Highly Consistent Sequential Segmentation[*]

Michael Donoser[1], Martin Urschler[2,1], Hayko Riemenschneider[1]
and Horst Bischof[1]

[1] Institute for Computer Graphics and Vision
Graz University of Technology, Austria
`{donoser,urschler,hayko,bischof}@icg.tugraz.at`
[2] Ludwig Boltzmann Institute for Clinical Forensic Imaging
Graz, Austria

Abstract. This paper deals with segmentation of image sequences in an unsupervised manner with the goal of getting highly consistent segmentation results from frame-to-frame. We first introduce a segmentation method that uses results of the previous frame as initialization and significantly improves consistency in comparison to a single frame based approach. We also find correspondences between the segmented regions from one frame to the next to further increase consistency. This matching step is based on a modified version of an efficient partial shape matching method which allows identification of similar parts of regions despite topology changes like merges and splits. We use the identified matched parts to define a partial matching cost which is then used as input to pairwise graph matching. Experiments demonstrate that we can achieve highly consistent segmentations for diverse image sequences, even allowing to track manually initialized moving and static objects.

1   Introduction

Unsupervised segmentation is one of the fundamental tasks in computer vision and is an important step for many high-level tasks including tracking, object recognition and 3D reconstruction. Despite the tremendous progress in the field of segmentation [5, 9, 1, 8], image segments have not been a popular choice as underlying representation in most areas of computer vision. This is due to the fact that segmentation is an ill-posed problem and results for images acquired under slightly different conditions (such as different lighting properties, viewpoint changes or movement in the scene) differ significantly. Mostly minor changes in image gradients lead to multiple splits and merges of neighboring regions and therefore most segmentation methods fail to provide consistent regions.

Analyzing region correspondences between two segmented images allows to improve the consistency [12] but since no a-priori information about the scene is used, results are still far from being acceptable for most vision tasks. The main

issue is that shape and appearance of regions change significantly for images of scenes obtained from different viewpoints, therefore matching regions in such images is a highly ill-posed problem.

In this paper we focus on the much easier problem of efficiently segmenting a sequence of images with the main goal of providing consistent segmentations throughout the sequence. This has many potential applications in computer vision like tracking, background substitution, object recognition or video editing. Please note, that we do not want to directly segment outlines of moving or static objects in the sequence, since we are not including any high-level cues or a-priori information. The results are solely intended to serve as pre-processing step for subsequent high-level vision applications.

In general there are two different directions of research for improving segmentation consistency in image sequences in the literature. First, using interest point tracks to find region matches and second, composite clustering of pixels from all frames in the sequence.

Feature-point based approaches use trajectories of interest points to merge regions with similar motions. Wills et al. [18] detected feature points using the Foerstner operator and matched them by comparing high-dimensional descriptors obtained by repeated filtering steps. The feature tracks are used to find similar motion layers in videos. In [11] feature points were tracked and grouped together by using a variant of the Expectation-Maximization method mainly for the task of video editing. Common to all these approaches is that interest point tracks do not directly define the spatial cohesiveness of the underlying objects. Mostly heuristics have to be applied to assign tracks to regions and strong assumptions on the scene have to be made to get reasonable results.

Clustering based approaches take all pixels from all frames at once and find the most appropriate spatial grouping. Each pixel is represented in the (x, y, t) space and compared to all others by analyzing local descriptors. For example in [6] the well known Mean Shift approach is used for hierarchical clustering, analyzing features consisting of color and motion cues. In [17] the video was represented as a graph and motion profiles were used to define the edge weights. Connected segments are found by a standard normalized cut method. Recently a hyper-graph cut method for video object segmentation was proposed [13], which allows to define more complex edge weights (beyond the pairwise setting) in graph matching, achieving promising results. But there are two major issues in these approaches. First, it is difficult to relate the spatial and temporal domain. Setting a suitable tradeoff between spatial and temporal changes again requires some knowledge about the scene, for example the amount and size of moving objects. Second, computation time is also a weak point since clustering in high-dimensional feature spaces becomes infeasible even for relatively short sequences.

We propose a method that achieves highly consistent segmentation results for image sequences by repeatedly matching regions between subsequent frames overcoming the aforementioned limitations. Our method can be applied online on a live stream from a camera, since we do not need to perform any post-processing over the entire sequence. Our work has three major contributions.

First, we propose a segmentation method for image sequences which provides highly consistent segmentations by using the result of frame t as initialization for segmenting frame $t+1$. The method makes use of a fast and accurate geodesic active contour algorithm formulated in the weighted total variation framework and significantly improves over single image consistency. Our second contribution is an adaption of a recently proposed partial shape matching method which can be used to obtain partial match costs between regions of subsequent frames. We propose a novel descriptor analyzing angles between chords connecting sampled points and corresponding horizontal lines. Finally, we show how to use the obtained partial matching costs in a spectral graph matching formulation. Graph matching returns a list of many to many correspondences between segments in consecutive frames and this list is used to merge and split segments. In such a way we obtain even more consistent results. This step is formulated in a general manner, and can therefore be applied to improve any of the available single image segmentation methods.

The most related work to ours is from Brendel and Todorovic [3] who recently proposed a video object segmentation method where the results of single frame segmentation were improved by also analyzing a partial match cost between segments. But there are some major differences to our work. First, they only use a single frame segmentation method like mean shift as input and neglect any analysis about how to improve the main segmentation consistency. For defining the partial match cost, they apply a partial matching method denoted as cyclic dynamic time warping which has a runtime of $200\,ms$. In our setup finding the partial match cost only requires a few milliseconds. Finally, for analysis of the identified costs for region matching they do a final clustering of all regions from the sequence by relaxation labeling which requires the entire sequence to be provided in advance and prevents online processing.

2   Consistent Image Sequence Segmentation

Unsupervised segmentation is one of the most intensively researched topics in computer vision and many different segmentation methods have been proposed achieving excellent results on reference data sets like Berkeley [5, 9, 1, 8]. But most of the methods are very sensitive to slight changes in image conditions and do not provide consistent segments.

The straightforward idea for obtaining image sequence segmentation is to apply any of the available segmentation methods independently to every frame. Of course, such an approach neglects that subsequent frames have many similarities that can be exploited to improve segmentation consistency. Another fundamental idea for image sequence segmentation is to use the segmentation of frame t as initialization for segmenting the frame $t + 1$. Unfortunately, it is often rather hard or even impossible to apply such a scheme to state-of-the-art segmentation approaches. Our approach utilizes this straight-forward idea and exploits a recently proposed segmentation method [8] which perfectly supports the initialization concept.

In [8] an unsupervised segmentation method was introduced which was based on the main idea of obtaining the final segmentation as a composition of several differently focused sub-segmentations. In a first step salient regions are extracted which highlight the main color and texture distributions of the image. Then, each of the salient regions is passed to a weighted total variation segmentation method (TV-Seg), that provides an accurate figure/ground segmentation for each salient region. Total variation segmentation minimizes a convex energy functional [4] and therefore returns a global optimal solution. It has proven to be one of the most accurate figure/ground segmentation methods available. Since each salient region provides one figure/ground segmentation, these results have to be merged to one composite image (see [8] for more details).

We extend this concept for segmentation of image sequences. Instead of applying a salient region detector to obtain the different initializations for the TV-Seg method, we use results of the previous frame. Each segmented region of frame t is mapped to the frame $t + 1$. In our experiments described in Section 4 this mapping is just a copy of the segment location from frame t to $t + 1$, but of course any motion model can be incorporated here. Each of the mapped segments is then used as salient region input to the total variation segmentation method for providing a figure/ground segmentation. One important property of the method proposed in [8] is that, after performing the segmentations for every salient region, the still unassigned areas in the image are automatically passed to TV-Seg to also get accurate segmentations in these areas. In such a way, for example newly appearing objects in the image sequence become also correctly segmented. Experiments demonstrate that these frame-to-frame segmentation approach improves the consistency compared to a single image based approach.

3   Frame-to-Frame Segment Matching

To further increase consistency in the sequences we additionally apply segment matching between subsequent frames. Finding such correspondences between regions allows to handle the frequently occurring splits and merges of regions as it is illustrated in Figure 1. Since the global properties of segments change drastically if merges and splits happen, they cannot be used directly for obtaining reliable matches.

Similar to [12] we compare similarities between segments in subsequent frames only in the portions that are common to the segments. To be able to identify the common parts of the two segments we exploit the properties of image sequences, assuming that the shape of object outlines does not change significantly from one frame to the next. Therefore, we use a partial shape matching method to identify the common parts between two segments. The goal of partial shape matching in this context is to find all boundary fragments that match between the segments, i. e. possess a high shape similarity. Please note, that for an entire sequence we have to compare a lot of different segments, therefore very high efficiency is required in this step.

This requirement prevents the use of most of the state-of-the-art shape matching methods like [2, 15, 10] since they require several hundreds of millisec-

onds per match. For example, the top performing shape matching method [10] on the well known MPEG-7 shape retrieval data set requires half a second per match which would be too inefficient for our scenario.

This was also outlined in the closely related video object segmentation framework of Brendel and Todorovic [3]. They therefore proposed a cyclic dynamic time warping (CDTW) method for obtaining the partial matches with quadratic complexity. The authors state that their method, excluding the segmentation itself, requires several seconds to process a frame, with a matching time of $200ms$ per segment pair despite using a C implementation. This runtime still seems to be too slow for an application like image sequence segmentation.

For efficient shape matching we adapt a recently proposed method [7] for this purpose. This paper introduced a partial shape matching method which uses sampled contour points as underlying representation and solves an order preserving assignment problem. Integral images are used as efficient underlying data structure which enables partial matching within a few milliseconds. This approach was designed for similarity transformation invariant matching, which is not directly applicable in our scenario, since the invariance properties lead to frequent unnecessary confusions in segment matches.

In [7] each shape with N sampled contour points is described by an $N \times N$ descriptor matrix containing angles between chords connecting the sampled points. The invariance of angles to translation and rotation leads to the aforementioned properties. To make the approach sensitive to rotation (which is desired in our scenario), one has to find a novel $N \times N$ descriptor. The rest of the method stays exactly the same (see [7] for details).

We propose a novel descriptor for matching which uses angles between sampled points on the segment outline and a hypothetical horizontal line. Let a segment outline sampled with N points be denoted as $B = b_1, b_2, \ldots b_N$. Our descriptor Ω is an $N \times N$ matrix where every entry ω_{ij} is defined by the angle between a line connecting the points b_i and b_j and a hypothetical horizontal line through b_j. Thus, the descriptor matrix Ω consists of angles computed by

$$\omega_{ij} = \sphericalangle \left(\overline{b_i \, b_j} \, , \, \overline{b_j \, h_j} \right) \quad \forall i, j = 1, \ldots, N \, , \tag{1}$$

where h_j is the hypothetical horizontal line through b_j. For each segment these angles are calculated over all possible point combinations. The resulting descriptor matrix is non-symmetric, translation invariant and encodes the rotational orientation, which is strongly in contrast to the descriptor of [7]. Based on this descriptor, we can apply the same, highly efficient matching scheme with encoded orientation information.

We now apply this method to obtain the partial matches between the segments. We use equidistantly sampled points along the segment boundaries as underlying representation. Since we only want rough partial matches between segments, it is not necessary to consider all boundary points for matching as it is also illustrated in the experiments. Matching returns a set of correspondences, where it is possible that several, not necessarily connected, boundary fragments are returned as result. Using the provided correspondences we can estimate any

type of transformation (e.g. a thin-plate spline transformation) between the matched regions which allows to register the segments to each other. Because we have an ordered sequence of boundary points, we can easily identify the areas of the boundaries that are not matched and estimate a connected region by just drawing a line between the endpoints of the fragments. In such a way we get the matched parts between the segments for both frames t and $t+1$ and can now easily estimate accurate partial match costs between the segments, for example by analyzing the color similarity. This step is also illustrated in Figure 1.
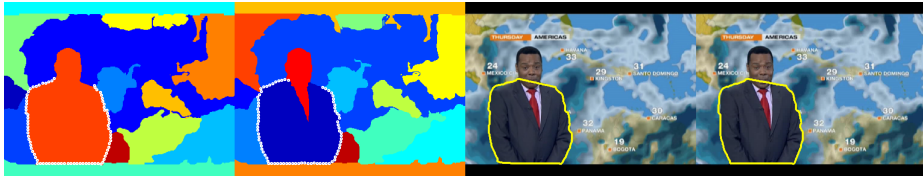

Fig. 1. First two columns show segmentation results in subsequent frames with occurring splits. Segments are matched by comparing shape, where matched parts are highlighted by white sampled points. Last two columns show partial match areas highlighted with yellow boundary, which are used to calculate the partial matching cost.

We use the obtained partial match cost to identify many to many correspondences between the segments in a graph matching scenario where we exploit unary (partial match cost) and pairwise (edge length) potentials.

Each segment is considered as a node of a graph, and neighboring regions are connected by an edge. Matching regions now equals to finding a binary assignment vector $\mathbf{x}^*$ of length $N_1 N_2$ where each entry x_{ia} should be one if segment i of the frame t matches to a segment a of the frame $t+1$ and N_1 and N_2 are the number of segments of frame t and $t+1$ respectively. The assignment vector is found in a quadratic assignment optimization procedure by maximizing

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}} \left(\mathbf{x}^T \mathbf{A} \, \mathbf{x} \right) = \underset{\mathbf{x}}{\operatorname{argmax}} \sum A_{ia,jb} \, x_{ia} \, x_{jb} \,, \qquad (2)$$

where A is a provided $N_1 N_2 \times N_1 N_2$ affinity matrix describing how well a pair of segments (i,j) in frame t agrees in terms of local descriptors and geometry with a pair of segments (a,b) in frame $t+1$. The affinity matrix A mainly contains the pairwise potentials and the unary potentials are placed in the main diagonal. Many different methods have been proposed to solve this NP-hard problem approximately, mostly by relaxing the discrete problem to the continuous domain and we apply a spectral method [14] for finding the principal eigenvector of the affinity matrix A to obtain the solution $\mathbf{x}^*$.

The most important step to obtain a reasonable matching result is the definition of an appropriate affinity matrix, containing the compatibilities of segments in subsequent frames. This segment compatibility is based on the identified partial matches between the segments. We define the entries of the affinity matrix

as a combination of unary potentials measuring the color similarity and pairwise potentials measuring the difference in edge lengths by

$$A_{ia,jb} = e^{-\left(w_1\,\epsilon(c_i,c_a) + w_2\,\epsilon(c_j,c_b) + w_3 \frac{d_{ij}-d_{ab}}{d_{ij}+d_{ab}} \right)} \,, \tag{3}$$

where c_i is a color descriptor of segment i, d_{ij} is the distance between the center points of the matched segments and w_1,w_2 and w_3 are manually selected weight parameters. We use n-dimensional color histograms as segment descriptor. Since the two segments to be compared are of approximately the same size after mapping, we do not need any normalization and can directly use a histogram intersection distance ϵ to define the unary potentials.

Graph matching returns a list of many to many correspondences and this list is used to merge and split segments in subsequent frames. Experiments prove that this step further improves segmentation consistency. Please note further, that the proposed graph matching method based on partial match costs is totally independent of the underlying segmentation method and can therefore improve any single image segmentation results.

4   Experiments

Experiments focus on demonstrating the improved consistency of our proposed method in comparison to the state-of-the-art. In Section 4.1 we first show the improved consistency by using our proposed image sequence segmentation methods in comparison to three state-of-the-art segmentation methods. In Section 4.2 we show that the proposed method is also applicable for tracking static and moving objects through sequences, based on a manual initialization of the object in the first frame.

In all experiments we used the same parameters as outlined in [8] for the total variation segmentation method TV-Seg. The distance between sampled points for partial shape matching was fixed to 10 and the graph matching weights w_1 to w_3 were all set to the same value 1/3. We implemented the proposed method in Matlab which enables segmenting an image in a few seconds independent of the type of initialization. All required frame-to-frame partial segment matches and the graph matching optimization together only require about $180ms$ per frame. Thus, the main computational bottleneck is the segmentation method, but it has to be pointed out that as e.g. shown by Pock et al. [16] it is possible to implement Total Variation methods on GPUs, thereby significantly reducing computation time.

4.1   Evaluating Consistency

Our first experiment shows the improved consistency of our proposed method in comparison to three state-of-the-art segmentation methods: Mean Shift [5], a graph based approach [9] and a saliency driven method [8]. We focused on single image methods because we are not aware of any publicly available approach for segmenting whole image sequences. Unfortunately, also for the two most related

Video / Method	GB [9]	MS [5]	Sal [8]	OurSeq	+ GM
Video News	70.23%	82.91%	84.96%	86.60%	**88.12%**
Video Flower Garden	57.64%	62.81%	60.51%	61.45%	**64.04%**
Video Daria Jack	54.52%	82.14%	92.66%	**96.88%**	**96.88%**
Video Dynamic Texture	52.50%	57.79%	76.83%	76.79%	**78.98%**
Video Cartoon	57.64%	72.59%	76.55%	81.07%	**82.06%**

Table 1. Comparison of image sequence consistency scores $f(S)$ on different test sequences. Best result per video is shown in bold.

methods [3, 12] no code is available. We selected five diverse image sequences for comparison: the well-known flower garden with a moving camera, a static camera news messenger sequence, a video from an action recognition data set, a dynamic texture example and a cartoon sequence.

For quantitative comparison of the obtained results we calculate a frame-to-frame consistency score $O(R, Q)$ by

$$O(R, Q) = \frac{1}{A} \sum_{R_a} \max_{Q_i} \frac{R_a \cap Q_i}{R_a \cup Q_i},$$ (4)

where R and Q are two segmentation results in subsequent frames, and R_a and Q_i are the corresponding segments, i. e. the consistency score is the mean overlap score between the segments, assuming R to be the reference segmentation. For a segmentation of an entire image sequence S it is possible to provide an overall score $f(S)$ by measuring the mean frame consistency score over the entire sequence. Of course, such an evaluation neglects the quality of the segmentations itself, for example always segmenting each image in one region would yield a perfect consistency score. Therefore, we parameterized all algorithms to provide approximately the same number of regions and as can be seen e. g. in Figure 2 all methods return reasonable segmentation results.

Table 1 summarizes the results on the five test sequences for Mean Shift (MS), Graph Based (GB), Saliency Segmentation (Sal), our extended segmentation method described in Section 2 (OurSeq) and the results for additionally activated graph matching (OurSeqGM) as described in Section 3. Results show that the saliency driven segmentation method [8] yields the most stable segments. Adding our initialization concept as described in Section 2 improves results on average about 2%. Additionally, activating graph matching as described in Section 3 provides another improvement of one percent. Although this quantitative improvement seems to be small, several wrong splits and merges are corrected, as it is also demonstrated in the tracking application presented in Section 4.2. Figure 2 furthermore shows selected frames from the news messenger sequence for the compared methods. Each segment is mapped to its mean RGB value to be able to identify wrong merges and splits easily. As can be seen the improved consistency using our proposed approach is visually much more appealing.
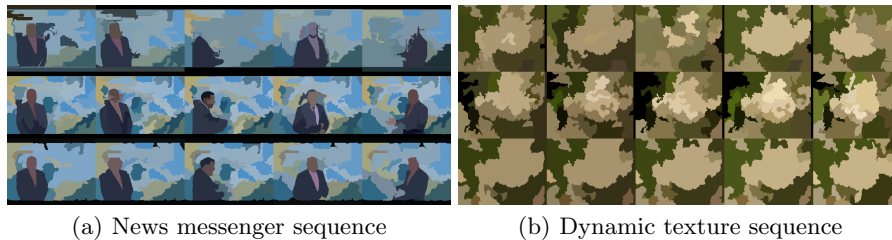
(a) News messenger sequence   (b) Dynamic texture sequence

Fig. 2. Direct comparison of image sequence segmentation results using Felzenszwalb's method [9] (first row), Mean Shift [5] (second row) and results of our proposed method (last row). Each obtained segment is mapped to its mean RGB color to be able to visually demonstrate improved consistency (best viewed in color).

Video / Method	GB [9]	+ GM	MS [5]	+ GM	OurSeq	+ GM
Video News	70.23%	70.99%	82.91%	83.72%	**86.60%**	**88.12%**
Video Flower Garden	57.64%	60.82%	**62.81%**	63.71%	61.45%	**64.04%**
Video Daria Jack	54.52%	56.04%	80.99%	82.96%	**96.88%**	**96.88%**
Video Dynamic Texture	52.50%	54.19%	57.79%	60.03%	**76.79%**	**78.98%**
Video Cartoon	57.64%	59.78%	72.59%	74.71%	**81.07%**	**82.06%**

Table 2. Comparison of image sequence consistency scores $f(S)$ on different test sequences. GM denotes activation of the proposed graph matching step which merges and splits several regions to improve consistency.

The graph matching approach presented in Section 3 is independent of the underlying segmentation. Therefore, we also made experiments concerning possible improvements using other underlying segmentations. These results are shown in Table 2 and again demonstrate improved consistency.

Since we only apply a frame-to-frame segment correspondence analysis, drifting might be an important problem. We do not notice severe drifting problems in our tested image sequences. A simple test to verify this visual insight is to apply our method also on the image sequence in reversed order (starting at the last frame) and compare the segmentation differences between the results for standard and reversed ordering. For quantitative evaluation we calculated the frame-to-frame consistency score as explained in Equation 4, this time comparing the segmentations of the same frame obtained in standard versus reversed ordering. On average we get a consistency score of 84.48%, which illustrates that no severe drifting effects take place.

4.2   Object Tracking

We further demonstrate the benefit of the proposed partial matching cost based graph matching step by using our method in a tracking scenario. We initialize a tracker by segmenting the first frame of the sequence and by manually merging regions that belong to the object-to-be-tracked.

Using this initialization as input, we apply the same methods as described in the previous sections, but this time constraining the graph matching to one-to-many matches. Our method finds the combination of segments in the subsequent frame that best fits to the initialization, which allows tracking the defined object throughout the sequence. To avoid major drifting effects we only update our reference segmentation when the graph matching step improves the frame-to-frame consistency.

Such a tracking approach enables quantitative evaluation of the segmentation accuracy by comparing obtained results to ground truth. We use the publicly available Weizmann database, which is commonly used to evaluate action recognition methods. This data set provides binary segmentations per frame for all videos. We selected three videos showing different actions and initialized our tracker in the first frame manually. On average, segmentation accuracy was 48.44% for Mean Shift [5], 53.37% for Felzenszwalb's method [9] and 70.12% for our proposed graph matching approach. The performance gain of approximately 20% comes from the fact that in single image based segmentation often the upper body part, the face and the legs are split which are correctly merged by our proposed method. In Figure 3 we directly compare results to the two most related image sequence segmentation methods of [12, 3] showing improved results due to our proposed graph matching verification. Further results and exemplary videos can be found at our homepage[1].
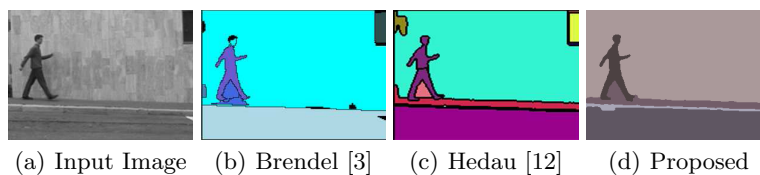

(a) Input Image   (b) Brendel [3]   (c) Hedau [12]   (d) Proposed

Fig. 3. Direct comparison on Weizmann sequence to most related methods of [12, 3].

5   Conclusion

This paper introduced an unsupervised method for segmenting image sequences. We first described a method obtaining highly consistent segments exploiting the similarities between subsequent frames. A second contribution showed that efficient partial shape matching exploiting a novel angle based descriptor allows finding similar parts between segments and the definition of a partial match cost. These costs are used to find correspondences between segments in subsequent frames in a pairwise graph matching step for handling the repeatedly occurring splits and merges of segments. This graph matching extension is formulated in a general way and can be applied to any available segmentation method, improving the overall sequence consistency. Experimental evaluation demonstrated the improved performance on diverse videos and an application for tracking manually initialized objects through sequences.

---

[1] `http://vh.icg.tugraz.at`

References

1. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: From contours to regions: An empirical evaluation. In: Proc. of Conf. on Comp. Vision and Pattern Recognition (CVPR) (2009)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. Trans. on Pattern Analysis and Machine Intelligence (PAMI) 24(4), 509–522 (2002)
3. Brendel, W., Todorovic, S.: Video object segmentation by tracking regions. In: Proc. of International Conf. on Comp. Vision (ICCV) (2009)
4. Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J.P., Osher, S.J.: Fast global minimization of the active contour/snake model. Journal of Mathematical Imaging and Vision 28(2), 151–167 (2007)
5. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. Trans. on Pattern Analysis and Machine Intelligence (PAMI) 24(5), 603–619 (2002)
6. DeMenthon, D., Megret, R.: Spatio-temporal segmentation of videos by hierachical mean shift analysis. In: Proc. of Conf. on Comp. Vision and Pattern Recognition (CVPR) (2002)
7. Donoser, M., Riemenschneider, H., Bischof, H.: Efficient part matching of outer contours. In: Proc. of Asian Conf. on Comp. Vision (ACCV) (2009)
8. Donoser, M., Urschler, M., Hirzer, M., Bischof, H.: Saliency driven total variation segmentation. In: Proc. of International Conf. on Comp. Vision (ICCV) (2009)
9. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. International Journal of Comp. Vision (IJCV) 59 (2004)
10. Felzenszwalb, P., Schwartz, J.D.: Hierarchical matching of deformable shapes. In: Proc. of Conf. on Comp. Vision and Pattern Recognition (CVPR) (2007)
11. Goldman, D.B., Gonterman, C., Curless, B., Salesin, D., Seitz., S.M.: Video object annotation, navigation, and composition. In: Proceedings of ACM symposium on User interface software and technology (UIST) (2008)
12. Hedau, V., Arora, H., Ahuja, N.: Matching images under unstable segmentations. In: Proc. of Conf. on Comp. Vision and Pattern Recognition (CVPR) (2008)
13. Huang, Y., Liu, Q., Metaxas, D.: Video object segmentation by hypergraph cut. In: Proc. of Conf. on Comp. Vision and Pattern Recognition (CVPR) (2009)
14. Leordeanu, M., Hebert, M.: A spectral technique for correspondence problems using pairwise constraints. In: Proc. of International Conf. on Comp. Vision (ICCV) (2005)
15. Ling, H., Jacobs, D.W.: Using the inner-distance for classification of articulated shapes. In: Proc. of Conf. on Comp. Vision and Pattern Recognition (CVPR). vol. 2, pp. 719–726 (2005)
16. Pock, T., Unger, M., Cremers, D., Bischof, H.: Fast and exact solution of total variation models on the gpu. In: CVPR Workshop on Visual Computer Vision on GPUs. pp. 1–8 (2008)
17. Shi, J., Malik, J.: Motion segmentation and tracking using normalized cuts. In: Proc. of International Conf. on Comp. Vision (ICCV) (1998)
18. Wills, J., Agarwal, S., Belongie, S.: What went where. In: Proc. of Conf. on Comp. Vision and Pattern Recognition (CVPR) (2003)