

# Unsupervised Mesh based Segmentation of Moving Objects

Andreas Kriechbaum  
JOANNEUM RESEARCH  
Institute of Information Systems & Information Management  
Steyrergasse 17  
8010 Graz, Austria  
+43 316 876 1241  
[andreas.kriechbaum@joanneum.at](mailto:andreas.kriechbaum@joanneum.at)

Roland Mörzinger  
JOANNEUM RESEARCH  
Institute of Information Systems & Information Management  
Steyrergasse 17  
8010 Graz, Austria  
+43 316 876 1125  
[roland.moerzinger@joanneum.at](mailto:roland.moerzinger@joanneum.at)

Georg Thallinger  
JOANNEUM RESEARCH  
Institute of Information Systems & Information Management  
Steyrergasse 17  
8010 Graz, Austria  
+43 316 876 1240  
[georg.thallinger@joanneum.at](mailto:georg.thallinger@joanneum.at)

## ABSTRACT

Multimedia analysis usually deals with a large amount of video data with a significant number of moving objects. Often it is necessary to reduce the amount of data and to represent the video in terms of moving objects and events. Event analysis can be built on the detection of moving objects.

In order to automatically process a variety of video content in different domain, largely unsupervised moving object segmentation algorithms are needed. We propose a fully unsupervised system for moving object segmentation that does not require any restriction on the video content. Our approach to extract moving objects relies on a mesh-based combination of results from colour segmentation (Mean Shift) and motion segmentation by feature point tracking (KLT tracker).

The proposed algorithm has been evaluated using precision and recall measures for comparing moving objects and their colour segmented regions with manually labelled ground truth data. Results show that the algorithm is comparable to other state-of-the-art algorithms.

The extracted information is used in a search and retrieval tool. For that purpose a moving object representation in MPEG-7 is implemented. It facilitates high performance indexing and retrieval of moving objects and events in large video databases, such as the search for similar moving objects occurring in a certain period.

## Categories and Subject Descriptors

I.4.6 [Image Processing and Computer Vision]: Segmentation - *Edge and feature detection, Region growing, partitioning*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AREA '08, October 31, 2008, Vancouver, British Columbia, Canada.  
Copyright 2008 ACM 978-1-60558-318-1/08/10...\$5.00.

I.4.8 [Image Processing and Computer Vision]: Scene Analysis - *Color, Motion, Shape, Tracking*.

H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing - *Indexing methods*.

## General Terms

Algorithms, Performance, Experimentation.

## Keywords

Event Detection, Moving Object Segmentation, Mean Shift, Tracking, MPEG-7, Unsupervised System, Occlusion, Spatial Segmentation, Motion Extraction, Clustering, Optical Flow, displaced frame difference.

## 1. INTRODUCTION

A critical task in event detection is the interpretation of semantically meaningful spatio-temporal objects in a large amount of video data. To achieve this task, the gap between pixel values and symbolic event description needs to be bridged. The successful application of object-based media description and representation depends largely on effective object segmentation tools.

Moving Object Segmentation (MOS) can be used for providing important spatio-temporal information about objects whose motion is more or less homogeneously at least over a certain period. The definition of moving object segmentation is the task of partitioning a block of frames based on the coherent motion of different objects in a dynamic scene [2]. Moving objects can be for example cars, planes or persons. A moving object contains local connected neighbour regions with similar velocities, which are expected to belong to one real object, e.g. a plane, car or a person.

Traditional approaches are trained to specific environments and fail to operate when applied to general video sources in an unsupervised way.

The challenge of an automatic, unsupervised system is to extract moving objects without any restriction on the content

and without any manual intervention in the moving object segmentation process.

Specific challenges in moving object segmentation are:

- Moving camera:, MOs move relative to the moving background
- Illumination variance: Illumination changes over short time period
- Shadows: Shadows are cast on the MO or by the MO
- Background clutter: Segmentation from background, e. g. swaying branches
- Multi object tracking: Objects move side by side and similarly
- Occlusions: Appearance, disappearance, objects move in front of others
- Computational complexity: Increases with scene dynamics, e.g. number of MOs
- Object behaviour: rigid/non-rigid objects, fast/slow moving objects

The aim of this work is to explore the feasibility of a fully automatic moving object segmentation system which tackles the above mentioned problems.

The rest of the paper is organized as follows: In Section 2 an overview of related work is given and Section 3 explains our algorithms for moving object segmentations in detail. Evaluation of experiments and results are presented in Section 4, and conclusions are presented in Section 5.

## 2. RELATED WORK

In literature different ways for classifying moving object segmentation approaches are discussed. A review of state-of-the-art techniques is presented in [27].

Generally, the algorithms can be divided into three main groups of moving object segmentation techniques. The following sections describe representative algorithms shortly.

### 2.1 Spatial Segmentation, Motion Extraction, Clustering

A Mean-Shift based algorithm proposed in [15] provides robust homogeneous colour regions according to dominant colours. Furthermore, frame intensity difference based motion detection is applied for motion extraction. The detected moving regions are analyzed by a region-based affine model and further tracked to increase the consistency of the extracted objects. A morphological open-close operator is used to remove gulfs and isthmi (narrow connection between two large regions) for object boundary smoothing. A shape coding optimization is done using boundaries of variable width. The algorithm is fast and highly accurate.

In [13] a two-dimensional feature vector is used for clustering in the feature space. The first feature is image brightness which reveals the structure of interest in the image. The second feature is the Euclidean norm of the optical flow vector. The optical flow field is computed using the Horn-Schunck algorithm [17]. By clustering the feature space, moving objects in the image are

detected. The algorithm has the advantage that it is robust regarding to background movement.

In [18] the moving object segmentation procedure is treated as a Markovian labelling process and is formulated by a hierarchical Markov random field (MRF) model. At the beginning of the process, a watershed algorithm is used. Afterwards foreground information is detected as an outlier of the estimated background motion in the initial motion classification stage. After that, the motion vector is estimated for each foreground region and is validated by an elaborate occlusion detection scheme. The initial object mask is segmented by the MRF model. A disadvantage of the approach is that it cannot deal properly with noise.

The PCA and GGM based algorithm proposed in [19] consists of three stages: the initial segmentation of the first frame using colour, motion, and position information, based on a variant of the K-Means-with-connectivity-constraint algorithm. Then a temporal tracking algorithm is applied, using a Bayes classifier and rule-based processing to reassign changed pixels to existing regions and to handle the introduction of new regions. Finally a trajectory-based region merging procedure is used that employs the long-term trajectory of regions to group them to objects with different motion. It is advantageous that the algorithm can handle fast moving of objects, new objects and disappearing objects.

### 2.2 Point Tracking, 2D Mesh

At the first frame of the video, an optimal number of feature points are selected as nodes of a 2D content-based mesh. These points are classified as moving (foreground) and stationary nodes-based on multi-frame node motion analysis described in [7], yielding a coarse estimate of the foreground object boundary. To extract the moving object, colour differences across triangles near the coarse boundary are employed. The boundary of the video object is refined by the maximum contrast path search along the edges of the 2D mesh. Next the refined boundary to the subsequent frame is propagated by using motion vectors of the node points to form the coarse boundary at the next frame.

This algorithm is able to detect occlusions but small objects cannot be found. The point tracker-based approach for moving object segmentation is a possible approach for a fully automatic moving object segmentation and similar to the approach developed by the authors.

### 2.3 Active Contours

The latest algorithms are often based on active contours. This is an iterative algorithm that produces a better contour-line description in every iteration step. It receives the previous contour line as input and uses some balancing constant factors (internal and external energy) to produce a new contour line description. Active contours (snakes) minimize the sum of the internal and external energy [24]. The graph cut algorithm [26] is an improvement of active contours, leading to a smooth contour free of self-crossing and uneven spacing problems. The internal force, which is used in the energy functions to control the smoothness, is no longer needed and the number of parameters is reduced.

An advantage of these algorithms is the ability to handle unknown noise, highly textured background, and partial object occlusions. The disadvantage of active contours based algorithms is the need for initial object segmentation or the requirement of initial seeds. Due to

the need for initial object segmentation this algorithm is often used for video object tracking and it is not considered in this work.

For our work we have selected and combined the most promising techniques in order to build a fully automatic system. A combination of spatial (Mean Shift Algorithm [8]) and temporal (KLT tracker [20]) information has been selected to be a good starting point for the automatic segmentation technique.

### 3. MESH-BASED MOVING OBJECT SEGMENTATION

The main goal of this work is to segment moving regions in videos without any restrictions on the content (like static or moving cameras). The development of the mesh based moving object segmentation algorithm should solve the problems in moving object segmentation.

Many approaches from literature suggest using dense motion fields or an optical flow field in which the video objects have to be segmented. However, the problem is that the method for optical flow field calculation for every pixel of the image is underspecified and that computing a velocity vector for every pixel of the images is often redundant because most pixels in an image have zero motion. In the proposed algorithm a similar, yet modified approach is chosen.

In general, the developed approach is based on the combination of pre-extracted features. These are feature points of a KLT feature tracker and regions extracted by colour segmentation. The used colour segmentation algorithm was developed using the specific Mean-Shift implementation described in [10]. There it was shown that the Mean-Shift algorithm is well suited for colour segmentation of image sequences, due to better results regarding the temporal stability of the segmentation compared to other approaches (for example Watershed algorithm). A drawback is that the Mean-Shift algorithm is computationally more expensive and the computational costs are even more an issue when the algorithm is applied to image sequences (videos). To overcome this several optimizations are proposed in [4].

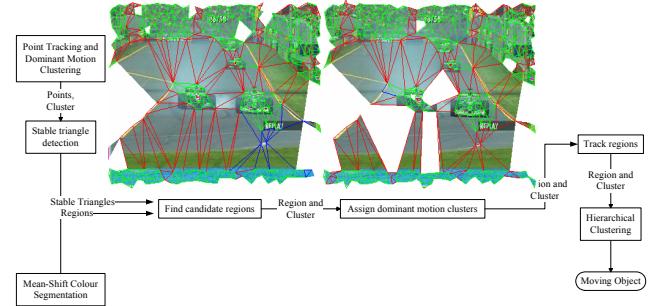
The MS produces a base-segmentation which is combined with the motion information generated by the KLT-tracker to get a moving object. The Mean-Shift algorithm provides the dominant motion cluster assignment as a new property of the LKT feature points. Using the ability of the MS algorithm to perform clustering in high dimensions, motion clusters are extracted. The inputs to the MS algorithm were motion-vectors and the pixel-distance between each feature point.

The moving object is extracted by combining the colour-segmented regions by dominant motion clusters. Details of the developed approach are described in the next sections.

#### 3.1 Mesh-based Moving Object Segmentation

The motivation of the mesh-based segmentation approach is the suggestion of optical flow based algorithms in the literature and the point tracker-based and mesh-based approach for moving object segmentation in [7]. In the developed approach the feature points are assigned to dominant clusters which are representing moving objects. Furthermore, this algorithm is based on the assumption that a colour-segmented region belongs to a single object. This object is either a foreground object (moving object)

or background object. This assignment is done similar to [7]. A workflow can be seen in Figure 1.

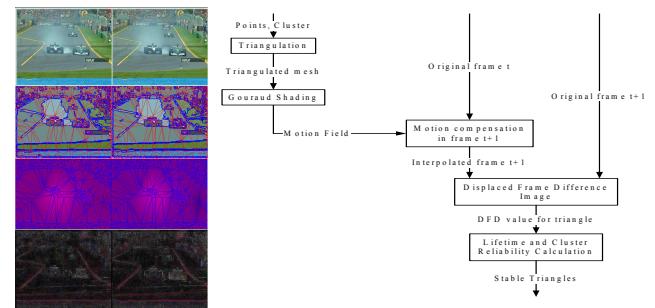


**Figure 1: Flow diagram in the mesh-based moving object segmentation.**

In this approach velocity stable triangles (build of the extracted feature points) which belong to one cluster are combined to represent a moving object. Triangles which are on the same object are from the same motion cluster. If not all points from the triangle are assigned to the same cluster the triangle is not used for the moving object segmentation. Triangles with points from the same cluster are assigned to colour segmented regions and so a stable skeleton of the object is formed. The stability of the triangles is determined by the interpolated motion field calculated for the feature points. The extracted moving object is reliable through the combination of the two base approaches (colour segmentation, feature point tracking) and further introduced quality measurements based on the motion field. The combination of these algorithms solves the problems of MOS with moving cameras. With this algorithm occlusion is not a problem; appearance and disappearance from objects in the image do not have any negative impacts to the algorithm.

#### 3.2 Stable Triangle Detection

This processing step is one of the most important steps in the mesh-based algorithm which leads to the desired quality. In this step reliable triangles are selected from all triangles for a further region assignment to the according dominant cluster. The feature points extracted by the KLT feature tracker are triangulated resulting in a mesh. The mesh extraction facilitates a fast computation of a dense optical flow field. This process is called motion estimation. The diagram in Figure 2 shows the process to make this clearer.



**Figure 2: Stable Triangle Detection with Motion interpolation.**

First, Delaunay's triangulation method [9] is applied to the extracted feature points containing their local motion information.

Several tests have shown that Delaunay's algorithm is very fast and has a high stability (linkage) of the extracted mesh over time which is a necessary precondition in the mesh-based moving object algorithm.

Second, a dense motion field is extracted. The motion field is calculated by using the Gouraud shading algorithm [16] on the velocity vectors of the extracted mesh. The assumption is that KLT feature points from the mesh which are assigned to one dominant motion cluster belong to a moving object. If these points which are on the moving objects have a correct assigned velocity vector and if there are many points inside the border of the moving object the Gouraud shading algorithm will provide a linear interpolated motion field of the image. The linear interpolation between the motion vectors is valid because of the minor error in the area of the moving objects and a larger error in the neighbourhood of the moving object. If too few points in the area of the moving object the algorithm provides worse results. Furthermore, if the points extracted by the KLT tracker have incorrect velocity the motion field will not be calculated properly. However, the Gouraud interpolation is only an approximation for a dense optical flow field. The motion field calculation is very fast in contrast to methods described in [6]. Using the extraction of the motion estimation, points from a frame can be found which are more reliable than other points in terms of temporal stability. This reliability can be used in the further processing steps.

Third, the displaced frame difference (DFD) image is extracted which shows the reliability of the extracted motion field. The DFD-image is calculated from the original image and an image predicted based on the calculated motion. Due to the interpolation of the motion field the predicted image is not identical (only motion compensated) to the original consecutive image. If the motion field is correctly calculated all values of the DFD-image are zero. Values which are not zero show pixels/regions which are unreliable calculated due to a false estimated motion field. The pixels/regions which are near zero have a high reliability.

Last, the reliability can be assigned to each triangle. One criterion is the DFD value for a triangle in relation to the number of points of the area of the triangle. A maximum DFD value is introduced as threshold to select triangles for further processing. Furthermore, at least two KLT feature points have to be assigned to one dominant motion cluster. The third point has not necessarily to be in the same cluster due to the possibility that the region is a border region of the object with eventually bad colour segmentation. Moreover, in the "reliable triangle search" triangles are discarded which are below a minimum temporal reliability to guarantee reliable moving object segmentation. This can be done based on the dominant cluster information over time. The dominant motion cluster is calculated over subsequent frames. Consequently each triangle has a lifetime due to the cluster assignment and tracking of the point. If the triangle has a longer lifetime than an adjustable parameter the triangle is a candidate for the further process. A problem of the algorithm is the dependency on the reliability of the KLT-tracker and on the stability of the triangulation over time. This dependency results in skipped frames of the tracked triangles over a certain time. These outliers are ignored by introducing a new parameter which defines the minimum appearance of the triangle. Otherwise the skipped frames would lead to discarded triangles.

Reliable triangles are selected by the help of previously introduced parameters, namely the DFD-depended triangle dependency, the

minimum lifetime and the minimum appearance percentage of that. In this process a number of reliable triangle candidates for each frame for further processing are extracted.

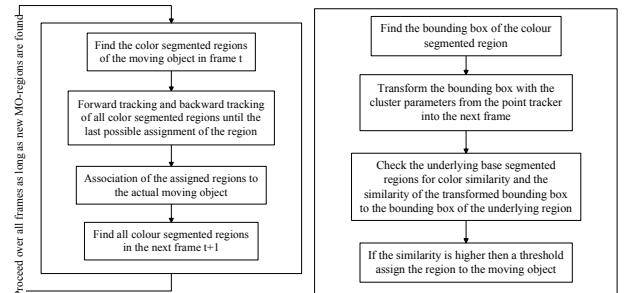
### 3.3 Cluster Assignment to Segmented Regions

After a set of stable triangle has been selected colour segmented regions can be assigned to the according triangles. The idea is that the current triangle is assigned to a dominant motion which means a possible region assignment overlaps the triangle. If the region is correctly assigned to a triangle and furthermore to a dominant cluster a skeleton of the moving object is extracted. The assignment to a dominant motion cluster is done with all segmented regions which are entirely or partially in triangle. A minimum adjustable threshold has to be exceeded. This threshold declares the minimum percentage of the area of a triangle which overlaps to a colour-segmented region. If the assignment of the clusters to the region is ambiguous, a further determination has to be done otherwise the multiple clusters could be assigned to the regions. If three points of the related tracking points inside the triangle belong to the same cluster it has a higher weight regarding the calculation then the triangles with only two points belonging to the same cluster. The triangle with the higher weight is assigned to the region. If the region contains more than one triangle with the same weight the triangle with the higher area overlap is selected.

All regions with the same dominant cluster assignment are selected to extract the skeleton of the moving object. In Figure 1 two further processing steps are depicted: tracking of the assigned regions and hierarchical clustering. These steps are used to get realistic moving objects of the extracted moving object skeletons.

### 3.4 Region Tracking

Region tracking is established to describe an approach to get first realistic moving objects from the initial moving object skeletons. To each extracted skeleton a cluster-parameter (velocity) from the dominant motion tracking is assigned. With the assignment of these parameters new regions according to the moving region of the actual frame are found in the prior frame and in the posterior frame. The new regions enlarge the moving object to become a more realistic one. In these frames it is possible to assign the newly found colour segmented regions to the skeleton to add further regions to the moving object. These new regions have to fulfil some criteria (see Figure 3):



**Figure 3: Region tracking.**

**Left:** tracking over the entire appearance of a skeleton of a moving region. The algorithm proceeds over all frames where colour segmented regions of the moving object are found.  
**Right:** backward/forward tracking algorithm to the previous and the next frame.

The new colour segmented regions have to achieve some thresholds like the mean RGB-colour threshold of the region, bounding box width, bounding box height and region area. These thresholds ensure that the corresponding new region of the next and previous frame is the same region of the moving object as in the actual frame. After region tracking the results are moving regions with each moving region containing points with a similar dominant motion due to an assignment of previous and next frames.

### 3.5 Hierarchical Clustering

To get more realistic segmentation results for non-rigid moving objects a further processing step is introduced. Using hierarchical clustering parts of non-rigid objects containing different dominant motion should be connected.

Hierarchical clustering is well known and commonly used in image processing. In hierarchical clustering some distances have to be introduced. By the help of these parameters hierarchical clustering can find the nearest nodes (represent features of moving objects) in terms of these parameters and combine them to a single node. The clustering is done by the single linkage clustering method which is described in [11]. The clustering is continued iteratively to find the next node until a hierarchical tree is extracted. In case of moving object segmentation the parameter nodes are features of moving regions. These are the motion trajectories and the distances of the centroids from the moving regions. The motion trajectories locate the clusters over several frames. If the motion trajectory of the compared regions is relatively similar and the distance is relatively short the compared regions belong to the same moving object. For a better understanding of the algorithm an example is given in the next paragraph.

The motion from all bones of a leg is not the same during a time period (for example in a walk of 4 gait cycles) but the overall motion trajectory is the same over this period and the distance of the bones is closely related.

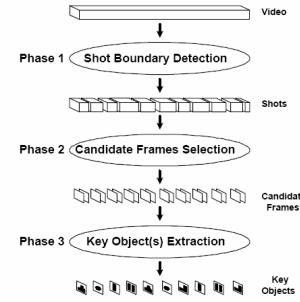
To exclude clusters that are too far away and that have motion vectors too different, a distance is introduced to get only close objects with similar motion in the correlation table of the hierarchical clustering. A threshold is introduced to cut off the tree to get real moving objects. The cut off threshold of the hierarchical tree was set to 0.7 as proposed in [1]. This value was also the result of several tests for the search for the best parameter adjustment.

### 3.6 Moving Object Segmentation in Videos

The input of moving object segmentation is usually a block of frames (BOF) which describes a limited amount of subsequent frames. In the proposed system, large videos are processed instead of a few frames thus a high amount of data has to be analyzed needing a considerable amount of time and storage. To keep these negative influences within a limit new approaches are needed. A common effective technique to extract moving objects in videos and films is the following, which is frequently described in many papers.

One of the most fundamental tasks in moving object segmentation for extracting a description of video and film is to find frames where the motion of the moving objects is high enough for

segmentation and the content is important related to the aspect of motion. In the literature there are many different techniques to get these frames, an example from [22] is shown in Figure 4. In the candidate frames the key objects are extracted which change significantly in their visual content. These important BOF's are usually after shot boundaries [15]. It is necessary to find these frames due to the limitation of memory and time. Recently many algorithms have been proposed to get the frames with the important content, a detailed description can be found in [22] and [19]. After the detection of shot-boundaries and key-frame-extraction the Mesh-based MOS approach can be applied.



**Figure 4: Moving Object Extraction Procedure in video and film [22].**

The reason to make shot boundary detection in moving object segmentation is the high content movement after such boundaries which results in the extraction of different moving objects. At shot boundaries many visual features changes and therefore it is crucial to detect the shot boundaries before doing further analysis like moving object segmentation [5].

### 3.7 Representation and Retrieval of Objects and Events

In the previous sections a way of extracting moving objects resulting in several moving object descriptions was described.

The retrieved moving objects and their trajectories are directly applicable to event analysis and retrieval. But how we can get events or actions out of the extracted moving objects? And how is it possible to represent or save the moving objects in an effective way so that the extracted moving objects can be compared to any other previously extracted moving objects?

A standardized way for describing the extracted moving objects is preferable, such as MPEG-7, which supports content-based video indexing and retrieval. An overview of MPEG-7 is discussed in [21]. The standardized format allows interoperability between applications. MPEG-7 predefines some features for moving object description. These features are low-level descriptions, describing elementary features like colour (e.g. Colour Layout, Colour Structure), texture and shape of regions. In this work, a moving object description structure with special focus on colour features has been developed based on the detailed audiovisual profile (DAVP) MPEG-7 profile [3].

Due to the vast amount of monitored data in surveillance systems and other archives, the extracted metadata can become very large which requires a special indexing step for ensuring fast search and retrieval for video objects. A solution is to import the MPEG-7

documents into a database and to establish an index of the previously mentioned MPEG-7 descriptions.

For that purpose we have developed a Search and Retrieval Tool [23] which is able to import MPEG-7 documents and formulate queries by a graphical user interface (GUI) and pre-defined SQL statements. Different videos can be opened, viewed and analyzed. After a definition of the video object the search tool builds automatically the query by a combination of predefined keywords (SQL statements) and the content-based extracted elements (MPEG-7 Descriptors). The used parameters (e.g. which descriptors should be combined) are defined by the type of analyzing process. The search result is represented in form of a list of references to the metadata descriptions of the matching moving objects, sorted by similarity.

In literature an event is defined as “something that happens at a given place and time”. Two types of events are possible: object domain events and frame or shot domain events. In the search tool these events are easily to retrieve. In the context of event retrieval, the most useful query parameters are the motion trajectories. The trajectories contain the information of primitive motion e.g. move left, move right. With SQL statements moving objects of the same motion can be searched for. Furthermore all moving objects within a certain period can be found. This search tool supports the user in bridging the gap between the numerical features and the symbolic description of the meaningful actions and events.

## 4. RESULTS

In general evaluation of automatic moving object segmentation is a complex process and different methods have been already been reported in literature.

Two important measurements are the misclassification penalty [12] and the precision-recall measures [25] for moving objects. The misclassification penalty is pixel based and leads to an evaluation of region segmentation only.

For event detection, it is crucial to extract motion trajectories from moving objects. For that purpose we need evaluation of the assignment of regions to moving objects (rather than region segmentation), so we decided to use the Precision/Recall approach. For computing the precision and recall ground truth data is required.

The ground truth data is extracted by Mean-Shift segmentation (colour segmentation). The colour segmented regions are candidates for the ground truth regions. The final ground truth regions (moving objects) were manually composed of a set of colour segmented regions. We adopt the precision and recall as follows:

$$\text{Precision: } p = \frac{n_t}{N_d} \quad \text{Recall: } r = \frac{n_t}{N_G}$$

$n_t$  number of correct segmented regions of all moving objects in frame  $t$

$N_d$  total number of segmented regions assigned to all moving objects in one frame by the algorithm.

$N_G$  total number of segmented regions assigned to all moving objects in one frame from the ground truth data.

In order to evaluate the specific challenges in moving object segmentation, we have selected sports video (skiing and car race) with dynamic scenes, multiple fast moving objects and occlusion.

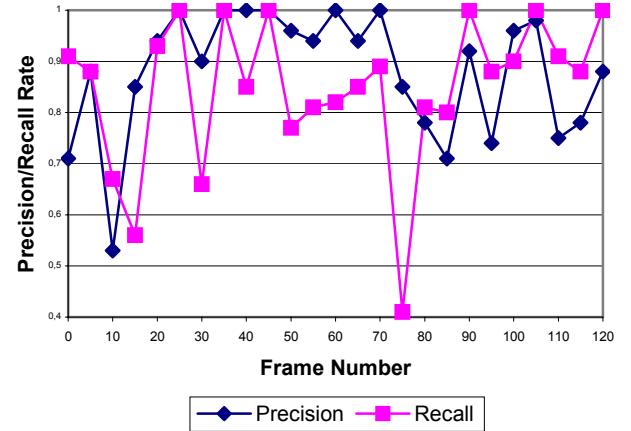


Figure 5: Precision and Recall values for 120 frames of the Formula-1 video. Average number of MO per frame is 2.3.

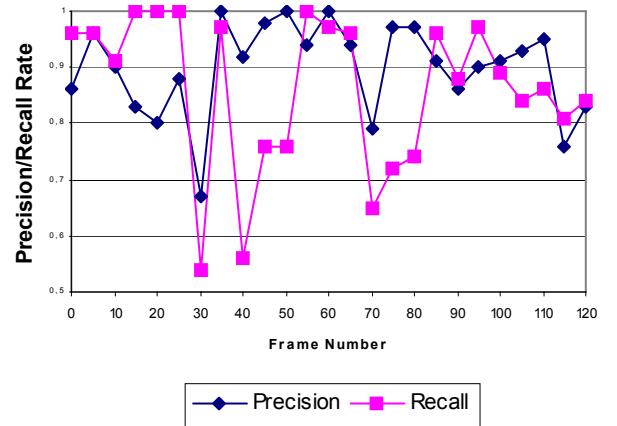


Figure 6: Precision and Recall values for 120 frames of the ski-race video. Average number of MO per frame is 1.2.

The precision/recall calculation shown in Figure 5 and Figure 6 indicate good segmentation results using our mesh-based algorithm. The outliers (worse moving object segmentation) are due to high motion of video objects and therefore worse feature tracking results. The precision and recall rates are high and similar (mean values about 0.85) for both videos. High precision values mean that nearly all found regions are correctly assigned (i.e. are part of) the real moving object. The lower recall value illustrates that a number of regions given in the ground truth are not segmented by the algorithm. This algorithm was designed to extract moving regions which are assured parts of the moving object, but the drawback is that fewer segmented regions are obtained. In the Formula-1 video more regions are found since the motion vectors can be calculated better on the rigid object (cars), in the ski-race fewer regions are found due to the high set of different motion (non-rigid) which is combined in one moving object.

The recall and precision values are similar to the results of the algorithm defined in [25]. Generally, the algorithm has problems

if not enough stable feature points are found by the tracker in relation to the number of segmented regions. This can happen if the object is too far away from the camera, the object has not enough corners or there is too much motion blur in the image.

In the following figures exemplary segmentation results are shown.



**Figure 7: Examples of correct MOS from ski-race video. Original images (top), colour segmented regions (centre), moving object segmentation (blue, bottom).**

Figure 7 visualizes correct moving object segmentation, which is in this case due to the good colour segmentation and the correct assigned tracking points to the objects.



**Figure 8: Examples of incorrect MOS from ski-race video. Original images (top), colour segmented regions (centre), moving object segmentation (blue, bottom).**

In Figure 8 incorrect examples of moving object segmentation are shown. The MOS results are false due to the incorrect assignment of tracking points to the object.

The analysis was done on an Intel Duo Processor (2.4GHz, 2MB L2 Cache, 800MHz FSB) and 2GB, 667MHz DDR2 SDRAM. The average operating time is 320 ms/frame with a resolution of 352x288, which is too slow for applications requiring real-time processing. However, it is possible to speed up the processing depending on the number of key-frames extracted per shot.

## 5. CONCLUSION

In the context of self configurable event detection, special focus is on unsupervised algorithms that are flexible enough for application in different domains.

In this work we presented a fully unsupervised mesh-based algorithm for moving object segmentation. The proposed system facilitates automatic moving object segmentation and is not restricted to pre-defined settings of the environment and therefore overcomes the limitation of many existing moving object segmentation tools.

The evaluation highlights that the quality of extracted moving objects of the mesh-based-algorithm has high precision and recall values of 0.85 on average and is therefore comparable with other state-of-the-art algorithms.

The results show that the algorithms are dependent on the base techniques namely Mean-Shift colour segmentation and KLT point tracking. The colour segmentation should separate regions of the foreground objects and the background objects. This was not always possible due the different light conditions and the similar colours between foreground and background. The point tracker has to generate enough stable points on these foreground objects. Another problem that limits the quality of the moving object segmentation is the fact that the foreground objects have less tracking points and they are usually smaller than the background.

Future work may be to restrict the application for a specified environment and implement self-adaptation. Further, improvement of run-time performance is necessary for being applied in real-time based systems, such as online event detection.

Generally, the results encourage a further development and application of the proposed system. Reasonable applications are semantic video indexing, content based video retrieval (e.g. search for similar moving objects), and compression algorithms of videos (e.g. the MPEG-4 format that contains a description of moving objects). This work has also proposed a compact and efficient representation of the content and moving objects using MPEG-7, including a database based indexing for retrieval of moving objects in large-scale video repositories.

## 6. ACKNOWLEDGEMENTS

The authors would like to thank Werner Haas, Werner Bailer and Peter Schallauer as well as several other colleagues at JOANNEUM RESEARCH, who provided valuable feedback. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 216465 (ICT project SCOVIS).

## 7. REFERENCES

- [1] G. Antonini, S. V. Martinez, M. Bierlaire, J. P. Thiran, “Behavioral Priors for Detection and Tracking of Pedestrians in Video Sequences Source”, International Journal of Computer Vision, Volume 69 , Issue 2, pp. 159 – 180, 2006.
- [2] A. Bab-Hadiashar, N. Gheissari, D. Suter, “Robust Model Based Motion Segmentation”, ICPR02, pp. 753-757, Aug. 2002.
- [3] W. Bailer , P. Schallauer, “Detailed audiovisual profile: enabling interoperability between MPEG-7 based systems”, International Conference on Multi Media Modelling, 2006.
- [4] W. Bailer, P. Schallauer, H. Bergur Haraldsson, H. Rehatschek, “Optimized Mean Shift Algorithm for Color Segmentation in Image Sequences”, Image and Video Communications and Processing, pp. 522-529, 2005.
- [5] W. Bailer, F. Höller, A. Messina, D. Airola, P. Schallauer, M. Hausenblas, “State of the Art of Content Analysis Tools for Video, Audio and Speech”, Deliverable Prestospace, Homepage:  
[http://www.prestospace.org/project/deliverables/D15-3\\_Content\\_Analysis\\_Tools.pdf](http://www.prestospace.org/project/deliverables/D15-3_Content_Analysis_Tools.pdf), 2005.
- [6] G. D. Borshukov, G. Bozdagi, Y. Altunbasak, A. M. Tekalp, “Motion segmentation by multistage affine classification”, IEEE transactions on Image Processing, Volume: 6, pp: 1591-1594, Nov 1997.
- [7] I. Celasun, A. M. Tekalp, M. H. Gökcetekin, D. M. Harmancı, “2-D Mesh-Based Video Object Segmentation and Tracking with Occlusion Resolution”, Signal Processing: Image Communication Volume 16, Issue 10 , August 2001.
- [8] D. Comaniciu, P. Meer, “Robust analysis of feature spaces: colour image segmentation”, Department of Electrical and Computer Engineering, 1997.
- [9] J. C. Davis, "Statistics and Data Analysis in Geology, Third Edition", John Wiley and Sons, 2002.
- [10] M. Donoser, “Object Segmentation in Film and Video”, Diploma thesis, TU-Graz, 2003.
- [11] R. O. Duda, P. E. Hart, D. G. Stork, “Pattern Classification”, J.Wiley, 2001.
- [12] C. E. Erdem, B. Sankur, “Performance evaluation metrics for object-based video segmentation”, Proceedings of the 10th European Signal Processing Conference (EUSIPCO '00), pp. 917-920, Tampere, Finland, September 2000.
- [13] S. Galić, S. Lončarić, “Spatio-temporal image segmentation using optical flow and clustering algorithm”, Proceedings of the First International Workshop on Image and Signal Processing and Analysis, 2000.
- [14] E. Galmar, “Survey on Video Object Extraction”, Multimedia Talks, Jan. 2005. URL:  
[http://www.eurecom.fr/~galmar/talks/MMtalk\\_Jan05\\_slides.pdf](http://www.eurecom.fr/~galmar/talks/MMtalk_Jan05_slides.pdf).
- [15] J. Guo, J. Kim, C.-C. Jay Kuo, “New Video Object Segmentation Technique with Color/Motion Information and Boundary Postprocessing”, Applied Intelligence Journal, March 1999.
- [16] W. Heidrich, H.-P. Seidel, “Realistic, Hardware-accelerated Shading and Lighting”, Proceeding of SIGGRAPH 99, 1999.
- [17] B. K. P. Horn, B. G. Schunck, “Determining optical flow”, Massachusetts Institute of Technology, 1980.
- [18] R. Lienhart, "Reliable Transition Detection In Videos: A Survey and Practitioner's Guide," International Journal of Image and Graphics (IJIG), vol. 1, No. 3, pp. 469-486, 2001.
- [19] L. Liu, G. Fan, “Combined Key-frame Extraction and Object-based Video Segmentation”, IEEE Trans. Circuits and System for Video Technology, 2005.
- [20] B. D. Lucas, T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision”, International Joint Conference on Artificial Intelligence, pages 674-679, 1981.
- [21] J. M. Martinez, “MPEG-7 overview”, International organisation for standardisation, 2002.
- [22] J. Oh, J. Lee, E. Vemuri, “An Efficient Technique for Segmentation of Key Object(s) from Video Shots”, ITCC '03: Proceedings of the International Conference on Information Technology: Computers and Communications, April 2003.
- [23] H. Rehatschek, P. Schallauer, W. Bailer, W. Haas, A. Wertner, “An innovative system for formulating complex combined content-based and keyword-based queries”, Proceedings of SPIE-IS&T, Electronic Imaging, vol. 5304, pp. 160-169, 2004.
- [24] G. Tsechpenakis, K. Rapatzikos, N. Tsapatsoulis, S. Kollias, “Object Tracking in Clutter and Partial Occlusion through Rule-driven Utilization of Snakes”, IEEE International Conference on Multimedia & Expo (ICME), 2003.
- [25] Z. Wei, D. Jun, G. Wen, H. Qingming, “Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model”, Real-Time Imaging, 2005.
- [26] N. Xu, N. Ahuja, R. Bansal, ”Object Segmentation Using Graph Cuts Based Active Contours”, CVPR03, pp. 46-53, 2003.
- [27] D. Zhang, G. Lu, “Segmentation of Moving Objects in Image Sequence: A Review”, Circuits, Systems and Signal Processing, vol. 20, nr. 2, pp. 143-183, 2001.