Journal of Information & Computational Science 9: 9 (2012) 2505–2514
Available at http://www.joics.com

A New Color Transformation Based Fast Outer Lip Contour Extraction ⋆

Jing Pan [a], Yepeng Guan [a,b,*], Shuangcheng Wang [a]

[a] *School of Communication and Information Engineering, Shanghai University Shanghai 200072, China*

[b] *Key Laboratory of Advanced Displays and System Application, Ministry of Education*

Abstract

Various methods for lip segmentation have been proposed, it still remains a challenging and difficult problem due to high variability of lip color and low chromatic contrast between the lip and skin. A novel automatic lip segmentation algorithm is proposed based on a novel color transformation in RGB space instead of complex color models. Comparative study with some existing lip segmentation algorithms has indicated the superior performance of the developed algorithm. The proposed algorithm produces better segmentation fast and efficiently.

Keywords: Color Transformation; Lip Contour Extraction; Component Labeling; Curve Fitting

1 Introduction

Lip contour contains a great deal of information. Lip extraction technique has great potential for human-computer interaction, and it has received much attention in recent years. Its application involves in science and social activities, such as audio-visual speech recognition [1]-[3], speaker identification, teaching deaf children pronunciation and so on. Moreover, speech recognition with visual features can greatly improve the accuracy in noisy environment. But it's faced with some difficult problems due to high variability of lip color and shape. The low chromatic contrast between the lip and skin is another problem for lip detection. It is a vital task to develop an accurate and efficient method for lip-contour extraction.

Many lip detection algorithms have been proposed in the past few years. In the color based method, a color transformation or color filter is used to enlarge the difference between the lip and skin [4]-[8]. Then perform post-processing to obtain the final lip segmentation. The processing time of these methods is low, however, these methods are sensitive to lighting and color contrast so that they cannot satisfactorily outline the lip. Some researchers try to apply clustering to solve this problem [7]-[10]. The number of clusters should be given in advance. The appearance

⋆Project supported by the National Nature Science Foundation of China (No. 11176016, 60872117).

*Corresponding author.
Email address: ypguan@shu.edu.cn (Yepeng Guan).

September 1, 2012

of tooth, tongue and facial hair requires the number to be an adaptive value. It is still a difficult problem to deal with. In order to solve this problem, another method is based on model.Active contour model (snakes) [11], deformable template [12], active shape model [13, 14] and several other parametric models [15, 16] are used to extract the lip contour. These methods can represent important lip features in a low-dimensional parameter space, and they are invariant to rotation, scaling and illumination. However, the major difficulty is that the construction of these models usually needs a large training to cover the variability of lips. The training process consumes too much time. Besides, many relative parameters should be initialized manually.

In this paper, a new scheme aiming for speed and accuracy is proposed to extract outer lip contour in color images. RGB color space is selected to segment lip from skin in this paper. R component is more prominent than other components in the lip region. A new color transformation is proposed to highlight the difference between the lip and the non-lip regions. The main contribution is that the developed color transformation simplifies the complex computation in RGB space instead of other complicated color models to improve the lip extraction. The second contribution is that the proposal is fast and robust by comparative study with some existing lip segmentation algorithms.

The paper is organized as follows: in Section 2, details of the lip segmentation algorithm are presented. A new measure that takes both the color information and the geometric characteristic into account is proposed. Section 3 presents the implementation of the whole lip segmentation process. Experimental results of the proposed algorithm are discussed in Section 4. The performances of the developed algorithm are compared with other lip segmentation methods. Finally, Section 5 draws the conclusions and further work.

2 Lip Segmentation

Lip segmentation is a difficult task because the color between the lip and skin is similar, and the lip edge is usually ambiguous. The most intuitional way to perform lip segmentation is by classifying the image's pixels based on their color properties. X. Zhang [17] used Hue of the HSV space for lip detection. Hue is free from the influence of illumination, but it is hardly reliable for low saturation pixels. Saturations of the lips and the skin are well separated. The S is defined as [18]:

$$S = \frac{Max(R, G, B) - Min(R, G, B)}{Max(R, G, B)} \tag{1}$$

Since R component is the greatest and B component is the smallest of RGB in the lip and skin region, $Max(R, G, B)$ is R and $Min(R, G, B)$ is B. Simplify the definition of S as following.

$$S = \frac{R - B}{R} \tag{2}$$

To analyze the difference between R-G and R-B, we extract 9817 pixels from the Aleix face database [19] to construct the following distributions shown in Fig. 1. One can see that the tongue has a serious influence on lip segmentation especially the overlap between the lip and tongue in Fig. 1 right map.

To reduce the impact of tongue, choose R-G to further analysis. Transform S to s_1, s_2, s_3 as

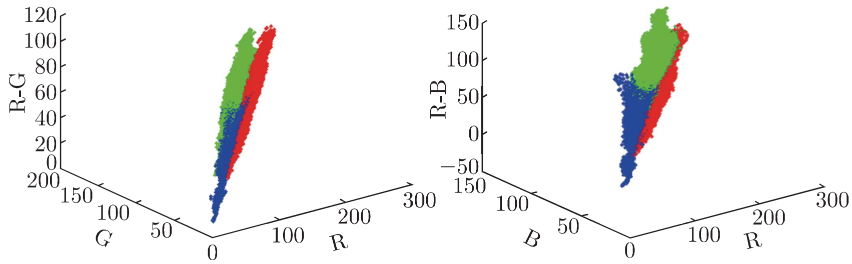

Fig. 1: R-G and R-B distribution (red represents lip, green is skin, and blue is tongue) from left to right, respectively

follows.

$$s_1 = \frac{R - G}{R}, s_2 = \frac{R - G}{G}, s_3 = \frac{R - G}{B} \tag{3}$$

Fig. 2 shows the distribution of above mentioned models respectively. From Fig. 2, it can be observed that overlap between the lip and non-lip region is smaller in Fig. 2 (a) than the others. It indicates that the distinction between the lip and skin of s_1 is more obvious than that of s_2 or s_3. In order to further overcome the effect of the face with cosmetic or beards, it is necessary to normalize s_1 to interval [0, 1] as follows.

$$s = 2 \times \arctan\left(\frac{R - G}{R}\right) / \pi \tag{4}$$

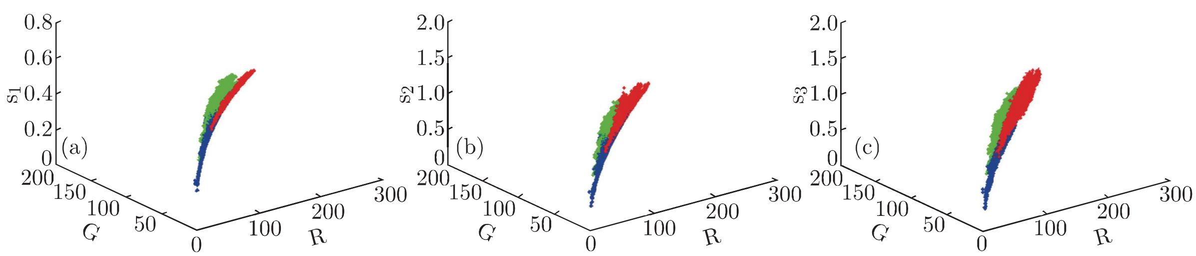

Fig. 2: Lip tone distribution (red is lip, green is skin, blue is tongue). s_1, s_2, and s_3 from left to right, respectively

3 Implementation of Outer Lip Contour Extraction

3.1 Lip Localization

In order to locate the lip from face images, it is necessary to perform face detection at first. Face detection technique developed by Viola and Jones [20] is adopted. To improve lip detection, set a mouth Region of Interest (ROI). Saeed and Dugelay in [21] proposed that the mouth ROI was the lowest one-third and the middle half of the detected face region. The method for lip localization is suitable for general face images as shown in Fig. 3.

In some cases, however, when mouth opens wide or the face is skew, the located face region will be not complete as shown in Fig. 4.

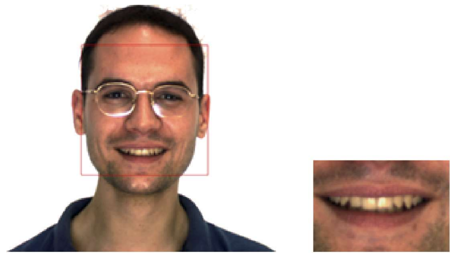

Fig. 3: Face detection and initial lip location by [21]. Face detection, initial lip location from left to right, respectively

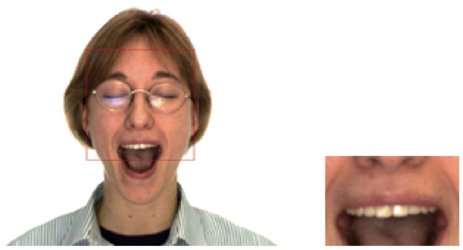

Fig. 4: Face detection and initial lip location by [21] when mouth opens widely. Face detection, initial lip location from left to right, respectively

To overcome the problem, it is necessary to expand appropriately mouth region. We find that when the detected height of facial region is expanded one-fifth downwards by the Aleix face database, it will get a complete mouth region as shown in Fig. 5.

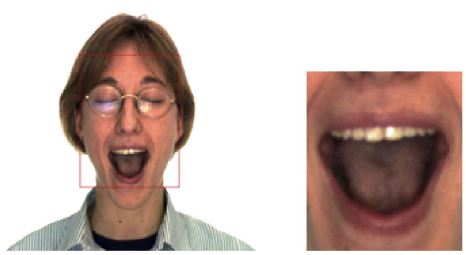

Fig. 5: Lip localization after expanding. The detected face region after expanding, located mouth ROI after expanding from left to right, respectively

3.2 Lip Segmentation

Assuming the real number of pixels belonging to the lip is N, compute s as (4) and compute the number of pixels of every s class. As mentioned above, the s value of the lip region is higher than non-lip region. Sum the number of the ordered pixels until it equals N, the corresponding s value is the segmentation threshold shown in Fig. 6. Set an adaptive threshold T to segment lip as (5).

$$g(x,y) = \begin{cases} 255, & s \geq T \\ 0, & s < T \end{cases} \tag{5}$$

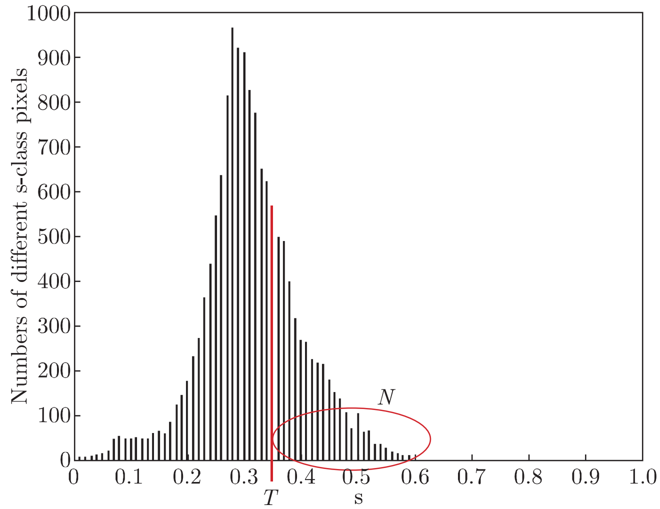

Fig. 6: The s histogram of Fig. 5 mouth ROI

Hamed and Khashayer [8] proposed that the N is within the range from 10% to 20% of the initial located mouth ROI. In order to determine an efficient value for different face images to robustly extract the outer lip from the mouth ROI, the Aleix face database and Caltech face database [22] are selected to analyze. Some results are given in Fig. 7.

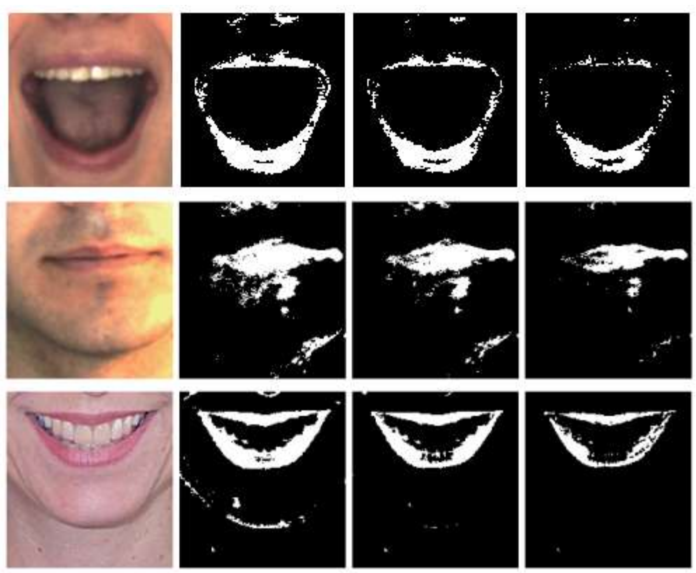

Fig. 7: Lip initial segmentation by different values from the Aleix face database and Caltech face database. Mouth ROI, lip segmentation using N =20%, N =15%, and N =10% from left to right, respectively. The images in the first two rows are from the Aleix face database, and the image in the third row is from Caltech face database.

It can be noted that the lip segmentation is distinct with different values of N by the proposed method. It indicates that the proposed color transformation can robustly discriminate the lip region from the mouth.

3.3 Post-processing

From Fig. 7, it can be seen that the extracted upper lip and lower lip are not connected especially when mouth opens wide. It means lip contour cannot be extracted simply by edge detection. A curve fitting method in [23] is applied to outline the lip contour. There are two common algorithms for this method including the nearest-neighbor interpolation and the bilinear interpolation. Take the computational complexity into consideration, the nearest-neighbor interpolation is selected. Points on the upper lip could be adequately represented by a fourth-order polynomial and points on the lower lip by second-order polynomial. The extracted outer lip contour is shown in Fig. 8.

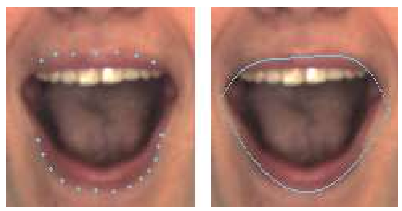

Fig. 8: The s histogram of Fig. 5 mouth ROI. The results of the outer lip. Some points on the outer lip edge, lip contour from left to right, respectively

4 Experimental Results

To demonstrate the effectiveness of the proposed approach for lip segmentation, some experiments with a large number of lip images from the Aleix face database and Caltech face database are selected to compare. In the experiments, the N is selected as 15% and keeps the same. The segmentation results are compared with the developed algorithms in [7] and [8]. Five different lip images are used to compare. The selected images are given in Fig. 9.

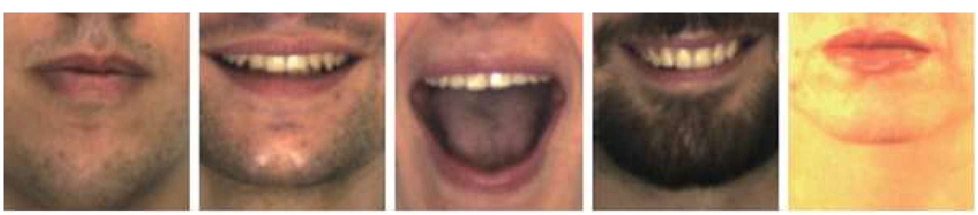

Fig. 9: Some selected lip images

One person shuts his mouth up in the first image (141×150 pixels), smiles with teeth in the second image (138×147 pixels), opens mouth wide in the third image (117×124 pixels), has dense beard in the fourth image (147×157 pixels), and is under different lighting in the fifth image (140×149 pixels).

Ghaleh and Behrad in [7] used RGB color space combined with Fuzzy C-means clustering (FCM) for lip segmentation. Hamed and Khashayar in [8] proposed a pseudo hue transform to segment the lip. The threshold is calculated by finding the maximum in the histogram of the transformed image. Some segmented results are shown in Fig. 10, Fig. 11, and Fig. 12, respectively.

From Fig. 10 one can see that the algorithm proposed in [7] extracts a lot of skin pixels as lip region. The algorithm in [8] extracts limited pixels as lip region in the first three images and is

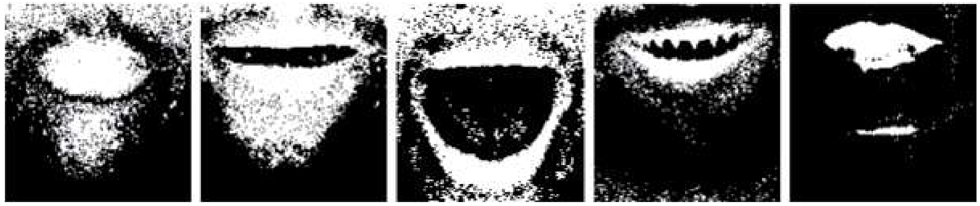

Fig. 10: Some lip segmentation in [7]

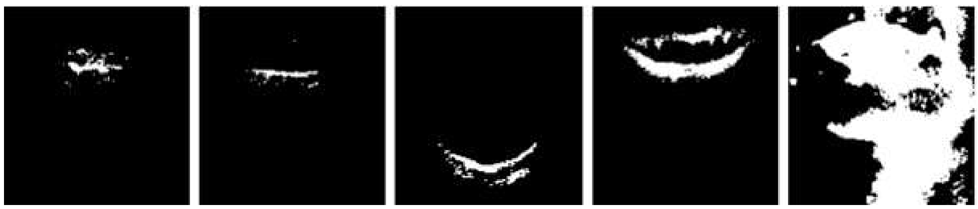

Fig. 11: Some lip segmentation in [8]

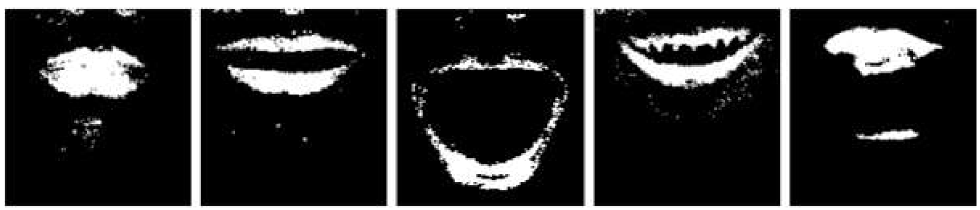

Fig. 12: Some lip segmentation by the proposal

poor under lighting condition from Fig. 11. The background produced by our developed algorithm is considerably clearer in comparison with the others, and the boundary is much smoother. Compared with the experimental results obtained by the methods in [7] and [8], it demonstrates the superiority of the proposed method in this paper.

In the following we apply a quantitative technique to evaluate the performance of our algorithm for lip segmentation. Since no ground truth is available, we manually draw the boundary of the lip for the lip images. The measure is the Segmentation Error (SE) defined as [9]:

$$SE = \frac{OLE + ILE}{2 * TL} \tag{6}$$

where, TL represents the actual lip region drawn up manually, OLE (Outer Lip Error) is the number of non-lip pixels extracted as lip pixels by mistaken, ILE (Inner Lip Error) is the number of lip pixels classified as non-lip pixels. If segment lip efficiently, the SE value will approach to 0%.

From Table 1, one can see that the SE percentages of ours are much smaller than that of all other methods investigated.

Another 30 images from the Aleix face database and Caltech face database have been selected randomly. The images of 85 speakers in the Aleix face database except those with scarf occlusion and 450 frontal face images of 27 people in the Caltech face database are used in the experiment. The selected images contain different lip shapes, lighting conditions and backgrounds. The average SE percentages of the three methods across these selected 35 images are given in Table 2 and our approach is proved again to have the smallest SE.

Table 1: Comparison of SE for lip segmentation in different methods

Images	Methods (%)		
	[7]	[8]	proposal
1st	137.8451	40.4262	9.2437
2nd	178.2474	45.7216	3.5051
3rd	124.5582	44.2003	7.5215
4th	87.5633	7.43671	25.6329
5th	10.4177	226.1766	7.6150

Table 2: Comparison of the average SE for lip segmentation in different methods

	Methods (%)		
	[7]	[8]	proposal
Average SE	153.2150	52.1617	9.1265

To further evaluate efficiency of our algorithm, we carry out lip segmentation experiments for the five images in Fig. 9 by a Pentium Dual T2390 2G 1.86GHz PC with C language. The results are given in Table 3. Obviously the proposed method is faster than the other two approaches.

Table 3: Comparison of executive time for lip segmentation in different methods

Images	Methods (ms)		
	[7]	[8]	proposal
1st	2680	31	31
2nd	1810	32	31
3rd	1520	32	32
4th	2890	47	32
5th	1350	47	32

Some lip segmentation examples from the Aleix face database and the Caltech face database are shown in Fig. 13. The first three rows images are from the Aleix face database, and the last row images are from Caltech face database.

5 Conclusions

A new color transformation is proposed to enlarge the difference between the lip and non-lip region, which is used to improve the lip segmentation. The introduced scheme for lip contour extraction reduces the computational complexity and performs well. Although the technique for lip extraction is continuously improved, it is still faced with some challenging situations, such as high levels of noise, more than one person. Further research should focus on these issues and develop novel algorithms for lip segmentation quality.

Fig. 13: Some color lip segmentation results obtained by the proposed algorithm

References

[1] X. Zhang, R. M. Merserratt, M. A. Clements, Audio-visual speech recognition by speech reading, Proceedings of IEEE International Conference on Digital Signal Processing, 2 (2002), 1069-1072

[2] N. Eveno, A. Caplier, P. Y. Coulon, Accurate and quasi-automatic lip tracking, IEEE Transactions on Circuits and Systems for Video Technology, 14 (2004), 706-715

[3] S. Alizadeh, R. Boostani, V. Adapour, Lip feature extraction and reduction for HMM-based visual speech recognition systems, Proceedings of International Conference on Signal Processing, 2008, 561-564

[4] Juan B. Gómez-Mendoza, Flavio Prieto, T. Redarce, Automatic lip-contour extraction and mouth-structure segmentation in images, Computing in Science and Engineering, 13 (2011), 22-30

[5] W. Gao, Y. Q. Chen, R. Wang, S. G. Shan, D. L. Jing, Learning and synthesizing MPEG-4 compatible 3-D face animation from video sequence, IEEE Transactions on Circuits and Systems for Video Technology, 13 (2003), 1119-1128

[6] U. Canzler, T. Dziurzyk, Extraction of non manual features for video based sign language recognition, Proceedings of the International Association for Pattern Recognition workshop on Machine Vision Applications, 2002, 318-321

[7] V. E. C. Ghaleh, A. Behrad, Lip contour extraction using RGB color space and fuzzy c-means clustering, Proceedings of IEEE 9th International Conference on Cybernetic Intelligent Systems, 2010, 1-4

[8] T. Hamed, Y. Khashayar, Automatic combined lip segmentation in color images, Proceedings of IEEE 3rd International Conference on Communication Software and Networks, 2011, 109-112

[9] A. W. C. Liew, S. H. Leung, W. H. Lau, Segmentation of color lip images by spatial fuzzy clustering, IEEE Transactions on Fuzzy Systems, 11 (2003), 542-549

[10] S. L. Wang, W. H. Lau, W. C. Liewa, S. H. Leung, Robust lip region segmentation for lip images with complex background, Pattern Recognition, 40 (2007), 3481-3491

[11] N. Eveno, A. Caplier, P. Y. Coulon, Key points based segmentation of lips, Proceedings of IEEE International Conference on Multimedia and Expo, 2 (2002), 125-128

[12] A. W. C. Liew, S. H. Leung, W. H. Lau, Lip contour extraction from color images using a deformable model, Pattern Recognition, 35 (2002), 2949-2962

[13] I. Shdaifat, R. Grigat, D. Langmann, Active shape lip modeling, Proceedings of International Conference on Image Processing, 2 (2003), 875-878

[14] X. Liu, Y. Cheung, M. Li, H. Liu, A lip contour extraction method using localized active contour model with automatic parameter selection, Proceedings of International Conference on Pattern Recognition, 2010, 4332-4335

[15] N. Eveno, A. Caplier, P. Y. Coulon, A parametric model for realistic lip segmentation, Proceedings of the 7th International Conference on Control, Automation, Robotics and Vision, 2002, 1426-1431

[16] M. -T. Yang, Z. -W. You, Y. -C. Shih, Lip contour extraction for language learning in VEC3D, Machine Vision and Applications, 21 (2009), 33-41

[17] X. Zhang, R. M. Mersereau, Lip feature extraction towards an automatic speech reading system, Proceedings of the International Conference on Image Processing, 3 (2000), 226-229

[18] S. H. Li, G. Z. Guo, The application of improved HSV color space model in image processing, Proceedings of the 2nd International Conference on Future Computer and Communication, 2 (2010), 210-213

[19] A. M. Martinez, R. Benavente, CVC Technical Report #24, http:// rvl1. ecn. purdue. edu/~aleix/aleix_face_DB. html, accessed June 1998

[20] P. Viola, M. J. Jones, Robust real-time face detection, International of Computer Vision, 57 (2004), 137-154

[21] U. Saeed, J. -L. Dugelay, Combining edge detection and region segmentation for lip contour extraction, Proceedings of the 6th International Conference on Articulated Motion and Deformable Objects, 2010, 11-20

[22] W. Markus, Caltech Face, http://www. vision. caltech. edu/html-files/archive.html, accessed 1999

[23] G. -S. Xu, Sub-pixel edge detection based on curve fitting, Proceedings of International Conference on Information and Computing Science, 2 (2009), 373-375