A Fast and Accurate Algorithm for Facial Feature Segmentation

B.K. Low and M.K. Ibrahim
DSP Systems and Applications Group
Department of Electronics and Electrical Engineering
School of Engineering and Manufacture, De Montfort University
The Gateway, Leicester LE1 9BN, UK
email:bklow@dmu.ac.uk, mki@dmu.ac.uk

Abstract

A facial feature segmentation algorithm for head and shoulder sequences is proposed. The method is based on simple spatial and temporal heuristics techniques for the purpose of practical implementation. Facial feature are segmented using two common cues i) regular and high image intensity variation in temporal domain due to various motion like talking and moving, ii) high edge density around the feature regions. Simple motion detection method and a generic spatial operator are applied to achieve the required segmentation. The paper addresses issue of how to combine both information to obtain final results and also assess the segmentation performance.

I. Introduction

The recent advance in face recognition technologies and others visual applications, has prompted an increase interest in face detection research. Various techniques have been proposed to achieve the segmentation and extraction task. In [4,8,20,25], color information is used to produce face candidates which are then used as features basis in face geometry and statistical analysis. Other methods include symmetry maps [10,14], integral projection [2,9], blob analysis [7,23], edge-based [6,9,11,26], subspaces/eigenspace [17,18]. Active shape models have also been applied mainly for facial features extraction. Deformable template is the most popular technique [5,7,9,27]. Snake and point-distributed model based techniques have also been presented in various applications [12,13,16]. Neural nets such as MLP are used in [3,22,24] to detect various-posed face in a complex scene. These methods although have achieved a considerable amount of success, most are still challenged by limitations of high computational requirement and complex solution (maximizing cost functions, solving partial equations etc.). In a practical system, it is important that the computational requirement and complexity of any part (especially pre-processing stage) are kept to minimal. In many practical/hardware implementations, simple and effective visual front-end processors are always preferred [20,25].

In this paper, we present a facial features segmentation framework for head and shoulder sequences using simple heuristics approach. The objective of the algorithm is to show that fast segmentation can be achieved by using only standard and simple complementary techniques. Two partial segmentation are acquired in parallel using both motion and spatial cue. These two solutions are then combined collaboratively to achieve a better segmentation performance. Motion cue allows the location of facial features to be inferred. On the other hand, spatial details around feature regions are provided by the spatial solution. This collaboration prevents the segmentation from relying on complicated one-solution method, for instance structure from motion (optical flow) and knowledge-based spatial methods.

II. Framework

The general structure of the algorithm is depicted in Figure 1. Each module of the framework will be explained in following sub-sections.

A. Locating Feature Using Motion Cue

Facial feature motion such as eye blinking and mouth movement always causes high image intensity change within a temporal period (high temporal change). The likelihood of the motion is high and natural (people are constantly moving, talking etc.) even in a short time period. Therefore, it is reliable and easy to extract facial

Proceedings of the 1997 International Conference on Image Processing (ICIP '97)
0-8186-8183-7/97 $10.00

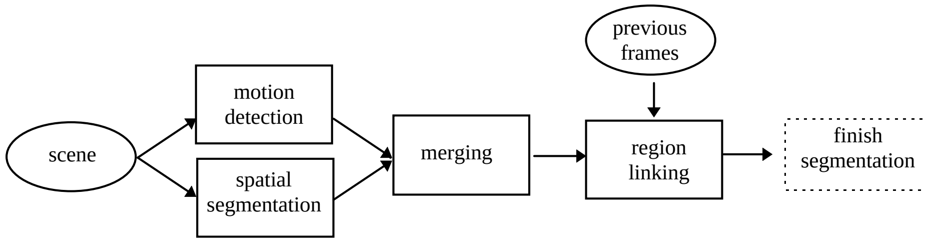

Figure 1: Framework of the segmentation algorithm

feature using motion cue. Another advantage is that complex background can be separated from the foreground much easier than using spatial methods.

A simple motion detection algorithm will be able to estimate the location of the fast motion region. Absolute interframe difference emphasizes image locations that change with time. With an appropriate high threshold value, binary motion image can be produced. The application of frame difference techniques have been reported in literature [1,23]. In these references, multiple interframe differences are fused together to find face silhouettes prior to facial feature location using spatial search. Our algorithm locates facial features directly using a single/short-term frame difference with a high threshold value [15].

B. Spatial Segmentation - Edge to Binary Region

The spatial solution provides the necessary spatial information for the merging stage. A spatial operator is used to extract image region of high edge density. This generic operator exploits the fact that facial feature edges are always dense while other moving edges such as head boundary are sparse. There are three basic operations in the process - *growing, smoothing* and *shrinking.* The growing and shrinking operations are similar to the edge growing techniques in [19] where the technique is used for connecting broken contour edges. The shrinking and growing mechanism on an image pixel are illustrated in Figure 2.

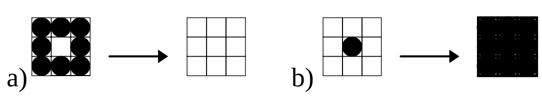

Figure 2: a) growing, b) shrinking

During the spatial operation, dense edges are fused together gradually to form various binary regions while solitary edges are being eliminated. Through interlacing the growing and shrinking operation in a non-linear and iterative manner [15], the fused regions resemble the actual physical appearance of the corresponding features and hence produce the required spatial information. Figure 3 gives some examples of the spatial segmentation on facial images.

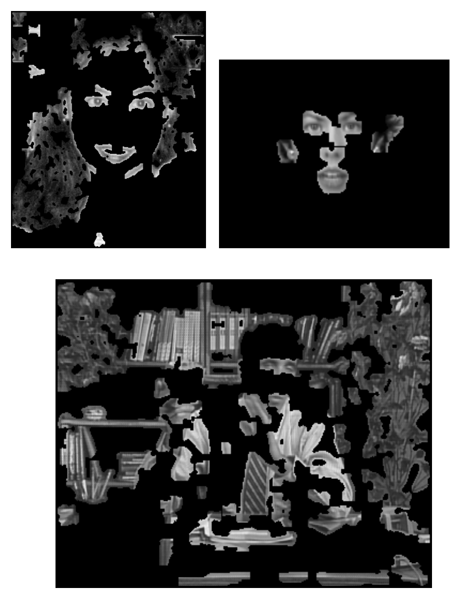

Figure 3: Examples of the spatial segmentation

C. Collaborative Merging

Both spatial and temporal solutions can be combined to achieve a better segmentation results. The combination routine which is termed as *collaborative merging* is described by $f(s,\tau)$ in equation 1. s and τ are the spatial and temporal solution respectively.

$$f(s,\tau) = T(R(s+\tau)) \qquad (1)$$

$R(\ )$ is an iterative region growing process and $T(\ )$ is a thresholding process. The whole process can be seen as

Proceedings of the 1997 International Conference on Image Processing (ICIP '97)
0-8186-8183-7/97 $10.00 © 1997 IEEE

expanding τ within the region boundaries of s and vice versa.

D. Region Linking

The algorithm suffers a drawback. Due to its time-dependent nature, continuous segmentation is not guaranteed through time. When facial feature motion is static or insignificant, the fast motion detector will fail to register the position of the features. In order to tackle this problem a region linking routine based on multiple hypothesis testing is applied. Feature regions from previous frame are projected onto current frame. The testing is carried out within the projected area. Two cases are considered

Case 1: Any region of current frame that coexist or overlap the projected area is put into correspondence with the corresponding region of previous frame. The new region then replaces the old region.

Case 2: The projected area is empty due to static and insignificant motion. Previous region information in this case, is used directly to recover the current feature position.

This linking routine tracks moving regions and at the same time fill in the non-motion gap between frames. A speedy global region growing and logical operation are applied to test all the regions simultaneously. No region labeling and local search for correspondence are performed due to the simplicity of the testing cases. The algorithm assumes the motion of features is not fast enough to result complete misalignment of corresponding region between frames. A drawback of this assumption is that segmentation error can be generated when regions are not linked correctly and subsequently causes 'leftover' regions throughout the sequence.

III. Performance on Video Sequence

A. Segmentation Measures

In order to study the performance of the algorithm on image sequences, we introduce two segmentation measures. We first segment salient features (eyes, mouth, nose and ear) from the sequence images manually and name the regions as *target regions*. Segmentation results that coincide with the target regions are termed *correct regions* while the residue are regarded as *segmentation error*. To measure the quality of a particular segmentation, we use i) hit ratio, area ratio between the correct and target region, ii) error ratio, area ratio between the segmentation error and the background.

B. Experimental Results

Experimental results for the video sequence 'Salesman' are presented in this section (see last example in Figure 3). In Figure 4a, the graphs of error ratio vs hit ratio for both initial spatial (x) and temporal (o) solutions are shown. Each point on the graph represents the segmentation ratio for one frame. The two different clusters indicate that both solutions are far from the ideal position of successful segmentation (1,0). This is because the segmentation methods are too simple to handle complex background and motion. Figure 4b shows that a slight improvement when both solutions are merged as more points are drawn near to (1,0). The segmentation inconsistency due to temporal dependency is also clearly evident with all point scatter around the y=0 line. Finally the performance of the algorithm has improved a great deal as shown in Figure 4c when the region linking routine is applied. Most of the points are now nearer to the ideal position. The percentage of complete capture (with background error tolerance) of eyes and mouth region for the salesman sequence are 100% and 86% out of 64 frames respectively. The figures are similar for another sequence 'Claire' (95% and 94% out of 166 frames).

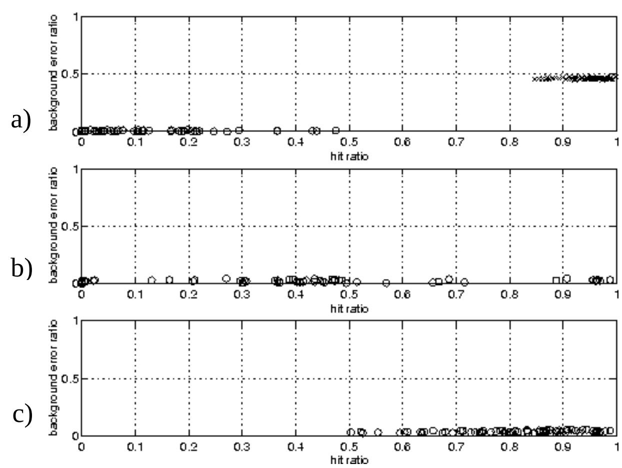

Figure 4: Error ratio vs hit ratio of a) motion and spatial solution respectively, b) collaborative merging, c) region linking

IV. Conclusion

In this paper, we introduce a segmentation framework for a cost-effective visual front-end processor. The algorithm takes into account several important issues such as complex background, orientation, etc. The results in the last section provides the evidence on the effectiveness of the algorithm in

Proceedings of the 1997 International Conference on Image Processing (ICIP '97)
0-8186-8183-7/97 $10.00 © 1997 IEEE

reducing background error and improving the overall segmentation performance. Since the algorithm makes use of standard and simple heuristic techniques, it is well-suited for real time and hardware implementation.

Reference

[1] P.J.L van Beek, M.J.T. Reinders, B. Sankur and J.C.A. van der Lubbe, Semantic segmentation of videophone image sequences, *Proc. SPIE Visual Communications and Image Processing 1992*, vol. 1818, Boston, MA. U.S.A, pp.1182-1193.

[2] R. Brunelli and T. Poggio, Face recognition: Feature versus templates, *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 15, pp. 1042-1052, 1993.

[3] G. Burel and D. Carel, Detection and localization of faces on digital images, *Pattern Recognition Letters*, pp. 963-967, Oct 1994.

[4] Q.W. Chen and M. Yachida, Face detection by fuzzy pattern matching, *Proc. of Int. Conf. on Computer Vision*, pp. 591-596, 1995.

[5] G. Chow and X. Li, Towards a system for automatic facial feature detection, *Pattern Recognition*, vol. 26, no. 12, pp. 1739-1755, 1993.

[6] I. Craw, H. Ellis and J.R. Lishman, Automatic extraction of face-feature, *Pattern Recognition Letters*, pp. 183-187, Feb. 1987.

[7] I. Craw, D. Tock and A. Bennett, Finding face features, *Proc. 2nd ECCV*, pp. 92-96, 1992.

[8] Y. Dai and Y. Nakano, Face-texture model-based on SGLD and its application in face detection in a color scene, *Pattern Recognition*, vol. 29, no. 6, pp. 1007-1017, 1996.

[9] L.C.De Silva, K. Aizawa and M. Hatori, Detection and tracking of facial features by using a facial feature model and deformable circular template, *IEICE Trans. on Information and Systems*, E78-D(9):1195-1207, Sep. 1995.

[10] S. Edelman, D. Reisfeld and Y. Yeshurun, Learning to recognize faces from examples, *Proc. of the 2nd ECCV*, pp. 787-791, Santa Margherita Ligure, Italy, May 1992.

[11] V. Govindaraju, S. Srihari and D. Sher, A computational model for face location, *Proc. of 3rd Int. Conf. of Computer Vision*, pp. 718-321, 1990.

[12] K.M. Lam and H. Yan, Facial feature location and extraction for computerized human face recognition, *Int. Symposium on information Theory and its Applications*, Sydney, Australia, Nov. 1994.

[13] A. Lanitis, A. Hill, T. Cootes and C. Taylor, Locating facial features using genetics algorithms, *Proc. of Int. Conf. on Digital Signal Processing*, Limassol, Cyrus, pp. 520-525, 1995.

[14] C.C. Lin and W.C. Lin, Extracting facial features by an inhibitory mechanism based on gradient distributions, *Pattern Recognition*, vol. 29, no. 12, pp. 2079-2101, Dec. 1996.

[15] B.K. Low, Facial features extraction using simple cue, *On-line technical report 1997 no.2 (tr97-2)*, DSP applications group, De Montfort University, Leicester, UK. (ftp site: /shore.eng.dmu.ac.uk/DSP/tr97-2.ps.gz)

[16] J. Luettin, N. Thacker and S. Beet, Speaker identification by lipreading, *Proc. of Int. Conf. on Spoken Language Processing*, Philadelphia, USA, 1996.

[17] S. Mckenna, S. Gong, et.al, Tracking facial feature points with Gabor wavelets and shape models, *Proc. 1st Int. Conf. on Audio- and Video-based Biometrics Person Authentication*, Lectures notes in Computer Science, Springer Verlag, 1997.

[18] B. Moghaddam and A. Pentland, Face recognition using view-based and modular eigenspaces, *SPIE Automatic Systems of the identification and Inspection of Humans*, vol. 2227, July 1994.

[19] O. Nakamura, S. Mathur and T. Minami, Identification of human faces based on isodensity maps, *Pattern Recognition*, vol. 24, no. 3, pp. 263-272, 1991.

[20] N. Oliver, F. Bérard and J. Crowley, LAFTER: Lips and face real time tracker, *M.I.T. Technical Report* No.396, Submitted to CVPR '97.

[21] D. Reisfeld and Y. Yeshurun, Robust detection of facial features by generalized symmetry, *Proc. of 11th IAPR Int. Conf. on Pattern Recognition*, pp. A117-120, The Hague, The Netherlands, August 1992.

[22] H. Rowley, S. Balaju and T. Kanade, Human face detection in visual scenes, *CMU-CS-95-158R*, Carnegie Mellon University, Nov. 1995.

[23] M. Turk and A. Pentland, Eigenfaces for recognition, *Journal of Neuroscience*, vol. 3, no. 1, 1991.

[24] J. Vincent, J. Waite and D Myers, Automatic location of visual features by a system of multilayered perceptrons, *IEE Proceeding-F*, vol. 139, no.6, Dec. 1992.

[25] C. Wong, D Kortenkamp and M. Speich, A mobile robot that recognizes people, *IEEE Int. Conf. on Tools with Artificial Intelligence*, 1995.

[26] C.Y. Wu and J.S. Huang, Human face profile recognition by computer, *Pattern Recognition*, vol. 23, no. 3/4, pp. 255-259, 1990.

[27] A. Yuille, D. Cohen and P. Hallinan, Facial feature extraction by deformable templates, *Technical reports CICS-P-124*, Centre for Intell. Control Systems, Cambridge, Ma. 02139, March 1988.

0-8186-8183-7/97 $10.00 © 1997 IEEE