Active Shape Models on adaptively refined mouth emphasizing color images

Axel Panning
University of Magdeburg, Germany
Axel.Panning@ovgu.de

Ayoub Al-Hamadi
University of Magdeburg, Germany
Ayoub.Al-Hamadi@ovgu.de

Bernd Michaelis
University of Magdeburg, Germany
Bernd.Michaelis@ovgu.de

Abstract

In this paper, we propose a hybrid method for lip segmentation based on normalized green-color histogram splitting and Active Shape Models (ASM). A new adaptive method for histogram splitting is applied in two steps. First, after defining a region of interest for mouth segmentation, a rough adaptive threshold selects a histogram region assuring that all pixels in that region are skin pixels. Second, based on these pixels, we build a Gaussian model which represents the skin pixels distribution and is used to obtain a refined optimal threshold for lip pixel classification. This process is used to refine the normalized green channel image for the elimination of inner distortions and gradients inside the lip region, which can misguide active contours (i.e. ASM) in the last step of the hybrid segmentation process. In the results, we present that the proposed method performed better than conventional ASM on unrefined color enhanced images or pure color-histogram based mouth segmentation.

Keywords: Feature extraction, Segmentation, Image processing, Application.

1 INTRODUCTION

The segmentation of mouth and lips is a fundamental problem in facial image analysis and is important for various applications. It can be utilized for lip reading, supporting speech recognition or expression analysis (i.e. facial expression, estimation of emotional state, pain recognition). Each application has its own limitation concerning speed, accuracy and robustness. The requirements for facial expression recognition can be very different depending on application context.

Often initially a color transformation is performed to exploit the different chromaticity of lips from skin. Basically, the segmentation approaches can be classified into two groups. The first group, Histogram based approaches, is a consequent continuation of the initial color transformation. The mouth region of interest (ROI) is binarized into lip and non-lip pixels, where non-lip pixels are mainly skin pixels. The crucial point in histogram based algorithms is the estimation of that particular threshold. A very easy approach, mostly used for first rough mouth segmentation is a fixed threshold, found by statistical average of numerous samples [8]. A more adaptive approach sets up a watershed like rule, which defines 15 percent of the darkest pixels in their color transformed mouth ROI as lip pixels[9]. Other works [5] assume a certain topology in the histogram. Following this idea they seek for a local minimum between a lip and a skin heap in the histogram and define the threshold here.

The second group, is focusing on detection of lip edges in the mouth ROI [4, 2]. They apply Active Contour Models (e.g. [4]) or deformable templates [2] to the mouth's ROI. Some approaches [7, 1] stabilize their Active Contours using support tracking points. The general assumption of edge based algorithm is, that the lips generate prominent edges at the skin-lip crossing. In monochrome images only a simple shadow casting can already cause serious problems. A hybrid of color and edge information is the usage of color images and their mouth-highlighting transformed representation (e.g as used in [4, 2]). This can suppress some issues like shadow casting. But still there is no guarantee the edges of the lips create significant edges here. This might happen for many cases. For people having Asian skin tone for example this rule holds true. However, for European/Caucasian this rule does not hold for all cases anymore, since the transition from skin to lip pixels does not form rough edges here for all subjects and conditions. Another usage of color and edge information is to align deformable templates or active contours using an energy minimization function, which refers to edge information and average color intensity inside of the template (or contour) as proposed in [2, 3].

In the proposed approach the advantages of both classes of algorithms (pure color based, and shape/edge based) shall be combined in another way. We chose *Active Shape Models (ASM)*, introduced by Tim Cootes [6], as representative for the edge/shape based algorithms. The idea is, that a color based approach can contribute to an

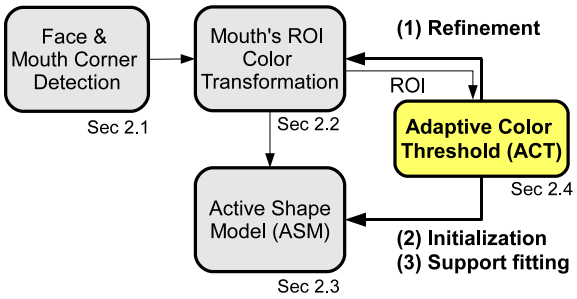

Figure 1: Flow chart of the proposed algorithm

edge and model based approach (ASM here) to improve its performance more than a simple prior transformation of color space, as most hybrid approaches do so far. In this context we propose a novel adaptive method for color based mouth segmentation. The rest of the paper is organized as follows. Section 2 describes the idea of combining histogram based thresholding and shape based extraction for mouth segmentation. The results of the proposed methods are presented in section 3. Section 4 gives a short summery and outlook.

2 MOUTH SEGMENTATION

The process chain is shown in Figure 1. All successive steps will be described in the following sub sections.

2.1 Locating Face and Mouth ROI

Object detection in image processing is always the search for a delimited area in which the targeted pattern is fitting. A general solution for this task has been developed by Viola and Jones [13]. They developed an algorithm, where a cascade of weak Haar-like features (see Fig. 2) is utilized to model image objects appearance. A Haar-like feature describes the difference of pixel intensities within similar sized sub regions of one rectangular region in an image. The most advantage, compared to other feature descriptors, is the fact, that they can be computed very fast using integral images. Once calculated, an integral image can provide the average intensity of any rectangular region of any size by one addition and two subtraction operations. This property is very important in context of applications, where speed issues are relevant. Another acceleration is provided by the cascaded structure of the classifier. During the search process not the whole classifier needs to be used at each potential position. Once one cascade step fails all successive cascade steps can be discarded, the current target region can be rejected and the search continues in the next potential region.

An implementation of the algorithm as well as face detection models can be found in the OpenCV c/c++ library, which are widely used. Also in this work, the available face models were used for face detection. Further two models for detection of mouth corners in facial regions were trained in order to define a region of interest (ROI) for further mouth segmentation processing. Database for the training was the *FGnet Database* from the Technical University of Munich [15]. To train the classificator 400 positive and negative samples were chosen. Positive samples were sub images where mouth corners were directly in center of sub images. Negative samples were chosen from randomly selected sub images where the mouth corner were not centred. [12].

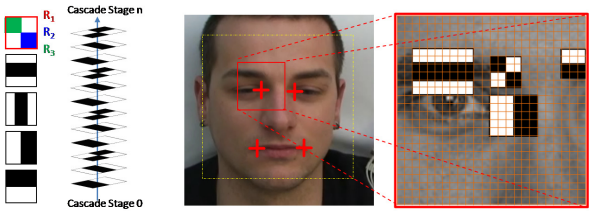

Figure 2: Left: base features for the cascade classifier and their cascadation. Middle: the face region, a result of the face detector is the search area for single facial features. Right: single weak features in their local arrangement forming a strong classifier.

2.2 Color Transformation

In common a color transformation is chosen converting the RGB from $\mathbb{R}^3$ to $\mathbb{R}^1$ exploiting the difference of lip and skin pixel colorness. Using the ground truth of our database, a comparative statistic was made to analyze their ability to separate lip from non-lip pixels, based on color information only. In result the green channel from normalized *rg* was superior to all others, which is defined by $nG = R/(R+G+B)$. We will refer to this in further context as nG color channel. The worst results were achieved by the $YCbCr$ based color transformations. Qualitative results of this prior study are given in table 1. The percentage is relative to the histogram of the complete ROI and outlines the false classified pixels using an optimal, FPR minimizing threshold found by the ground truth. Under advantageous conditions lips and skin pixel form two well noticeable bell curves in the histogram with a noticeable local minimum in between (Fig. 3 left). This can motivate approaches like [5], searching for this minor local minimum. However, these optimal cases cannot be assumed in general. The general structure of the histogram can vary in different scenarios (Fig. 3). More complex situations can create numerous minor local minima instead of only one major minimum. In other cases the smaller bell curve related to the lip pixels can be directly attached to the larger bell which represents the skin pixels without producing any local minimum (Fig. 3 middle). This multiple behavior can be observed independently from the chosen color transformation. Intersection of skin and lip color in the mouth ROI with respect to differen

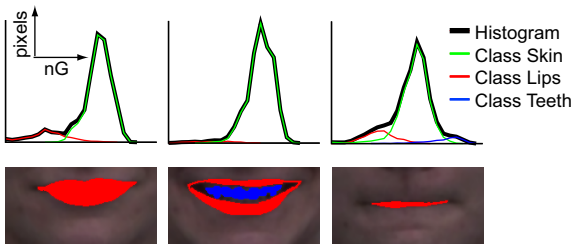

Figure 3: (Top Row): Histograms. The real histogram is the fat black line. The colored lines represent skin/teeth (no-lip) and lips. These information are only gathered by ground truth here and are a-priori unknown in application case, (Bottom Row): ground truth. The three samples show three states normal mouth state(left), open mouth with appearing teeth (middle), pressed lips with almost none lip pixels left(bottom).

Used in	Transf.	With teeth	No Teeth
[2]	$u(Luv)$	4.61 %	3.16%
[10]	G/B	5.97 %	2.59%
[3]	G/R	2.30 %	1.45%
[4]	Cr^2	11.75 %	10.97%
[4]	Cr/Cb	13.08 %	11.16%
[9]	$R/(R+G)$	2.36 %	1.48%
not found	nG	**0.09%**	**0.38%**

Table 1: Intersection for different color transformations

color transformations was analyzed with and without teeth appearance. However, the appearance of the teeth had just a minor impact to the separability (see Table 1) using a histogram threshold.

2.3 Active Shape Model

Active Shape Models (ASM) combine assumptions about specific shape behaviour and image signal response at the model points of the shape. Base of the ASM is a set of model points forming one or more contours, which are stored in the mean shape $\overline{m}$. The modelized shape variance is stored in a vector matrix S. A weighting vector $\overrightarrow{w}$ applies the different shape variations to the mean shape. The fitting process alternates two steps until convergence:

```
(0)     initialize mean shape
        near the object.
do
{
    (1) search for special image
        signal near model points
        (gradients, pattern)
    (2) find a shape, based on S,
        fitting best to the (image
        signal based) model points,
        found in step 1.
}
until(convergence)
```

Step (2) in the algorithm results in the final shape m, by applying the following equation

$$m = T(\overline{m} + S\overrightarrow{w}) \tag{1}$$

where $\overline{m}$ is the mean shape of the mouth (a vector containing all x- and,y-coordinates of the shape points one below the other), S is the matrix of column wise aligned shape variation vectors, $\overrightarrow{w}$ is the vector, containing the weights for each shape variation of S, and T is a affine transformation including x- and y-translation, scaling and rotation. The unknown $\overrightarrow{w}$ and T are estimated by solving

$$\delta = S\overrightarrow{w} \tag{2}$$

with

$$\delta = T^{-1}(m^*) - \overline{m} \tag{3}$$

where m^* are the associated landmark points based on any measurement in the image data. The estimation of T is described in [6].

The used shape model for the mouth consists of 22 contour points (see Fig. 4). Only the outline of the mouth will be addressed here. The mean shape $\overline{m}$ was found by average of 20 samples. Classical ASM as introduced in [6] define the shape variation matrix S by calculating eigenvectors from the covariance matrix based on size normalized samples. This method has some drawbacks. It demands very exactly and equidistantly picked landmarks for all samples. Further a few number of samples with less variations can cause wrong mutual dependencies. To resolve semantically and technically clean modes, the shape modifier vectors for S were created manually with expert knowledge. Five different modes were defined (see Fig. 4).

The edge fitting has two main parameters. a) the method of edge detection and b) the range of edge detection. Cootes [6] suggests statistical patterns here. In case of mouth shape this results, more or less, in a kind of gradient detection. The manifold of profile structures is considerable. Only the lineup of all 57

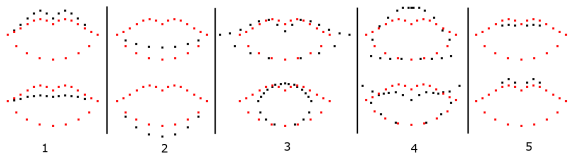

Figure 4: The five mouth modi and their behaviour. The red dots show the mean shape $\overline{m}$. The two stacked images of a mode show the impact of negative (lower row) or positive (upper row) weighting. Each mode represents one column of S.

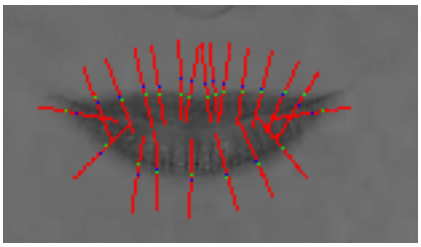

Figure 5: Red lines are the profiles taken during the edge fitting process. Blue dots mark the starting points from the last model-aligned instance respective the initial mouth model. The green dots represent that point in the profile, where the largest gradient is found.

samples does not show some special pattern, different from gradients, which could be fitted in a definite pattern. For edge fitting a profile P_i of the normal to the shape boundary is collected for each model point. The normals are defined by the neighbour points in the contour (Fig. 5). The profile contains the information of interpolated sub-pixels along the profile line. A simple concatenation of full pixels along a Bresenham based line was distorting. Profiles always are collected with a width of three pixels, where outer pixels got lower weight than inner pixels. As feature for model point detection gradient function was used, which is defined as follows:

$$p_i^* = \underset{t}{argmax} \left(\sum_{j=0}^{t} p_{i,j} - \sum_{j=t+1}^{k} p_{i,j} \right) \qquad (4)$$

where p_i^* is the found point to profile P_i with maximum gradient. The length of the profiles is an important parameter here. In the current version a length of 30% of mouth width is used (in average 30 pixels for the used samples). All Points p_i^* build the next instance of m^*.

2.4 Adaptive Lip Pixel Enhancement

Color enhancement for application of Active Contours in general (e.g. Snakes, Active Shape Models etc.) has been introduced already in previous works. But the lips of the subjects not always have uniform color. So inside of the lip itself distortions (causing gradients) can occur, which are more prominent than the outer (targeted) edge. These gradients can attract the contours falsely and thus misguide the whole active contour. To avoid this an in-between-step is suggested, which is performed after color transformation but before the application of ASM. The idea of the Lip-Pixel-Refinement is to flatten the lip pixels, in order to weaken the edges inside the lips. Therefore an adaptive histogram based algorithm (which is a considerable segmentation method itself already) will determine a threshold to define the lip pixels class in the ROI. This

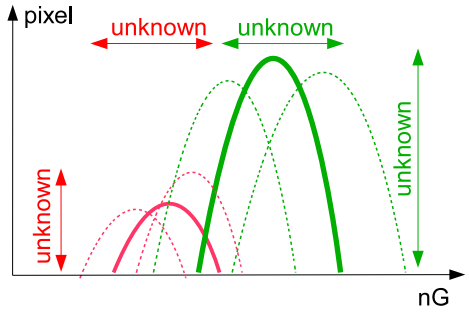

Figure 6: Model Assumption. There are more skin than lip pixels. Lip pixels have lower intensity than skin pixels. Unknown is their exact centering, scattering, distribution and intersection.

pre-segmentation is used to equalize and flatten distortions inside the (so far known) lip segment.

2.4.1 ACT: Adaptive Color Threshold

Basically the ROI contains two classes of pixels: *Lip-Pixels* and *Non-Lip-Pixels*. A general description of a statistically based model for lip or skin pixel distributions is not reliable, due large variance among subjects and illumination conditions. Thus an adaptive histogram based approach was developed to separate skin from lip pixels. As color transformation for lip pixel enhancement the green channel of the *rg* has been chosen (in further context referred as *nG*). The approach makes following assumptions:

- Skin pixels are Gaussian distributed in the histogram
- Lip pixels have lower intensities than Skin pixels (in *nG*)

Skin pixels and lip pixels can be mixed in the histogram (see Fig. 6), thus there is not always a perfect threshold to separate skin pixels by color information only. The algorithm prefers wrong positive skin pixels rather than wrong positive lip pixels. Latter case produces a kind of flow out which causes more damage to the segmentation than lip pixels which are classified as skin pixels. Wrong positive lip pixels are caused by a threshold greater than the optimum, with respect to the chosen *nG* transformation. With increasing intensity of *nG* also the probability of adding a high amount of pixels to the lip pixel class in one single step is increasing (see Fig. 7). Knowledge about the skin pixel distribution can provide a threshold that most likely avoids wrong positive lip pixels (see Fig. 6). The target threshold should be the foot-point of the skin pixel distribution, in order to avoid wrong positive lip pixels.

Basically a Gaussian distribution is estimated by a set of samples, calculating σ and μ, which are represented by single skin pixels here. A-Priori it's unknow which

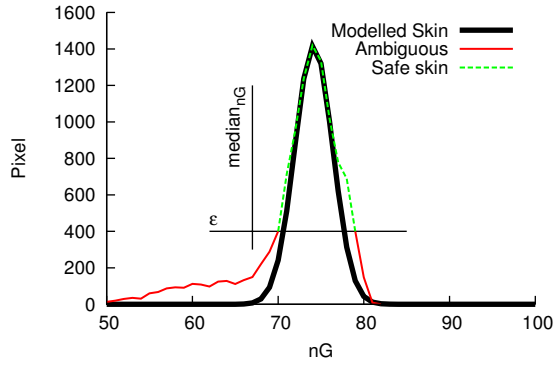

Figure 7: Based on initial guess the upper part of the histogram is defined as *safe*. This part is base for the approximation of skin pixel distribution

pixels belong to skin and to lip class. The idea is to select a part of the skin pixels, which can be assumed to be 'safely' part of skin pixel class. The condition is made simply by its number of occurrence in histogram. In other words, it is an 'initial guess' avoiding participation of lip pixels.

Let $h(x)$ be the value to the x^{th} slot of the and h_{max} be the global maximum of the histogram. To calculate σ_s and μ_s in first step σ^* is calculated for all pixels satisfying following condition:

$$h(x) > \varepsilon \tag{5}$$

$$x > median_{nG} \tag{6}$$

Condition in Eq.(5) represents an expected maximum ratio of mouth size to ROI, where ε is adjusted using a parameter α ([1]) with $\varepsilon = h_{max}/\alpha$. Additionally a median constraint, relative to all occurring intensity values in nG, was introduced to avoid disturbances from peaks of very low intensities in the histogram. Both very conservative conditions formulate a reasonable 'initial guess'. However, only the median constraint itself is a decent classification, which can compete with classic watershed method (See 3.2, Fig. 12).

The Gaussian distribution has scatters less than the original skin pixel distribution. However there is a correlation between α and the ratio of σ^*/σ_s. The unknown σ_s can be approximated following equation

$$\sigma_s = \sigma^* \cdot \left(1 + \frac{1}{\alpha}\right) \tag{7}$$

The larger α, the smaller the part left out from the histogram. This will raise the quality of approximation. If $\alpha \to \infty$ the whole histogram is used. But in case of application the lower intensity edge parts of skin pixel distribution is mixed with lip pixels. Therefore the choice of α is a trade off between approximation accuracy and risk to include lip pixels to the initial guess. Estimation of the threshold is done using the cumulative distribution function of $\mathcal{N}(\sigma_s, \mu_s)$ applying a low border ($\lambda = 0.01$).

$$th = \underset{x \in \mathbb{R}}{argmax}(\lambda < F(x)) \tag{8}$$

with

$$F(x) = P(X \leq x) \tag{9}$$

where x is the intensity value of possible thresholds in nG.

2.4.2 Combining ACT and Active Shape Models

The result from section 2.4 contribute in three ways to the problem of ASM fitting:

1. more accurate initialization
2. fixing corner points of the model
3. better gradient fitting due refined base image (distortion reduction)

As seen in the ASM algorithm in section 2.3 the ASM need to be initialized near by the image object. The quality of initialization can effect the result enormously. For initialization the mean shape $\overline{m}$ needs only a affine transformation $T_{tx,ty,scaling,rotation}$. The result BLOB obtained in section 2.4.1 provides such corner points, which are more accurate than the points provided by the method outlined in section 2.1. Furthermore the BLOB can be used to derive the weighting of the first shape mode (mouth opening-closing). Thus the ASM can be initialized very close and in appropriate scaling and shape to the image.

Naturally ASM suffer problems, when model points correspond to object corners with acute angles. Mouth corners represent such special case. Once the analyzed profile does not hit the narrow object, the gradient operator will not find any reasonable gradient. An additional problem appears since this weak model points represent in case of the mouth model the only forces drawing or pushing the whole model in horizontal direction. Replacing the sensor function for this model points can counteract this issue. Instead applying gradient operators the model points for mouth corners are set to the corner points found by BLOB using the ACT algorithm in section 2.4.1. These points will not change anymore during the fitting process, so they can be seen as *fixed*.

[1] In the current experiments a value of 3 was chosen. Basically values between 2 and 4 gave good results.

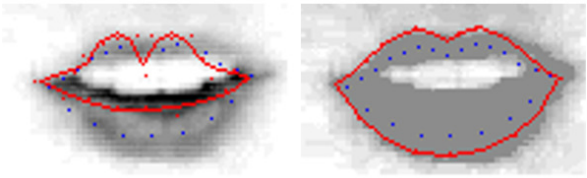

Figure 8: Sample28, impact of Adaptive Color Threshold. Left: the original *nG* converted image. Right: The ACT fixed image. Blue Pixels represent the initialization of the ASM. The red dots are the edge fitting points. The red line is the resulting ASM. (REMARK:BILDER LIEGEN IN ORDNER 'asmExtended/asm/'

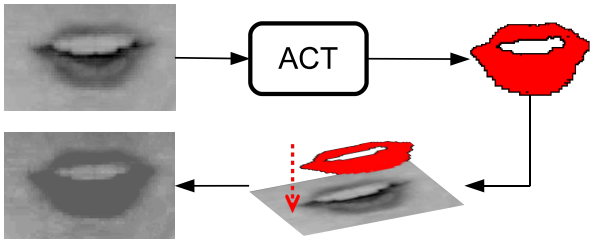

Figure 9: Scheme of fusing ACT result and *nG* transformed image

Figure 10: Combining Binary image information and ASM. Top Left: RGB, Top Right: binary result from ACT, Bottom Left: *nG*, Bottom Right: Fusion

One more application of the ACT is the elimination of inner gradients and distortions in the lip segment, in order to optimize the ASM algorithm. The threshold obtained in Eq. 8 which is used to refine the *nG* ASM work channel (See Fig.9) by applying following equation

$$I^*_{x,y} = \begin{cases} th & , I_{x,y} < th \\ I_{x,y} & , I_{x,y} \geq th \end{cases} \quad (10)$$

The fusion of binary image and original transformed *nG* channel has some advantage. As outlined in section 2.4.1 the method of ACT prefers a minimization of false positive lip pixels. To it is most likely it is missing a part of the mouth. Therefore in case of incompletely allocated lip pixels the soft edges at lip borders still remain, where in the binary image no border could be found there (see Fig. 10). This of course can also cause inner gradients, but have significantly smaller impact.

3 EXPERIMENTAL RESULTS

In the experiments the algorithm was tested on 57 images partially from the Faces Database from CIT [11] and partially from own recored data covering a wide range of illumination and saturation (see Table 2). The resulting mouth ROI had a size range of approximately 160x80 pixels. The mouth sizes varied in width between 120 and 150 pixels. The mouth height varied between 20 and 70 pixels. The higher variance is due to the opening of the mouth as greater impact to the height than the e.g. smiling has impact to the width. For each of the 57 images a ground truth was created consisting of a binary blob for lip pixels and a contour (which is equal to the outline of the binary blob).

The following sub sections will present the results and quality of the single process steps (Mouth corner point detection, lip pixel classification using ACT and Mouth contour detection using ASM).

Channel[Range]	*H[0,360]*	*S[0,1]*	*I[0,1]*
Mean	129.6	0.37	0.53
Variance	3666.5	0.02	0.04
Max	70.9	0.19	0.24
Min	247.6	0.72	0.91

Table 2: Image Conditions (in ROI), H=Hue, S= Saturation, I=Intensity

3.1 Detection Rate of Feature Points

Deviation for measuring detection quality of single feature points inside the face (mouth corners in this work) is given in relation to inter-ocular distance of the person (distance of both eye centers). Though this value, in relation to the face size, suffers inter-individual variations it is commonly used in shortage of better alternatives. We provide the results relative and additionally as pixel error (normal and squared) in table 3. The results with an accuracy of less than 10% relative error are good compared to other works [14]). Since the detection points are primary used to determine the ROI it was important to detect the mouth corners at all with sufficient accuracy.

Error Type	*Left*	*Right*	*Overall*
Relative	7.0%	6.5%	6.8%
Pixel Error	$7.05px$	$6.41px$	$6.73px$
Sqr Pixel Error	$100.04px^2$	$86.32px^2$	$93.17px^2$

Table 3: Error of mouth corner detection using method described in section 2.1

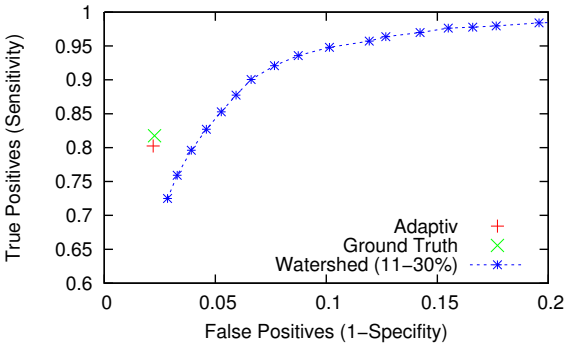

Figure 11: The cross on the left marks the quality of our proposed algorithm. The curves represent the ROC plots for different watershed percentages ranging from 10 to 30% in different color spaces. It's obvious here, that our adaptive threhsold is superior to threhsolds selected by a general watershed percentage. This is independent from the choosen color space or the choosen percentages for the watershed

3.2 Quality of ACT

The accuracy of the lip pixel classification is crucial for the idea of the proposed method. To show the good performance the results were put in context to an Watershed classification method (e.g. used in [9]), which is adaptive to the coloring but not to the ratio of mouth size to ROI size. Referring to the binary blob of the ground truth the adaptive classification of lip pixels, using the method proposed in in section 2.4.1, reaches a hitrate of 80.24% (True Positives - TP). However, this is no perfect result. When selecting a threshold from ground truth the rate is only insignificantly better (81.75%) (see Table 4). This is the best available hitrate aiming on low False Acceptence Rate (a maximum FAR of 5% was defined when selecting threshholds using groud truth) and False Rejection Rate. As mentioned in section 2.4.1 the primary aim is to avoid falsely accepted lip pixels (FAR). The *optimal* results in Table 4 represent results for thresholds which were chosen with respect to the ground truth and a ROC-Plot.

3.3 Improvement of ASM

In section 2.4.2 several improvements of the classical ASM algorithm were introduced. This subsection describes the impact of the different improvements. In first stage each of the three improvements were applied independent from each other, to analyze their individual impact to the algorithm. The error is calculated as average of all model-points. Ground truth was the outline of the ground truth blobs (See Fig. 3). So to each point of the ASM the distance to the closest point of the outline was calculated. The results are listed as normal and square error in Table 5. The algorithm is parametrized as outline in section 2.3. However, the second improvement of height fixing before ASM initialization is based on the information of the ACT generated blob. To measure the impact of this improvement the ASM was initialized using ACT but applied than to the unfixed *nG* image. The best refinement of the results is achieved by the initial height fixing. This finding should be considered in context of the chosen profile length in the ASM algorithm. Longer profiles could supersede the height initialization. On the other hand the ASM could get attracted by far objects like nose or eventually even by the chin (longer profiles of course would demand larger regions of interest). In non frontal views too long profiles also could touch regions outside the face, with unpredictable behavior. To avoid this, the algorithm would need a (likely on skin color based) good face segmentation. The ACT refinement and fixing of the ASM at the initialization points have only little but noticeable effect to the results.

Figure 12: Results of different illumination and mouth poses.

When the ASM is initialized there are two options to chose the initialization points: a) the corner points which were the base for the ROI, found by the method outlined in section 2.3; b) the corner points, based on the blobs found by the adaptive threshold defined in 2.4.1. To measure the impact of different sources for initialization of ASM both available options a) and b) were exploited and additionally c) a run utilizing the ground truth points for mouth corners. These runs were done using all optimization methods listed above (ACT Refinement, Anchored Corners and ACT based height refinement). Apparently the start points taken by ACT result in a similar quality as the points chosen by ground truth in average. The facial feature points found by the AdaBoost trained Haar-Like features are sufficient to define a ROI. For the further steps if ASM initialization they lead to less accuracy.

4 SUMMARY AND CONCLUSION

In this paper new modifications for Active Shape Models based on an adaptive color based method for lip

Result Base	*TP-μ*	*TP-σ^2*	*FAR-μ*	*FAR-σ^2*
Optimal (by ground truth)	81.75%	3.88%	2.262%	0.002%
Proposed	80.24%	2.55%	2.200%	0.031%

Table 4: Results for Lip Pixel Classification (Histogram based using ACT)

Method	*Error*	*Square Error*
Classic ASM	$3.21px$	$18.24px^2$
(1) ACT Refinement	$2.88px$	$19.40px^2$
(2) Height Fix	$2.08px$	$10.19px^2$
(3) Anchored	$2.89px$	$19.86px^2$

Table 5: Impact of separately introduced improvements to ASM Algorithm

Init Method	*Error*	*Square Error*
Haar-Like	$2.75px$	$18.24px^2$
ACT	$2.10px$	$10.17px^2$
Groundtruth	$2.08px$	$9.06px^2$

Table 6: Impact of different ASM Initialization

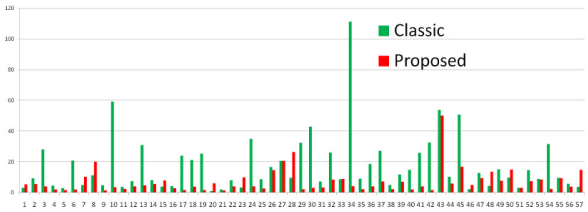

Figure 13: Image wise squared error of the proposed algorithm compared to classical method.

pixel classification were introduced. In contrast to other methods using color emphasizing of lip pixels, this method incorporates a refinement step. This refinement step eliminates edges inside the lip pixel segments, which can mislead the borders during the edge fitting step. The refining of *nG* image for edge fitting mainly helps to detect the lower mouth border. In this areas the crossover from skin to lip pixels often does not create a significant edge (for women this occurs more rarely due usage of lipsticks). Further the lip pixel classification creates a rough mouth blob. Based on this blob the shape model can be initialized better and closer to the real shape. The lip pixel classification performs good and is adaptive to various image conditions and skin tones. This skin vs lip color model assumption is designed and limited for Caucasian, European and Asian skin types. Further this method will suffer problems for very dark colored subjects respectively less illuminated scenes. Also the problem of bearded people was not addressed here. The more the beard color is different from general skin tone(light-gray, black) the greater the chance that this method fails. But this problem remains to all so far known methods and need further investigations and other solutions. Compared to classical shape models the presented method performs more accurate. In future works we will try to incorporate more shape modes and add a inner contour for opened mouth.

ACKNOWLEDGEMENTS

This work was supported by DFG-Schmerzerkennung 473 (FKZ: BR3705/1-1), Innovationsfond der Universität Magdeburg, CBBS C4 M: (FKZ: UC4 -3704M) and DFG-Transregional Collaborative Research Centre SFB/TRR 62

REFERENCES

[1] A. Al-Hamadi, A. Panning, R. Niese, and B. Michaelis. A model-based image analysis method for extraction and tracking of facial features in video sequence. In *ICSIT 2006, Amman,Vol.3*, pages 499–509, 2006.

[2] S. Arca, P. Campadelli, and R. Lanzarotti. A face recognition system based on local feature analysis. In *Audio- and Video-Based Biometric Person Authentication*, pages 182–189, 2003.

[3] C. Bouvier, P.Y. Coulon, and X. Maldague. Unsupervised lips segmentation based on roi optimisation and parametric model. In *IEEE International Conference on Image Processing*, pages IV: 301–304, 2007.

[4] Jingying Chen, Bernard Tiddeman, and Gang Zhao. *Advances in Visual Computing*, volume 5359/2008 of *LNCS*, chapter Real-Time Lip Contour Extraction and Tracking Using an Improved Active Contour Model, pages 236–245. Springer, 2008.

[5] P. Cisar and Zelezny M. Using of lip-reading for speech recognition in noisy environments. In *Speech Processing*, pages 137–142, Prague, 2004. Academy of Sciences of Czech Republic.

[6] T.F. Cootes, D. Cooper, C.J. Taylor, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.

[7] N. Eveno, A. Caplier, and P.Y. Coulon. Accurate and quasi-automatic lip tracking. *Circuits and Systems for Video Technology*, 14(5):706–715, May 2004.

[8] E. A. Ince and S. A. Ali. An adept segmentation algorithm and its application to the extraction of local regions containing fiducial points. In *ISCIS*, pages 553–562, 2006.

[9] J.Y. Kim, S.Y. Na, and R. Cole. Lip detection using confidence-based adaptive thresholding. In *International Symposium on Visual Computing*, pages I: 731–740, 2006.

[10] Trent W. Lewis and David M.W. Powers. Lip feature extraction using red exclusion. In Peter Eades and Jesse Jin, editors, *Workshop on Visual Information Processing*, volume 2 of *CRPIT*, pages 61–67, Sydney, Australia, 2001. ACS.

[11] California Institute of Technology. Faces 1999 (front). http://www.vision.caltech.edu/archive.html, 1999.

[12] A. Panning, A. Al-Hamadi, R. Niese, and B. Michaelis. Facial expression recognition based on haar-like feature detection. *Pattern Recognition and Image Analysis*, 18(3):447–452, 2008.

[13] Paul Viola and Michael Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001.

[14] Danijela Vukadinovic and Maja Pantic. Fully automatic facial feature point detection using gabor feature based boosted classifiers. In *International Conference on Systems, Man and Cybernetics*, volume 2, pages 1692–1698, Hawaii, 2005.

[15] Frank Wallhoff. Facial expressions and emotion database. http://www.mmk.ei.tum.de/ waf/fgnet/feedtum.html, Technical University of Munich, 2006.