MATCHING ERROR BASED CRITERION OF REGION MERGING FOR JOINT MOTION ESTIMATION AND SEGMENTATION TECHNIQUES

Markus Schütz and Touradj Ebrahimi

Signal Processing Laboratory
Swiss Federal Institute of Technology at Lausanne (EPFL)
CH-1015, Lausanne, Switzerland
Phone: (+41 21) 693 70 88
Fax: (+41 21) 693 76 00
E-Mail: schutz@ltssg4.epfl.ch

ABSTRACT

This paper describes a region merging method for joint motion estimation and segmentation of digital video sequences. The region merging criterion is based on the measure of the matching error for a region when applying a previously estimated motion to it. A region adjacency graph is used for data representation, which allows a scan independent processing and gives a high-level view. The method is simple-shape object-oriented and starts from a block-based segmentation. The aim of the proposed technique is to define simple shaped objects in a scene using motion information and a simple test.

1. INTRODUCTION

In the framework of video coding or video manipulation, segmentation and motion estimation are two strongly related key points. A precise segmentation is needed to perform accurate motion estimation and an accurate motion estimation is needed to perform precise segmentation. This is due to the properties of the human visual system (HVS) which not only takes into account spatial properties of an image sequence, but is also sensitive to temporal variations [1].

To bypass this chicken and egg problem, various techniques have been proposed to perform joint spatio-temporal segmentation [2, 3, 4, 5]. The idea behind these techniques is to perform spatial segmentation on one hand and temporal segmentation on the other, and then to merge both informations to obtain a precise final segmentation.

The proposal we make here focuses on the temporal segmentation process. We assume that the spatial segmentation part is precise enough due to over-segmentation. After a general description in Sec. 2, we will describe a new region merging criterion in Sec. 3. The technique used to perform region merging is described in Sec. 4. Finally, Sec. 5 gives some simulation results.

2. TECHNIQUE DESCRIPTION

The technique proposed in [2, 3] is based on two main independent processing. First, a spatial segmentation of the image is performed using a watershed algorithm. This leads to a so called "over-segmentation", which is very precise, but cannot be used to perform easily object tracking for example.

The second processing constructs a segmentation based only on motion information such as that described in [4], using an affine motion model. The first step of this process is to build a block based segmentation with block sizes of 4×4 or 8×8 pixels. Then, according to a merging criterion, these blocks are merged together to form regions. This merging process is performed iteratively until the merging criterion cannot be further satisfied or a given number of regions is reached.

These two processing lead to two different segmentations, one based on spatial informations and one on temporal informations. None of the two is really satisfactory. But, if they are linked together a good result can be obtained as shown in [2, 3]. The principle is the following: The regions $\mathcal{S}_i$ obtained with the spatial segmentation processing are linked to the regions $\mathcal{T}_j$ obtained with the temporal segmentation processing, with a weight w_{ij} for each link. This link typically is

the mean square error (MSE), where motion coefficients V_j are applied to region $\mathcal{S}_i$. Figure 1 illustrates the complete processing for a simple image.

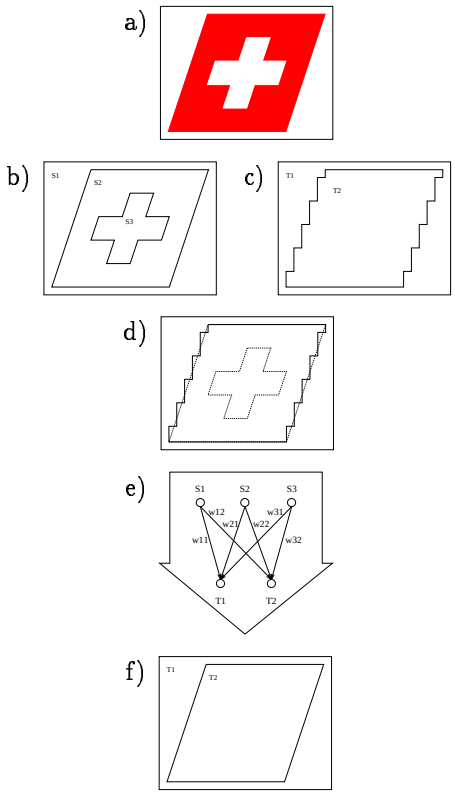


Figure 1: An example of joint spatial-temporal segmentation: a) original moving image, b) boundaries of spatial segmentation, c) temporal segmentation based on motion information, d) spatial boundaries superposed on temporal boundaries, e) assignment of each spatial region to one of the temporal regions, f) result of the joint spatial-temporal segmentation.

3. REGION MERGE CRITERION

If the algorithm description can be done simply, there are still some problems to solve, especially in the temporal segmentation. In Sec. 2, a region merging criterion was discussed, but not specified. Several possibilities exist. In [4] the criterion was based on similarity of motion coefficients. Actually, to merge two regions, they had to have the same motion coefficients.

The drawback of this merging criterion is the estimation of the motion vectors and the previously described chicken and egg problem. One cannot assume that the computed motion coefficients are accurate enough due to the approximative block-based segmentation, specially on boundary regions. Thus, a better criterion has to be defined.

We propose here to use the mean absolute error (MAE) information as a criterion measure. We can assume that if we take two regions A and B, at least one of the two estimated motions will be right. The idea is to exchange the estimated motions between the two regions and to check if the MAE stays in an acceptable range. Thus, for two regions A and B with their estimated motions $\vec{M}_A$ and $\vec{M}_B$, we compute the MAEs for region $A \cup B$ with motion $\vec{M}_A$ and region $B \cup A$ using motion $\vec{M}_B$.

The MAE ($\mathcal{E}$) for a given region A of size N_A pixels is defined as:

$$\mathcal{E}(A,\vec{M}_A) = \frac{1}{N_A} \sum_{i=0}^{N_A-1} \left| I(\vec{r}_i,t) - I(\vec{r}_i - \vec{M}_A, t-1) \right|, \quad (1)$$

where $I(\vec{r},t)$ is the gray level value of pixel at position $\vec{r}$ at time t, and $\vec{M}$ the estimated motion.

To avoid computing MAEs for the complete $A \cup B$ and $B \cup A$ regions and thus top speed up the algorithm, we can use the fact that for region $A \cup B$ (resp. region $B \cup A$), $\mathcal{E}(A,\vec{M}_A)$ (resp. $\mathcal{E}(B,\vec{M}_B)$) is already known. From Eq. 1 it is easy to define the MAE for a merged region:

$$\mathcal{E}((A,\vec{M}_X) \cup (B,\vec{M}_Y)) = \frac{1}{N_A+N_B} \left(N_A \mathcal{E}(A,\vec{M}_X) + N_B \mathcal{E}(B,\vec{M}_Y) \right). \quad (2)$$

Thus, if we want $\mathcal{E}((A,\vec{M}_A) \cup (B,\vec{M}_A))$, only $\mathcal{E}(B,\vec{M}_A)$ has to be computed. The result can then be obtained by applying Eq. 2.

4. REGION MERGE ALGORITHM

To avoid scanning dependency during region merging in the motion based segmentation process, a high-level view of the segmentation to merge is needed. This high-level view must indicate which regions have to be merged at a given step. To achieve this goal, the starting block based segmentation is stored in a region adjacency graph (RAG) [6].

Fig. 2-a gives an example of a simple RAG. The nodes of the graph are the regions of the segmentation and the links between the nodes represent a neighboring relation between the regions. Such a RAG is not oriented.

To use the merge criterion described in Sec.3 we need an oriented graph. Indeed, it can easily be shown using Eqs. 1 and 2 that:

$$\mathcal{E}((A,\vec{M}_A)\cup(B,\vec{M}_A)) \neq \mathcal{E}((B,\vec{M}_B)\cup(A,\vec{M}_B)).$$

We modify the RAG of Fig. 2-a to obtain a weighted and oriented RAG (WORAG) as shown in Fig. 2-b. Thus we have a weighted neighbor relation between regions.

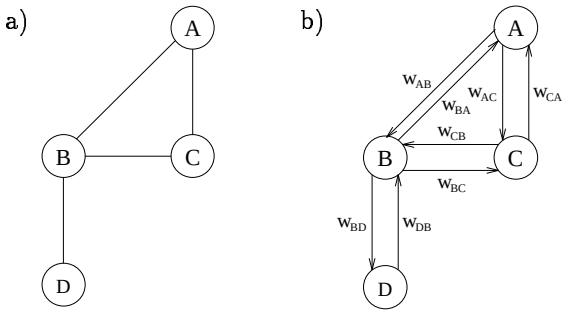


Figure 2: a) Region adjacency graph, b) Weighted oriented region adjacency graph.

The link weights of the WORAG should reflect the modification of MAEs for regions as if they were merged, using the estimated motion for the region of the starting node. Thus, the weights are defined as the difference of MAEs of the union of two neighboring regions using different motions for the arrival node. The weight relations are given in Eqs. 3 and 4.

$$\mathcal{W}_{AB} = \mathcal{E}((A,\vec{M}_A)\cup(B,\vec{M}_B)) - \mathcal{E}((A,\vec{M}_A)\cup(B,\vec{M}_A)) \quad (3)$$

$$\mathcal{W}_{BA} = \mathcal{E}((B,\vec{M}_B)\cup(A,\vec{M}_A)) - \mathcal{E}((B,\vec{M}_B)\cup(A,\vec{M}_B)) \quad (4)$$

Once the graph is built and all the link weights are computed, the algorithm chooses the link with the biggest weight (e.g: the strongest link) and merges the two linked regions to one. The motion of the starting node of the link is then applied for the newly created region.

To avoid re-computing the complete sum of differences on the new region and to speed up the algorithm, the MAE for this region can be found using already known information. We are looking for $\mathcal{E}((A,\vec{M}_A)\cup(B,\vec{M}_A))$ if region B was merged to region A, and $\mathcal{E}((B,\vec{M}_B)\cup(A,\vec{M}_B))$ if region A was merged to region B. Using Eqs. 3 and 4, it is easy to define:

$$\mathcal{E}((A,\vec{M}_A)\cup(B,\vec{M}_A)) = \mathcal{E}((A,\vec{M}_A)\cup(B,\vec{M}_B)) - \mathcal{W}_{AB}, \quad (5)$$

and

$$\mathcal{E}((B,\vec{M}_B)\cup(A,\vec{M}_B)) = \mathcal{E}((B,\vec{M}_B)\cup(A,\vec{M}_A)) - \mathcal{W}_{BA}, \quad (6)$$

where $\mathcal{E}((A,\vec{M}_A)\cup(B,\vec{M}_B))$ is computed as in Eq. 2, with $\mathcal{E}(A,\vec{M}_A)$ and $\mathcal{E}(B,\vec{M}_B)$ being known.

Finally, the old links from or to regions A and B are updated for the new $A\cup B$ region, using the new MAE value and Eqs 3 and 4.

The merging process is performed iteratively. The stoping condition of the iteration is either given by the desired final number of regions or by the minimal value a weight must have. For the latest, typical values lay around -0.1. In theory, to be coherent with the minimization of MAE used for motion estimation, it should be zero, but to avoid problems related to computation precision, it had to be chosen slightly lower than zero.

5. RESULTS

Simulations have been performed using a translational motion model with half-pixel accuracy motion vectors. A bilinear interpolation was used to get the gray level values of non-integer coordinates, using the four nearest known gray level values [7]. Blocks of size 8×8 were used as starting segmentation and support for the motion estimation. The minimal weight value was set to -0.1.

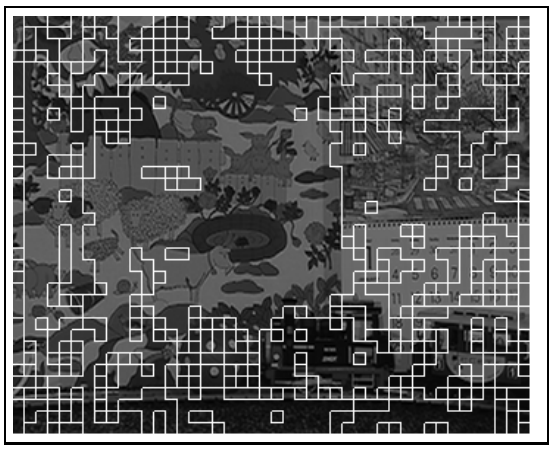
Figure 3 shows segmentation results on the CIF (288×352) sequence "Mobile and Calendar", using frame 146 as frame t and frame 145 as frame $t-1$. Fig. 3-a shows the segmentation obtained using only equality of motion vectors as region merging criterion and Fig. 3-b shows the segmentation obtained after applying the merging criterion based on the matching error measure.

This example shows that the proposed technique is able to detect objects moving in the scene using rough simple-shaped contours. The objects "train" and "calendar", which are moving in this part of the scene are seen as single objects with coherent motion.

6. CONCLUSION

The described technique tries to improve the joint spatial-temporal segmentation of image sequences

a)



b)

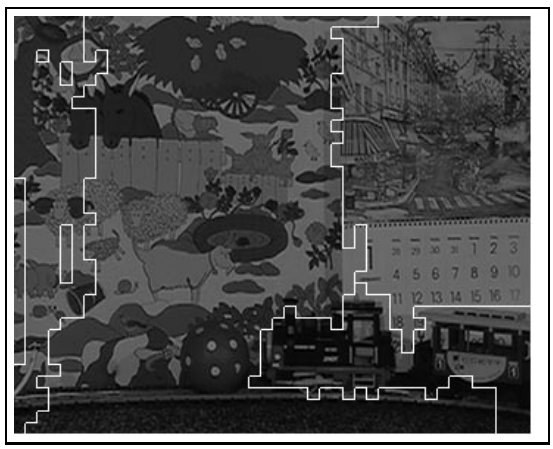


Figure 3: CIF (288×352 test sequence "Mobile and Calendar": a) Segmentation using motion equality, b) Segmentation using the matching error criterion and minimum link weight of -0.1.

by enhancing the motion based segmentation processing. This is done on one hand by using a region-merging criterion based on the measure of the mean absolute error (MAE) or matching error instead of the similarity between motion coefficients, which allows insensitivity to badly estimated motion in blocks where motion estimation fails. On the other hand, a region adjacency graph (RAG) technique is used to give a high-level representation of regions and their possible links, which allows a scan independent processing. Simulation results show that this technique allows to segment image sequences on a simple shaped object oriented basis.

7. REFERENCES

[1] M. Kunt, A. Ikonomopoulos, and M. Kocher. "Second generation image coding techniques". In *Proceedings of the IEEE*, Vol. 73, pp. 549–575, April 1985.

[2] C. Gu, T. Ebrahimi, and M. Kunt. "Morphological moving object segmentation and tracking for content-based video coding". MPEG N0319, ISO/IEC JTC1/SC29/WG11, November 1995.

[3] C. Gu, T. Ebrahimi, and M. Kunt. "Morphological spatio-temporal segmentation for content-based video coding". In *International Workshop on Coding Techniques for Very Low Bit-rate Video*, Tokyo, Japan, November 1995. VLBV'95.

[4] T. Ebrahimi, Homer Chen, and B. G. Haskell. "A region based motion compensated video codec for very low bitrate applications". In *IEEE International Symposium on Circuits and Systems*, Vol. 1, Seattle, Washington, April 30 – May 3 1995. ISCAS'95.

[5] F. Dufaux, F. Moscheni, and A.Lippman. "Spatio-temporal segmentation based on motion and static segmentation". In *IEEE International Conference on Image Processing*, Vol. I of III, pp. 306–309, Washington, D.C, October 1995. ICIP'95.

[6] A. Rosenfeld and A. C. Kak. *Digital Picture Processing*, Vol. 2. Academic Press, Orlando, Florida, second edition, 1982.

[7] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*, chapter 3: Interpolation and Extrapolation, pp. 105–128. Cambridge University Press, second edition, 1992.