GENERATION OF LONG-TERM COLOR AND MOTION COHERENT PARTITIONS

Camilo C. Dórea, Montse Pardàs, Ferran Marqués

Technical University of Catalonia, Barcelona, Spain
{camilo, montse, ferran}@gps.tsc.upc.es

ABSTRACT

This paper describes a technique for generating partition sequences of regions presenting long-term homogeneity in color and motion coherency in terms of affine models. The technique is based on region merging schemes compatible with hierarchical representation frameworks and can be divided into two stages: *Partition Tracking* and *Partition Sequence Analysis.* Partition Tracking is a recursive algorithm whereby regions are constructed according to short-term spatio-temporal features, namely color and motion. Partition Sequence Analysis proposes the Trajectory Adjacency Graph (TAG) to exploit the long-term connectivity relations of tracked regions. A novel trajectory merging strategy using color homogeneity criteria over multiple frames is introduced. Algorithm performance is assessed and comparisons to other proposals are drawn by means of established evaluation metrics [1].

Index Terms— Sequence analysis, tracking, region merging

1. INTRODUCTION

The ability to generate video segmentations based on scene content is of fundamental importance to a diverse set of services and applications. Several video segmentation proposals have focused on spatio-temporal strategies relying on region clustering schemes. Starting from an over-segmentation of the current frame, [2] proposes the use of graph-based clustering rules using luminance and motion features to form objects. In [3], a General Merging Algorithm operating over a Region Adjacency Graph (RAG) iteratively merges neighboring regions according to similarity measures based on color and motion distances. The algorithm has been applied to segmentation of images as well as sequences. In [4], objects are extracted at a semantic level by merging tracked regions according to motion similarity calculated over a set of previous frames. The use of multiple levels of scale for segmentation and hierarchical representations are further discussed in [5]. Recently, graph-based object tracking using hierarchies of nested partitions has been presented in [6]. Our method is mostly related to [3] and [5] in their use of hierarchical representations. In contrast to [4], our use of extended temporal supports comprehends the analysis of the entire partition sequence.

The objective of this work is to partition image sequences into their constituent objects. We define objects as regions which present long-term color homogeneity while undergoing coherent motion in terms of affine models. In achieving segmentation, a hierarchical approach is proposed whereby merging criteria of progressively greater complexity form partitions of increasing levels of spatio-temporal consistency. The results yield a temporally stable yet detailed support over which other higher-level applications may operate.

As illustrasted in Fig. 1, the approach contains 3 main levels within a partition hierarchy constructed bottom-up via region merging. In level 1, a pure intra-frame approach is adopted using only color homogeneity within regions as the merging criterion [3] to form fine color partitions (typically 2000 regions). In level 2, the estimation of motion between consecutive frames allows short-term spatio-temporal connectivity to be imposed. Color homogeneity and affine motion coherence are combined in a Partition Tracking (PT) scheme responsible for labeling regions across frames. Finally, level 3 is created in a Partition Sequence Analysis (PSA) stage which considers spatio-temporal connectivity features of the entire tracked partition sequence of level 2 to propose region mergings based on long-term color homogeneity criteria.

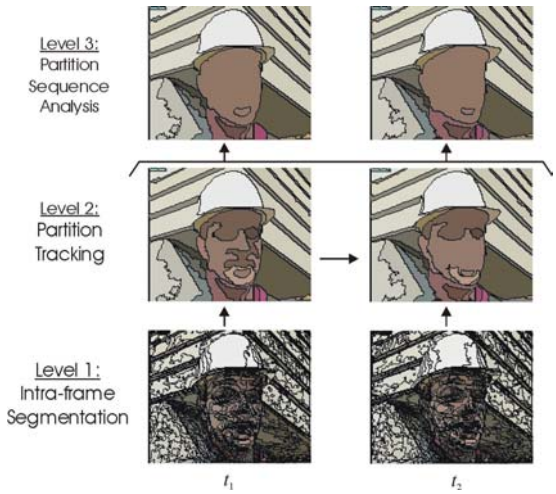

Fig. 1 Partition hierarchy levels of the proposed approach.

This paper is organized as follows. The PT and PSA stages are discussed in Sections 2 and 3, respectively. Experimental results, including novel and established evaluation metrics, are presented in Section 4. Conclusions are drawn in Section 5.

2. PARTITION TRACKING

The Partition Tracking stage is a recursive algorithm responsible for motion projecting a given partition P^{t-1} and determining P^t as outlined in the block diagram of Fig. 2. The *Motion Compensation* and *Marker Fitting & Validation* blocks form the

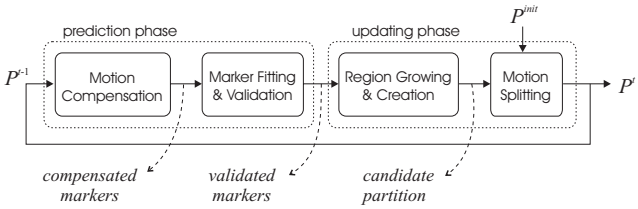

Fig. 2 Overview of the Partition Tracking algorithm.

prediction phase of the algorithm while the *Region Growing & Creation* and *Motion Splitting* blocks form the updating phase.

An *initial partition* P^{init} is built exclusively through intra frame mode, relying only on color information. For such, the General Merging Algorithm of [3] is applied with similarity measure between two regions R_1 and R_2 given by

$$S_c(R_1,R_2) = A_{R_1}\left\|\mathbf{C}_{R_1} - \mathbf{C}_{R_1\cup R_2}\right\|_2 + A_{R_2}\left\|\mathbf{C}_{R_2} - \mathbf{C}_{R_1\cup R_2}\right\|_2 \quad (1)$$

where $\mathbf{C}_R$ are vectors containing the average color components $(\overline{Y},\overline{U},\overline{V})_R$ for a given region R and A_R is the area. The termination criterion is either a final number of regions or the PSNR measure of the resulting partition.

2.1. Partition Prediction

Regions of P^{t-1} are forward motion compensated and establish a set of markers, forming a prediction of P^t. Areas not covered by any marker or covered by multiple markers are labeled as uncertainty. The compensated markers define the inital spatio-temporal connectivity constraints used in grouping the regions of the fine color partition (see Fig. 1, level 1) at t. Affine models, determined via least-squares fitting over dense optical flow, are used in compensation.

Each compensated marker is accommodated to the current image in the *Marker Fitting & Validation* block. They are superimposed on the fine color partition and each marker is reduced to the set of fine regions which it covers in at least 55%. Markers are also tested for color homogeneity in terms of PSNR and further reduced if necessary.

2.2. Partition Update

Region Growing & Creation is responsible for forming a candidate partition through color-based region merging. The fine color regions are superimposed on the uncertatiny areas between validated markers and can either merge to a validated marker, allowing projected regions to grow, or merge amongst themselves. Once a termination PSNR is reached for the partition, any relevant remaining regions (typically larger than 50 pixels) are assigned new labels.

Color features alone may not guarantee the segmentation of a moving object. The *Motion Splitting* block checks the angular error between the given optical flow and the affine model of each region of the candidate partition. Motion incoherent regions are split into their fine components and re-segmented with the merging algorithm and the following similarity measure

$$S_m(R_1,R_2) = \left\|(\overline{d}_x,\overline{d}_y)_{R_1} - (\overline{d}_x,\overline{d}_y)_{R_2}\right\|_2 \quad (2)$$

where $(\overline{d}_x,\overline{d}_y)_R$ are the translational models computed from dense motion fields for region R.

3. PARTITION SEQUENCE ANALYSIS

The Partition Sequence Analysis (PSA) stage is a batch processing procedure which exploits the color and connectivity features of all tracked partitions within an analysis window to propose new segmentations of greater temporal stability through *trajectory merging*.

3.1. Trajectory Merging Algorithm

In order to analyze long-term connectivity relations, a new representation capable of synthesizing the structural properties of partition sequences is needed. The Trajectory Adjacency Graph (TAG) is proposed. Analogously to RAGs, TAGs are well suited for the implementation of merging algorithms.

First, let R_i^t be a region of label i contained in partition P^t and $t \in [t_{min}, t_{max}]$. A *trajectory* (T_i) is the set of regions R_i^t for all t in which R_i^t exists:

$$T_i = \left\{ R_i^t \mid R_i^t \neq \varnothing, \forall t \right\} \quad (3)$$

Trajectories are assumed to be temporally connected. Two trajectories T_i and T_j are *adjacent* if and only if R_i^t and R_j^t co-exist in at least one frame and are neighbors in every frame they co-exist in.

The TAG is composed of nodes representing all the trajectories of a partition sequence and links connecting adjacent trajectories. Note that for a given partition sequence the TAG may be disjoint (see Fig. 3a). A temporal splitting procedure may thus be necessary (Section 3.3). The trajectory merging algorithm proceeds by removing the link of the TAG with the largest similarity (Section 3.2), merging R_i^t and R_j^t for every t in which they co-exist, and updating the partition and TAG accordingly (see Fig. 3c and 3d).

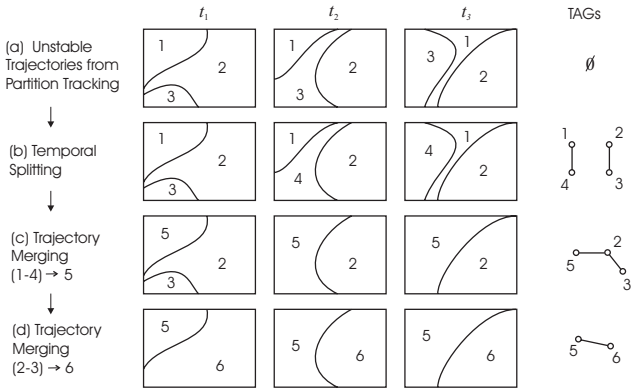

Fig. 3 (a) Tracked partition, (b) temporal splitting, (c), (d) iterations of the trajectory merging algorithm over TAGs.

3.2. Long-term Color Similarity

The spatio-temporal connectivity constraints imposed by recursive tracking schemes may lead to sub-optimal segmentations. The temporal evolution of these areas is particularly unstable since clear color boundaries are not available to constrain the current segmentation. See, for example, the evolution of the regions within the color-homogeneous face area of Fig. 1 (level 2). In order to exploit long-term color features, the color homogeneity criterion is relaxed and a color similarity between border areas is proposed:

$$S_b(R_i^t, R_j^t, t) = \left\| \mathbf{C}_{B_{R_i^t, R_j^t}} - \mathbf{C}_{B_{R_i^t, R_j^t}} \right\|_2 \quad (4)$$

The border pixels $B_{Ri,Rj}$ of region R_i relative to R_j are the set of all pixels of R_i whose distance from R_j is $\leq n$ as illustrated in Fig. 4. $\mathbf{C}_B$ is the vector of average YUV color components within the border area B.

Fig. 4 Tracked regions and corresponding border pixels (n=3).

The long-term color-based similarity measure between adjacent trajectories T_i and T_j is given by the temporal average over all t for which both R_i^t and R_j^t exist:

$$S_T(T_i, T_j) = \sum_t S_b(R_i^t, R_j^t, t) / \sum_t 1 . \quad (5)$$

3.3. Sequence Instability

The use of long-term similarity measures such as (5) will prioritize the fusion of any trajectories presenting similar border color. However, not all trajectories of similar color require merging. An unconstrained merging scheme would ultimately lead to loss of representation detail. Temporal stability criteria are used in determining which regions must be processed.

A trajectory T_i is considered unstable if its *lifespan* or $\{t \mid R_i^t \neq \varnothing, \forall t\}$ is small in size or if relative temporal area differences (RTAD) among its regions are large. RTAD is defined as

$$RTAD(R_i^t, R_i^{t+1}) = \frac{| A_{R_i^t} - A_{R_i^{t+1}} |}{\min(A_{R_i^t}, A_{R_i^{t+1}})} . \quad (6)$$

Trajectory merging is restricted to the set of trajectories deemed unstable. Instability, however, can significantly alter connectivity relations producing disjoint TAGs (see Fig. 3a). Thus, trajectories are *temporally split* or relabeled whenever the RTAD exceeds a specified threshold (typically 2) as shown in Fig. 3b.

4. RESULTS

In [1], Erdem et al. propose a set of evaluation metrics for video objects without ground-truth segmentations:

D_{color} - the color contrast between pixels just inside and just outside of an object boundary.

D_{motion} - the difference between average motion vectors in squares just inside and outside of an object boundary. This metric is dependent upon optical flow accuracy.

D_{histo} - the color histogram difference between object instances and a reference frame.

All metrics in [1] are normalized from 0 to 1, the worse score being 1, and temporally averaged. Application of these metrics requires our trajectories to be (manually) grouped into video objects.

Unlike other object-oriented segmentation techniques, we are also interested in the accurate and temporally consistent description of an entire partition sequence of, for example, 40 regions per frame as opposed to the 2 or 3 regions of interest common in many video sequences. Thus, besides using subjective quality, we propose 2 more metrics:

Lifespan Ratio (R_{life}) - measures the average lifespan of all trajectories in a sequence relative to the sequence duration. The ratio is a value from 0 to 1. If no objects are known to enter or exit the scene, R_{life} should be close to 1.

Area Ratio (R_{area}) - characterizes the average stability of trajectories. Large lifespans alone may be meaningless unless the transitions between R_i^t and R_i^{t+1} are acceptable. A transition is considered stable if RTAD < 20% (see Equation 6). R_{area} measures the number of stable transitions relative to the total number of transitions.

Both R_{life} and R_{area} are relative to the total number of regions in all frames of the partition sequence. This allows comparisons between sequences of different lengths or different number of regions per frame (NRPF).

Results obtained from PT are compared with two other tracking schemes: General Merging Algorithm of [3] and graph-based tracking of [6] (initialized with 40 and with 10 regions). Evaluation scores are summarized in Tables 1 and 2 for the Foreman QCIF and Coastguard CIF sequences. With Foreman, GMA performs slightly worse than PT while with Coastguard, significant temporal instability increases D_{histo} and decreases R_{area} of GMA. Graph-40 presents larger R_{area} at the cost of smaller R_{life} for Foreman and generally worse performance when compared to PT for Coastguard. Note that the approach in [6] involves no motion estimation and results are dependent upon adequate initialization. As observed in the metrics for Graph-10, tracking of fewer but larger regions produces superior scores, however, visual inspection of the partitions indicate occasional loss of motion boundaries.

Improvements introduced by PSA when applied to PT are demonstrated by the increases in R_{life} and R_{area} of Tables 1 and 2. PSA was also seen to be successful in improving sequences from GMA [3] and Graph [6]. Resulting partitions from PT and PT+PSA are presented in Figs. 5 and 7. For both sequences, the same threshold set is used. The lifespan graph for the entire Foreman sequence [0, 299] generated with PT is presented in Fig. 6a. The abrupt appearance and disappearance of regions around frame 180 corresponds to the entrance/exit of regions due to camera panning. In Fig. 5, as well as in the lifespan graphs of Figs. 6b and 6c, we appreciate the stabilization introduced by PSA. The number of trajectories is reduced while maintaining significant regions of color and motion. The face, previously over-segmented and displaying non-rigid motion, is grouped as a single region by PSA (Fig. 5, 3rd col.). The holes correspond to the chin and lips which were stable enough to be considered as objects during a limited interval. In Fig. 7, the PT algorithm is capable of introducing new objects (large boat) which have been preserved in PSA. Column 3 contains the trajectories (manually selected) corresponding to the large and small boats. Complete video segmentation results are also available at http://gps-tsc.upc.es/imatge/_Camilo/icip06.

5. CONCLUSIONS

Experimental results corroborate the performance of both the Partition Tracking and Partition Sequence Analysis algorithms. The use of novel long-term color similarity measures and TAGs has provided temporal stabilization within a hierarchical segmentation approach. Future work includes the application of motion-based measures for trajectory merging and the development of a representation framework for the entire shot.

Algorithm	frames	Average NRPF	D_{color} x10^{-1}	D_{motion} x10^{-1}	D_{histo} x10^{-2}	R_{life}	R_{area}
GMA	[0,50]	53.0	9.03	7.16	3.0	0.81	0.76
Graph-40	[0,49]	45.4	9.08	6.96	2.8	0.33	0.84
Graph-10	[0,23]	9.5	8.72	7.36	4.3	0.79	0.99
PT	[0,50]	46.7	8.96	7.16	2.7	0.81	0.81
PT+PSA	[0,50]	41.3	8.95	7.16	2.7	0.88	0.83

Table 1. Evaluation scores for Foreman sequence.

Algorithm	frames	Average NRPF	D_{color} x10^{-1}	D_{motion} x10^{-1}	D_{histo} x10^{-2}	R_{life}	R_{area}
GMA	[0,50]	92.4	8.51	8.34	13.3	0.60	0.47
Graph-40	[0,29]	50.2	8.01	7.28	4.4	0.31	0.78
Graph-10	[0,29]	13.7	7.75	7.79	3.2	0.47	0.96
PT	[0,50]	59.8	7.90	6.89	5.3	0.49	0.81
PT+PSA	[0,50]	53.0	7.90	6.91	5.4	0.74	0.82

Table 2. Evaluation scores for Coastguard sequence.

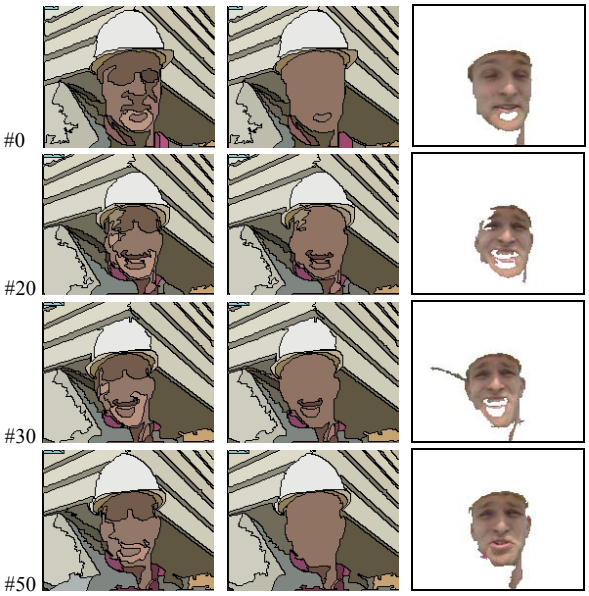

Fig. 5 Results from PT, PT+PSA and face region for Foreman.

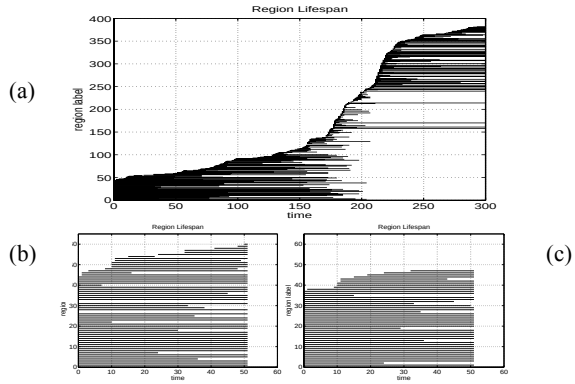

Fig. 6 Lifespan graphs for Foreman [0,299] and [0,50].

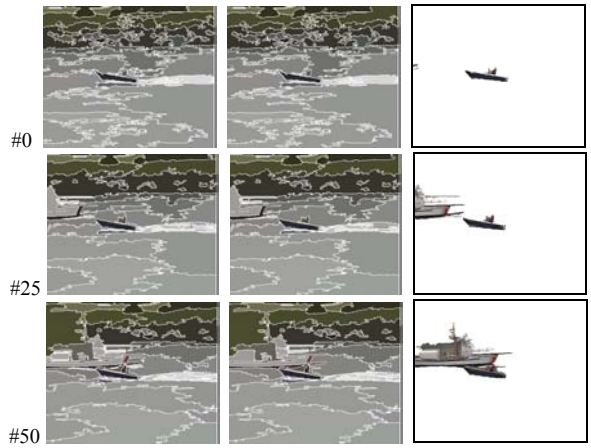

Fig. 7 Results from PT, PT+PSA and boats for Coastguard.

6. ACKNOWLEDGEMENTS

We thank Beatriz Marcotegui from the CMM of the Ecole des Mines de Paris for providing the processed sequences of [6]. This work has been partly supported by the EU project NoE MUSCLE FP6-507752, by grant TEC2004-01914 of the Spanish Government and by CAPES - Brazilian Government.

7. REFERENCES

[1] C. E. Erdem, A. M. Tekalp and B. Sankur, "Performance for video object segmentation and tracking", *IEEE Trans. Image Proc.*, vol. 13, no. 7, pp. 937-951, 2004.

[2] F. Moscheni, S. Bhattacharjee and M. Kunt, "Spatial-temporal segmentation based on region merging", *IEEE Trans. Patt. Anal. Mach. Intell*, vol. 20, pp. 897-915, 1998.

[3] L. Garrido and P. Salembier, "Region based analysis of video seuences with a general mergin algorithm" *EUSIPCO'98*, Greece, vol. 3, pp. 1693-1696, 1998.

[4] V. Mezaris, I. Kompatsiaris and M. Strintzis, "Video object segmentation using Bayes-based temporal tracking and trajectory-based region merging", *IEEE Trans. Circuits Sys. Video Tech.*, vol. 14, pp. 782-795, 2004.

[5] P. Salembier and L. Garrido, "Binary Partition Tree as an efficient representation for image processing, segmentation and Information Retrieval", *IEEE Trans. Image Processing*, vol. 9, pp. 561-576, 2000.

[6] C. Gomilla and F. Meyer, "Tracking objects by graph matching of image partition sequences," 3rd IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition, Italy, pp. 1-11, 2001.