Optimized Mean Shift Algorithm for Color Segmentation in Image Sequences

Werner Bailer[a*], Peter Schallauer[a], Harald Bergur Haraldsson[b], Herwig Rehatschek[a]

[a]JOANNEUM RESEARCH, Institute of Information Systems and Information Management, Steyrergasse 17, 8010 Graz, Austria

[b]Tokyo Institute of Technology, Precision and Intelligence Laboratory, R2-13, 4259 Nagatsuta, Midori-ku, Yokohama, 226-8503 Japan

ABSTRACT

The application of the mean shift algorithm to color image segmentation has been proposed in 1997 by Comaniciu and Meer. We apply the mean shift color segmentation to image sequences, as the first step of a moving object segmentation algorithm. Previous work has shown that it is well suited for this task, because it provides better temporal stability of the segmentation result than other approaches. The drawback is higher computational cost.

For speed up of processing on image sequences we exploit the fact that subsequent frames are similar and use the cluster centers of previous frames as initial estimates, which also enhances spatial segmentation continuity. In contrast to other implementations we use the originally proposed CIE LUV color space to ensure high quality segmentation results. We show that moderate quantization of the input data before conversion to CIE LUV has little influence on the segmentation quality but results in significant speed up. We also propose changes in the post-processing step to increase the temporal stability of border pixels.

We perform objective evaluation of the segmentation results to compare the original algorithm with our modified version. We show that our optimized algorithm reduces processing time and increases the temporal stability of the segmentation.

Keywords: segmentation, mean shift, image sequence, objective evaluation

1. INTRODUCTION

The application of the mean shift algorithm to color image segmentation has been proposed in 1997 by Comaniciu and Meer [2]. Since then it has become a widely used method for color image segmentation, as it provides significantly better segmentation results as other approaches (e.g. the Watershed algorithm), for example less oversegmentation and robustness against illumination changes.

We apply the mean shift color segmentation to image sequences, as the first step in a moving object segmentation algorithm. Previous work has shown that the mean shift algorithm is well suited for segmentation of image sequences, because it provides better temporal stability of the segmentation result than other approaches (cf. for example [6]). However, the drawback is that the mean shift algorithm is computationally more expensive, and computational costs are even more an issue when the algorithm is applied to image sequences. This means that there is need for optimization in order to use the algorithm in applications where performance is an issue.

One obvious optimization is to exploit the temporal redundancy in the image data when applying the algorithm to image sequences. It can be expected that the algorithm converges faster. This optimization has already been mentioned in [8] and [9]. Other researchers advocated the use of another color space than CIE LUV to reduce the complexity of the algorithm, for example, YUV [9] or HSV [10] have been proposed. Another reported optimization avoids applying the mean shift procedure to pixels that lie on the mean shift trajectory of other similar pixels which have already been processed [1]. The hierarchical mean shift proposed in [5] has been proposed to reduce complexity in spatio-temporal

* werner.bailer@joanneum.at; phone +43 316 876-1218; fax +43 316 876-1191; http://iis.joanneum.at

video object segmentation. Recently, the adaptive mean shift [7] has been proposed, which is optimized for the application of the mean shift algorithm to high dimensional feature spaces.

In this work, we study approaches to performance optimization that do not impair the quality of the segmentation result. One such idea is the propagation of cluster centers, which also has some related issues, such as the treatment of abrupt scene changes (e.g. cuts and fast camera motion). As the runtime of the algorithm depends on the population of the feature space, we propose pruning the feature space by quantizing the input data as a novel optimization approach.

We perform objective evaluation of the segmentation quality of our proposed optimized algorithm to show the effects of these propagating cluster centers and quantization.

2. PROPOSED OPTIMIZATIONS

Mean shift is a technique for analysis of feature spaces. When used for color image segmentation, the image data is mapped into the feature space, resulting in a cluster pattern. The clusters correspond to significant features in the image, namely dominant colors. Using the mean shift procedure, we can locate these clusters and therefore extract the dominant colors from the image to use for segmentation.

The clusters are located by applying a search window in the feature space, which shifts towards cluster center. The magnitude and direction of the shift in feature space are based on the difference of the center of the window and the local mean value inside the window. When the magnitude of the shift becomes small (according to a threshold) the center of the search window is declared as a cluster center and the algorithm is said to have converged for one cluster. This procedure is repeated until all significant clusters have been extracted.

2.1. Propagation of Cluster Centers

The number of shifts needed to locate a cluster center in feature space depends on where the search begins. Finding the best starting location for the search window in feature space is critical to reduce the number of shifts. When applying the mean shift algorithm to color segmentation of a single image, several randomly chosen locations in feature space are considered and the one with the highest density of feature vectors is selected. This is done to make sure that the search starts in a high-density region and thus reduces the number of shifts needed to reach convergence.

When applying the mean shift algorithm to image sequences, we exploit the temporal redundancy in subsequent frames by using the cluster centers of previous frames as initial estimates. This approach significantly reduces the number of iterations until the algorithm converges and thus speeds up processing on image sequences, since the location of cluster centers in the previous frame corresponds to a high-density region in the current frame, under the assumption of similarity of color content between subsequent frames.

When the assumption of similarity of subsequent frames does not hold, e.g. due to abrupt scene changes, using the previous cluster centers is not an optimal way to estimate the starting position in feature space, and can result in impairment of segmentation quality. This is due to the fact that in cases of very quickly varying content, locations of previous clusters can correspond to low-density regions in feature space of the current frame, causing the algorithm to require more shifts to converge, or converging in a low-density area where the located cluster does not correspond to a significant feature of the input data. The algorithm checks when this occurs by comparing the number of feature vectors contained in the recently found cluster to a threshold. When this comparison implies that the procedure converged at a region with unsatisfactory density of feature vectors, the cluster is not accepted and a new search begins at a location determined by using the conventional method for single image segmentation as described in [2]. This procedure is done for every cluster collected from previous frame. Some of the clusters may still correspond to a dominant color in the current frame (e.g. background color may be the same) even though propagating other clusters is not applicable

2.2. Color Space Quantization

To ensure high quality segmentation results we use the CIE LUV color space, which has been suggested as an appropriate feature space for color segmentation applications because of its isotropic properties [2]. In most cases, the input data is given in the traditional RGB color space, which is related to CIE LUV by nonlinear conversion.

Moderate quantization of the input data before conversion to the CIE LUV color space does not have significant effects on the quality of segmentation. When the quantized data is mapped to feature space, the resulting sparse feature space contains similar cluster pattern compared to the full resolution feature space. Nearby feature vectors have actually been grouped into one coordinate due to the quantization; most of these feature vectors grouped together would likely have been contained in the same cluster after cluster estimation in a feature space originating from unquantized input data.

The quantization of input data significantly reduces the runtime of the algorithm, since computations that require iteration through distinctive set of data, both in image domain and feature space, have now considerably less data to go through. When computing the mean shift, i.e. the translation of the search window in feature space, there are fewer distinctive feature vectors inside the search window, and since the algorithm needs to access all the data inside the window, there is a performance gain in using this sparser feature space. The high-density regions appear at similar locations as in the full resolution feature space, and therefore the mean shift computation delivers comparable results, with a similar number of shifts and estimated clusters.

2.3. Post-processing Steps

In the original mean shift paper, post-processing steps for removing small regions – which very often lie along the borders of larger regions – from the segmentation result are described.

We propose the following optimizations of these post-processing steps. Starting with the smallest existing region, the contour of the region is recursively merged to the appropriate neighboring region until all pixels of the region have been moved to other regions. This process is repeated for all remaining small regions until they have been successively merged to larger regions. Opposed to scanline approach in the original implementation, this method is independent of region orientation and will therefore improve temporal stability at the region borders.

During post-processing, some pixels have to be reassigned to the cluster which is closest to the pixel color in feature space. In the proposed optimization, this computation is performed for each existing color of input data, instead of each pixel, avoiding repetition of computations, since in natural images more than one pixel will usually share the same color value. This is especially efficient when color quantization is used and the number of distinctive colors is reduced significantly below the total number of pixels in the image.

3. RESULTS

3.1. Evaluation Methodology

We show the results of the proposed optimizations by comparing both the runtime performance and the segmentation quality of the original and the proposed algorithm. As test material four well known image sequences (coastguard, flower garden, foreman and tennis) have been used, all of them in SIF resolution.

The performance is compared in terms of the runtime of the algorithms on a defined system and by comparing the number of mean shift iterations required until convergence of the algorithm.

For objectively comparing the quality of the resulting color segmentations, we use some of the metrics proposed in [4]. We perform standalone evaluation, i.e. the results are not compared against a ground truth, but the results of applying the metrics to the different segmentations are compared. We have selected a subset of the metrics proposed in [4], as not all of them are applicable to a segmentation only based on color. For example, we cannot the boundaries in the motion vector field to coincide with the region boundaries of the color segmentation. We have thus selected the metrics listed below. Most of them are calculated from single regions, so in the results given below we state the mean of the calculated metric over all regions of a frame.

3.1.1. Spatial Metrics

Number of regions: The number of resulting regions in the segmentation result. To few regions may cause loss of details, too many regions indicate over-segmentation.

Sequence	**Coastguard**	**Flower Garden**	**Foreman**	**Tennis**
Original	3.82	7.33	3.84	5.02
Proposed	2.71	3.49	2.87	3.89
Reduction (%)	29.05	52.39	25.26	22.51

Table 1: Average number of mean shift iterations per cluster using the original mean shift algorithm and the proposed optimizations.

Compactness and Circularity: Both relate the perimeter of a region to its area and are thus measures of the complexity of the region shape.

Elongation: The region area related to the number of morphological erosion steps until a region disappears.

Spatial perceptual information (SI): A measure of the edge energy inside a region.

Texture variance: The intensity and color variance inside a region.

3.1.2. Temporal Metrics

The temporal metrics are calculated between corresponding regions in subsequent frames. The corresponding region in frame t+1 to a region in frame t is determined as the one with the most similar mean color. All of the temporal metrics described below thus evaluate the temporal stability of the segmentation, as segmentation errors, that will lead to a change in the mean color (such as merging with a neighbor region) will result in higher differences of these metrics.

Difference in elongation: Change of the elongation metric in subsequent frames.

Difference in size: Change of the region size in subsequent frames.

3.2. Propagation of Cluster Centers

To evaluate the effects of the propagation of cluster centers, we have applied the original algorithm and the proposed one on the same image sequences. We compare the number of mean shift iterations and the segmentation quality.

Table 1 shows the average number of mean shift iterations per cluster on the four test sequences. In all cases the propagation of the cluster centers reduces the number iterations needed, up to more than 50%. As can be expected, the performance gain depends on the visual activity in the image sequence. For example, among the four test sequences, the highest performance gain is achieved in the flower garden sequence, where the color content of the background and the objects remains stable throughout the sequence. Nonetheless a performance gain of around 25% can also be achieved in the other sequences. These results show that the value given in [8] can be reached if the visual content changes only moderately over time.

The evaluation of segmentation quality (cf. Figure 1 and Figure 2) shows that our proposed algorithm when applied to image sequences produces fewer regions than the original one. As a consequence, the resulting regions have more complex shape. This is especially expressed by the circularity measure, which is lower for our algorithm. It can be seen that curves are similar, but there is an offset of about 0.2 for all four sequences. The compactness measure supports this observation, as it is a bit higher throughout the sequences, especially in parts with lower visual activity.

The SI and texture variance are measure of the uniformity within the segmented regions in terms of color. The results show that although the regions have become more complex, the SI and texture variance are equal or only slightly higher. Only in sections just after high visual activity (such as fast pans) both values are higher than that of the original version.

The temporal measures show a higher temporal stability of our algorithm. The mean elongation difference is generally lower. In parts with lower visual activity it is significantly more stable than that of the original algorithm. The mean difference of region sizes is similar in sequences with more visual activity, but like the elongation difference, it turns out to be more stable in parts with continuous and slowly changing motion, such as in the flower garden sequence.

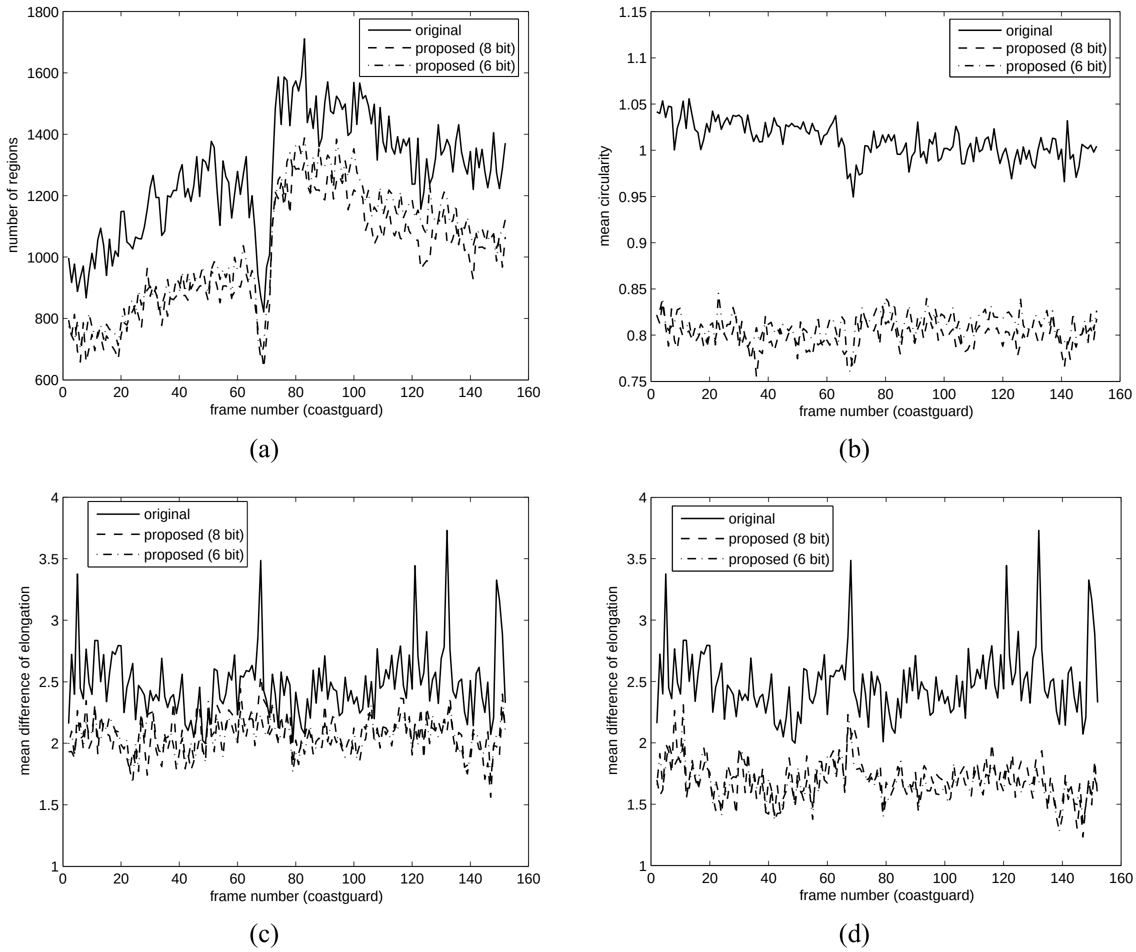


Figure 1: Comparison of segmentation quality measures between original and proposed algorithm with different quantization. First 150 frames of Coastguard sequence, the segmentation in (d) is with only minimal post-processing.

The question is how much of the observed changes in the segmentation quality measures are due to the modifications in post-processing. In fact, when we look at the results of our algorithm with only a simple post-processing step of removing regions of size 1 and 2 pixels, we observe a significantly higher number of regions and lower values for compactness and elongation and higher circularity. SI and texture variance are also lower, but to a much lesser extent. The temporal measures show the same trend than those of the results with post processing (cf. Figure 1(d) and Figure 2(d)): The differences in elongation and size are generally lower, the more when the visual activity in the content is only moderate. This proves that the gain in temporal stability is mainly due to the propagation of cluster centers and partly due to improved post-processing.

3.3. Color Space Quantization

We assume that moderate color quantization of the input data will not impair the segmentation quality, but speed up calculation. To evaluate this assumption, we compare the runtime and the results of our modified mean shift algorithm with quantization 4, 6 and 8 bit (i.e. no quantization).

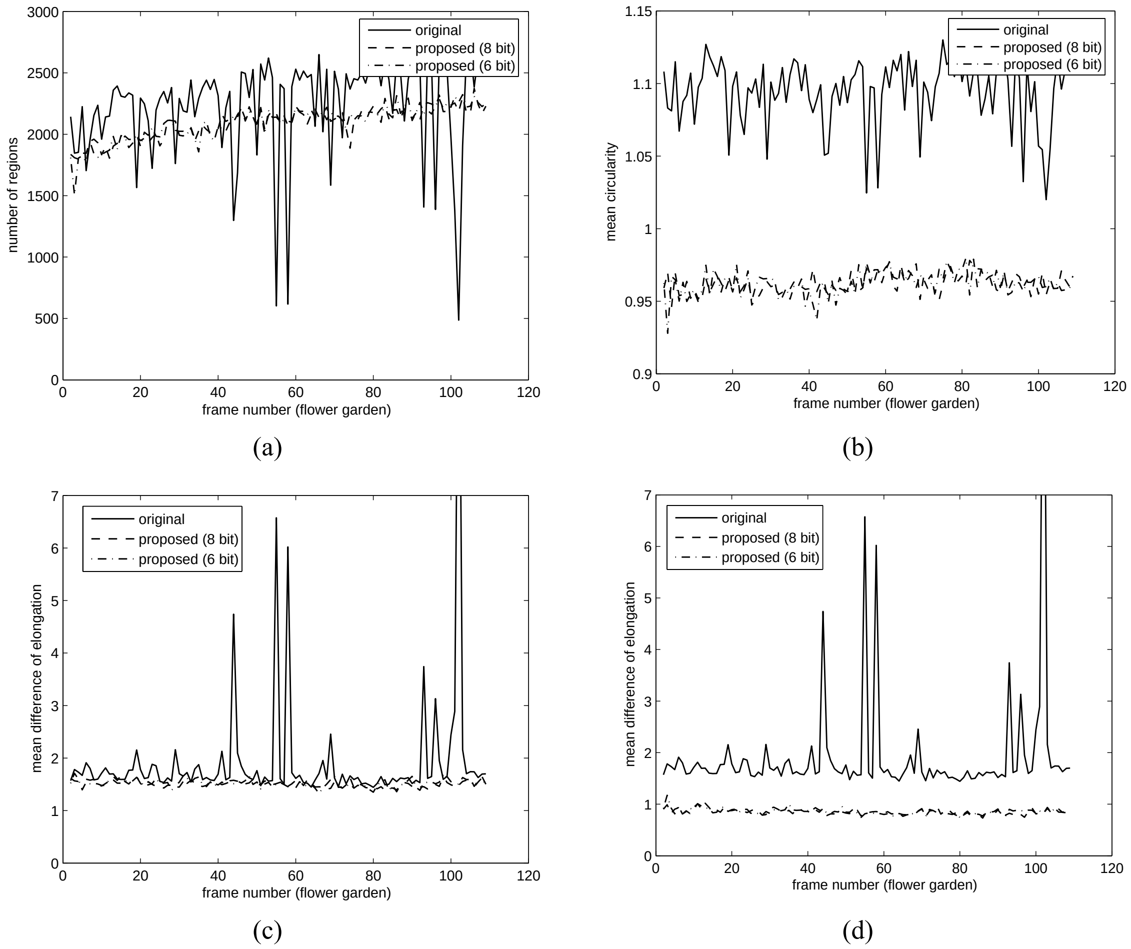


Figure 2: Comparison of segmentation quality measures between original and proposed algorithm with different quantization. Sequence Flower Garden, the segmentation in (d) is with only minimal post-processing.

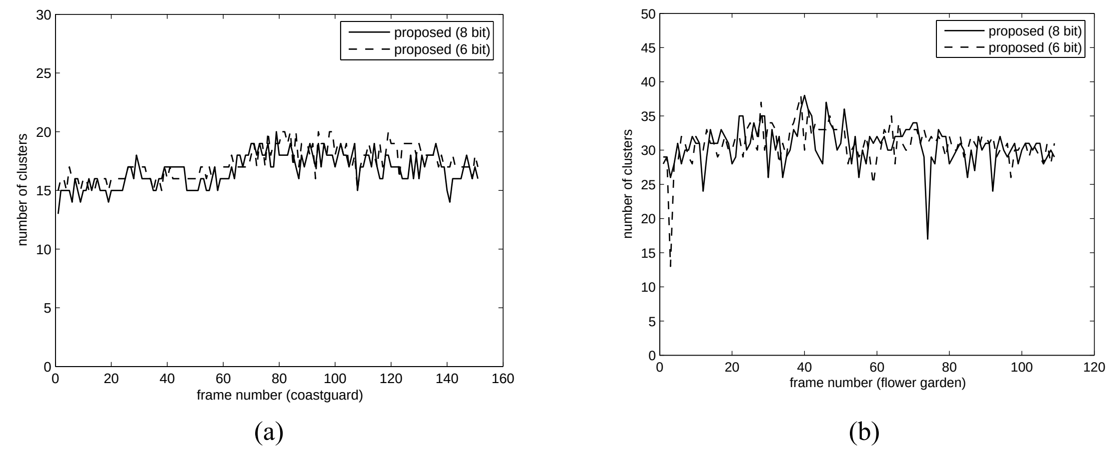


Figure 3: Number of clusters per frame with different quantization: (a) first 150 frames of the Coastguard sequence, (b) Flower Garden sequence.

Sequence	**Coastguard**	**Flower Garden**	**Foreman**	**Tennis**
Acc. nr. clusters (8)	4891	3340	6049	1906
Acc. nr. clusters (6)	5055	3372	6111	1931
Performance Gain	35.62%	26.91%	34.05%	43.39%

Table 2: Accumulated number of clusters throughout the sequence using different quantization and performance gain when using 6 bit quantization of the input data instead of 8 bit quantization.

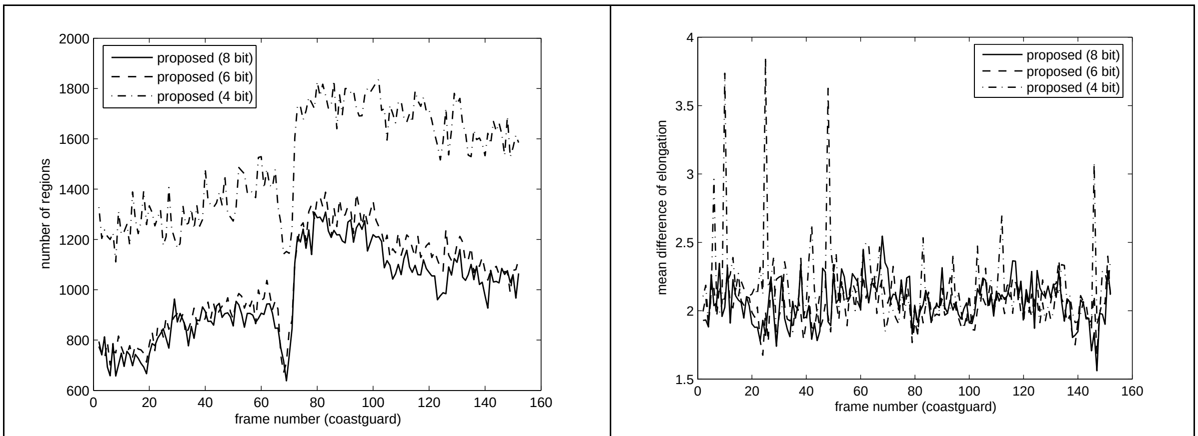
Table 2 shows the gain in runtime performance of the algorithm when 6 bit quantization is used compared to using unquantized input data (i.e. 8 bit). The number of shift operations needed is nearly identical and thus independent of the quantization of the input data. The number of clusters differs slightly in single frames (Figure 3), but the accumulated number of clusters over time is similar, as can be seen from Table 2. The performance gain results from reduced pre-processing time, i.e. building the feature space, and from faster calculation of the shift direction due to the sparser feature space.

The segmentation quality measures show that the resulting segmentations are very similar when 6 or 8 bit quantization is used. The number of regions is in some cases slightly higher when 6 bit quantization is used and we observe a drastic increase in the number of regions when using 4 bit quantization (cf. Figure 4). The reason is that regions with similar color may be broken apart when quantization is applied. The other spatial and temporal measures used do not show significant differences between 6 and 8 bit quantization.

4. CONCLUSION

The mean shift algorithm is well suited for color segmentation of image sequences, as it provides better results than comparable algorithms. The drawback is higher computation cost which demand for optimization of the algorithm.

In this work we have proposed an optimized mean shift algorithm for color segmentation of image sequences. We follow the idea of using the resulting cluster centers of a frame as initialization for the subsequent frame, in order to exploit the temporal redundancy and reduce computation time by 20-50%. The originally proposed CIE LUV color space has been replaced in other works for performance reason. We propose to use this color space and to moderately quantize the input data to prune the feature space to be segmented.



(a) (b)

Figure 4: Comparison of effects of different quantization of input data on the segmentation result of the first 150 frames of the Coastguard sequence: (a) number of regions, (b) mean difference of elongation in subsequent frames.

Using objective evaluation of segmentation quality we have shown that our proposed algorithm produces less and thus more complexly shaped regions. The temporal stability of our algorithm is higher, especially in scenes with fairly low visual activity. We have also shown that moderate quantization of the input data has no significant effect on the quality of the resulting segmentation, but speeds up computation by over 40%.

Recently, extensions to the mean shift algorithm have been proposed in order to make it efficiently usable on higher dimensional feature spaces. It has to be evaluated if the optimizations proposed in this work can also be applied to segmentation of image sequences using higher dimensional feature spaces, e.g. texture segmentation or joint color and texture segmentation of image sequences.

ACKNOWLEDGEMENTS

The work described in this paper has been done during the time Harald Bergur Haraldsson was working at the Institute of Information Systems at JOANNEUM RESEARCH. The authors would like to thank Dorin Comaniciu from Rutgers University for publishing the code of the original mean shift implementation, as well as several colleagues at JOANNEUM RESEARCH who supported this work, especially Michael Donoser for adapting the original implementation, and Roland Mörzinger for helping with the evaluation. This work has been funded partially under the 5th Framework Programme of the European Union within the IST project "MECiTV" (IST-2001-37330, http://www.meci.tv).

REFERENCES

1. C. M. Christoudias, B. Georgescu and P. Meer, "Synergism in Low Level Vision", *Proc. 16th International Conference of Pattern Recognition*, Quebec City, Canada, August 2001.
2. D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation", Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, June 1997, pp. 750-755.
3. D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, nr. 5, May 2002, pp. 603–619.
4. P. L. Correia and F. Pereira, "Objective Evaluation of Video Segmentation Quality", *IEEE Trans. Image Proc.*, vol. 12, nr. 2, Feb. 2004, pp. 186–200.
5. D. DeMenthon and R. Megret, "Spatio-temporal Segmentation of Video by Hierarchical Mean Shift Analysis", Technical Report LAMP-TR-090, Univ. Maryland, 2002.
6. M. Donoser: *Object Segmentation in Film and Video*, Diploma Thesis, Technische Universität Graz, 2003.
7. B. Georgescu, I. Shimshoni and P. Meer, "Mean shift based clustering in high dimensions: a texture classification example", *Proc. Ninth IEEE International Conference on Computer Vision*, Oct. 2003, pp. 456–463.
8. J. Guo and C.-C. Jay Kuo, *Semantic Video Object Segmentation for Content-Based Multimedia Applications*, Kluwer Academic Publishers, 2001.
9. J. Guo, J. Kim and C.-C. Jay Kuo, "Fast Video Object Segmentation Using Affine Motion and Gradient-based Color Clustering", IEEE Second Workshop on Multimedia Signal Processing, 1998.
10. H. Wang and D. Suter, "Color Image Segmentation Using Global Information and Local Homogeneity", *Proc. VIIth Conf. Digital Image Computing: Techniques and Applications*, Sydney, Dec. 2003.