SEGMENTATION AND MOTION ESTIMATION OF MOVING OBJECTS FOR OBJECT-ORIENTED ANALYSIS-SYNTHESIS CODING

Jae Gark Choi, Si-Woong Lee and Seong-Dae Kim

Dept. of E.E. KAIST 373-1
GuSeong-Dong YuSeong-Gu Daejeon Korea
Email : cjg@sdvision.kaist.ac.kr

ABSTRACT

This paper presents a segmentation and motion estimation method for object-oriented analysis-synthesis coding. A major difficulty in estimating general motion is that it requires a large area of support in order to achieve a good estimation. Unfortunately, when the supporting area is large it is very likely to have multiple moving objects. To solve this problem, we propose a multi-stage segmentation method which is based on optical flow. The basic concept is to group homogeneous subregions with respect to simpler mapping model into large homogeneous regions with respect to more complex mapping model. By applying a hierarchy of mapping parameter model progressively, we can segment the whole changed region into several parabolic patches. Especially person's face in head-and-shoulder images can be described as one object.

1. INTRODUCTION

Recently new algorithms called object-oriented analysis-synthesis coding have been studied to avoid annoying artifacts like blocking[1][2]. These algorithms are based on object region extraction and object- based motion compensation. One of the essential problems of object-oriented analysis-synthesis coding is segmentation and motion estimation of moving objects, i.e. the image analysis.

In [3] a segmentation and motion estimation algorithm is represented that formulates the analysis task as a hierarchical application of object motion estimation and boundary detection. A similar segmentation method is given in [4]. In contrast to [3] it uses a more general model, taking curved surfaces into account. Also intra-frame segmentation is used to make the segmentation more reliable. Hötter[3] and Diehl[4] assume that, although not strictly true, the whole region can be correctly described by one mapping parameter set which is estimated. Then the estimated mapping parameter set primarily describes the motion of the dominant(normally the largest) object in the region because this object mainly determines the absolute minimum of the error function. However if there is not a dominent object in the region, these techniques are bound to fail to extract objects. On the other hand a segmentation method based on the optical flow is presented[5]. Here the flow field is divided into segments which are consistent with moving rigid objects with roughly planar surface. Then these segments are grouped under the hypothesis that they are induced by the same mapping parameters of a rigid object. For this purpose, a generalized Hough transform is used. However it needs tremendous computaional complexity and memory occupation due to Hough transformation.

The more complex the mapping model is, the larger the region which can be described by one parameter set will be. However the more complex the mapping model is, the larger area of support is required in order to achieve a good estimation. Unfortunately, when the supporting area is large it is very likely to have multiple moving objects. This is a major difficulty in estimating object motion and boundary.

To solve the problem, we propose a multi-stage segmentation scheme which is based on three main stages. The approach is to segment the dense motion field into arbitrarily shaped regions corresponding to the moving objects. The basic concept is to group homogeneous subregions with respect to simpler mapping model into larger homogeneous region with respect to more complex mapping model progressively. To do this, a hierarchy of mapping parameter models is defined. In the first stage, the flow field is partitioned into segments having similar motion vectors. Each segmented region has 2-D translational motion, and will be called as a 2-D translational patch. The second stage is to group 2-D translational patches into segments which is consistent with a rigid motion of a roughly planar surface. We call the segmented region a planar patch. In the final stage, planar patches are grouped into a homogeneous region which is generated by the roughly parabolic rigid objects with the general 3-D motion. We call the homogeneous region a parabolic patch. Hence we can segment the image into several parabolic patches.

2. A HIERARCHY OF MAPPING PARAMETER MODELS

To describe the relation of temporal luminance changes to motion, the mapping parameter model is necessary. a hierarchy of mapping parameter model for multi-stage segmentation can be defined as:

This work was supported by Daewoo Electronic co.

2.1. Level 1: 2-D translation model

The model describes 2-D translation motion. It is a simple transformation:

$$u(x,y) = a_1, \qquad v(x,y) = b_1 \tag{1}$$

where (u,v) is the displacement vector and (x,y) is a image-space coordinate of a point.

2.2. Level 2: Planar patch model

The planar patch model describes the case of the parallel projection, a rigid planar surface and 3-D general motion. The model is an affine transformation which contains 6 parameters. Its displacement vector becomes

$$\begin{aligned} u(x,y) &= a_1 + a_2x + a_3y, \\ v(x,y) &= b_1 + b_2x + b_3y. \end{aligned} \tag{2}$$

2.3. Level 3: Parabolic patch model

The parabolic patch model supports parabolic surfaces, parallel projection and 3-D general motion. The model is a quadratic transformation which has 12 parameters. Its displacement vector becomes

$$\begin{aligned} u(x,y) &= a_1 + a_2x + a_3y + a_4x^2 + a_5y^2 + a_6xy, \\ v(x,y) &= b_1 + b_2x + b_3y + b_4x^2 + b_5y^2 + b_6xy. \end{aligned} \tag{3}$$

3. ALGORITHM DESCRIPTION

In this section, we describe our multi-stage segmentation algorithm. Fig.1 illustrates the whole process. First, the changed region and dense motion field are estimated. Then multi-stage segmentation is done for changed region. Each process is described in the followings.

3.1. Change detection

The aim of the change detection is to distinguish between temporally changed and unchanged region of two successive images. This change detection algorithm is similar to the algorithm given in [3].

3.2. Dense motion field estimation

The motion vector estimation procedure consists of two parts: the motion vector detection and the motion vector correction parts. For motion vector detection, we use the brute-force search method of BMA with the 9x9 measurement window. Its maximum displacement is ±10 pixels and the distance of measurement pixels is 4. For motion vector correction, a median filter is applied at the motion vector field. After the correction, we use bilinear interpolation to obtain a motion vector per pixel.

3.3. The first segmentation stage

Neighboring pixels of similar motion vectors are grouped together to form a segmented region. The segmented region has homogeneity with respect to 2-D translational motion. Here the region-growing method by pixel-aggregation[6] is used for the segmentation. In region growing, the dense motion field is scanned in a row-wise manner. Each point is assigned to a segment during the scanning procedure. Let us suppose that we are currently at the point (k,l). Then the point (k,l) has some adjacent segments in its neighborhood. First, we try to merge a point (k,l) with one of its adjacent segments. If this merging fails, the point is assigned to a new segment. For this task the method employs the non-symmetric half plane neighborhood (k-1,l-1), (k-1,l), (k-1,l+1), (k,l-1). The merging rule is based on the similarity measurement M_i, defined as

$$M_i = |d_{x,R} - d_{x,i}| + |d_{y,R} - d_{y,i}| \tag{4}$$

where $d_{x,R}$ and $d_{y,R}$ are the x and y components for the local representative motion vector of a segment, and $d_{x,i}$ and $d_{y,i}$ are the x and y components of the motion vector d_i respectively. If M_i is less than a given threshold, the neighboring segment becomes the merging candidates. We assign a point (k,l) to a segment which has the smallest error of the candidates.

3.4. The second segmentation stage

The second stage is to group 2-D translational patches into segments which is consistent with a rigid motion of a roughly planar surface. The grouping problem can be solved by applying the proposed region-growing technique. The procedure can be described in the following way:

1. Find a seed region for growing. we select as a seed region the largest one R_i among the regions which are not yet assigned to any of the already created segments. By this selecting method, we can obtain the more reliable mapping parameters because the seed region has a largest area of support within homogeneity.
2. Find the merging candidates $\{R_j : j = 1, \ldots, n\}$ of the seed region R_i. Only adjacent regions which are not yet assigned to any of the already created segments are considered as candidates
3. Test the merging of the candidates into the seed region. The merging test is done by computing, using the least-squares technique, mapping parameters and related error values over $R_i \cup R_j$.
   (a) Computing an optimal affine transformation: To compute the mapping parameter, the error function to be minimized is

   $$\begin{aligned} &E_{ij}(a_1, \ldots, b_3) = \\ &\textstyle\sum_k [(u(x_k,y_k) - a_1 - a_2x_k - a_3y_k)^2 \\ &\quad + (v(x_k,y_k) - b_1 - b_2x_k - b_3y_k)^2] \end{aligned} \tag{5}$$

   where k is an element of region R_i or region R_j. Taking partial derivatives with respect to $a_1, \ldots, b_3$ and equating to 0, a set of six linear equations is obtained. If these equations are independent, their solution, denoted by $a_1^*, \ldots, b_3^*$ represents the optimal affine transformation.

   (b) Criteria for a merging decision: Substituting this optimal solution and the motion vectors contained in R_j into (6) and using a normalization equation, a new error function σ_j is obtained for the candidate R_j:

$$\sigma_j = \sqrt{E_j(a_1^*, \ldots, b_3^*)/\sum_j 1}. \qquad (6)$$

σ_j is the standard deviation of the motion vectors in R_j from the optimal solution. If σ_j is less than given threshold, the candidate R_j is merged into the seed region R_i.

4. Repeat the previous steps 2-3 recursively until no more merging exists around the seed region. If no more merging, we find a new seed region and repeat the above steps 1-4.

3.5. The third segmentation stage

In the third stage, planar patches which are consistent with the quadratic transformation are merged into parabolic patches. It is similar to the previous stage except for using a quadratic transformation. Now the whole changed region is segmented into several parabolic patches.

4. SIMULATION RESULTS

Two consecutive frames of "Claire" sequence in CIF, which is shown in Fig.2, are used for the simulations. Fig.3 shows the changed region and motion vector field for images in Fig.2. We simulated the proposed method and the conventional method by Adiv[5]. Fig.4 shows the results of our multi-stage segmentation method. The results of the three stages of the segmentation is shown in Fig.4(a)-(c). Note that the whole changed region is segmented into 3 parabolic patches in the final stage. Especially person's face in head and shoulder images is segmented as one object. Each parabolic patch is motion-compensated with the quadratic transformation (3). The result of object-based motion compensation is shown in Fig.4(d). Motion-failure(MF) parts are detected by thresholding the absolute difference between the original image and the object-based motion compensated image. Fig.4(e) shows the motion-failure parts. The motion failure occurrs in regions such as mouth, eyes and some of uncovered background. However the motion-failure parts are very small compared with the entire image. Fig.5 shows the results of the conventional method by Adiv. Fig.5(a) shows the segmentation result where the changed region is segmented into 7 patches. Here Each patch is consistent with moving rigid object with roughly planar surface. Fig.5(b) shows the object-based motion compensation using transformation (7).

$$\begin{aligned} u(x,y) &= a_1 + a_2x + a_3y + a_4xy, \\ v(x,y) &= b_1 + b_2x + b_3y + b_4xy \end{aligned} \qquad (7)$$

Fig.5(c) shows the motion-failure parts. The proposed method is compared with the conventional method by Adiv in Table 1. Comparative items are numbers of parameter, the ratio of motion-failure parts and PSNR of reconstructed image. As shown in Table 1, our algorithm has smaller parameters to be transmitted and motion-failure parts. Moreover it has better performance in PSNR.

5. CONCLUSION

This paper presents a new segmentation scheme for object oriented coding. The method is based on a multi-stage segmentation that groups homogeneous subregions into homogeneous regions with respect to more complex model progressively. Simulation results for head-and-shoulder images show that the changed region is segmented into several parabolic patches. Since the efficiency of object-oriented coding depends on the size of the objects, a high coding efficiency is expected.

6. REFERENCES

[1] H. G. Musmann, M. Hötter and J. Ostermann, "Object-Oriented Analysis-Synthesis Coding of Moving Images", Signal Processing: Image Communication, Vol. 1, No. 2, October 1989

[2] Y. Yokoyama, Y.Miyaamoto and M. Ohta, "Very Low Bit-rate Video Coding with Object-based Motion Compensation and Orthogonal Transform", SPIE, Vol.2094, pp.12-pp.23, 1993

[3] M. Hötter and R. Thoma, "Image Segmentation Based on Object Oriented Mapping Parameter Estimation", Signal Process., Vol.15, No.3, October 1988

[4] N. Diehl, "Object Oriented Motion Estimation and Segmentation in Image Sequences", Signal Processing: Image Communication, Vol.3, No.1, Feburary 1991

[5] Gilad Adiv, "Determining Three-Dimensional Motion and Structure from Optical flow Generated by Several Moving Objects", IEEE. Trans.PAMI, Vol.7, No.4, July 1985

[6] Ioannis Pitas, Digital Image Processing Algorithms, pp. 264-266, Prentice Hall(1993)

Table 1. The comparison of the proposed method and Adiv's method

item —— method	# of parameter (# of object x model complexity)	ratio of MF region (%)	PSNR (dB)
proposed method	36 (3 x 12)	0.848	36.069
method by Adiv	56 (7 x 8)	1.069	35.639

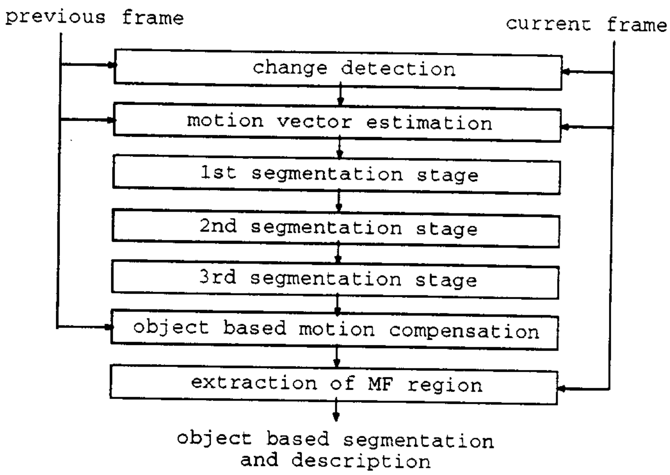

Figure 1. Structure of the multi-stage segmentation algorithm

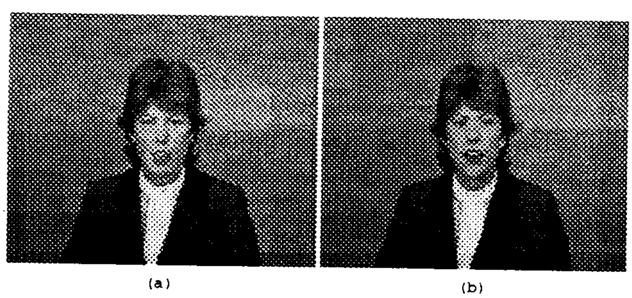

Figure 2. Two successive images (a) the first image (b) the second image

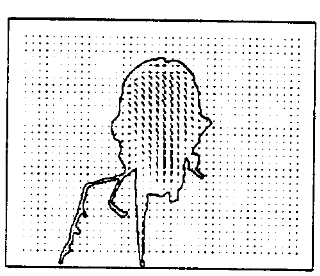

Figure 3. Changed region and motion vector field

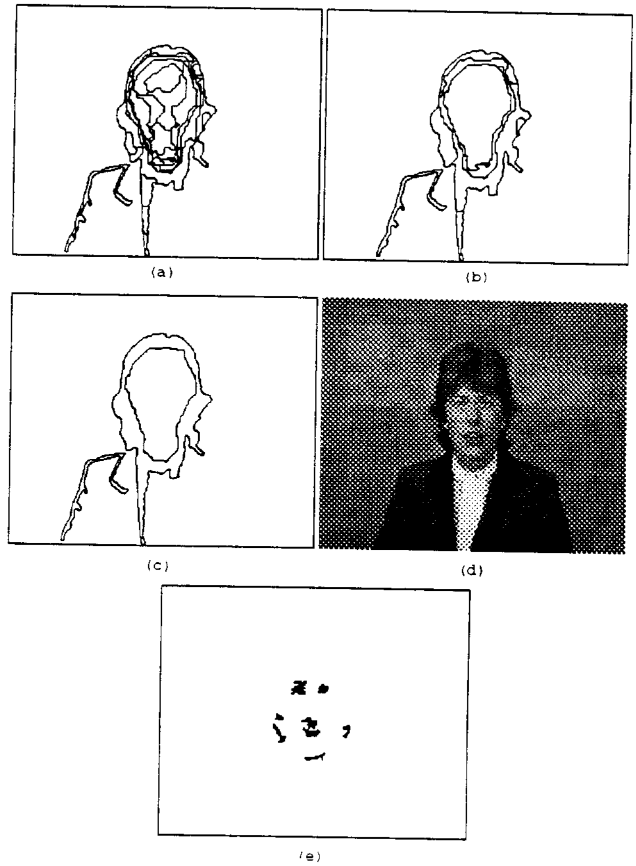

Figure 4. Results by the proposed method (a) the 1st stage segmentation result (b) the 2nd stage segmentation result (c) the 3rd stage segmentation result (d) reconstructed image (e) MF region

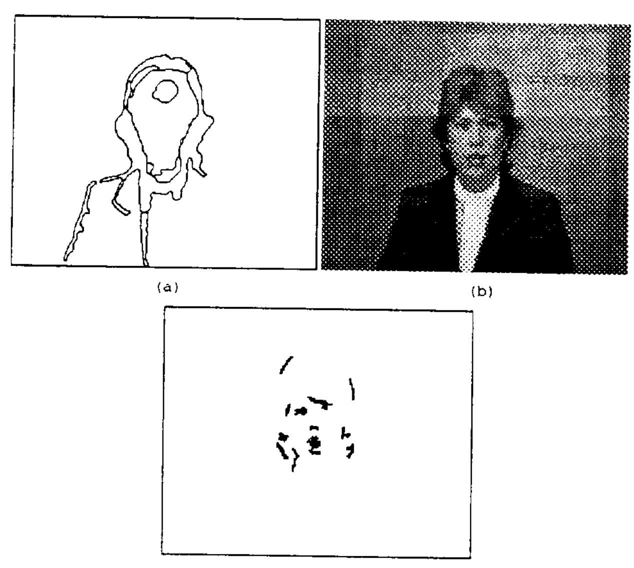

Figure 5. Results by Adiv's method (a) the segmentation result (b) reconstructed image (c) MF region