Region-Based Super-Resolution Aided Facial Feature Extraction from Low-Resolution Video Sequences

T. Celik, C. Direkoglu, H. Ozkaramanli, H. Demirel, M. Uyguroglu

Advanced Technologies Research and Development Institute,
Department of Electrical and Electronic Engineering, Eastern Mediterranean University,
email: turgay.celik@emu.edu.tr

ABSTRACT

Facial feature extraction is a fundamental problem in image processing. Correct extraction of features is essential for the success of many applications. Typical feature extraction algorithms fail for low resolution images which do not contain sufficient facial detail. In this paper, a region-based super-resolution aided facial feature extraction method for low resolution video sequences is described. The region based approach makes use of segmented faces as the region of interest whereby a significant reduction in computational burden of the super-resolution algorithm is achieved. The results indicate that the region-based super-resolution aided extraction algorithm provides significant performance improvement in terms of correct detection in accurately locating the facial feature points.

1. INTRODUCTION

Facial feature extraction has been a topic of extensive research for several decades [1, 2]. This preprocessing is essential for the success of many applications such as teleconferencing, animation, facial expression analysis, man machine interfacing, face recognition and lip reading for the deaf etc. In facial expression analysis even the smallest errors can be interpreted differently leading to wrong facial expression recognition. In 3D modeling the correct extraction of features is imperative for the model adaptation, texture mapping and the subsequent animation.

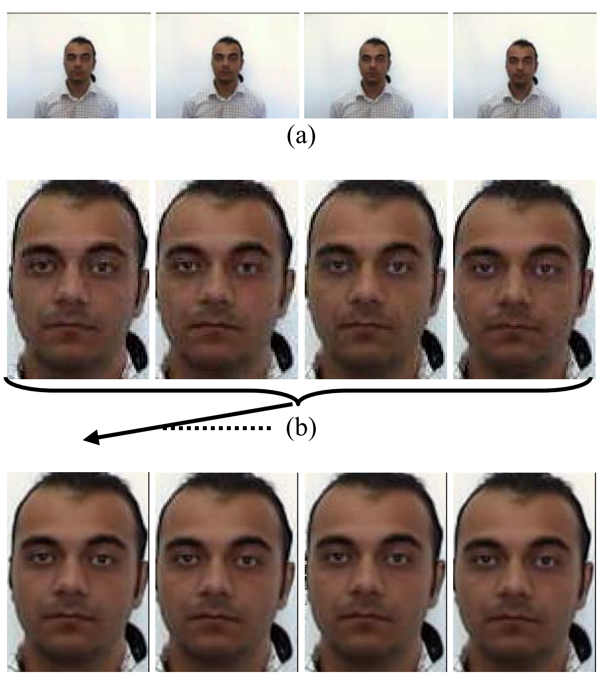
Most facial feature extraction methods are sensitive to various non-idealities such as variations in illumination, noise, orientation, color space used. The problem becomes more challenging for adverse conditions where the resolution of the camera limits the quality of the images that one has to work with. Inserting the crucially needed details alleviates the difficulties. As a technique which can produce a high resolution image from a sequence of low resolution images, super-resolution [3] fills this gap very nicely. Thus, in this paper we propose a region-based super-resolution aided facial feature extraction technique for video sequences that may not contain enough facial details. The region of interest, the human face in this paper is segmented from the incoming video. This region of interest is then super-resolved, improving the quality of the region from the facial features is extracted, using a frequency domain registration method and the subsequent interpolation [4]. Here the aim in region-based super-resolution is to decrease the processing time of the super-resolution algorithm, due to the application of the super-resolution algorithm to a much smaller region. Thus the super-resolution algorithm may become feasible for real-time implementation. The facial feature extraction part involves a position, scale, shape and skin color invariant face detection part and the subsequent extraction of feature points based on an efficient combination of methods such as morphology, median filtering, adaptive intensity clustering, edge detection, circle and ellipse fitting. For comparison purposes correct location of facial features are manually extracted for both the low resolution and super-resolution images. Our results indicate that the region-based super-resolution aided method provides significant improvement in correctly detecting locations of the facial features in video sequences taken in adverse conditions.

2. REGION-BASED SUPER-RESOLUTION

The idea of super-resolution was first introduced in 1984 by Tsai and Huang [5] for multi-frame image restoration of band-limited signals. A good overview of existing algorithms is given by Borman and Stevenson [3]. Recently super-resolution image reconstruction has been a topic of active research. It promises to be a good tool which can improve the performance of many image processing applications. Super-resolution is comprised of two main steps. First, the registration of the low resolution images onto a high resolution grid and second the super-resolution image formation. Correct registration is critical for the success of super-resolution image reconstruction. If the images are inaccurately registered, the high resolution image is reconstructed from incorrect data and it is not a good representation of the original image.

0-7803-8874-7/05/$20.00 ©2005 IEEE

Popular methods for the registration are the spatial domain method of Karen [6] and the frequency domain approach as described in [4]. The subsequent reconstruction of the high resolution image can be accomplished by interpolation or iteratively as described by Peleg et. al., [7]. In this work, the incoming video is segmented to isolate the face region from every frame. Following the segmentation a frequency domain registration method and the subsequent interpolation using four consecutive frames as described in [4] is employed. It should be pointed out that the segmentation of the face from the incoming video also helps the registration process since it limits the amount of vertical and horizontal shifts in the segmented face sequence. Fig. 1a shows the original low-resolution frames, the segmented frames which are input to the super-resolution algorithm are depicted in Fig. 1b, and the super-resolved face images are shown in Fig. 1c.



(c)

Fig. 1: (a) The original low-resolution frames - 240 x 320 Pixels, (b) Segmented face images - 100 x 73 Pixels,
(c) Super-resolved face images - 200 x 146 Pixels

3. FACIAL FEATURE EXTRACTION

The facial feature extraction method uses the segmented and super-resolved faces for the processing. Fig. 2 shows the flowchart of the overall facial feature extraction algorithm. The first step is the face detection part. The feature invariant approach with multiple attributes such as skin color and shape is a popular method for this purpose [1]. YCbCr skin color modeling is used to obtain a binary image with a set of connected components. Face candidate regions from these components are prepared by using median filtering and morphological operations to remove non face regions. Several heuristic criteria such as holes inside the face candidate region together with orientation

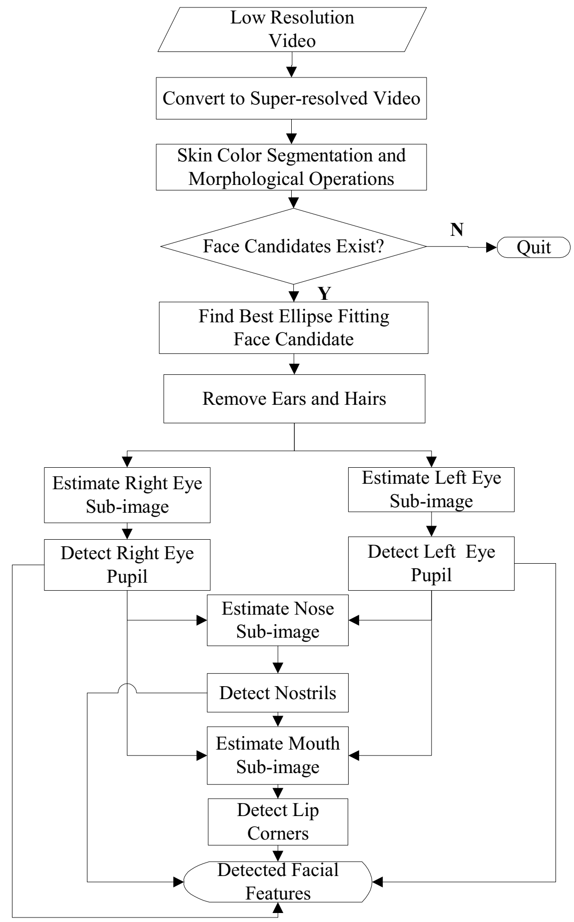
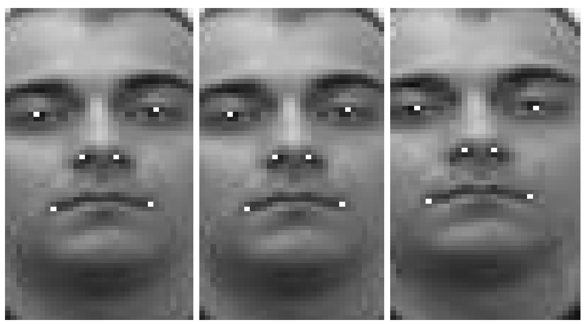


Fig. 2: Flowchart of the facial features detection algorithm

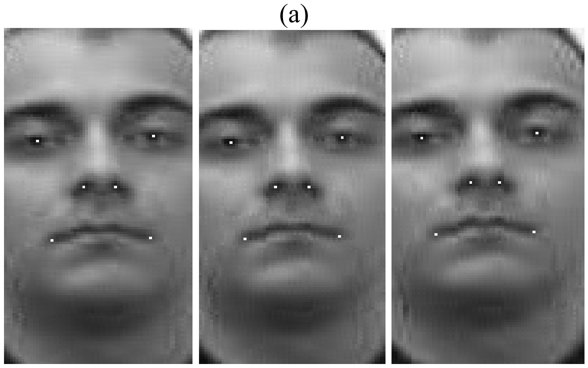
and a best ellipse fitting approach constitute the two independent face candidate elimination procedures. Once the face image is obtained we proceed by detecting the facial feature sub-images for the left and right eye, the nose and the mouth from which the feature points are extracted.

The detection of pupils involves an intensity based adaptive clustering on the gray level eye sub-images. As a result of this adaptive clustering, pupils and eyebrows remain as the two largest white regions. The novelty we introduce in the process of extracting pupil locations is achieved at this stage by the method of circle fitting. A quality measure of the fit is introduced and the region with the highest circle quality is declared as the pupil. Nostrils are also found by applying intensity based adaptive clustering to nose sub-image. The biggest two regions in the clustered binary image are the nostrils. More challenging problem is to find the lip corners. Lip cut formed by applying Sobel vertical derivative operator to the mouth sub image is the most important cue [2]. It limits the vertical position of lip corners. Again intensity based adaptive clustering is applied to limit the horizontal position of right and left lip corners. A combination of lip cut information and clustered binary sub-image is used for extracting the lip corners.

The pictorial results of the above method for three consecutive frames of the low resolution video sequence are depicted in the top row of Fig. 3. The white dots indicate the location of the detected features. In the bottom row of Fig. 3, the detected facial feature points for the same frames of the super-resolved sequence are depicted.



(a)



(b)

Fig 3: Detection of feature points, (a)Frames 5,6 and 7 of the low resolution sequence, (b)Frames 5,6 and 7 of the super-resolved sequence.

4. RESULTS AND DISCUSSIONS

The detection algorithm presented above is applied to the low resolution and the super-resolved video sequences. The original low resolution video sequences used in this comparison are 9 seconds records of a single slowly moving approximately frontal face images in constant background. A single video sequence contains 90 frames in total. A total of 2 video sequences were tested. The frames are of size 240x320. The face region constitutes a very small portion of the whole frame with approximately 100x73 pixels. This makes the facial features barely recognizable. In order to assess the effectiveness of the super-resolution aided feature extraction the feature points are manually extracted from both the low and high resolution images. Table 1 shows the results of this comparison. If a facial feature is detected to be within three pixels of the actual location (determined manually) it is declared to be a correct detection. In both video sequences tested the correct detection rate increases significantly with the super-resolution aided approach. Furthermore as can be observed from the Fig. 3 the feature points detected by the aid of the super-resolution approach converges to the exact points. This may prove to be very valuable in applications like facial expression analysis where the accurate locations of facial feature points are of utmost importance.

	Low Resolution	Super Resolution
	Correct Detection Rate (%) (3 Pixel Range)	Correct Detection Rate (%) (3 Pixel Range)
Left Eye	53.33	57.78
Right Eye	64.44	73.33
Left Nostril	90.00	98.89
Right Nostril	76.67	95.56
Left Lip Corner	78.89	95.56
Right Lip Corner	82.22	94.44

(a)

	Low Resolution	Super Resolution
	Correct Detection Rate (%) (3 Pixel Range)	Correct Detection Rate (%) (3 Pixel Range)
Left Eye	98.89	98.89
Right Eye	93.33	95.56
Left Nostril	97.78	98.89
Right Nostril	95.56	97.78
Left Lip Corner	94.44	97.78
Right Lip Corner	95.56	98.89

(b)

Table 1: Performance of the super-resolution aided feature extraction method (a) Subject A, (b) Subject B.

Subject A was chosen as a difficult subject in that the facial features are obstructed by illumination. He had a light beard and moustache which obstructed the nostrils and the lip corners. It should be noted that there is almost no skin color region between the eyelashes and the eye. On the other hand, Subject B is an easy subject with a distinct rounded face where the facial features are not obstructed. He had no beard or moustache. The skin color region between the eye and eyelashes can be distinctly observed in Fig. 4b. Furthermore no wrinkles exists on the face of subject B

As shown in Table 1 the results for the low resolution sequences are quite different for the two subjects. The feature extraction algorithm performs very well for Subject B. However it does very poorly for Subject A.

(a) (b)

Fig. 4: Subjects used in tests (a) Subject A (b) Subject B

For the super-resolution aided approach, improvement is achieved for both subjects. Up to 20% improvement in correct detection is achieved for subject A. The improvement for subject B is more moderate and up to 3%. The large improvement in the correct detection of facial features for subject A is a good indication of the usefulness of super-resolution aided approach.

5. CONCLUSIONS

A region-based super-resolution aided automatic facial feature extraction method is presented. First the segmentation is carried out and head is extracted from the incoming frames. Then the consecutive frames are used to reconstruct the higher resolution image frames via a frequency domain registration algorithm and the subsequent cubic interpolation on a higher resolution grid. The facial feature extraction method involves a position, scale, shape and skin color invariant face detection part and the subsequent extraction of facial features based on an efficient combination of geometry based methods. The results indicate significant improvement in the performance of the region-based super-resolution aided extraction algorithm.

6. REFERENCES

[1] M.H. Yang, D.J., Kreigman, and N. Ahuja, "Detecting Faces In Images: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, 2002.

[2] R. Lanzarotti, P. Campadelli, and N.A. Borghese, "Automatic Features Detection for Overlapping Face Images On Their 3D Range Models," *Proceedings of the IEEE 11th International Conference on Image Analysis and Proceeding (ICIAP2001)*, Palermo, Italy, 26-28, pp.316-321 Sept. 2001.

[3] S. Borman, and R. Stevenson, "Spatial resolution enhancement of low-resolution image sequences - a comprehensive review with directions for future research," *Tech. Rep.*, University of Notre Dame, 1998.

[4] P. Vandewalle, S. Süsstrunk, and M. Vetterli, "Double resolution from a set of aliased images," *SPIE/IS&T Electronic Imaging Conference*, vol. 5301, Jan. 2004.

[5] Huang, T.S., *Advances in Computer Vision and Image Processing*, JAI Press, vol. 1, ch. 7, pp. 317–339, 1984.

[6] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacement," *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp. 742–746, Jun. 1988.

[7] M. Irani, and S. Peleg, "Improving resolution by image registration," *CVGIP:Graphical Models and Image Processing,* vol. 53, no. 3, pp. 231–239, May 1991.