Region of Interest Generation in Dynamic Environments Using Local Entropy Fields

Luciano Spinello and Roland Siegwart

Autonomous Systems Lab, ETH Zurich, Switzerland
email: {luciano.spinello, roland.siegwart}@mavt.ethz.ch

Abstract. This paper presents a novel technique to generate regions of interest in image sequences containing independent motions. The technique uses a novel motion segmentation method to segment optical flow using a local entropies field. Local entropy values are computed for each optical flow vector and are collected as input for a two state Markov Random Field that is used to discriminate the motion boundaries. Local entropy values are highly informative cues on the amount of information contained in the vector's neighborhood. High values represent significative motion differences, low values express uniform motions. For each cluster a motion model is fitted and it is used to create a multiple hypothesis prediction for the following frame. Experiments have been performed on standard and outdoor datasets in order to show the validity of the proposed technique.

1 Introduction

Region of interest (ROI) generation in image sequences has a considerable importance in a number of mobile robotics applications using vision, especially in object detection and tracking in dynamic and outdoor environments. The aim of this paper is to segment regions of interest that contain independent relative motions in the optical flow and to use these moving regions to further enhance the segmentation in the following frames.

Several approaches can be found in literature to segment motions in image sequences including layered based analysis [1][2][3], multi-body factorization [4], motion eigenvectors analisys [5][6][7] or methods that exploit the 3D information contained in n-image sequences [8][9] to obtain precise results with computationally demanding algorithms.

The technique explained in this paper generates a motion segmentation within two frames and it is designed to work in sequences with moving observers and moving objects. Such kind of segmentation can be used in mobile robotics where waiting for n-frames for an hypothesis can represent a drawback.

The novelty of this paper is given as follows:

– Neighborhood definition in the motion vector field through the use of a Delaunay triangulation: the arcs of the graph define the neighbor search path.

– Generation of a local entropy field to compute motion difformities in the optical flow: for each node of the graph a local entropy value measures the information quantity of the vector's neighborhood motion.
– Use of a two states Markov Random Field (MRF) for defining uniform motion boundaries: the class $\mathcal{H}$ define the boundaries of the clusters that are then labeled using a graph cut approach.
– Segmentation enhancement in following frames using a multiple hypothesis motion prediction scheme.

The algorithm is also built considering a practical implementation with compact memory usage, suitable for mobile robotics.

The structure of the paper is the following. The motion field neighbourhood and topology is explained in Sec. 2. The local entropies field is described in Sec. 3. The design of the Markov random field for the problem is explained in Sec 4. The multiple hypotheses motion prediction is explained in the following Sec 5 and the experimental results are shown in Sec 6. A schematic explanation of the technique presented in this paper is given in fig. 1.

2 Motion Field Neighbourhood

A sparse optical flow method based on the work of Kanade, Lucas and Tomasi [10] is here used (KLT) in order to overcome disadvantages of dense optical flow techniques in outdoor environments In the sparse optical flow the velocity vectors

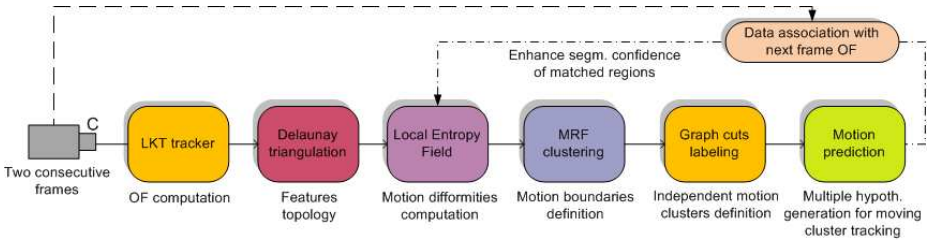


Fig. 1. Overview of the proposed technique.

are scattered in the image plane. The neighborhood information is expressed through the use of a Delaunay triangulation. The origins of the velocity vectors of the computed optical flow are used as the set P of points that define the vertices of the triangulation. The edges of the graph defined by the Delaunay triangulation define the search path among features. Delaunay triangulation has been selected for its low computational complexity and unique tessellation of the space.

3 Local entropies field

In this paper the optical flow is considered for a local entropies field computation. The dual of a Delaunay triangulation, the Voronoi decomposition, is used to tessellate the optical flow in cells. A local entropy value is computed for each cell's neighborhood, defined by the adjacent cells using the information theory.

Let's consider an information source emitting n symbols $A = \{a_1, a_2, \ldots, a_n\}$ with probability $\mathbf{u} = [P(a_1), P(a_2), \ldots, P(a_n)]$. Having modeled the information source, we can develop the input-output characteristics of the information channel. Because we modeled the source as a discrete random variable, the information transferred to the output of the channel is also a discrete random variable. Given the channel symbol alphabet $B = \{b_1, b_2, \ldots, b_n\}$ the probability of the event that symbol b_k is presented to the information user is $P(b_k) = \sum_j^n P(b_k|a_j)P(a_j)$. The probability distribution for the complete output alphabet can be computed from:

$$\mathbf{v} = Q\mathbf{z}; \qquad Q = \begin{bmatrix} P(b_1|a_1) & P(b_1|a_2) & \cdots & P(b_1|a_n) \\ P(b_2|a_1) & \vdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ P(b_k|a_1) & P(b_k|a_2) & \cdots & P(b_k|a_n) \end{bmatrix} \tag{1}$$

The matrix Q is referred to as *forward channel transition matrix*. In our case the alphabet A considered is defined by the difference between salient values for motion segmentation. Salient values in the optical flow for relative motion segmentation are defined by the vector angle α, the vector norm m, and the L_1 distance between the origin of two vectors l. Values of α and m give an information about motion difference between two vectors, l is useful to avoid grouping of distant vectors in the motion field.

For each cell an histogram is built using the salient value differences between the reference vector and the neighborhood. The probability set $\mathbf{u}$ is computed using this histogram. High uncertainty is given to conditional probability relative to symbols expressing large differences in the matrix channel Q. This means that probabilities emitted from the source expressing large differences are considered more noisy, therefore they contribute more to the channel output entropy value $H(\mathbf{v}) = -\sum_j^n P(b_j) \log P(b_j)$, effectively highlighting motion difformities in the cell's neighborhood.

For each cell $c(i)$ its local channel output entropy value is computed using the salient values α, l, m in the neighborhood $N(c(i))$:

$$H_c(i) = c_1 \underbrace{\sum_{j \in N(c(i))} P(b_j)^\alpha \log_2 P(b_j)^\alpha}_{angle} + c_2 \underbrace{\sum_{j \in N(c(i))} P(b_j)^m \log_2 P(b_j)^m}_{norm} + c_3 \underbrace{\sum_{j \in N(c(i))} P(b_j)^l \log_2 P(b_j)^l}_{dist} \tag{2}$$

where c_1, c_2, c_3 represent the weighting factors. The cell local entropy value expresses the amount of information present in its neighborhood, using the cell as reference value. The smaller the value $H_c(i)$ is the stronger the evidence is that the cell is contained in a similar motion region. Cluster boundaries are cells in which the entropy is substantially bigger. An example of a computed local entropy field is shown in Fig.9. Intuitively, the motion segmentation algorithm has to segment local minima in the local entropy field to obtain a correct clustering.

4 Clustering with a two state MRF

We model the clustering problem of motion segmentation with a pairwise Markov random field (MRF). Well known methods already exist to achieve clustering with MRF, but most of them are computationally expensive or memory demanding. In Fig.2 the circles represent network nodes and the lines indicate the statistical dependencies among nodes. With x_i we indicate the states and with y_i the observations. The problem is formalized using only two states instead of defining one for each cluster. This improves the convergence speed and the memory compactness of the algorithm. The considered state space is defined by:

$$x_i \in \{\mathcal{H},\ \mathcal{L}\} \tag{3}$$

where $\mathcal{H}$ defines a cluster boundary (high entropy) and $\mathcal{L}$ a cluster element (low entropy). Instead of retrieving from the clustering phase the segmented clusters, as the classical cluster approach, we retrieve the boundaries of the clusters defined by the state $\mathcal{H}$. We let the observations y_i be proportional to

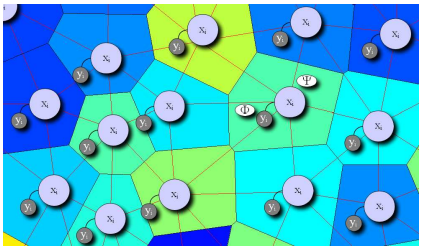


Fig. 2. Pairwise Markov Random Field connected through Delaunay triangulation graph.

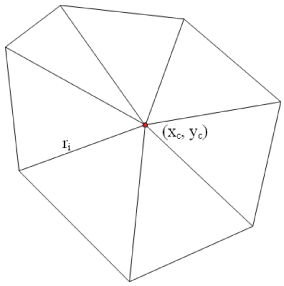


Fig. 3. Shape descriptor.

the observed local entropy values $H_c(i)$ explained in the previous section. For this network, the overall joint probability of the nodes x_i and the observations y_i is proportional to the product of all sets of compatibility matrix Ψ, relating the possible states of each pair of neighboring hidden nodes, and the vectors Φ, relating each observation to the underlying hidden states:

$$p(\{x\},\{y\}) = \frac{1}{Z}\prod_{(ij)} \Psi_{ij}(x_i,\ x_j)\prod_i \Phi(x_i, y_j) \tag{4}$$

where Z is a normalization constant, and the first product is over all neighboring pairs of nodes, i and j.

The compatibility function $\Psi_{ij}(x_i,\ x_j)$ is specified to have an high compatibility if the states are the same and a low value if they are different. It is defined by a distance measure:

$$\Psi_{ij}(x_i,\ x_j) = e^{\left(-\frac{d_{ij}(x_i,\ x_j)}{2\sigma^2}\right)} \tag{5}$$

where d_{ij} is defined with respect to the mean entropy value. This MRF model has similarities with the case of image noise removal with MRF [11]. The method aims to suppress boundary cells that have just small support from their neighbors. This compromises the segmentation of clusters having a very small area, because the MRF considers them as noise. The compatibility function $\Phi(x_i, y_j)$ enhances high values of $H_{c(i)}$ as strong evidences of $\mathcal{H}$ and low values of $H_{c(i)}$ as strong evidences of $\mathcal{L}$.

Finding the exact solution of Eq.4 can be computationally intractable, but good results are obtained using an approximate solution based on a fast, iterative algorithm called loopy belief propagation (BP) [12]. The standard belief-propagation algorithm updates messages m_{ij} from node i to node j, which are vectors of two elements (the dimensionality of the state).

The messages are determined self-consistently by the message update rules:

$$m_{ij} = \sum_{x_i} \Phi(x_i, y_j)\Psi_{ij}(x_i,\ x_j) \prod_{k\in N(i)/j} m_{ki}(x_i) \tag{6}$$

where $N(i)$ denotes the nodes neighboring i. Furthermore the belief at node i is proportional to the product of the local compatibility function at that node and all the messages coming into the node i:

$$b_i(x_i) = \mu\Phi(x_i, y_j) \prod_{j\in N(i)} m_{ij}(x_i) \tag{7}$$

where μ is the normalization constant (the beliefs must sum 1). If we consider the observation y_i fixed and we focus on the joint probability distribution of x_i in Eq.4, the belief propagation in fact gives the exact marginal probabilities for all the nodes in any singly-connected graph. Even though the convergence of belief propagation algorithm is not guaranteed [13] in graph with loops, it is often used in literature to solve marginalization problems (i.e. [14]). Markov random field marginalization defines only boundaries thus a labeling procedure is needed to define motion blobs in the image. A *marching line*-like algorithm is run among border nodes in order to distinct the different hulls.

5 Multiple hypothesis motion prediction

The motion segmentation obtained using two frames is used to augment the information of the following frames. This is achieved using a multiple hypothesis motion prediction. The idea is to use the motion clusters at time t to enhance the motion segmentation in the following frame, matching clusters with similar motions. An affine motion model is fitted to the motion vectors contained in each blob using a least squares fitting method. A shape descriptor is used to describe the convex hull of a cluster:

$$s_d = \{x_c, y_c, r_1, \ldots, r_n\} \tag{8}$$

where (x_c, y_c) is the centroid of the convex hull and r_i the radii connecting it to the boundary points, as shown in Fig.3. The set of possible shape hypotheses N_h we want to generate for each cluster is created adding to each radius a Gaussian noise. The affine motion model of each cluster is applied to the centroid coordinates of the N_h shapes generated. The multiple hypotheses set of time t is then compared with the unsegmented optical flow of frame $(t+1)$. A greedy nearest neighbor data association technique is applied to check which hypothesis fits better to the data. If the data association problem is satisfied the old motion model reinforces the local potentials $\Phi(x_i, y_j)$ of the nodes covered by the predicted cluster, increasing the value of the $\mathcal{L}$ state.

6 Experimental Results

The algorithm was tested on several image sequences taken from PETS2000-PETS2001 standard dataset [15] and a real-world outdoor dataset.

Image sequences which show close relative motion have been selected from PETS2000 and PETS2001 datasets in order to show significative and not trivial results. 800 tracking features are computed between two consecutive images. In fig. 4(a), fig. 4(b), fig. 5(a) and fig. 5(b) moving objects are correctly segmented but due to lack of stable tracking features some parts of the objects are outside of the convex hull. In order to obtain a quantitative result a sequence detection rate is evaluated. The sequence detection rate is given by the sum of detections rate for each frame divided by number of frames. A correct detection is defined as that segmented moving region bounding box that overlaps at least 70% of the annotated one. 15 frames are considered for each of the sequences shown in fig. 6; this frame quantity has been selected in order to complete a crossing between shown objects. Detection rate for 4(a) is 80.0%, for 4(b) is 93.3%, for 5(a) is 93.3% and 5(a) is 86.6%. We did not consider a false positive rate analysis because this algorithm is designed to be an hypothesis generator therefore false positives (a clustered moving region that do not correspond to an annotated moving part) have to be considered in later stages (i.e. object detection) and are not included in the scope of this paper.

The mobile platform Smartter [16], used to acquire the outdoor datasets, is based on a Daimler-Chrysler Smart vehicle equipped with several active and

passive sensors (fig. 10). The camera used is a Sony 910XCR firewire camera mounted behind the windscreen equipped with a wide field of view lens.

A moving dataset and a static dataset are retrieved from our mobile platform Smartter to show the performance of the algorithm in different real-world conditions.

In the dataset 1, the Smartter is moving and several cars are passing in the other lane. The cars are segmented together due the similar motion due to the settings in the algorithm that prefers to cluster together similar motion than doing overclustering. In a bigger cluster it is more probable to find objects than in many fragmented blobs. In fig. 6(a) the Voronoi tessellation and local potentials are shown, the overlayed motion clusters are depicted in fig. 6(b).

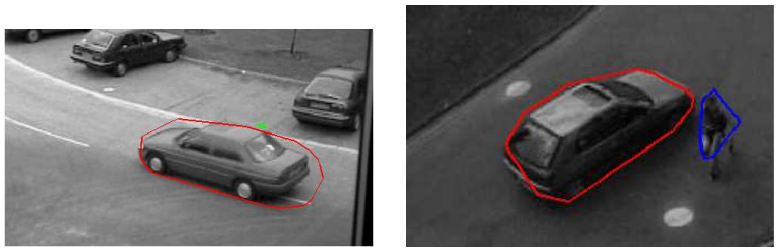
In the dataset 2, the Smartter is still in front of a red traffic light. The car in the other lane is correctly segmented. In fig. 7(a) the optical flow, the triangulation and local potentials are shown, the overlayed motion clusters are depicted in fig. 7(b).

In both datasets Markov random field steady state is always reached less than 30 message propagations. Another experiment was performed to test the

(a) PET2000 - frame 1218 (b) PET2001 - frame 424

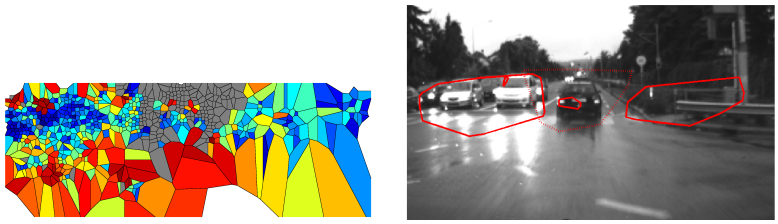
Fig. 4. a) Two semented pedestrians are moving in the street in different directions. Due to the lack of features the moving convex hull does not cover completely the moving pedestrians b) A car and a pedestrian are moving in opposite directions; well defined separation is shown.

scalability of the algorithm with respect to the number of extracted features. The optical flow information is consequently decreased but, as shown in fig. 8(a), the motion segmentation scales gracefully with respect to the features quantity. The motion segmentation algorithm here described has been also used as a constraint for an AdaBoost cascade based car detection. An Haar feature based Adaboost classifier is trained with a car data set (trunk/front) to obtain car detection. Classically, the trained classifier searches all over the image for classified features at different scales [17]. This extensive search is now constrained in the segmented motion clusters. This enhances the execution speed and potentially reduces the



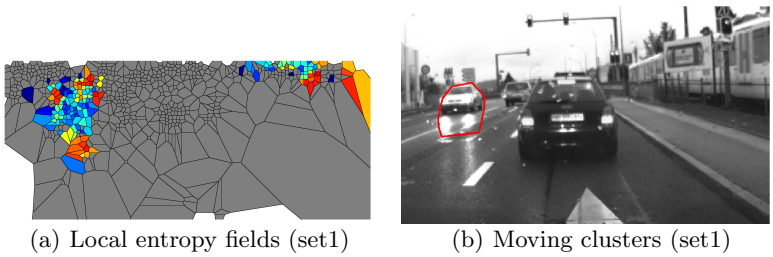
(a) PET2000 - frame 129 (b) PET2001 - frame 563

Fig. 5. a) A car is moving on a street. A small incorrectly segmented moving region is found (green area - false positive). b) A car and a pedestrian are moving in opposite directions; well defined separation is shown.



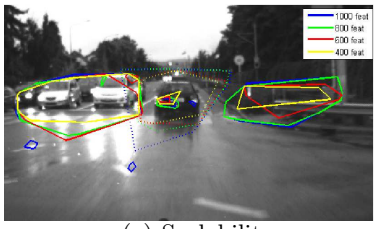
(a) Local entropy fields (set1) (b) Moving clusters (set1)

Fig. 6. [Outdoor set 1: moving observer/moving objects] a) Local entropies field and Voronoi tessellation. Two low entropy zones are in the field (in blue), in the gray zone a below than 1px optical flow is detected. b) Moving clusters detected. Two cars in the left lane are segmented together; the right cluster segments the features of the right part of the road. The dotted central cluster depicts a static cluster in which the optical flow is negligible.



(a) Local entropy fields (set1) (b) Moving clusters (set1)

Fig. 7. [Outdoor set 2: static observer/moving objects] a) Local entropies field and Voronoi tessellation. Only the cells associated with left car show a coherent motion. b) Moving clusters detected. The car in the other lane is correctly segmented.



(a) Scalability

(b) Application example: AdaBoost Haar features cascade based car detection.

Fig. 8. a) Scalability of the proposed technique with respect to the number of tracked features. b) The classification algorithm is run only in the segmented moving regions. The overlayed yellow boxes depict a car detection.

false positive rate because the car is only searched in a relative motion area. The area constrained detection is shown in fig. 8(b).

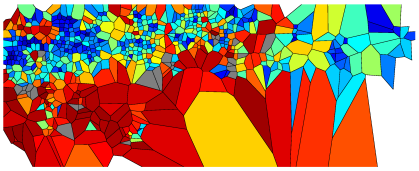


Fig. 9. A local entropies field. The magnitude of the entropy in the image is depicted using a color space from blue (low value) to red (high value).

Fig. 10. Smartter platform: mobile platform for autonomous navigation, mapping and perception.

7 Conclusions

This paper presented a novel technique to segment region of interest based on independent motion detection. Optical flow is segmented using local entropy values for each motion vector computed on a Delaunay graph and it is then clustered using a two states Markov Random Field. The clusters motion information is used to create a multiple hypothesis prediction for the following frames in order to influence the clustering.

Experiments have been conducted on different datasets to show that independent motion regions are correctly segmented. Moreover a scalability test shows that the segmentation detail scales gracefully with the number of features. This

technique has been already used to constraint an Adaboost based method for car detection with promising results.

The method presented in this paper does not use heavily computationally expensive or memory demanding techniques therefore it represents a suitable choice for mobile robotics applications.

References

1. J. Wang and E. Adelson, "Representing moving images with layers," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 625–638, 2004.
2. Y.Weiss and E. H. Adelson, "A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 1996, pp. 321–326.
3. J. Xiao and M. Shah, "Accurate motion layer segmentation and matting," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
4. Q. Ke and T. Kanade, "A subspace approach to layer extraction," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001, pp. 255–262.
5. J. Shi and J. Malik, "Motion segmentation and tracking using normalized cuts," in *6th Intl. Conference on Computer Vision*, 1998, pp. 1154–1160.
6. D. Cremers and S. Soatto, "Motion competition: a variational approach to piecewise parametric motion," *Intl. Journal of Computer Vision*, vol. 62, no. 3, pp. 249–265, May 2005.
7. F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, "Segmenting, modeling, and matching video clips containing multiple moving objects," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.
8. W.-S. Tong, C.-K. Tang, and G. Medioni, "Simultaneous two-view epipolar geometry estimation and motion segmentation by 4d tensor voting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1167–1184, 2004.
9. R. Vidal, "Multi-subspace methods for motion segmentation from affine, perspective and central panoramic cameras," in *IEEE International Conference on Robotics and Automation (ICRA05)*, 2005.
10. J. Shi and C. Tomasi, "Good features to track," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, Seattle, 1994.
11. P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.
12. J. Yedidia, W. Freeman, and Y. Weiss, *Exploring Artificial Intelligence in the New Millennium*. Morgan Kaufmann, 2003, ch. Understanding Belief Propagation and Its Generalizations, pp. 236–239.
13. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
14. W. Freeman, T. Jones, and O. Carmichael, "Example-based super-resolution," *IEEE Image-Based Modelling, Rendering, and Lighting*, vol. 21, March-April 2002.
15. D. P. Young and J. M. Ferryman, "Pets metrics: On-line performance evaluation service," in *ICCCN 05: Proceedings of the 14th International Conference on Computer Communications and Networks*, 2005.
16. S. Kolski, K. Macek, and L. Spinello, "Secure autonomous driving in dynamic environments: From object detection to safe driving," in *Workshop on Safe Navigation in Open and Dynamic Environments (IROS2007)*, 2007.
17. P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, 2002.