

Constrained Region-Growing and Edge Enhancement Towards Automated Semantic Video Object Segmentation

L. Gao¹, J. Jiang², and S.Y. Yang¹

¹Institute of Acoustics, Chinese Academy of Sciences, China

²School of Informatics, University of Bradford, UK

Future_Gao@hotmail.com, J.Jiang1@brad.ac.uk,
S.Y.Yang@hotmail.com

Abstract. Most existing object segmentation algorithms suffer from a so-called under-segmentation problem, where parts of the segmented object are missing and holes often occur inside the object region. This problem becomes even more serious when the object pixels have similar intensity values as that of backgrounds. To resolve the problem, we propose a constrained region-growing and contrast enhancement to recover those missing parts and fill in the holes inside the segmented objects. Our proposed scheme consists of three elements: (i) a simple linear transform for contrast enhancement to enable stronger edge detection; (ii) an 8-connected linking regional filter for noise removal; and (iii) a constrained region-growing for elimination of those internal holes. Our experiments show that the proposed scheme is effective towards revolving the under-segmentation problem, in which a representative existing algorithm with edge-map based segmentation technique is used as our benchmark.

1 Introduction

Video object segmentation consistent with human visual perception has long been identified as a difficult problem since it requires characterization of semantics to define the objects of interest. Unique definition of such semantics is not possible as the semantics are often context dependent and thus low-level segmentation has been focusing on segmentation of regions rather than objects. As video processing and coding is moving towards content-based approaches, object segmentation becomes an important research topic. The content-based video compression standard, MPEG-4, stands as a most representative scenario for such content-based approaches. As a result, research on this subject has been very active and many algorithms have been reported in the literature. Existing video object segmentation can be roughly classified into two categories according to their primary segmentation criteria. One category is represented by those regional segmentation techniques [1-5], where spatial homogeneity is primarily used as the criteria to develop rules towards the segmentation design. As these techniques are rooted among low-level image processing and essentially data-driven, precise boundaries of the segmented regions can often be obtained. However the computation incurred is normally high since iterative operations are often required. Examples of such techniques include watershed, snake modeling, and region-growing [11] etc. The other category of segmentation can be characterized by detection of changes [6-15], where motion information is utilized to segment those

moving objects together with other spatio-temporal information. In this category, the objects segmented are close to semantic video objects and thus provide promising platforms for further research and development.

Specifically, existing research on video object segmentation is built upon change detection assisted by other sideline information including spatial segmentation, edge detection and background registration etc. [6-15]. In [6], Kim et. al. described a spatio-temporal approach for automatic segmentation of video objects, where hypothesis test based on estimated variances within a window is proposed to exploit the temporal information, and spatial segmentation is included to assist with detection of object boundaries. The final decision on foreground and background objects is made in combining the spatially segmented object mask with the temporally segmented object mask, in which a two-stage process is designed to consider both the change detection and the historical attributes. In [7], another similar approach was described towards a robust or noise-insensitive video object segmentation, which follows the idea of combining spatial edge information with motion-based edge detection. Based on these algorithms, we carried out a series of empirical studies and testing. Our experiments reveal that, while the algorithms perform well generally, there exist an under-segmentation problem, where parts of the object region are missing or there exist holes inside the object region. This is a serious problem especially when the background has similar intensity values to those inside objects or at the boundary of the object region. To rectify this problem, we propose a new scheme of automatic semantic object segmentation, where elements of edge enhancement and constrained region growing are proposed, combining the strength of change detection with the strength of spatial segmentation (region-grow). We also illustrate via extensive experiments how our proposed algorithm could achieve this objective in comparison with the existing VO segmentation algorithm reported in [7]. The rest of the paper is organized into two sections. One section is dedicated to our proposed algorithm design, and the other is dedicated to experiments and presentation of their results. Finally some concluding remarks will also be made in the same section.

2 The Proposed Algorithm Design

2.1 Edge Enhancement and Linear Filtering

In practice, when the grey level difference between the object and the background is small, part of the object at its boundary will have similar intensity values to that of background. In this circumstance, edge detection could fail to detect all the edges of the object, and thus some parts inside the objects become missing. To reduce such a negative effect upon object segmentation, we propose a simple linear transformation as part of the pre-processing to enhance the contrast of the luminance component of the video frame before edge detection is performed. Although there exist many contrast enhancement algorithms that may provide better performances, our primary aim here is not only improving the segmentation accuracy, but also maintaining the necessary simplicity for real-time applications. Considering the fact that increase of

contrast will inevitably introduce additional noise, we also designed a simple 2D filter to remove the noise. By combining these two elements together, we achieve the objective that edges are enhanced to enable edge detection to extract the boundaries of moving objects and hence ready for semantic object segmentation.

Given an input video frame, $I_n(x, y)$, assuming that their intensity values are limited to the range of $[a, b]$, its transformed video frame can be generated as:

$$g_n(x, y) = a' + \frac{b' - a'}{b - a} \times (I_n(x, y) - a) \quad (1)$$

Where $a' = 0, b' = 255$ and $a = 70, b = 180$.

Following the contrast enhancement, we then apply Canny edge detector [7] to extract edges from the video frames to characterize the semantic objects to be segmented. This is essentially a gradient operation performed on the Gaussian convoluted image. Given the n th video frame I_n , the Canny edge detecting operation can be represented as:

$$\Phi(I_n) = \theta(\nabla G * I_n) \quad (2)$$

where $G * I_n$ stands for the Gaussian convoluted image, ∇ for the gradient operation, and θ for the application of non-maximum suppression and the thresholding operation with hysteresis to detect and link the edges.

Following the spirit of the work reported in [7], we extract three edge maps: (i) the difference edge map $DE_n = \Phi(I_{n-1} - I_n)$, (ii) the current edge map $E_n = \Phi(I_n)$, and (iii) background edge map E_b , which contains background edges to be defined by manual process or by counting the number of edge occurrences for each pixel through the first several frames [7]. Our implementation adopted the latter option.

From these three edge maps, a currently moving edge map, ME_n^{change} , representing the detected changes can be produced as follows:

$$ME_n^{change} = \left\{ e \in E_n \left| \min_{x \in DE_n} \|e - x\| \leq T_{change} \right. \right\} \quad (3)$$

where e stands for edge pixels inside the moving edge map, and T_{change} for a threshold empirically determined as 1 in [7]. Essentially, (3) describes an operation in selecting all edge pixels within a small distance of DE_n .

Further, a temporarily still moving edge map ME_n^{still} can also be produced by considering the previous frame's moving edges. This edge map is used to characterize those regions that belong to the moving object but temporally no change is incurred between two adjacent frames. Such an operation can be described as given below:

$$ME_n^{still} = \left\{ e \in E_n \left| e \notin E_b, \min_{x \in ME_{n-1}} \|e - x\| \leq T_{still} \right. \right\} \quad (4)$$

where T_{still} is another threshold, which is also empirically determined as 1 in [7]. As indicated in (3), temporarily still moving edge map contains edge pixels that they are part of current edge map but not part of background edge map, and they also satisfy the condition: $\min_{x \in ME_{n-1}} \|e - x\| \leq T_{still}$.

After the identification of those moving edges by (3) and (4), the remaining operation for extracting video objects is to combine the two edge maps into a final moving edge map: $ME_n = ME_n^{change} \cup ME_n^{still}$, and then select the object pixels via a logic AND operation of those pixels between the first and the last edge pixel in both rows and columns [7].

As the contrast enhancement introduced by (1) often produce noise, the edge maps produced could be affected. As a result, to remove the additional noise introduced by the linear transformation, we adopt the method of 8-connected linking region sign to design a filter and apply the filter to both the motion edge map ME_n and the extracted object sequences.

Given an edge pixel at (x, y) whose value is 1 in the binary map, we examine its 3×3 neighborhood to produce a set of all connected points $N = \{A_1, A_2 \dots A_k\}$. If $k \leq T$, a threshold for noise removal, the set of points N will be regarded as noise and thus deleted.

2.2 Constrained Region Growing

Most existing segmentation algorithms often include post-processing to improve the final segmented objects [7-15]. The change detection based techniques such as the one reported in [7] tends to design post processing based on morphological operations, which proved to be an effective towards removal of small holes inside the object regions. When part of the moving object is relatively still across a few frames, however, the edge maps in both (3) and (4) will fail to include those pixels, and thus create large holes inside the segmented object, for which the morphological operations will not be able to recover those missing parts inside or at the boundary of the segmented object. To this end, we propose a constrained region-growing technique to recover those missing parts. Unlike the normal region growing used by those spatial segmentation techniques for still images, our proposed region-grow is under certain constraints to reflect the fact that object segmentation has been done by change detections across adjacent frames. Therefore, the constraints include: (i) the seed selection is fixed at those edge points at the boundary of the final edge map; (ii) the number of pixels outside the first and the last edge points must be smaller than a certain limit. In other words, if the majority of the pixels on any row are outside the boundary of the edge map, the constrained region grows will not be applied.

Given the final edge map ME_n , we examine those pixels outside the first and the last pixel in each row to see if any further growing can be facilitated by using the pixel at the boundary of the edge map as seeds. Specifically, given the set of pixels outside the first and the last edge points in the i th row: $PO_i = \{PO_1, PO_2, \dots PO_k\}$, we decide whether the region of those edge points should be grown into any of the points inside PO_i or not by the following testing:

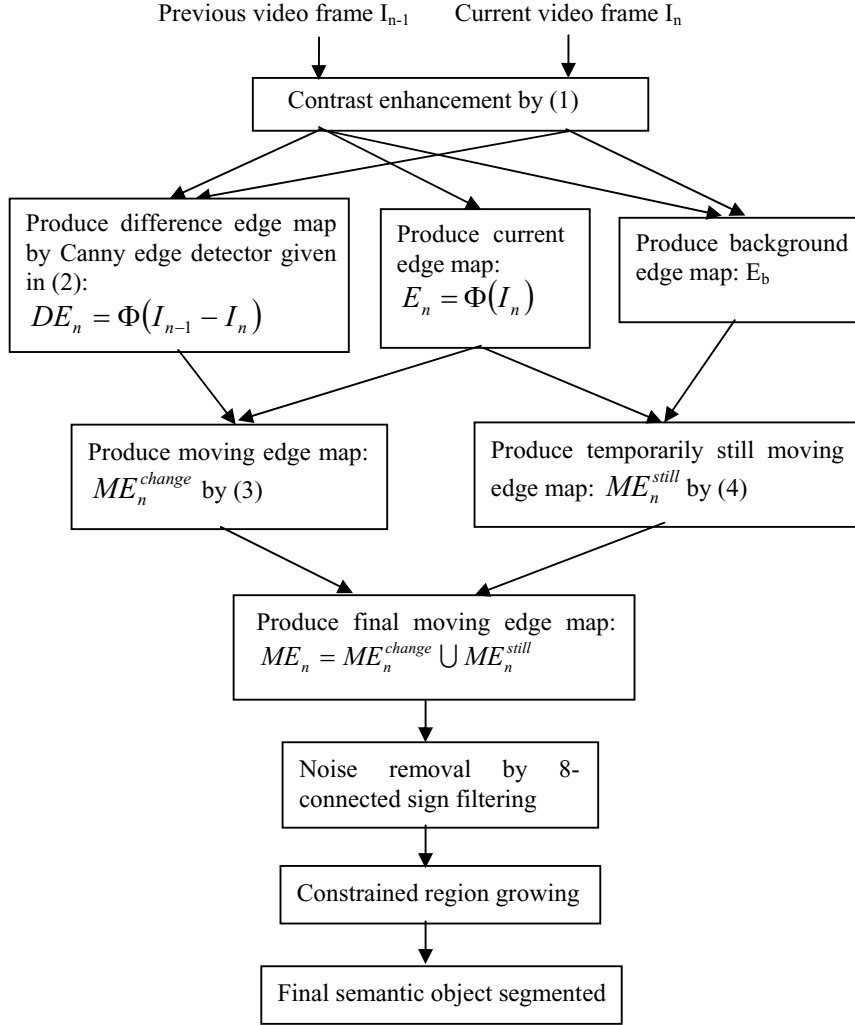


Fig. 1. Summary of the proposed algorithm

$$PO_i = \begin{cases} e_i & \text{if } \|PO_i - e\| \leq T_e \\ PO_i & \text{else} \end{cases} \quad (5)$$

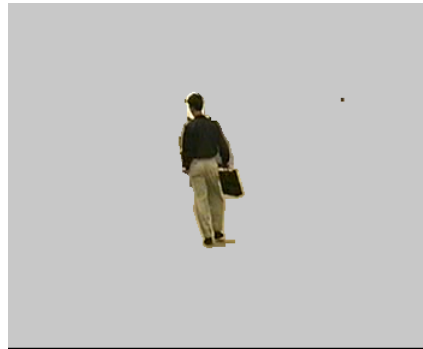
where T_e is a threshold indicating that the pixel tested is very similar to e , which is the first or the last edge pixel depending on which of these two edge points is closest to the position of PO_i . If the condition is satisfied, the PO_i will be grown into the edge points. Otherwise, they will stay as they are outside the edge map.

The above post processing will also apply to those points along the columns. After the post processing, the VO is extracted by logic AND operation of both row and column candidates as described in [7].

In summary, the proposed segmentation algorithm is highlighted by Figure 1.



(a): segmentation by benchmark for Hall-monitor frame 71



(b): segmentation by the proposed for Hall-monitor frame 71



(c): segmentation by benchmark for Clair



(d): segmentation by the proposed for Clair



(e): segmentation by benchmark for Mother-daughter (frame304)

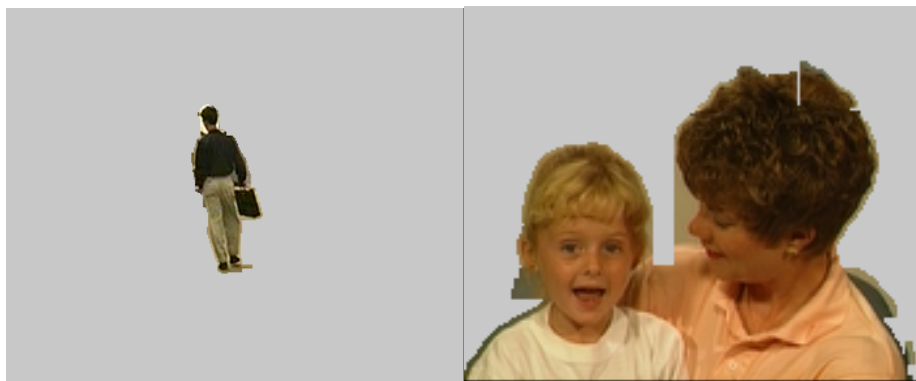


(f): segmentation by the proposed for Mother-daughter (frame 304)

Fig. 2. Comparison of segmentation results between the benchmark and the proposed with linear transform given in (1)

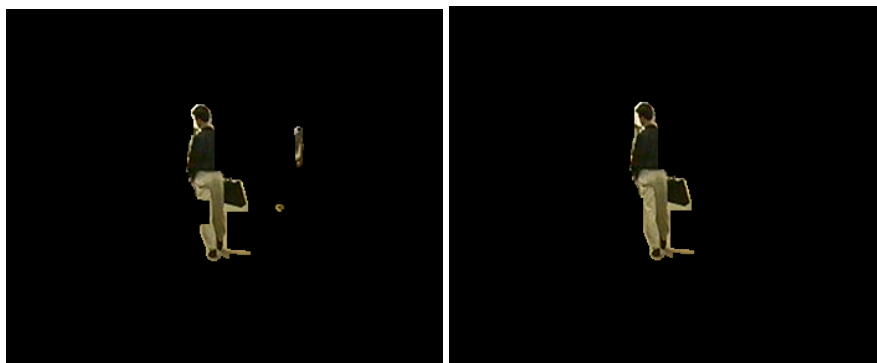
3 Experimental Results and Conclusions

To evaluate the proposed algorithm, we used a set of three video clips: *Hall-monitor*, *Claire*, and *Mother-daughter*, all of which are publicly available and *Hall-monitor* is the same as that used in [7]. In order to enable detailed analysis of how each element of the proposed algorithm actually contributes to the effect of final video object segmentation, we implemented the VO segmentation algorithm as in [7] as our benchmark, and carried out experiments each time one element of the proposed algorithm is added. These elements include: (i) linear transform for contrast enhanced edge detection; (ii) filter design for noise removal; and (iii) constrained region-growing.



(a): Segmentation result by the proposed for *Hall-monitor* (frame71) (b): Segmentation result by the proposed for *Mother-daughter* (frame304)

Fig. 3. Illustrations of segmentation by the proposed with both linear transform and noise removal filtering



(a): Segmentation by benchmark(frame 73) (b): segmentation by the proposed (frame 73)

Fig. 4. Comparison of segmented results by benchmark and the proposed, where only the constraint region growing is considered

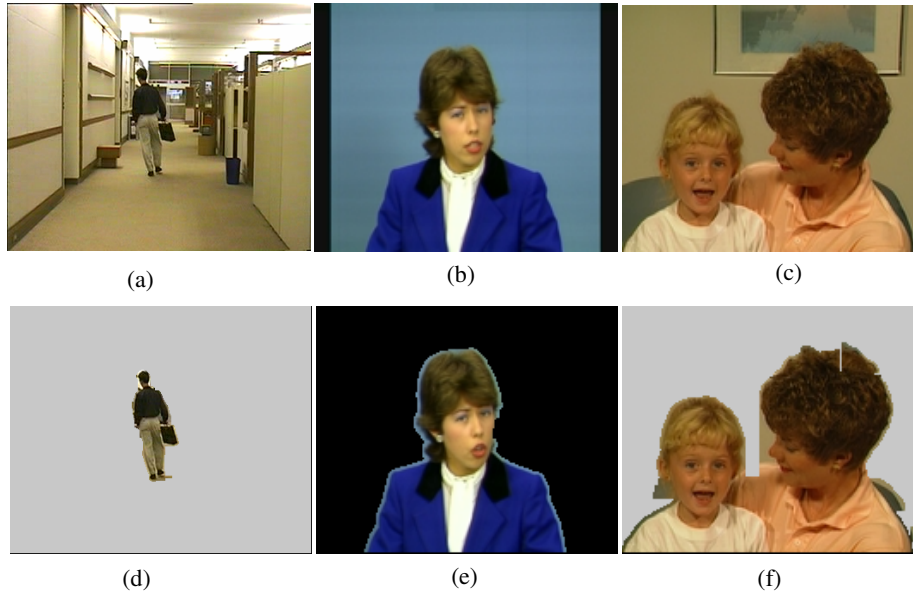


Fig. 5. Final segmentation results by the proposed algorithm: (a)-(c) are originals, and (d)-(e) are the segmented video objects

Figure 2 illustrates the comparison of the segmented results, where part-(a), (c), (e) and (f) are the segmented objects by the benchmark and part-(b), (d), (f) and (g) are the segmented objects by the proposed algorithm, where only the linear transform is included. As seen, the proposed linear transform introduced additional noise while the segmentation accuracy is improved.

Figure-3 illustrates the segmented results by the proposed algorithm, where both linear transform and the noise removal filter are considered, from which it can be seen that the noise introduced is effectively removed.

Figure-4 illustrates the comparison of segmented results between the benchmark and the proposed, where part-(a) and (c) represents the results of the benchmark, and part-(b) and (d) the results of the proposed with only the element of constrained region-growing. Although the proposed constrained region-growing can not recover all the missing parts, it can still be seen that the proposed algorithm does recover the missing part inside the left leg, which has achieved significant improvement compared with the benchmark.

Finally, by gathering all the elements, the full segmented video objects by the proposed algorithm can be illustrated in Figure 5. Note that all the figures illustrated here are much larger than those given in references [6-15]. If we make the pictures smaller, the segmentation results will look better as those boundaries will look smoother.

In this paper, we proposed an automatic semantic object segmentation scheme to provide a possible solution for the under-segmentation problem experienced by most existing segmentation techniques [6-15]. From the experimental results shown in Figure 2 to Figure 5, it can be seen that, while the proposed algorithm can effectively recover those missing parts inside the video object, it inevitably introduces some of the back-

ground points into the object region, which can be referred to as over-segmentation, for which further research is being organized around: (i) looking for other cues to provide additional semantic information for segmentation; (ii) combinational approaches in both change detection and background registration[9]; and (iii) inclusion of further spatial segmentation elements such as snake modeling, and water shed [1-5] etc.

Finally, the authors wish to acknowledge the financial support from the Chinese Academy of Sciences and European Framework-6 IST programme under the IP project: Live staging of media events (IST-4-027312).

References

1. K. Harris S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Trans. on image processing*, Vol.7, No.12, pp. 1684-1699, 1998
2. L. Vincent, P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.13, Issue 6, pp. 583-598, 1991
3. R. J. O'Callaghan, and D. R. Bull, "Combined Morphological-Spectral unsupervised image segmentation," *IEEE Trans. on image processing*, Vol.14, No. 1, pp. 49-62, 2005
4. Tsai YP, Lai C.C. Hung YP et al. A Bayesian approach to video object segmentation via merging 3-D watershed volumes *IEEE Trans. On Circuits and Systems for Video Tech.* 15 (1) pp. 175-180 Jan. 2005
5. Jung CR, Scharcanski J 'Robust watershed segmentation using wavelets' *Image and Vision Computing* 23 (7): 661-669 Jul. 1 2005
6. Kim M. et al. 'A VOP generation tool: automatic segmentation of moving objects in image sequences based on spatio-temporal information', *IEEE Trans. Circuits, Systems for Video Technology*, Vol 9, No 8, 1999, pp. 1216-1227
7. Kim C. and Hwang J.N. 'fast and automatic video object segmentation and tracking for content-based applications', *IEEE Trans. Circuits and Systems for Video Technology*, Vol 12, No. 2, 2002, pp. 122-128
8. Salembier P. and Pardas M. 'Hierarchical morphological segmentation for image sequence coding', *IEEE Trans. Image Processing*, Vol 3, No 5, 1994, pp. 639-648
9. Chien S.Y. et. Al. 'Efficient moving object segmentation algorithm using background registration technique', *IEEE Trans. Circuits and Systems for Video Technology*, Vol 12, No 7, 2002, pp. 577-589
10. Shamim A. and Robinson A. 'Object-based video coding by global-to-local motion segmentation', *IEEE Trans. Circuits, Systems for Video Technology*, Vol 12, No 12, 2002, pp. 1106-1115
11. Feng G.C. and Jiang J "Image segmentation in compressed domain" *Journal of Electronic Imaging*, Vol .12, No 3, SPIE, 2003, pp. 390-397
12. Toklu C. et. Al. 'Semi-automatic video object segmentation in the presence of occlusion', *IEEE Trans. Circuits and Systems for Video Technology*, Vol 10, No 4, 2000, pp. 624-635
13. Kervrann C. and Heitz F. 'Statistical deformable model-based segmentation of image motion', *IEEE Trans Image Processing*, Vol 8, No 4, 1999, pp. 583-594
14. Meier T. and Ngan K. 'Automatic segmentation of moving objects for video object plane generation', *IEEE Trans. Circuits, Systems for Video Tech.*, Vol 8, No 5, 1998, p. 525
15. Xu Y. et. Al. 'Object-based image labeling through learning by example and multi-level segmentation', *Pattern Recognition*, Vol 36, pp. 1407-1423, 2003