An Efficient 3D Facial Landmark Detection Algorithm with Haar-like Features and Anthropometric Constraints

Martin Böckeler, Xuebing Zhou

CASED - Center for Advanced Security Research Darmstadt
martin.boeckeler@googlemail.com
xuebing.zhou@cased.de

Abstract: In the last few years 3D face recognition has become more and more popular due to reducing cost of scanners and increasing computational power. The crucial and time-consuming step is landmark localization and normalization of facial surface. Due to acquisition, noise and other artifacts like spikes and holes occur. Most systems require computational intensive preprocessing steps to eliminate these artifacts. As a consequence, a trade-off between runtime or detection accuracy must be made. In contrast, we propose a landmark detection algorithm which uses the Viola & Jones classifier on gradient images. The algorithm is able to reliably detect landmarks in raw 3D data without complicated preprocessing. Additionally, selection of sub regions is exploited to limit search regions. It further reduces false detection rate and improves significantly detection accuracy.

1 Introduction

The field of biometric includes a wide range of applications driven approaches to automatically identify or verify a person. Different biometric characteristics like fingerprint, vein, hand geometry, iris, etc. can be used. The disadvantages of these modalities are that the subject usually has to cooperate with the system to gather a useful probe. Among those, the face is a very important biometric modality. The main advantages are that the subject doesn't have to make direct physical contact with system and that face recognition is well accepted by users. Common 2D face recognition is sensitive to illumination changes and variation of pose and expression. 3D face recognition is more robust to illumination, because the surface of the face is more illumination invariant than the texture. Moreover, it can better overcome pose changes than 2D face recognition by correcting poses into a normalized position.

The accuracy of normalization has a strong influence on the recognition performance. The localization of landmarks is the key point in this process. Because of sensor noise and acquisition artifacts, the first step in 3D face recognition is a complex preprocessing to filtering noise and smooth facial surfaces. The preprocessing can increase detection accuracy, however it also enlarges the runtime. Therefore, many systems have to make a compromise between runtime and accuracy. This paper introduces a new landmark detection algorithm that achieve a high accuracy without complicated preprocessing.

This paper is organized as follows: Section 2 gives an overview of the existing techniques in the field of 3D landmark detection. Section 3 shows the details of the algorithm and also provides some background knowledge. Section 4 analyses the evaluation results. Section 5 gives conclusions of the paper and an outlook to the future.

2 The Existing Techniques

Falling sensor prices and increasing computational power have made the domain of 3D face recognition more and more popular in the past few years. Since long, 3D face recognition is an active research area. An early approach to recognize faces in 3D data was already done by using profile planes of the face in 1989 [CLR89]. In this early attempt, an iterative process extracted the profile planes from the range data using Gaussian curvature analysis. The reported performance reaches 100% with 18 datasets. Dibeklioglu et al. [DSA08] presented an algorithm based on statistical and heuristic localization. The statistical method using local features to determine the most likely location for each landmark. The localization of the nose tip was done in a heuristic way via curvature analysis. The cross database accuracy was between 48% and 100%. Perakis et al. [PPTK10] presented various methods for 3D landmark detection that were suitable to detect landmarks from frontal and side facial scans. They used local shape descriptors based on the shape index, the extrusion map and spin images. Detected landmarks were classified and labeled with help of a facial landmark model that was derived from the statistical mean shape of manually annotated landmarks. The reported cross database performance was between 83% and 97%.

The algorithm proposed in this paper is based on an existing approach of Ajmal Mian [Mia11]. Mian used gradient images and range images and detect landmarks with the Viola & Jones detection algorithm. The gradient images were directly generated from the raw 3D data. Classifier detected landmarks in the x-gradient, the y-gradient and the range images separately. Multiple landmark candidates were detected and clustered. The remaining candidates were further filtered due to a priori information of facial topology, namely the triangle of the two outer eye corners and the nose tip. Anthropometric constraints of the face were used to eliminate incorrect triangles. The performance was evaluated on 4007 face scans and reaches 99.9%.

3 An Efficient Algorithm based on Gradient Images

In this section we will show the details of the proposed algorithm. The algorithm aims to detect 10 important facial landmarks. These are the outer eye corners, the center of the pupils, the inner eye corners, the sides of the nose and the mouth corners, see figure 1.

The basic idea of this algorithm is to train the classifier for different landmark with the Viola & Jones object detection approach [VJ01]. Because the Viola & Jones is based on Haar liked features, the 3D data first has to be transformed into special gradient images to

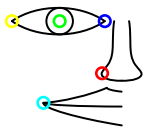

Figure 1: The detectable landmarks from the left to the right. Eye: exocanthion right, centre of pupil right, endocanthion right. Nose: alare right. Mouth: cheilion right. The figure only shows the right landmarks of all 10 detectable.

enhance Haar like structures.

3.1 Gradient image

A gradient image is a special representation of the raw 3D data computed from depth information [Cro84]. The z-value of adjacent measure points will be compared either from left to right (x-gradient) or from the top to the bottom (y-gradient). If the second of the compared measure points is closer to the scanner than the first one, the pixel will be colored brighter. If the second measure point is farther away from the scanner, than the pixel will be colored darker. Figure 2 shows a plotted 3D face and the corresponding x- and y-gradient images. The advantages using gradient images are that edges on the 3D surface are enhanced and changing of 3D information are better described. It also allows to use image processing methods to detect landmarks.

3.2 Viola & Jones object detection

The Viola & Jones classifier is one of the most successful object detection algorithms. Its success relays on three main contributions. Firstly an efficient machine learning algorithm combining Adaboost with Haar like features is used. Haar like features describes relative illumination changes of adjust blocks. A vast number of Haar like features can be generalized. The Adaboost algorithm selects the most significant Haar-like features, which best describe the searched object.

Secondly a special image representation, called integral image, also known as summed-area table is used. An integral image contains the sum of gray scale values belonging to pixels. The average intensity of a certain rectangle can be easily calculated with only four values in an integral image. The last contribution of the Viola & Jones object detection is to use complex classifier as a multiple stages cascade. If the similarity between Haar like features derived from an inspected area and these from a classifier reaches a defined threshold, the area passes to the next stage of the classifier. Once the similarity is below the threshold, the area is rejected. Only when the inspected area passes all stages of the classifier, it is marked as a candidate of the searched object. The thresholds in the various

stages are determined during the training process.

3.3 The Proposed Algorithm

The presented algorithm is implemented in C++ and uses the Open Source Computer Vision library (OpenCV). As mentioned before, the raw 3D data has to be transformed into gradient images. In this approach the detection of the alare, the endocanthion and the cheilion is done on the y-gradient. The detection of the right exocanthion is done on the x-gradient, the detection of the left exocanthion is done on the mirrored-x-gradient and the detection of the pupils is done on a special gradient visualization, defined as x-abs-gradient. To generate the x-abs-gradient, the absolute values of the x-gradient plus an inversion of all black pixels are taken. Figure 2 visualizes different gradient images used for the landmark detection.

The eyes are the regions where artifacts often occur. Especially, regions of pupils cannot be captured and results holes. This property can be exploited to detect pupils inside the x-abs-gradient as the good visible, white filled disks in a dark region.

After the creation of the gradient images, the nose detector is applied. The reason for choosing the nose as the first detected modality is based on the facts, that the nose changes its shape only slightly due to expression. Additionally, the nose is normally not unintended covered by any other body part such as the pupils due to the closing eyes. Therefore, the nose detection reaches the highest detection rate with the trained classifier in comparison with other landmarks.

If the nose is detected, sub regions for the detection of the landmarks can be determined. The sub region for the pupils laying in a small rectangle above the detected nose. Its height is 130% of the detected nose. The preliminary sub region for the cheilions laying right under the detected nose with the same height as the detected nose. In the case that no nose was found, both sub regions have the size of the whole image. The landmark detection starts with the right and the left alare in the region of the detected nose. The second pair of landmarks, which are searched in its sub region, are both pupils. Every other sub region for the remaining landmarks is created dynamically. That means that their width is calculated as a result of already detected landmarks. In the following we describe

Figure 2: from the left to the right: the plotted raw-3D data (just for better imagination of the real face), the y-gradient, the x-gradient, the mirrored-x-gradient and a new developed gradient visualization named x-abs-gradient.

the detection process for the landmarks on the right side. The same process can be applied for the landmarks on the left side. The sub region for the endocanthions has always the same height as the sub region for detecting the pupils. The smallest region can be created, when the right pupil and both alares are detected. In this case, the width of the sub region shrinks between these two points. Figure 3 illustrates these process for all alternatives of already detected landmarks. The creation of sub regions for the exocanthions and cheilions follows the same scheme.

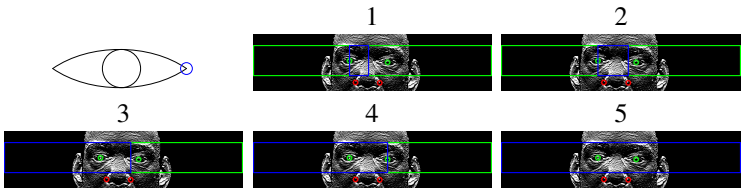

Figure 3: sub regions for the right endocanthion. 1: from the right pupil to the midpoint of the alares, 2: from the right pupil to the left alare, 3: from the right edge of the image to the left alare, 4: from the right edge of the image to the left pupil, 5: whole search region as before for the pupils.

4 Experimental Results

The proposed algorithm is tested with the Face Recognition Grand Challenge (FRGC) v2 database [Nat11]. The database contains 3D face scans with a resolution of 640×480 pixels. 500 scans are chosen to train the classifier and another 500 datasets are chosen for testing.

A big advantage by using sub regions is the lower runtime of the algorithm. It takes only 409ms to detect all 10 landmarks with sub regions and about 4900ms without sub regions. This implicit a speed up by nearly factor of 12. The measurements are the average of 100 program throughputs and taken with an Intel Core 2 Duo @ 2.16 GHz. But not only the runtime is reduced, additionally detection accuracy is improved with the use of sub regions. A single classifier detects many false positives on the whole image as shown in figure 4. After clustering, the resulting two landmarks are still far away from their actual position. By using sub regions, the false positive rate reduces and detected candidates are close to each other as well as to the actual position.

Individual classifier are derived from the training process. An important step in the training process is to prepare positive and negative samples. Positive samples are images that only contain the searched object while the content of negative samples can be arbitrary and only has the restriction, that it does not contain the searched object. The classifier are trained with the first 500 datasets out of the FRGC database. To double the number of positive samples, each classifier is trained for the right landmark representation and its mirrored

Figure 4: classifier output after detecting the cheilion. On the left side the detection on the whole image and the final clustering. On the right side the detection in the created sub regions.

counter piece. So each classifier has nearly 1000 positive samples. Since the training data is also noisy or includes closed eyes, not every classifier has the amount of 1000 positive samples. Table 1 gives an overview of the sample sizes used in the training process.

classifier	Nose	Alare	Pupil	Endocanthion	Exocanthion	Cheilion
# pos	1000	944	964	952	762	892
# neg	2000	3776	3856	3808	4572	3568
# pos : # neg	1:2	1:4	1:4	1:4	1:6	1:4

Table 1: the number of positive and negative samples that are used due the training process for each classifier. To have a minimum amount of training samples, the number of negative images were increased in cases of a lower count of positive samples.

The evaluation of classifier performance is shown in figure 5. Performance criteria is the distance, which was calculated out of the raw 3D data, between a detected landmark and its annotated ground truth counter-piece. The performance on testing data is nearly as good as the performance on training data. Only the detection accuracy of the exocanthion is significantly lower. This is caused by the training process of the classifier and the high deformation of the exocanthion during expression. The training of the exocanthion classifier was only done with neutral expression, so the classifier is unable to detect exocanthions that are highly distorted. This aggravation marks a big disadvantage in the general use of classifier that are based on the object detection approach. A classifier can only detect the objects which it was trained for. As soon as the object highly changes its appearance, the classifier is unable to detect the object in the picture.

The preexisting work of Ajmal Mian [Mia11] detected both exocanthions and the nose tip, so performance comparison can only be done for the exocanthion classifier. Figure 5 shows the performance comparison of Mians exocanthion classifier and the one generated in this approach. It must be mentioned, that the curve of Mians classifier was manually reconstructed out of the original paper. Up to a detection error of 5mm, the sub region approach proposed in this paper outperforms the one from Mian that uses anthropometric constrains. For detection faults bigger than 5mm, Mians approach has an obvious higher performance by nearly 20%.

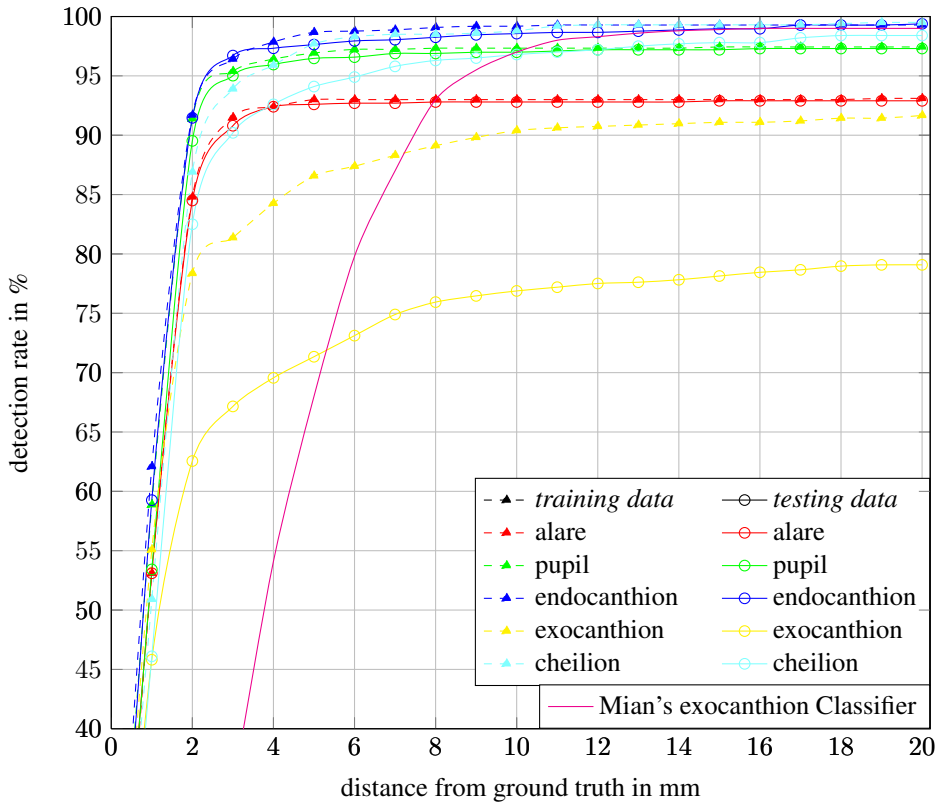

Figure 5: precision comparison between training data, testing data and the reconstructed performance curve of Ajmal Mian's exocanthion classifier.

5 Conclusions and Future Work

In the paper we proposed a reliable landmark detection algorithm with low detection time. The usage of the x-abs-gradient image overcomes the well known pupil detection problem in 3D face recognition. The creation of sub regions not only dramatically reduce the detection time by nearly the factor of 10, they also increase the whole performance of the system. The designed algorithm has a big potential with the opportunity of further improvements. At first an enhancement of the classifier training should be done. An increase of training data would directly result a higher detection performance and would also boost the robustness against expression. Also a cross database development can be considered, because right now the algorithm runs only on the FRGC v2 database. In order to reduce the algorithm runtime time farther, a parallel detection can be implemented. The last improvement that will be listed here is an increase of the number landmarks to detect. That makes an implementation of an identification or verification system possible.

References

[CCS06] Alessandro Colombo, Claudio Cusano, and Raimondo Schettini. 3D face detection using curvature analysis. *Pattern Recognition*, 39:444 – 455, 2006.

[CLR89] J.Y. Cartoux, J.T. LaPreste, and M. Richetin. Face authentication or recognition by profile extraction from range images. *Proceedings of the Workshop on Interpretation of 3D Scenes*, pages 194 – 199, 1989.

[Cro84] Franklin Crow. Summed-Area Tables for Texture Mapping. *Computer Graphics*, Volume 18, Number 3:207 – 212, 1984.

[DSA08] Hamdi Dibeklioglu, Albert Ali Salah, and Lale Akarun. 3D Facial Landmarking under Expression, Pose, and Occlusion Variations. *Journal of Physics D-applied Physics*, 2008.

[Mia11] Ajmal Mian. Robust Realtime Feature Detection in Raw 3D Face Images. *Applications of Computer Vision WACV, IEEE Workshop*, pages 220–226, 2011.

[Nat11] National Institute of Standards and Technology. Overview of the FRGC, Feb 2011. Last visited on 9th of April 2013.

[PPTK10] Panagiotis Perakis, Georgios Passalis, Theoharis Theoharis, and Ioannis A. Kakadiaris. 3D Facial Landmark Detection & Face Registration: A 3D Facial Landmark Model & 3D Local Shape Descriptors Approach. Technical report, Computer Graphics Laboratory, University of Athens, 2010.

[SLSORPBQ10] Maurício Pamplona Segundo, IEEE Luciano Silva, Member, IEEE Olga Regina Pereira Bellon, Member, and Chauã C. Queirolo. Automatic Face Segmentation and Facial Landmark Detection in Range Images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Volume 40 Issue 5:1319–1330, 2010.

[Spr11] Luuk Spreeuwers. Fast and Accurate 3D Face Recognition. *International Journal of Computer Vision*, 93:389 – 414, 2011.

[VJ01] Paul Viola and Michael Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. *Computer Vision and Pattern Recognition*, 1:511 – 518, 2001.