Robust Localization of Ears by Feature Level Fusion and Context Information

Anika Pflug, Adrian Winterstein, Christoph Busch
da/sec - Biometrics and Internet Security Research Group
Hochschule Darmstadt, Germany
{anika.pflug, adrian.winsterstein, chrostoph.busch}@cased.de

Abstract

The outer ear has been established as a stable and unique biometric characteristic, especially in the field of forensic image analysis. In the last decade, increasing efforts have been made for building automated authentication systems utilizing the outer ear. One essential processing step in these systems is the detection of the ear region.

Automated ear detection faces a number of challenges, such as invariant processing of both left and right ears, as well as the handling of occlusion and pose variations. We propose a new approach for the detection of ears, which uses features from texture and depth images, as well as context information. With a detection rate of 99% on profile images, our approach is highly reliable. Moreover, it is invariant to rotations and it can detect left and right ears. We also show, that our method is working under realistic conditions by providing simulation results on a more challenging dataset, which contains images of occluded ears from various poses.

1. Introduction

The observation of the outer ear, which is frequently referred to as auricle, is an emerging biometric method, which has drawn the attention of research during the last years. In forensic investigation, ear prints on doors and windows can be collected and are used as a means for identifying perpetrators [1].

The outer ear is believed to be just as unique as the face. An extensive study of Iannarelli [6] and a more recent study from India [12] show, that the outer ear possesses numerous characteristics, which make each ear unique. In Fig.1, we have annotated a number of features, which are used by the German Criminal Police Office for identifying a subject.

Forensic investigators do not only value the uniqueness of the outer ear, but also its permanence. In contrast to the face, the structure of the outer ear remains stable after the 6th month of life [6]. A recent study by Ibrahim *et al*. [7] confirms that the recognition rate of a biometric system is not affected considerably over eleven months.

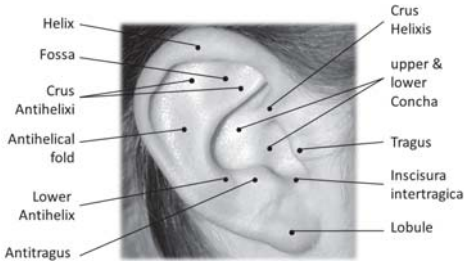

Figure 1. Selected features of the outer ear.

Due to the proximity of the observed physiological characteristics, ear recognition may be considered as a valuable extension to face recognition systems. Ear and face images can be collected with the same hardware. Especially in unconstrained scenarios, such as video surveillance, the outer ear can contribute additional features, which increases the chance of identifying a person in off-pose images.

The contribution of this paper is a novel ear detection algorithm, which uses texture and depth images for localizing the ear in full profile as well as in off-pose images. We utilize the rich details on the ear surface and of edge images for determining the ear outline in an image. We present a set of flexible rules, which allow us to distinguish between the ear outline and other objects in the image. These rules describe an abstract ear model and include context information. Our algorithm is invariant to rotations and it can detect left and right ears with the same parameter set. Moreover it is robust to pose variations and occlusion. The feasibility of the proposed ear detection system is shown by providing simulation results on the UND-J2 database [13] as well as UND-NDOff-2007 [5]. A detection rate of 99% on the UND-J2 dataset shows that our approach is outperforming other recent work. Moreover, we also show that our method has the ability to detect ears under realistic conditions, where it has to handle occlusions and pose variation.

In the upcoming section, we describe the state of the art in ear detection in 3D images and describe related work.

Subsequently, we introduce the proposed algorithm. In section 4 we point out the experimental setup and present simulation results using the previously mentioned datasets. Finally, conclusions are drawn and an outlook on future work is given in Sec. 5.

2. Related Work

In contrast to 3D meshes, depth images have a matrix-like data structure and they can be acquired along with the texture image with a single capture process. If depth images are recorded under controlled conditions, they are co-registered with the texture image, which makes it easy to combine texture and depth information. This fact has inspired a number of researchers to come up with different approaches, which use depth and texture information for ear detection and recognition. Many of these early methods are developed and tested on the public dataset collected by the University of Notre Dame, such as the collections F and G, and especially collection UND-J2 [13] which is the largest publicly available database for ear recognition. It consists of texture and depth images of left profiles. In Tab. 1, we compare the detection rates of recent ear detection algorithms, which were tested on the UND-J2 dataset.

Yan and Bowyer [13] propose a full biometric ear recognition system based on the profile images of the UND-J2 collection. They first locate the nose tip and then use skin color from the texture image for narrowing the search space for the ear region. Subsequently the lower concha is detected by using curvature information from the corresponding depth image. The final ear region contour is fixed by using an active contour approach.

Chen and Bhanu [4] also combine texture and depth images in their detection approach. First, they create different image masks from skin color and depth information. Then this mask is used for selecting edges from the depth and the texture image. The ear detection is completed by clustering these edges and selecting the largest cluster of edges in the superimposed image.

In the ear recognition system proposed in [8], Islam *et al*. use Haar-like features for building an ear classifier for 2D images. Because the texture and the depth images are co-registered, the detected ear region in the texture image and in the depth image have the same position.

More recently, several ear detection approaches were proposed, which are exploiting the properties of the detailed surface structure of the auricle. Zhou *et al*. [14] extract local histograms of categorized shapes from a sliding window and use a SVM for deciding whether a local histogram represents an ear or not. Subsequently the largest cluster of positive detections is selected as the ear region. In their paper, Zhou *et al*. only provide results on a subset of UND collection F, where this algorithm achieved a detection rate of 100%. However, the detection rate drops significantly, when when the ear is rotated by more than 10 degrees [9].

Another class of ear detection algorithms are approaches, which use specific line-like patterns for localizing the outer ear. Prakash *et al*. [11] define edges in the 3D image as regions with maximum depth variation. Based on his assumption, an edge map is created from the depth image. Subsequently, the local orientations of the edges are approximated with vectors. These vectors serve as the edges of an edge connectivity graph. Subsequently their algorithm generates potential ear candidates from these graphs and then selects the final ear region by comparing the each candidate with a reference template.

In [10] ears are detected using the specific distribution of surface curvatures in the ear region. This results into convex and concave edges, which are then processed in a number of subsequent steps for combining the multiple neighboured edges to ear candidates. According to its proportion, size, redundancy and cumulated slope, a score is assigned to each candidate. The final ear region is defined by the circumference of the ear candidate with the highest score. This approach, however, has some limitations if the depth image is noisy or if the ear contour is on the verge of the depth image. In these cases, no meaningful curvature information can be extracted from the depth image and the detection fails. In Tab.1, the reported detection accuracy of the cited previous work is summarized and compared to the proposed detection method. With a detection accuracy of almost 99% on the UND-J2 dataset, the proposed ear detection approach is not only competitive to the graph based approach by Prakash *et al*. with respect to the detection performance - moreover the proposed method has no need to exclude challenging samples from the evaluation as it was done in [11].

3. Ear Detection System

The proposed ear detection algorithm utilizes the fact, that the surface of the outer ear has a delicate structure with high local curvature values. In some depth images, however, some parts of the ear are missing, because curvature can only be measures between neighbouring points. However, many depth images have holes next to the outer helix, which results in missing curvature values [10] (see Fig. 2 for an example). We solve this problem by fusing co-registered edge and depth images.

Our ear detection approach consists of four different steps, which are illustrated in Fig.2. We start with a preprocessing step, where edges and shapes are extracted from the texture and the depth image. Subsequently, the edges and shapes are fused in the image domain. In the next step, the components are combined with each other to ear candidates and a score for each ear candidate is computed. Finally, the enclosing rectangle of the best ear candidate is returned as the ear region.

Table 1. Comparison between the proposed method and previously achieved results on UND-datasets.

Author	Performace	Remarks
Yan and Bowyer[13]	97.38% rank-1	UND-J2 collection, depth and color images, no details about detection accuracy
Chen and Bhanu [4]	87.11%	UND collection F and subset of collection G, depth and color images
Prakash *et al*. [11]	99.38%	Subset of UND-J2 collection, depth images only, 10% of images excluded
Zhou *et al*. [14]	98.22%	Results on UND-J2 based on re-implementation in [9], depth images only
Islam *et al*. [8]	100%	203 profile images from UND-J2 database, 2D images only
Pflug *et al*. [10]	95.65%	UND-J2 collection, depth images only
Proposed method	99%	UND-J2 collection, depth and color images

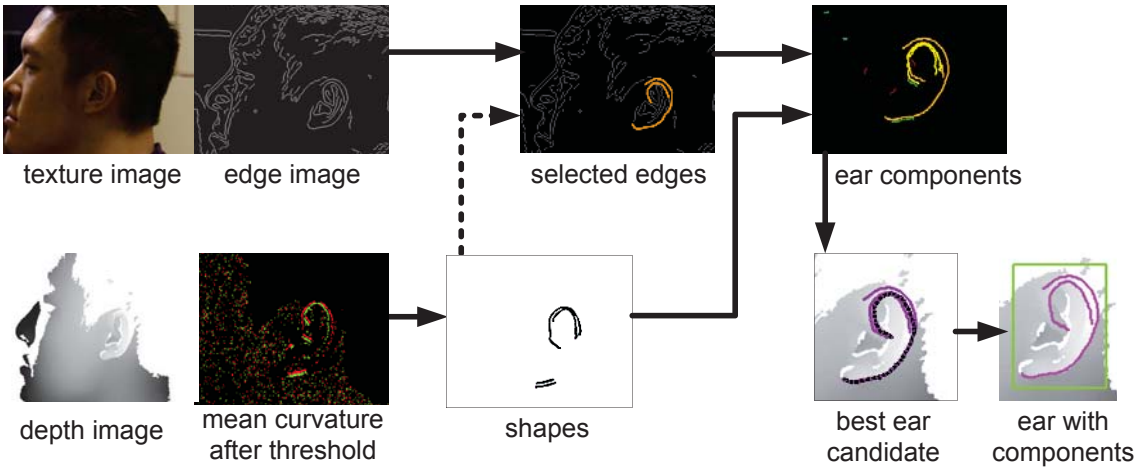

Figure 2. Illustration of the processing steps of the proposed ear detection system (database images taken from [13]).

3.1. Preprocessing

The outcome of the preprocessing step is an edge image from the texture, which is created using a Canny edge detector [3]. Furthermore, we determine a number of shapes from the depth image, which serve as the basis for the subsequent fusion step.

We first calculate the mean curvature [2] of the depth image. The key points on the ear surface have large convex an concave curvature values. Therefore, we apply a threshold to the mean curvature image in order to keep only large convex and concave curvature values. We now delete all connected components from the image, which are smaller than 3 pixels. As a result, we get a number of blobs, which are mainly located in the ear region. These blobs are thinned to a width of 1 pixel and subsequently re-connected using the method proposed in [10]. After the reconnecting step, a limited number of lines is left from the mean curvature image. For the remainder of this paper, these lines are referred to as shapes.

3.2. Fusion of Texture and Depth Information

In the fusion step, we select a number of edges from the edge image, based on the vicinity to the position of the most prominent shapes and other criteria. First, we select the ten longest shapes from the mean curvature image.

If the sign of the local orientation of one of the selected shapes changes, the shape is divided into two shorter shapes. Shapes can also be split up if they contain a corner. The two resulting shapes may be declined, if there exists another shape in the curvature image, which is longer than any of the two divided shapes. The result of this procedure is a set of the longest shapes, which are smoothly curved.

Each shape in the set is dilated with a circular structuring element. After the dilation, each shape represents a region, which is not used for selecting appropriate edges from the edge image. As shown in Fig.3, we select all edges, that exceed a predefined threshold with their intersecting points in the dilated region. For edges, which have an endpoint inside the dilated region, the minimum number of intersecting pixels should be smaller than for other edges. This is

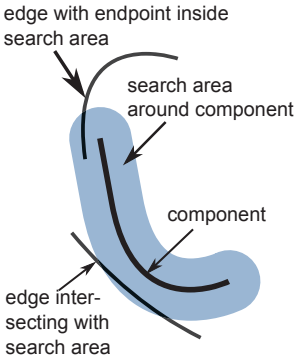

Figure 3. Fusion of 2D and 3D data. Shapes from the depth image are dilated and intersecting edges are selected.

due to the fact, that edges with intersecting endpoints play a special role in the upcoming combination step and also lead to better results in the scoring step.

All intersecting edges are added to the set of components. In case, an intersecting edge has a corner or if the sign of the local orientation changes, it is divided into to smaller edges. A divided edge, that does not have an intersecting point with the dilated region, is removed from the set of components.

3.3. Combination of Components

In the combination step, all components are combined with each other or smaller shapes nearby in order to obtain complete ear outlines. In a first combination round, we only combine components with each other. In a second round, we also allow other shapes and edges, that were not selected as components to be combined with ear candidates. This is done by picking a component and then combine it with other components. A component, that has been combined with other components is considered an ear candidate. Thus we create a new ear candidate, each time a component is linked with one of the existing ear candidates.

During the creation of new ear candidates, a component can be adapted to the ear candidate in a number of ways. The component can either be translated, pruned or interpolated in order to fit to the ear candidates. In Fig. 4, an example for the combination of an ear candidate and two components is shown. When $comp_1$ is combined with the ear candidate, we have to translate the contributing component $comp_1$ and fill remaining gaps. In the second step, $comp_2$ is translated and pruned.

Each time, a new ear candidate is created, we compute a score, that describes the fitness of this candidate (see Sec. 3.4.3 for more details on the scoring system). However, without any additional constraints, we would have to combine all components with each other. Many of the created ear candidates would be redundant and the detection would

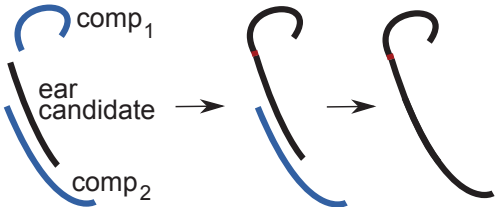

Figure 4. Iterative creation of an ear candidate though stepwise combination and adaptation of components.

be inefficient. Therefore, we introduce a terminating criterion, which prevents the algorithm from doing an exhaustive search through all possible combinations between ear candidates and components.

Let $maxScore$ be the best score, that has been achieved by any ear candidate for all components. A newly created candidate will only be used in subsequent combination steps, if its score is higher than $0.7 * maxScore$. An ear candidates with a lower score will be discarded and is not used in subsequent iterations. In the first iteration, it is very likely that an ear candidate will satisfy this condition and a large number of new candidates will be creates. However, as the combination step proceeds further, an increasing number of new candidates will be discarded. The more iterations have been completed, the higher the probability, that an ear candidate with a high score has been created and the more candidates are dropped.

3.4. Scoring system for ear candidates

Each time an ear candidate and a component are combined, we assign a fitness score to the newly created ear candidate. The score reflects the similarity of the ear candidate with an idealized ear outline. This similarity is expressed through a number of criteria, a good ear candidate has to comply with. The fitness score is composed of three different components, which reflect different properties. We distinguish between the individual score, that is computed for each ear candidate, a relative score and a context score.

After the combination step, we select the ear candidate with the highest score return its bounding box as the detected ear region. If we cannot find an ear candidate with a larger score than 0.5, we consider the ear to be occluded and do not return an ear region.

Let I_i be the individual score of the ith ear candidate, R_i the relative sore and C_i the context score. Furthermore, let ω_1, ω_2 and ω_3 be weighting factors for each of the scores. The total fitness score of the ith ear candidate, denoted by S_i can be expressed as follows:

$$S_i = \omega_1 I_i + \omega_2 R_i + \omega_3 C_i \quad (1)$$

The values for ω should be adapted according to the variance of each partial score I, R and C. The larger the vari-

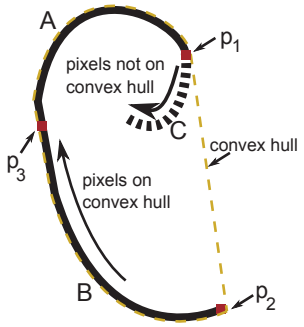

Figure 5. Calculation of the sum of local orientations using the convex full of an ear candidate.

ance between single ear candidates, the higher the weight of the according partial score.

3.4.1 Individual Score

The individual scores consists of three components, which measure the cumulated sum of local orientations, the contribution to the sum of local orientations in corners and the proportion.

The sum of local orientations reflects the expectation, that an ear candidate should be convex and it should have a gap (connecting line between $p1$ and $p2$ in Fig. 5). We first compute the convex hull of an ear candidate and define two points p_1 and $p2$. The points p_1 and $p2$ are the points in $A \cup B$ with the maximal distance, such that all pixels on the connecting line $\overline{p_1 p_2}$ between them are on the convex hull but not in $A \cup B$.

As shown in Fig. 5, p_1 and $p2$ are on the opposite side of the outer helix. We can define a third point p_3, which divides the ear candidate pixels on the convex hull into two equally sized subsets A and B. Finally, we distinguish between those pixels of the ear candidate, that are on the convex hull and those that are not. The set of pixels, that are not on the convex hull is denoted as C.

Let sum_{hull} be the sum of all local orientations in $A \cup B$. If we have a good ear candidate, this sum is expected to be larger than 1.5π.

$$onHull = \begin{cases} \frac{sum_{hull}}{1.5\pi} & sum_{hull} \leq 1.5\pi \\ 1 & otherwise \end{cases} \quad (2)$$

We also expect a good and complete ear candidate to be symmetric. This can be expressed by comparing the sum of local orientations in A, denoted as sum_A and B, denoted as sum_B, with each other. The division by π is necessary for getting a normalized result between 0 and 1.

$$ratio = \begin{cases} \frac{|sum_A - sum_B|}{\pi} & |sum_A - sum_B| \leq \pi \\ 1 & otherwise \end{cases} \quad (3)$$

Based on the curvature sum of all pixels in $A \cup B$, we also calculate a weighting factor, denoted as λ, for the score contribution of $onHull$ and $ratio$.

$$\lambda = \begin{cases} \frac{1}{onHull} & onHull \leq \frac{1}{3} \\ 1 & otherwise \end{cases} \quad (4)$$

The outer helix score I_o, that measures the fitness of an ear candidate with respect to its cumulated sum of local orientations is composed of the previously defined coefficients $onHull$ and $ratio$, the weighting factor λ and a fourth component, which is a penalty score. For a good ear candidate, the sum of local orientations in C, denoted as sum_C, should be as small an possible. We hence subtract $\frac{sum_C}{20}$ from the total fitness score. The value of the denominator has been obtained empirically.

$$I_o = (1 - \lambda) * onHull + \lambda * ratio - \frac{sum_C}{20} \quad (5)$$

The larger the different between sum_A and sum_B, the higher higher the influence of $ratio$ and the lower the influence of $onHull$. This reflects the fact, that incomplete ear candidates with a small sum of local orientations should get a better score than those with sum of local orientations close to 2π. The algorithm will hence be less strict with incomplete ear candidates than with complete ones.

In addition to the measure of the distribution of local orientations on the ear candidate, we also require, that a good ear candidate should have as few corners as possible. Corners are an indication for jagged components or failures during the combination step. We hence compare the ratio between the sum of local orientations in $A \cup B \cup C$, denoted as sum_{ABC}, and the sum of local orientations at all corners of the component θ. For a good ear candidate, θ should be as small as possible. In order to increase the impact of this criterion in cases where θ is large, we use the quadric function.

$$I_c = 1 - \left(\frac{\theta}{sum_{ABC}} \right)^2 \quad (6)$$

In the last criterion for the individual score, we assure that the ratio between the major and the minor axis of a surrounding ellipse should be between $2:1$ and $3:1$. Let ρ be the ratio between the major and the minor axis of an ear candidate. The proportion score I_p decreases faster with larger deviations from the ideal ratio.

$$I_p = \begin{cases} 1 - \frac{(\rho-2)^2}{4} & 0 < \rho < 2 \\ 1 & 2 \leq \rho \leq 3 \\ 1 - \frac{(\rho-3)^2}{4} & 3 < \rho < 5 \\ 0 & otherwise \end{cases} \quad (7)$$

All components of the individual score are normalized values between 0 and 1, whereas higher scores represent better ear candidates. The individual score for the ith ear candidate I_i is the mean value of the three components for this candidate I_{o_i}, I_{c_i} and I_{p_i}.

3.4.2 Relative Score

The relative score compares different ear candidates with each other and is calculated in two steps. This score rewards ear candidates, if they are composed of long neighboured components. First, we calculate a base score, called η. The base sore s based on the total length of the ear candidate in pixels l. Because this candidate was built by reconstruction the ear outline, we subtract the number of pixels that had to be filled in during the combination step g and the sum of all distances between all components the current ear candidate is composed of. This sum is denotes as m.

$$\eta = l - g - m \tag{8}$$

Subsequently, we normalize this score by dividing it by the maximum value of η for all the ear candidates in the image. The usage of the quadric function ensures, that ear candidates, that only have a slightly smaller score η than the maximum η are rejected. The relative score for the ith ear candidate R_i in the fused image is defined as follows:

$$R_i = \begin{cases} (\frac{\eta_i}{\max(\eta)})^2 & \eta_i > 0 \\ 0 & otherwise \end{cases} \tag{9}$$

3.4.3 Context Score

The scoring system is completed with an estimation of the ear candidate's size in relation to the size of the silhouette of the head in the depth image. We assume that a good ear candidate is located in the head region of the image and that the ratio between the height of the ear and the diameter of the face should be within fixed bounds. These bounds depend on the image scale and should be set individually for each capture device. We denote the lower bound for the ratio between h and d as τ and the upper bound as υ, respectively.

Moreover, let h be the largest distance between any two points of the ear candidate. Furthermore, d is the distance between the edge of the silhouette and the point in the middle of line segment $\overline{p_1 p_2}$ (see Sec.3.4.1). As shown in Fig.6, the face diameter d is measured orthogonally to the line segment $\overline{p_1 p_2}$. For all yaw poses, the outer ear is located at the back of the head and d should be pointing towards the nose.

We hence assume that for a good ear candidate, d_i should be among the longest diameters for all ear candidates, denoted by D. If d_i is smaller than the mean value of the larger half on D, the ear candidate is rejected. If not, we compute the ratio between h_i and d_i.

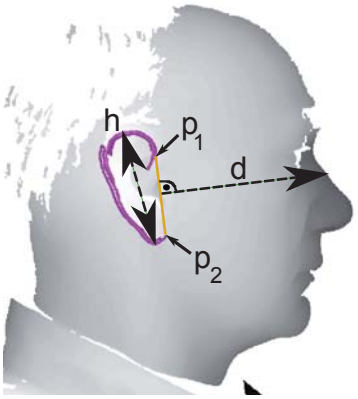

Figure 6. Estimation of optimal size using context information from the silhouette (Depth image taken from [13]).

$$cr_i = \frac{h_i}{mean(\{d_i \in D \quad | \quad d_i > 0.5 \max(D)\})} \tag{10}$$

Dependent on cr_i, we can now compute the context score C_i according to equation 11. Note that the context score decreases faster if cr_i is too small. This reflects that the outer ear can be relatively large, compared to the face diameter. We hence prefer to keep large ear candidates. If the ear candidate is too small, however, it should receive a low score.

$$C_i = \begin{cases} 1 & \tau \leq cr_i \leq \upsilon \\ 1 - 2(\tau - cr_i)^2 & cr_i > \tau - \frac{1}{2} \\ 1 - \sqrt{cr_i - \upsilon} & cr_i < \upsilon + 1 \\ 0 & otherwise \end{cases} \tag{11}$$

4. Experimental Setup and Results

For obtaining the detection performance of our approach, we conducted two experiments on two different datasets. In the first experiment, we evaluate the impact of the image domain fusion. This experiment is conducted on the UND-J2 dataset [13] and on the UND-NDOff-2007 dataset [5]. In the second experiment, we show the robustness of our approach to rotation and flip.

The UND-J2 collection contains 1776 unique left profile images from 404 different subjects (-90 degrees yaw pose). Four images had to be removed from the database, because their texture and the associated depth images did not belong to the same subject. The UND-J2 dataset also contains some images, which are off pose. However, we did not exclude these images from our test set.

The UND-NDOff dataset was originally collected for the purpose of face recognition research. For an ear detection

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

Figure 7. Subfigures (a)-(e) show examples for successful detections for left and right ears in images with pose variation, partial occlusion, missing depth data and rotation. Subfigure (f) shows an example, of an ear candidate with missing components and subfigure (g) shows a detection failure (Original images were taken from[13] and [5])

.

system, it represents a more realistic, but also more challenging scenario, than the profile views of UND-J2. We selected 2785 images with yaw poses between ± 90 and $\pm$ 30 degrees, whereas yaw poses of ± 90 are profile images and yaw poses of ± 30 are half profiles. In some cases the data collection contains different combinations between yaw and pitch poses. If different pitch poses are available for a given yaw pose, all pitch poses were included to the test set. The UND-NDOff-2007 dataset contains images where the ear is partly or fully occluded by hair and some subjects are wearing earrings. There is also a small number of images, were the subject has moved during the acquisition, which results in poor image quality.

The detection rates are calculated on the basis of manually created ground truth data. For creating the ground truth, we marked the ear region with a surrounding rectangle, and stored the coordinates of the upper left and the lower right corner. A detection for a given image pair is considered to be successful, if the overlap O between the ground truth pixels G and the pixels in the detected region R is at least 50%.

$$O = \frac{2|G \cap E|}{|G| + |E|} \tag{12}$$

4.1. Impact of Image Domain Fusion

In the first experiment, we show the impact of the image domain fusion and the context score on the detection results. In Fig.7(a) - 7(c), some examples for detected ears with partial occlusion and different poses are shown. The detection rates in Tab. 2 show, that the proposed ear detection algorithm is robust to pose variation. Although the detection rate drops, with increasing deviation from ± 90 degrees, it still detects more than 75% of the ears correctly, if the yaw pose is ± 45 degrees. This also includes images, where the ear is partially occluded, as shown in Fig.7(b) and images, where the algorithm correctly recognizes, that the ear is occluded.

Image domain fusion and the usage of context information generally improve the detection rate of the proposed method. The improvement gets more significant with larger deviations from ± 90 degrees yaw pose. We can extract

Table 2. Comparison between the detection rates with and without image domain fusion on UND-J2 [13] and UND-NDOff-2007 [5].

dataset	**yaw pose**	**depth only, no context**	**fused with context**
UND-J2			
	-90	92,9%	99%
UND-NDOff-2007			
	-90	86,9%	96,5%
	-60	70,9%	83,5%
	-45	50,5%	76,5%
	-30	23,7%	58,9%
	30	19,8%	42,7%
	45	49,4%	76%
	60	86,1%	85%
	90	91,8%	93,5%

good ear candidates form profile images, even without image domain fusion. Moreover, the number of ear candidates, that get rejected though the context score is small. With larger pose variations, the probability increases, that the 3D data in the ear region is of low quality and hence that many shapes from other image regions are selected. Further, we get an increasing number of false ear candidates from the depth image. By using image domain fusion an context information, we can give preference to the correct ear candidate. From this we an conclude, that the usage of context information substantially contributes to the algorithm's robustness to pose variation.

In some cases, the detected ear region is too small (see Fig.7(f)), because the algorithm fails to find all necessary ear components. This happens, when the number of shapes from the depth image is not sufficient or if the edges in the texture images are interrupted. This issue, however, can be addressed by allowing the algorithm to choose more shapes from the depth image before starting the image domain fusion.

Especially for images of yaw poses ± 30 degrees, there

Table 3. Results on rotated and flipped images from UND-J2 [13].

image orientation	detection rate
No rotation	99%
180 degrees	98,6%
Mirrored	99%
90 degrees clockwise	99%
90 degrees anticlockwise	98,8%

is another common type of error. Fig.7(g) illustrates an example. This error is mainly responsible for the drop in the detection rate at ± 30 and occurs if the ear is not visible in the image. In these cases, the algorithm selects shape from the nose or the eye region and creates ear candidates from them. Often, these ear candidates are rejected because of their low context score, but it happens that they are good enough for not being rejected. If there is no better ear candidate available, the algorithm will then mark it as the ear region.

4.2. Rotation Invariance

For evaluating the rotation invariance, we rotated the images from the UND-J2 dataset by 90 degrees clockwise and anticlockwise. Furthermore, we have run simulations on images rotated by 180 degrees and on images, that have been mirrored (also referred to as vertical flip). As it can be seen in Tab.3, the detection rate stays stable for all rotations. Two examples for successfully detected ears in rotated images are shown in Fig.7(d) and 7(e).

The detection rates for left and right profiles (yaw pose ± 90 degrees) in Tab.2 in connection with the results on the mirrored images, stresses, that the proposed method can be used for left and right ears, without changes in the parameter set.

5. Conclusion

In this paper, we have presented a new approach to ear detection, which proposes a scoring system for ear components, that are derived from co-registered texture and depth images. The proposed method utilizes the distribution of local orientations, the length of components and context information for detection of the outer ear in images from multiple poses and for left and right ears. Moreover, we have shown that our algorithm is invariant to rotation and robust to partial occlusion, while maintaining the same detection accuracy than previous work.

Our algorithm does not only localize ears, but also estimates their orientation, which is important for normalization. In the future, we plan to use the ear outlines, which are a side-product of the localization approach as a basis for normalization and feature extraction. At the same time, we are planning to improve the robustness to pose variation, by conducting more experiments on context information. Additionally, we plan to improve the throughput by exploiting the fact, that the combination step can easily be parallelized.

References

[1] I. Alberink and A. Ruifrok. 'Performance of the FearID earprint identification system'. *Forensic Science International*, 166(2-3):145 – 154, 2007.

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

[3] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679 –698, nov. 1986.

[4] H. Chen and B. Bhanu. 'Human Ear Recognition in 3D'. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):718 –737, April 2007.

[5] T. Faltemier. Rotated profile signatures for robust 3d feature detection. In *8th IEEE International Conference on Automatic Face & Gesture Recognition*, 2008.

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

[7] M. Ibrahim, M. Nixon, and S. Mahmoodi. 'The effect of time on ear biometrics'. In *International Joint Conference on Biometrics (IJCB)*, pages 1 –6, October 2011.

[8] S. Islam, M. Bennamoun, and R. Davies. 'Fast and Fully Automatic Ear Detection Using Cascaded AdaBoost'. In *Applications of Computer Vision, 2008. WACV 2008. IEEE Workshop on*, pages 1 –6, jan. 2008.

[9] A. Pflug, P. Back, and C. Busch. Towards ear detection that is robust against rotation. In *The 46th Annual IEEE International Carnahan Conference on Security Technology*, 2012.

[10] A. Pflug, A. Winterstein, and C. Busch. Ear detection in 3d profile images based on surface curvature. In *Proceedings of IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2012.

[11] S. Prakash and P. Gupta. 'An Efficient Technique for Ear Detection in 3D: Invariant to Rotation and Scale'. In *The 5th IAPR International Conference on Biometrics (ICB)*, 2012.

[12] P. Singh and R. Purkait. 'Observations of external earAn Indian study'. *HOMO - Journal of Comparative Human Biology*, 60(5):461 – 472, 2009.

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

[14] J. Zhou, S. Cadavid, and M. . Abdel-Mottaleb. 'Histograms of Categorized Shapes for 3D ear detection'. In *International Conference on Biometrics: Theory Applications and Systems*, 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.