Automatic Face Feature Points Extraction

Dominik Rupprecht, Sebastian Hesse, and Rainer Blum

Hochschule Fulda – University of Applied Sciences,
Marquardstr. 35, 36039 Fulda, Germany
{Dominik.Rupprecht,Sebastian.Hesse,Rainer.Blum}@
informatik.hs-fulda.de

Abstract. In this paper we present results of finding a way to automatically equip a three-dimensional avatar with a model of a user's individual head. For the generation of the head model certain so-called face feature points must be extracted from a face picture of the user. A survey of several state-of-the-art techniques and the results of an approach for the extraction process are given for the points of the middle of each iris, the nasal wings and the mouth corners.

Keywords: Avatar, Face Feature Points, Integral Projection, Circle Detection, E-Commerce.

1 Introduction

High quality three-dimensional virtual human bodies (avatars) are playing an increasingly important role in the areas of information technology, for example in networked 3D worlds like Second Life[1] and Twinity[2], in the simulation of processes or in virtual try-ons of clothing.

In the field of creating avatars, and most of all customized ones, especially in the context of an e-commerce application the customers' acceptance of their personal avatar is very important. On possible solution to increase the users' acceptance of their avatar might be to equip it with an individual, three-dimensional model of their individual heads. For this challenge the automatic face modeling software "FaceGen" [1] provides a basis. Offering frontal human face pictures as input, a textured three-dimensional face model is generated. Unfortunately, the software requires the user to manually identify several so-called face feature points, first. But, as the context of this work calls for a fluent web shopping experience, it is advisable to completely automate the head model generation process, including the face feature extraction. This would keep the customers from any tedious extra work.

The approach presented here combines several state-of-the-art techniques. It differs from similar face recognition techniques, where often detection of regions and template matching are applied. In contrast, for the context at hand, it is essential to identify accurate points in order to obtain a good model and texture as result of the automatic head generation process.

[1] http://secondlife.com/
[2] http://www.twinity.com/

V.G. Duffy (Ed.): Digital Human Modeling, HCII 2011, LNCS 6777, pp. 186–194, 2011.

1.1 Objectives and Significance

In this paper we present intermediate findings of the ongoing research project KAvaCo[3]. Focus of its research activities is to improve customized avatars for the usage in an innovative, interactive system for the support of apparel retail via Internet. The results reported here were obtained in the Bachelor thesis of this paper's co-author Sebastian Hesse. To interpret our findings correctly and to transfer them to other research contexts, we provide details on the study's context.

1.2 The KAvaCo Project

On the one hand, the aim of the ongoing research project KAvaCo is to develop an interactive system for generating humanoid, anatomically correct, animatable, virtual humans. Secondly, the avatars are used as representations of individual customers in e-commerce and are evaluated as such. Imitation refers both to geometric properties such as body sizes and shapes as well as to the visual appearance, such as skin or hair color.

All of the necessary information can be entered by the users themselves. For a realistic representation of the head, photos are processed, which can be provided by the customers with little effort using conventional digital cameras. The optimal design of the user interaction is a central goal for both the creation and the use of the avatars.

In addition to the technical progress the project also investigates the until now only partially clarified scientific question, what demands are made by customers on their personal virtual twins in the context of e-commerce. Concerning the representation, for example, it has to be clarified which level of realism is desired by customers, whether individual ”problem areas“ and characteristics should be concealed or whether an idealized representation achieves greater acceptance.

The results of the research will be applied and tested in the context of the clothing retail. Here comes into play that the developed avatars can directly be used in systems for virtual-try on for made-to-measure and ready-made clothing. But also the use for other applications such as in the area of public health, for example the visualization of weight gain and loss, is planned.

2 Method

The face modeling software employed by the project, ”FaceGen”, needs eleven points from the frontal face image for computation. In this work techniques and approaches for the points of the eyes (feature point: middle of each iris), the nose (nasal wings) and the mouth (mouth corners) were considered.

The approach which is finally used to extract the facial feature points is divided into two sections. The first is to reduce the image data to small areas of interest for later detection. This means, that the feature points do not have to be searched in the whole face image, but in small zones. Afterwards, in the second section the required points are searched through several extracting techniques. The techniques are applied to regions where the feature points are expected. The detailed procedure is shown in the following sections.

---

[3] `http://www.kavaco.info/`

To evaluate the results of each concept a set of 100 frontal face images has been acquired from [2]. To be as representative as possible the persons on the photos were chosen from different nationalities, different age and gender. Each of the approaches was tested with all pictures and documented. Therefore, the detected sections and features were logged to validate them against human specified ones. Classification numbers were defined to decide how robust the region detection and the feature extraction are. For the region detection the classification numbers are false-positive-, false-negative- and right-positive-rate in percent, for the feature extraction it is the variance between the identified and the ideal position in pixel.

3 Region Detection

3.1 Detection Concept

As said before, the first part of the identified solution reduces the area that is subject to the later detection, to small regions. The detection area is constrained in a "top-down" manner to so-called regions of interest (ROI), here, first the face, then eye, mouth and nose. The "region-detection" makes use of the Haar-Like Feature Extraction with AdaBoost (see e.g. [3]) at the beginning. The Haar-Like Feature Extraction is an object recognition technique which uses weak features to detect domain objects. With the learning algorithm AdaBoost a classifier is constructed for the detector.

The identification of the detection area starts by searching the part of the image that contains the frontal face. This area is then divided into three further ones:

1. Area – upper face third, for searching the eyes-region
2. Area – middle face third, for searching the nose-region
3. Area – lower face third, for searching the mouth-region

This division of the face is done by intuition, because the position of the searched regions in normal faces is given by the human anatomy.

These new regions are treated as search candidates for the remaining three ROI. Due to this division into smaller areas the detector needs less performance and it also reduces the detection of false regions.

A test iteration with the 100 human faces showed, that the detection of the face- and eyes-region has a good detection rate. Only two percent of these both regions were not detected. But the false-positive rate of the other two regions was too high (30 percent for the nose and 47 percent for the mouth). So there is a need to extend this concept to counteract the high false-positive rate.

3.2 Extended Detection Concept

In a second concept some extensions are made to the first one reported in the previous section in order to decrease the false-positive-rate of the mouth and nose region. Therefor a validation was implemented which checks the regions with a set of rules. The detected regions are related to each other and the face area. Primary point of reference is the region of the eyes which has a very good detection rate. The violation of a rule excludes the region from the list of potential candidates. The rules were set up under the condition that they are not too detailed and not too general to avoid excluding too many right or too little wrong detected regions.

Even though the eye-regions have good results there are rules implemented to prevent a false detection here. This can occur in the area of the nostrils because the integral projection of two eyes and nostrils could be very similar. Because the nostrils are less distant than the eyes, the width of the eyes-region is related to the width of the face. If the width of the eyes region falls below 30 percent of the face region width it will be excluded.

For the validation of the nose region it is related to the eyes region. Two rules for the nose region are used to check that the nose region's horizontal position is within the eyes region. Another rule ensures that the nose region is not above the eyes.

The mouth region uses the same rules as the nose region but the rule for the vertical position is not applied, because the mouth is only looked for in the lower face third.

Using these rules the false-positive rate could be reduced in the mouth region by 10.03 percent and the nose region by 11.43 percent. But there were many areas that were too small to be correct ones. Thus, the rules were adjusted by analyzing the smallest region in context of other regions. It turned out that the mouth and nose region fits 60 times into the facial region.

An extension of the rules checking this threshold decreased the false-positive rate for the mouth region by 23.22 percent and for the nose region by further 6.82 percent. The false-positive rate was now low enough to develop and use some techniques to extract the desired feature points in the detected regions.

4 Feature Extraction

Facial feature points can be identified through their miscellaneous properties. Their extraction has also a high relevance in domains like gesture detection, emotion recognition and the transfer of real facial expressions onto an avatar. Such feature points are often used to identify a face in an image but, vice versa, to detect the face region at first with a different, global approach could help making a following feature extraction more robust and effective.

The feature extraction approach presented here makes use of several image processing techniques which are described in the following sections.

4.1 Extraction of the Center of the Eye

The eye is one of the most distinctive regions in the human face. Two well recognizable attributes are the elliptic geometry of eye and iris and its brightness distribution.

One possibility is to use the geometric attribute of the eye to detect its center. To detect the iris the Circular Hough Transformation can be applied to the respective region. Another technique to find the iris is the integral projection. With the integral projection it is possible to extract the brightness distribution of the eye region. The following sections will analyze both techniques.

Circle Detection. To detect the iris with the Circular Hough Transformation (cf. [4]) it is necessary to transform the region from a colored to a gray scale image. Most image processing algorithms are applied on gray scale or binary images because they contain less information than colored images. Hereafter, possible image artifacts are removed with a Gaussian filter. Now, the iris can be highlighted by an edge-detector, here the Canny filter. The result of the Canny filter has the benefit that edges are represented as one pixel wide lines (also called binary edges). The image delivered by this pre-processing step forms the basis for the subsequent circle detection.

To prevent false-detected circles, a smallest and largest acceptable radius is calculated dynamically relative to the height of the eye region. The minimum radius has to be greater than one tenth of the eye region height and the maximum radius has to be smaller than half of the region height.

Fig. 1. Results of the Circular Hough Detection

The circle detection has good results for many eyes regions (see Fig. 1, part a). But too much result images show that this method produces also a lot false-detected circles (see Fig. 1, part b). This is the result of a too dark illumination of the images or not wide enough opened eyes. An increase of the contrast or the change from binary to gray scale images did not produce better results. So, this approach was no longer analyzed.

Integral Projection. As already mentioned in section 4.1 another method to detect the eye is the integral projection. This method uses the brightness distribution in the eyes region. To detect the iris it is necessary to calculate the vertical and horizontal integral projection. As shown in [5] the coordinates of the iris can be extracted from the two integral projections. The x-coordinate can be derived from the horizontal and the y-coordinate from the vertical projection.

The calculation of the projection can be disturbed by several factors such as dark bushy eyebrows, eyelashes and dark shadows. So, a pre-processing step is needed to decrease the effect of these factors.

After transforming the image into a gray-scale one, a median filter is used to eliminate image artifacts. To highlight the iris and remove reflection in the eye, morphological operations are used. A method called "erosion" enlarges the dark areas in an image. That is how the iris can be highlighted but it also highlights shadows, eyebrows and eyelashes. This can be contracted by increasing the contrast (Fig. 2 a) and using "dilation" (Fig. 2 b) to shrink dark areas before applying the erosion method (Fig. 2 c).

(a) increasing contrast (b) dilation (c) erosion

Fig. 2. Preprocessing the image for the integral projection

The used pre-processing method delivers good results so that the integral projection can extract the darkest area. The large part of disturbing factors can be eliminated through the contrast enhancement and use of dilation. Mostly all of the extracted x-coordinate from the horizontal and the y-coordinate from the vertical integral projection were inside the iris.

Only bad exposure and big shadows disturbs the extraction too much. But such face images will also result in bad textures and cannot be used for the three-dimensional head-generation. Therefore the exposure problem is not analyzed anyway.

At the end the results of this method were usable. For the head generation process with “FaceGen” feature points with a variance less than 20 pixels from the ideal position can be used. Only 12.24 percent of the detected points were too far away and could not be used for the automated head generation process.

4.2 Extraction of the Nasal Wings

In contrast to the eyes, the nose does not have such good attributes to recognize. Most noticeable is the tip of the nose, the nostrils and the contours of the nasal wings. In the following two approaches to track the nose are described.

Contour tracking. The pre-processing of the contour tracking is similar to the circle detection of the eyes. Often, the nasal wings have differences in brightness to the surrounding face so that they can be extracted as edges. The result of the Canny-filter is the basis for this method. An algorithm is applied onto the edge image which follows the edges to find contours. The contours that are farthest to the right and left potentially represent the nasal wings.

Inspecting maps of edges for the 100 test faces, it is obvious that this method cannot deliver useful results. The differences in brightness from the nasal wings and the surrounding face were not big enough. Often, it was even not possible to mark the nasal wings manually because of a fluent crossover between the nasal wings and the surrounding face region (Fig. 3 a). Only if the exposure is good enough the edges are recognizable. Another problem is the falsification of the edge image by facial hair like beards (Fig. 3 b). Brunelli and Poggio [6] had used this method to find the nose but it is not useful to extract the nasal wings.

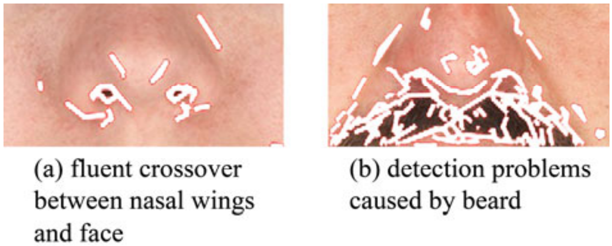

Fig. 3. Results of the contour tracking of the nose

Integral Projection. As Hua et al [5] describe the integral projection can be used to find the tip of the nose or the nostrils. So, for pre-processing, the region has only to be transformed in a gray-scale image and possible image artifacts have to be removed by using the median-filter. The nose region is then analyzed with the horizontal projection. In this projection the recognized low points are the nostrils and the amplitude in between represents the tip of the nose. The vertical position of the nostrils can be extracted by calculating the vertical integral projection. Here the lowest point mostly describes the dark nostrils. The effect of the nostrils in the integral projection cannot be highlighted by using the erosion. Sometimes the tip of the nose is beneath the nostrils so that they are not visible and hence not detectable.

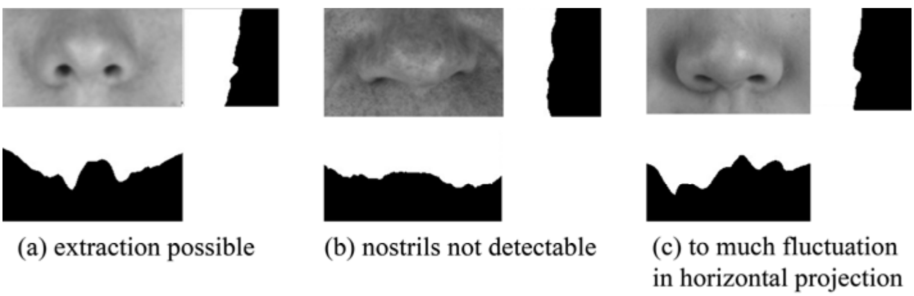

Fig. 4. Results of the integral projection to identify the nostrils

In some cases this method could deliver acceptable results (Fig. 4 a). But, depending on the face tilt and the shape of the nose the nostrils may not be identified (Fig. 4 b). A good and consistent exposure of the nose is also important. Otherwise the horizontal projection is useless (Fig. 4 c). But another problem is that the position of the nostril does not imply the position of the respective wing of the nose, which is needed for the head generation process. So, this method could not be used to detect the wings of the nose.

4.3 Extraction of the Corners of the Mouth

The mouth has some distinctive attributes to recognize. In most cases the mouth is darker than the face. This characteristic can be used to extract the contour of the mouth and its width and position. At first the contour tracking algorithm like section

4.2 was used to extract the mouth but the results were not usable. There are too many disturbing factors like beards, small lips or makeup. These circumstances make an extraction with this method impossible.

As already mentioned the mouth has a darker brightness as the surrounding face. So, it is possible to extract it by applying a vertical and horizontal integral projection in the mouth region. In the pre-processing step the region is transformed into a gray scale image, then using dilation to reduce the effect of dark hairs and afterwards erosion to mark the dark lips. In the horizontal projection the width of the mouth can be extracted by searching the two outer flanks and the vertical projection is useful to detect the vertical position.

Here it was also not possible to extract the feature points exactly enough. The vertical position is often recognizable as a down point in the vertical integral projection. But a dark beard or bad lighting conditions disturb the extraction massively. In [6] it is described how to extract the width of the mouth with the horizontal projection but the effect of the disturbing factors is too big. The shape of the mouth often varies a lot so that lips were not recognizable. So this method can also not be used to extract the required feature points.

5 Conclusion

This paper shows that there are many different ways to try to extract the several required feature points from a face photo. The chosen approach with detection of the face first and then separation of the regions of interest for a subsequent feature detection appears to be a good choice.

The Haar-Like Feature Extraction with AdaBoost identifies the face with a very high right-positive rate. The implemented rules, which identify the detected ROI as valid, helped to reduce the initial high false-positive rate when selecting the three regions of interest for the eyes, the nose and the mouth.

The quality of the analyzed approaches for the feature extraction varies a lot. The integral projection is the best choice for finding the center of the eyes. Almost 90 percent of the given eyes could be detected. But for the other two feature points the algorithms are too vulnerable to disturbing factors like facial hair or bad exposure. So, the differences in brightness between nose and face are mostly too small to detect the edges of the nose. The extraction of the mouth has similar problems. The shape of the mouth that greatly varies from face to face as well as beards impede useful results.

As future development it is planned to investigate further procedures to get better results. One possible solution may be the combination of different approaches for example template matching with transformable templates, to find the edges of mouth and nose. These edges can then be used to extract the desired feature points. Another possible attempt might be to separate the mouth from the rest of the face not only by its brightness distribution but by its color and subsequently detect its corners.

Acknowledgments. This research was financially supported by the German Federal Ministry of Education and Research within the framework of the program “Forschung an Fachhochschulen mit Unternehmen (FHprofUnd)”.

References

1. Singular Inversions Inc. - FaceGen, http://facegen.com/
2. SmartNet IBC LTD - Human references for 3D Artists and Game Developers, http://www.3d.sk/
3. Viola, P., Jones, M.J.: Robust real-time object detection. International Journal of Computer Vision 57, 137–154 (2004)
4. Gupta, P., Mehrotra, H., Rattani, A., Chatterjee, A., Kaushik, A.K.: Iris Recognition using Corner Detection. In: Proceedings of 23rd International Biometric Conference (IBC 2006), Montreal, Canada (2006)
5. Hua, G., Guangda, S., Cheng, D.: Feature points extraction from faces. In: Proceedings of the Image and Vision Computing, New Zealand (2003)
6. Brunellin, R., Poggio, T.: Face recognition: Features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 15, 1042–1052 (1993)