one criterion extremely well, no score for a single criterion may be larger than 150% of the lowest score among the four. In this case, A_i is set to the value of the lowest score.

$$A_i = \min(B_{P_i}, B_{K_i}, B_{R_i}, B_{C_i}) \tag{6}$$

The ideal proportion of an ear is measured by calculating the ratio between the major and the minor axis of an ellipse, that encloses the current combined shape. This ratio should be between $2/1$ and $3/1$. Any deviation from these ratios decreases the total proportion score B_{P_i} for a combined shape i.

$$B_{P_i} = \begin{cases} 1 - \frac{(2 - proportion_i)^2}{4} & 0 < proportion_i < 2 \\ 1 - \frac{(proportion_i - 3)^2}{4} & 3 < proportion_i < 5 \\ 1 & 2 \leq proportion_i \leq 3 \\ 0 & else \end{cases} \tag{7}$$

Additionally, a complete outline of an ear should have an accumulated curvature of approximately 2π. Values below 2π indicate that the outline is not complete, whereas values larger than 2π indicate that the currently evaluated shape is not an ear. The accumulated curvature K_i of a combined shape i is defined as the sum of all curvature values c_j of the n pixels on the shape's outline. The curvature values on the outline of the shape are calculated by using the method proposed in [6].

$$curvSum_i = \sum_{j=1}^{n} |c_j| \tag{8}$$

Then the total curvature score is defined as

$$B_{K_i} = \begin{cases} 1 - \frac{|2\pi - curvSum_i|}{1,5\pi} & 0,5\pi < curvSum_i < 3,5\pi \\ 0 & else \end{cases} \tag{9}$$

The third criterion B_{R_i} measures the number of pixels on the outline of the combined shapes, that have parallel shapes in their neighborhood. As stated earlier, the presence of parallel lines in the projected image is an important property for the ear outline, due to of the outer ear's unique shape. The ratio of parallel components is the number of pixels contained in components that have a parallel line $p_{parallel}$ divided by the total number of pixels in the current combined shape p_{total}.

$$B_{R_i} = \frac{p_{parallel}}{p_{total}} \tag{10}$$

The last absolute criterion is the number of corners in the combined shape. This criterion is used to exclude jagged

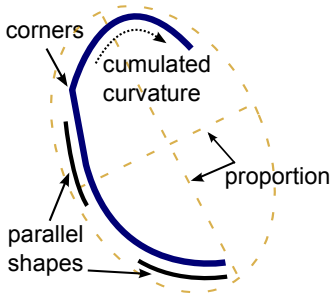


Figure 4. Visualization of the criteria for absolute score computation

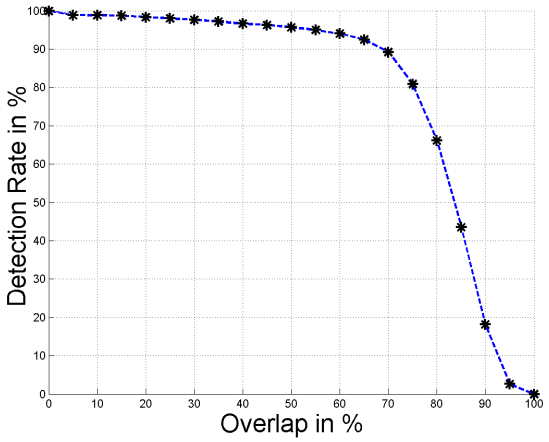


Figure 6. Detection rates for different grades of overlap between ground truth and detected region

and noisy lines, which often represent hair or clothes. The outline of the helix should be smooth and hence, no corners should be present in an optimal combined shape representing an ear, and their contribution to the total curvature of the shape is 0. In order to find corners in a given combined shape, we use the corner detector described in [6]. In order to determine B_{C_i}, the ratio between the total angle accumulated in corners θ_i and the total accumulated angle of the whole combined shape Θ_i is calculated.

$$B_{C_i} = 1 - \left(\frac{\theta_i}{\Theta_i}\right)^2 \tag{11}$$

2.3.2 Relative Criteria

Due to the ear's self-similarity, the absolute criteria can also be fulfilled by combined shapes, that only cover part of the ear. To overcome this, the algorithm additionally compares combined shapes, using relative criteria in order to select the largest and most complete combined shape. The relative score is composed of a bonus for large shapes l_i, which is the total number of pixels the connected shape consists of, and two non-linear penalty scores. g_i reflects the total distance in pixels, that had to be bridged during the reconstruction step, and m_i is the distance, in pixels, between the single components the combined shape is composed of. The exponent λ adapts the penalty score to the resolution of the projected 2D curvature image and can be a value between $[1, 2]$. For the models in UND-J2 a value of 1.2 proved to be a good choice for λ.

The non-linear weighting of the distance penalty makes sure that small distances are strongly preferred to high distances between the components. This makes the algorithm pick the largest but also most compact component in the image. Taking all these factors into account, the relative score N_i for each combined shape is

$$N_i = l_i - g_i^\lambda - m_i^\lambda \tag{12}$$

In order to be able to compare the relative scores with each other and to use them in the final evaluation of the combined shape, N_i is normalized to a value between $[0\ 1]$. $max(N)$, here, is the maximum score among all relative scores N_i for the current combined shape i.

$$R_i = \begin{cases} \frac{N_i}{max(N)} & N_i > 0 \\ 0 & N_i \leq 0 \end{cases} \tag{13}$$

Finally, each combined shape i is evaluated by combining its absolute score A_i and the relative score R_i to create the final score S_i. A_i and R_i can be weighted to change the influence of each of the scores. A_i is more sensitive in differentiating between ear-like and other combined shapes, whereas R_i makes the algorithm choose the largest shape with its components close to each other. We assigned larger weight to A_i as this criterion appeared to be the stronger of the two.

$$S_i = \omega_1 A_i + \omega_2 \frac{1}{4} R_i \tag{14}$$

3 Detection Performance

We tested our approach on the 3D models contained in the UND-J2 collection [8]. The dataset consists of 404 subjects with up to 22 samples (3D models) per subject, with a total number of 2414 images. Figure 5 shows some examples for successful and unsuccessful detection results using our algorithm. As a result of our tests, we measured a detection rate of 95.65%.

In our test setup, the detection rate is defined as the percentage of overlapping pixels between a ground truth and the ear region marked by the algorithm. The ground truth

(a) (b) (c) (d) (e) (f)

Figure 5. Examples for successful(subfigures 5(a) and 5(b)) and unsuccessful detections.

was generated by manually cropping the ear region from the images. If more than 50% of the pixels in the ground truth and the detected ear region from the algorithm overlap, we consider the detection as successful. In Figure 6 the detection rate is plotted against the minimum grade of overlap between the ground truth and the detected region. Even if the constraint for a positive detection were set to an overlap of at least 70%, 90% of all images would still comply.

In Figure 5(a) and 5(b), two examples of successfully detected ears are given. In general, our algorithm successfully detected the ear region, occasionally when it was partially occluded by hair. Our approach, however, had problems distinguishing long strains of hair, located near the ear region from the actual outline of the auricle (see Figure 5(c) and 5(d)). In the binarized image, long hair is reflected by prominent and smoothly curved lines, which look similar to the shape of the outer helix. For example, in 5(c), the marked region is too large because a strain of hair is included in the set of shapes, that represent a typical ear. This behavior is caused by the relative score, which tends to favor the largest ear-like shape in the image. If there is a strain of hair, that fits the absolute criteria and expands the marked region, the ear detection considers it as a part of it.

In some cases the algorithm considers clothes, such as scarves and collars, as the ear region. This happens if the auricle's outline is not sufficiently represented in the curvature image, which is mostly caused by occlusion or missing data. In these cases, the detector incorrectly uses smoothly curved and parallel lines from other parts of the image, such as long hair strains or cloth.

4 Conclusion

Our method for ear detection based on surface curvature delivered promising results and outperformed comparable edge-based methods, such as the one proposed by Chen and Bhanu. The ear detection works reliably and is also capable of delivering valuable information for a subsequent feature extraction step. We are optimistic, that the performance of our method can be further improved by including the Tragus location and edge images, which are calculated from the corresponding 2D image.

Furthermore we plan to refine the measure for the accumulated curvature for more accurate curvature estimation. We will also conduct more experiments on additional databases in order to evaluate the robustness against pose variations. In addition, a feature extractor which uses the helix outline, which is returned by our ear detection algorithm, will be developed.

References

[1] S. Ansari and P. Gupta. 'Localization of Ear Using Outer Helix Curve of the Ear'. In *International Conference on Computing: Theory and Applications (ICCTA)*, pages 688 – 692, March 2007.

[2] S. Attarchi, K. Faez, and A. Rafiei. 'A New Segmentation Approach for Ear Recognition'. In *Advanced Concepts for Intelligent Vision Systems*, volume 5259 of *LNCS*, pages 1030 – 1037. Springer Berlin / Heidelberg, 2008.

[3] P. J. Besl. *'Surfaces in Range Image Understanding'*. Spinger, 1988.

[4] B. Bhanu and H. Chen. *Human Ear Recognition by Computer*, chapter 3D Ear Detection from Side Face Range Images, pages 21 – 59. Springer, 2008.

[5] M. Choras. Image pre-classification for biometrics identification systems. In *Advances in Information Processing and Protection*, pages 361 – 370. Springer US, 2008.

[6] X. He and N. Yung. 'Corner Detector based on Global and Local Curvature properties'. *Optical Engineering*, 47(5), 2008.

[7] A. V. Iannarelli. *'Ear identification'*. Paramont Publishing Company, 1989.

[8] P. Yan and K. Bowyer. 'Biometric Recognition Using 3D Ear Shape'. *Pattern Analysis and Machine Intelligence*, 29:1297 – 1308, August 2007.

[9] J. Zhou, S. Cadavid, and M. . Abdel-Mottaleb. 'Histograms of Categorized Shapes for 3D ear detection'. In *Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, November 2010.

Acknowledgement

This project is funded by the Federal Ministry of Education and Research (BMBF) of Germany in the context of the research programme for public safety and security.