Region and Graph-Based Motion Segmentation

Fernando C. Monteiro[1,2] and Aurélio Campilho[1,3]

[1] INEB - Instituto de Engenharia Biomédica
[2] Instituto Politécnico de Bragança, Portugal
[3] FEUP - Faculdade de Engenharia, Universidade do Porto, Portugal

Abstract. This paper describes an approach for integrating motion estimation and region clustering techniques with the purpose of obtaining precise multiple motion segmentation. Motivated by the good results obtained in static segmentation we propose a hybrid approach where motion segmentation is achieved within a region-based clustering approach taken the initial result of a spatial pre-segmentation and extended to include motion information. Motion vectors are first estimated with a multiscale variational method applied directly over the input images and then refined by incorporating segmentation results into a region-based warping scheme. The complete algorithm facilitates obtaining spatially continuous segmentation maps which are closely related to actual object boundaries. A comparative study is made with some of the best known motion segmentation algorithms.

1 Introduction

Motion segmentation is basically defined as grouping pixels that are associated with a smooth and uniform motion profile. The segmentation of an image sequence based on motion is a problem that is loosely defined and ambiguous in certain ways. Though the definition says that regions with coherent motion are to be grouped, the resulting segments may not conform to meaningful object regions in the image. Recent applications such as content-based image/video retrieval, like MPEG-7 [5], and image/video composition, require that the segmented objects are semantically meaningful. Indeed, the multimedia standard MPEG-4 [9] specifies that a video is composed of meaningful video objects. In order to obtain a content-based representation, an image sequence must be segmented into an appropriate set of semantically shaped objects or video object planes. Although the human visual system can easily distinguish semantic video objects, automatic video segmentation is one of the most challenging issues.

There is a strong interdependence between the definition of the spatial support of a region and of its motion estimation. On one hand, estimation of the motion information of the region depends on the region of support. Therefore, a careful segmentation of the regions is needed in order to estimate the motion accurately. On the other hand, a moving region is characterized by coherent motion characteristics over its entire surface (assuming that only rigid motion is permitted). Therefore, an accurate estimation of the motion is required in order to obtain an accurate segmentation of the region.

A. Campilho and M. Kamel (Eds.): ICIAR 2008, LNCS 5112, pp. 609–618, 2008.

In this paper, a hybrid framework is proposed to integrate a differential optical flow approach and a region-based spatial segmentation approach to obtain an accurate object motion. Motion information will be initially represented through a dense motion vector field. For the task at hand we adopt a high accuracy optical flow estimation based on a coarse-to-fine warping strategy proposed by Brox et al. [3] which can provide dense optical flow information. This method accelerates convergence by allowing global motion features to be detected immediately, but it also improves the accuracy of flow estimation because it provides better approximation of image gradients via warping.

To partitioning each frame into a set of homogeneous regions we used a variation of the rainfalling watershed implementation [7]. The proposed method performs rainfall only within the regions of interest in which a pixel shows variation in gradient magnitude. The set of neighbour pixels with constant gradient magnitude, i.e. within a flat region, are desert regions where rain rarely falls or, to be more precise, where only a raindrop falls.

Handling spatial and temporal information in a unified approach is appealing as it could solve some of the well known problems in grouping schemes based on motion information alone [14, 15]. Brightness cues can help to segment untextured regions for which the motion cues are ambiguous and contour cues can impose sharp boundaries where optical flow algorithms tend to extend along background regions. Graph based segmentation is an effective approach for cutting (grouping) sets of nodes and its extension to integrate motion information is just a matter of adding a proper similarity measure between nodes. The assignment of motion to regions allows the elimination of optical flow errors (outliers).

The remainder of this paper is organized as follows: in Section 2, motion estimation algorithm is presented. In Section 3, we build the region-based motion graph. The proposed motion segmentation algorithm is presented in Section 4. In Section 5, experimental results are analysed and discussed. In Section 6, a comparative study is made, and, finally, conclusions are drawn in Section 7.

2 Variational Methods

Optical flow is defined as the 2-D vector that matches a pixel in one image to the warped pixel in the other image. In other words, optical flow estimation tries to assign to each pixel of the current frame a two-component velocity vector indicating the position of the same pixel in the reference frame. Given two successive images of a sequence $I\left(x,y,t\right)$ and $I\left(x,y,t+1\right)$ we seek at each pixel $\mathbf{x} := \left(x,y,t\right)^{\mathrm{T}}$ the flow vector $\mathbf{v}\left(\mathbf{x}\right) := \left(v_x, v_y, 1\right)^{\mathrm{T}}$ that describes the motion of the pixel at $\mathbf{x}$ to its new location $\left(x+v_x, y+v_y, t+1\right)$ in the next frame.

Differential methods, and in particular variational methods based on the early approach of Horn and Schunck [6] are among the best performing techniques for computing the optical flow [3, 4, 10]. Such methods determine the desired displacement field as the minimiser of a suitable energy functional, where variations from model assumptions are penalised. In general, this energy functional consists of two terms: a data term and a smoothness term. While the data term represents the assumption that certain image features do not change over time and thus allow for a retrieval of corresponding objects in subsequent frames, the smoothness term stands for the assumption that neighbouring pixels most probably belong to the same object and thus undergo a similar type of motion. Due to the smoothness constraint which propagates information from textured areas to nearby non-textured areas the resulting flow field is dense i.e. there is an optical flow estimate (vector) available for each pixel in the image. Brox et al. [3] proposed a variational method that combines brightness constancy with gradient constancy assumptions and a discontinuity-preserving temporal smoothness constraint. In order to allow for large displacements, this technique implements a coarse-to-fine warping strategy. Applying non-quadratic penaliser functions to both the data and the smoothness term and also integrating the gradient constancy assumption, results in the optical flow model described by the following functional:

$$E\left(v_x, v_y\right) = E_D\left(v_x, v_y\right) + \alpha E_S\left(v_x, v_y\right) \quad , \tag{1}$$

where α is some positive regularisation parameter which balances the data term E_D with the smoothness term E_S: Larger values for α result in a stronger penalisation of large flow gradients and lead to smoother flow fields.

The minimization of $E\left(v_x, v_y\right)$ is an iterative process, with external and internal iterations [3]. The external iterations are with respect to scale. The internal iterations are used to linearise the Euler-Lagrange equations and solve the resulting linear set of equations. The reader is referred to Thomas Brox's PhD thesis [2] for a solution to minimize this functional.

3 Building the Region-Based Motion Graph

The definition of the region similarity which involves not only motion information but also spatial characteristics is a challenging issue. All the available information should be put to work in order to robustly define the objects present in the scene. We propose a region similarity measure that exploits both spatial similarity $w_s\left(i, j\right)$ and motion similarity $w_m\left(i, j\right)$:

$$W\left(i, j\right) = \varphi \cdot w_m\left(i, j\right) + \left(1 - \varphi\right) \cdot w_s\left(i, j\right) \quad , \tag{2}$$

where φ is a regularisation term that reflects the importance of each measure. Spatial similarity measure is obtained by

$$w_s\left(i, j\right) = w_{ic}\left(i, j\right) \cdot w_I\left(i, j\right) \quad , \tag{3}$$

with w_{ic} as the similarity obtained by intervening contours and w_I as the intensity similarity described in [7]. The role of w_s is only to be a refinement measure. Therefore, in our experiments φ was set to 0.95.

3.1 Motion Similarity Measure

Using atomic regions implicitly resolves the problems identified earlier which requires smoothing of the optical flow field since the spatial (static) segmentation process will group together neighbouring pixels of similar intensity, so that all the pixels in a area of smooth intensity grouped in the same region will be labelled with the same motion. We thereby presume two basic assumptions: i) all pixels inside a region of homogeneous intensity follow the same motion model, and ii) motion discontinuities coincide with the boundaries of those regions.

For region-based motion segmentation, we assign a unique motion vector to each region given by the peak in the optical flow histogram distribution. The idea here is to represent a motion vector $\mathbf{v} = (v_x, v_y)$ in a (U_x, U_y) plane with radius ρ and the argument θ given by:

$$\rho(\mathbf{v}) = \log\left(1 + \beta\left(v_x^2 + v_y^2\right)^{1/2}\right) \qquad \theta(\mathbf{v}) = \tan^{-1}\left({}^{v_y}\!/_{v_x}\right) \quad , \tag{4}$$

where β is a positive parameter included to reflect the variation in the similarity judgement of motion from person to person.

The motion information of each region are computed in reference to different points - the centroids of the regions. We define a motion distance $d_m(i,j)$ expressing the degree of similarity between the motion fields of two regions R_i and R_j in reference to the centroid of R_i which can be expressed as:

$$d_m(i,j) = \sqrt{(\Delta^2 U_x + \Delta^2 U_y)} \quad , \tag{5}$$

with $\Delta U_x = \rho_i \cos\theta_i - \rho_j \cos\theta_j$ and $\Delta U_y = \rho_i \sin\theta_i - \rho_j \sin\theta_j$, where ρ_i, ρ_j, θ_i and θ_j are calculated by (4). In fact, this motion distance expresses how well the motion model of region R_j can also fit the motion of region R_i.

As the distance measures have their own range it is desirable to normalize their values. The parameter σ_m in (6) is used to normalize the distance measure to a range $[0,1]$.

$$w_m(i,j) = \exp\left(-d_m(i,j)^2 \big/ \sigma_m^2\right) \quad . \tag{6}$$

4 Motion Segmentation Algorithm

In this section, we describe the fusion of spatial segmentation and motion information for high quality motion segmentation. If it is true that, for synthetic sequences, flow field values can be computed exactly, that is not the typical scenario where flow field is estimated from a sequence of images. Then, our approach should be robust against inaccuracies in the motion information. We used the implementation of Brox et al. [3] which produces results that are among the best of all the currently available methods for optical flow estimation [3, 4, 10].

We assume that a region of uniform motion (rigid motion) will be composed of one or more atomic regions each of which possessing uniform intensity. Consequently, the motion boundaries will be a subset of the intensity boundaries determined at this stage. We refer to this assumption as *segmentation assumption.* Our choice of this assumption is supported by the following fact: the atomic regions resulting from the spatial pre-segmentation are usually small enough to justify the assumption of piecewise constant intensity and motion.

The proposed algorithm can be summarized in the following steps:

Step 1: Spatial pre-segmentation: frames are partitioned into homogeneous atomic regions based on their brightness and gradient properties (watershed).
Step 2: Motion estimation: estimates the dense optical flow field with the variational scheme proposed by Brox et al. [3].
Step 3: Region motion extraction: extracts the highly reliable optical flows for each atomic region. It selects from the dense flow field the dominant motion vector according to the directions and magnitudes of the optical flows. This step eliminates the influence of noise and outliers.
Step 5: Region-based motion graph: builds the region-based motion graph where the nodes correspond to regions.
Step 6: Graph partitioning: multiclass spectral based graph partitioning using the normalized cut approach [8].

Figure 1 illustrates the intermediate and final results of the method. The input is represented by two consecutive frames of the Ettlinger Tor sequence (available at *http://i2iwww.ira.uka.de/image_ sequences/*). The sequence consists of 50 frames of size 512 × 512 and depicts a variety of moving cars (up to 6 pixels

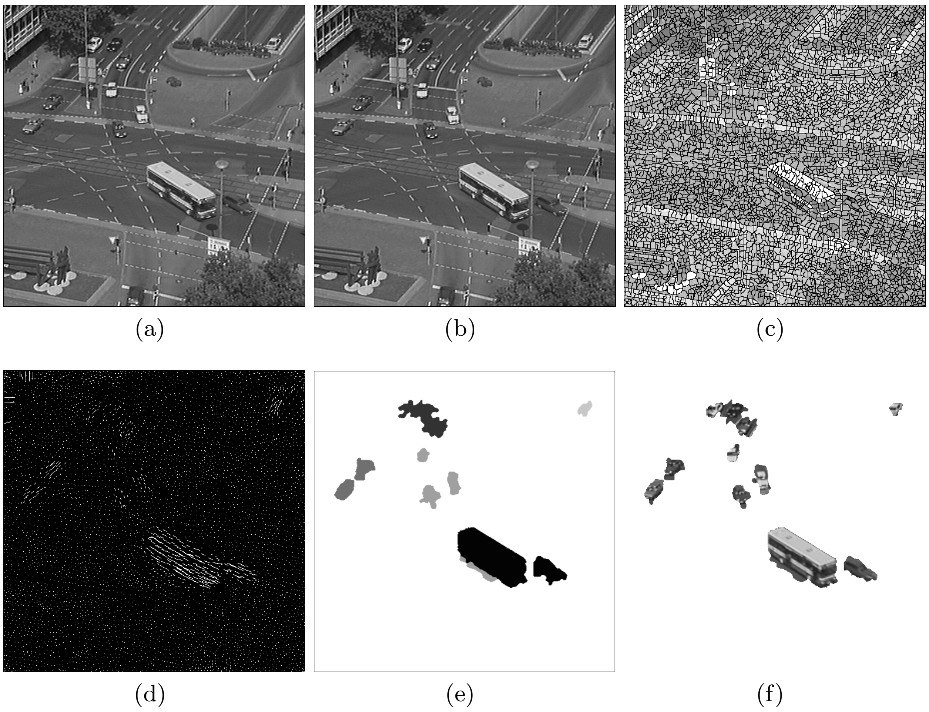
(a) (b) (c)
(d) (e) (f)

Fig. 1. Illustration of the proposed motion segmentation algorithm. (a)-(b) Frame 5 and 6 of the Ettlinger Tor sequence. (c) Atomic regions. (d) Region-based vector field scaled by a factor of 2. (e) Motion segmentation. (f) Moving regions.

per frame). Thereby five groups of cars can be formed according to their velocity and direction: 1) a bus and a car in the foreground are moving fast to the right; 2) in the middle area three cars are moving in a similar direction of group 1 but slower; 3) two cars on the left are moving to the left; 4) in the upper middle area three cars are moving slowly to the left; 5) on the upper right area a car is moving up.

In the first step, an initial segmentation of the frames is achieved with watershed-based segmentation. The result is a fine partition of the image into regions with similar intensity where region size is kept small. Motion estimation between the frames is obtained with the variational method described in Section 2. In the following, a dominant motion vector is associated with each region produced in step 1. Figure 1.d) shows the resultant flow vectors scaled by a factor of 2. Figure 1.e) shows the result of the motion segmentation where different kind of motions are represented by different grey-scale intensities in accordance with the five groups upper referenced.

Using spatial information reduce the "halo" originated by the smoothness term used in the motion estimation process allowing to obtain a more accurate segmentation. Even more, the segmentation effectively separates the groups of cars according to their type of motion.

The area under the bus was labelled as belonging to group 2 and not to group 1 as a consequence of the brightness similarity between the bottom of the bus and the ground. In other words, since the smoothness term expands the optical flow along areas of homogeneous intensity it has also expanded the bus motion to the ground. However, the optical flow of the ground has a lower magnitude which makes it more similar to the motion of the cars in group 2 than to the motion of the bus. This shows the accuracy of the motion segmentation algorithm which separates the ground region from the bus.

5 Experimental Results

The motion segmentation algorithm was tested using several benchmark test sequences: *Salesman* and *Flower Garden*. These two are among the sequences widely used by authors for testing video segmentation and coding applications. Figure 2 shows the segmentation result with the *Salesman* sequence.

The Salesman sequence does not possess any global motion, but the motion of the non-rigid object (salesman) is significant in this sequence, especially in respect to the arm movements. It can be seen in Figure 2.e) that our proposed algorithm yields satisfactory multiple motion segmentation where different colours represent different movements. Regions such as the arm of the Salesman and his hand, which moves with motion involving rotation, are correctly segmented. Also the shirt, that is divided in two by the arm, is correctly merged.

Figure 3 shows the segmentation result with the *Flower Garden with Car* sequence. In this experiment a moving car was included in the scene. The sequence was shot by a camera placed on a driving car, and the image motion is related to distance from the camera. Thus the tree, which is closest to the camera moves

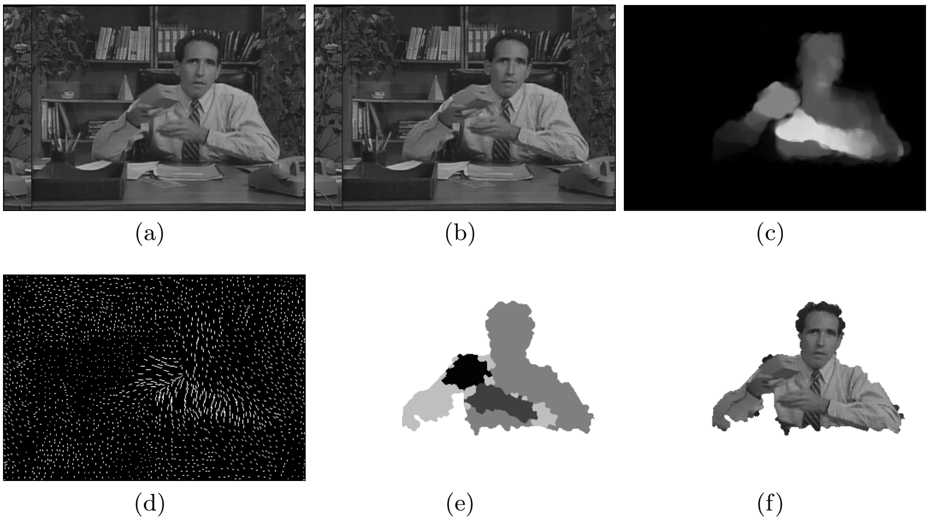

Fig. 2. Salesman sequence. (a)-(b) Frames 14 and 15. (c) Computed dense optical flow. (d) Region-based vector field. (e) Motion segmentation. (f) Moving objects.

fastest. The inter-frame difference detects motion at every image pixels. Flower Garden sequence contains many depth discontinuities, not only at the boundaries of the tree but also in the background. In this sequence, the camera captures a flower garden with a tree in the centre. Also, the flower bed gradually slopes toward the horizon showing the sky and far objects. Semantically, this sequence has four layers: the tree, the flower bed, the house and the sky.

Although the tree divides the flower bed the algorithm merges the two parts in one only segment. This happens also in the house layer. Note that in the area that contains the tree's branches, only one segment is chosen since the sky area has no brightness variation. From Fig. 3.d) it looks like as the bottom of the flower bed, the tree and the sky have the same motion information. However, the segmentation algorithm making use of the intensity information, correctly divides these parts. The region-based approach extracts the tree's edges accurately along major part of the trunk, even in similar textured area of the flower bed, but less well in other areas. The fine detail of the small branches cannot be well represented by image regions, and these are segmented poorly.

6 Comparative Results

As demonstrated by the results shown in this paper, motion segmentation is a difficult task. It is also difficult to assess, in quantitative terms, the accuracy of a segmentation. It is therefore instructive to compare the results generated by

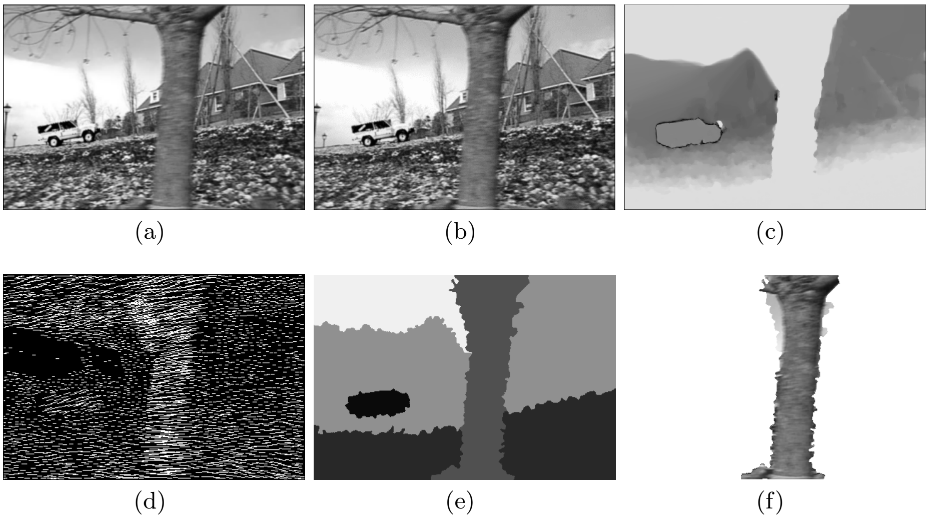

Fig. 3. Flower Garden with Car sequence. (a)-(b) Frames 5 and 6. (c) Computed dense optical flow. (d) Region-based vector field scaled by a factor of 2. (e) Motion segmentation. (f) Tree segment.

this region-based system with work published by other authors over recent years; this gives an indication of the relative success of the region-based approach.

A comparison with a number of authors who have also analysed the Flower Garden sequence is realised in Fig. 4. In this comparison we analyse the accuracy of the resulting tree segment. The results are extracted from the published papers. Although each author displays their results differently it is not difficult to compare them. Again, with no accepted quantitative measure of segmentation performance, a qualitative comparison is made between results.

Wang and Adelson [14] used this sequence in their paper introducing the layered representation. The use of normalized cuts for motion segmentation was introduced in [11], in which graph cutting techniques are used to obtain a motion related set of patches in the image sequence. Comparisons with Ayer and Sawhney [1], Vasconcelos and Lippman [13] and Weiss and Adelson [15] are also presented in Fig. 4. These authors' results show some outlying pixels or regions that are absent in our approach which gives the system presented in this paper a more pleasing appearance. Figure 4.c) shows the result of the edge-based motion segmentation scheme from Smith [12]. The area at the bottom of the tree is correctly segmented only in our approach and in the method of Ayer and Sawhney and in the method of Smith.

The segmentation of the tree in the Wang and Adelson estimate is to be too wide, while the edge-based approach of Smith and the method of Shi and Malik misses a few sections. Ayer and Sawhney, and Vasconcelos and Lippman are better outline, but there is more noise in the background.

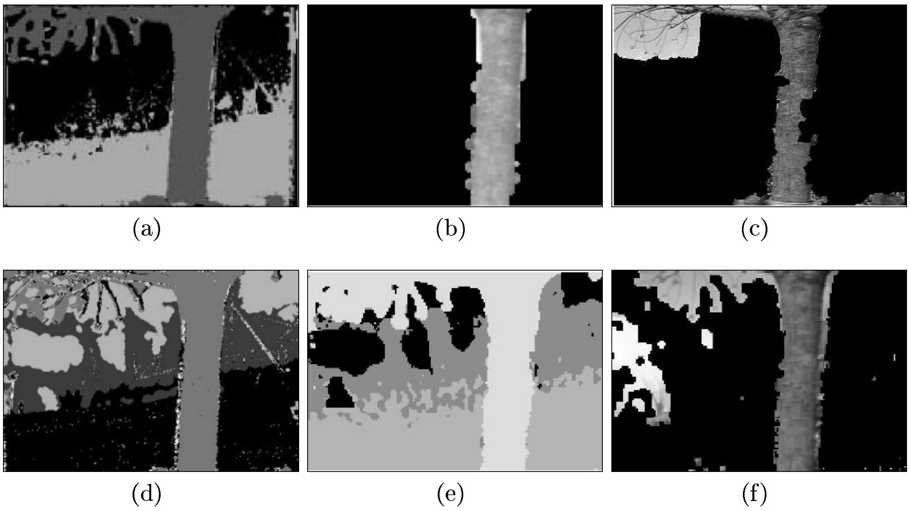

Fig. 4. Comparative results with the Flower Garden sequence. Results presented by (a) Ayer and Sawhney in [1], (b) Shi and Malik in [11], (c) Smith in [12], (d) Vasconcelos and Lippman in [13], (e) Wang and Adelson in [14], (f) Weiss and Adelson in [15].

7 Conclusion

A method for multiple motion segmentation was presented, relying on a combined region-based segmentation scheme. The spatial pre-segmentation of the frames in homogeneous intensity regions by the watershed algorithm results in an oversegmented partition. A grouping step is the performed using a region-based motion graph built on the partition obtained in the pre-segmentation stage. The derivation of a motion-based partition of the images was achieved through a graph labelling process in a spectral-based clustering approach. To achieve this aim an appropriate similarity function (energy function) was defined. Links weights now denote a similarity measure in terms of both spatial (intensity and gradient) and temporal (flow fields) features. To compute the flow field we use a high accuracy optical flow method based on a variational approach. The region-based graph-labelling principle provides advantages over classical merging methods which by operating a graph reduction imply irreversibility of merging. Moreover, spectral-based approach avoids critical dependency in the order in which regions are merged. The proposed approach successfully reduces computational cost, while enforcing spatial continuity of the segmentation map without invoking costly Markov random field models. By simultaneously making use of both static cues and dynamic cues we are able to find coherent groups within a variety of video sequences. The experimental results presented in this paper show that the proposed method provides satisfactory results in motion segmentation from image sequences.

Acknowledgements

The authors would like to acknowledge to Thomas Brox for providing optical flow software of Brox et al.'s algorithm [3].

References

[1] Ayer, S., Sawhney, H.S.: Layered representation of motion video using robust maximum-likelihood estimation of mixture models and mdl encoding. In: Proc. IEEE International Conference on Computer Vision, June 1995, pp. 777–784 (1995)

[2] Brox, T.: From pixels to regions: Partial differential equation in image analysis. PhD thesis, Department of Mathematics and Computer Science, Saarland University, Germany (2005)

[3] Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)

[4] Bruhn, A., Weickert, J., Schnörr, C.: Lucas/Kanade meets Horn/ Schunck: combining local and global optic flow methods. International Journal of Computer Vision 61(3), 1–21 (2005)

[5] Chang, S.-F., Sikora, T., Puri, A.: Overview of the MPEG-7 standard. IEEE Transactions on Circuits and Systems for Video Technology 11(6), 688–695 (2001)

[6] Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence 17(1-3), 185–203 (1981)

[7] Monteiro, F.C.: Region-based spatial and temporal image segmentation. PhD thesis, Faculdade de Engenharia da Universidade do Porto, Portugal (2007)

[8] Monteiro, F.C., Campilho, A.: Spectral Methods in Image Segmentation: A Combined Approach. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005. LNCS, vol. 3523, pp. 191–198. Springer, Heidelberg (2005)

[9] MPEG4. MPEG-4 video verification model, version 15.0. ISO/IEC/JTC1/SC29/ WG11 N3093 (1999)

[10] Papenberg, N., Bruhn, A., Brox, T., Didas, S., Weickert, J.: Highly accurate optic flow computation with theoretically justified warping. Int. Journal of Computer Vision 67(2), 141–158 (2006)

[11] Shi, J., Malik, J.: Motion segmentation and tracking using normalized cuts. In: Proc. of IEEE Int. Conference on Computer Vision, pp. 1154–1160 (1998)

[12] Smith, P.: Edge-based motion segmentation. PhD thesis, Department of Engineering, University of Cambridge (2001)

[13] Vasconcelos, N., Lippman, A.: Empirical bayesian motion segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 217–221 (2001)

[14] Wang, J., Adelson, E.: Representing moving images with layers. IEEE Transactions on Image Processing 3(5), 625–638 (1994)

[15] Weiss, Y., Adelson, E.H.: A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models. In: Proc. of IEEE International Conference on Computer Vision and Pattern Recognition, San Francisco, USA, June 1996, pp. 321–326 (1996)