

# Another Paradigm for the Solution of the Correspondence Problem in Motion Analysis

Ayoub Al-Hamadi, Robert Niese, and Bernd Michaelis

Institute for Electronics, Signal Processing and Communications (IESK)

Otto-von-Guericke-University Magdeburg

D-39016 Magdeburg, P.O. Box 4210 Germany

ayoub@iesk.et.uni-magdeburg.de

**Abstract.** This paper demonstrates a technique of analysing the following three problems: automatic extraction of moving objects, suppression of the remaining errors and solution of the correspondence problem for the video sequences motion analysis. Here we use a new *paradigm* for solving the correspondence problem and then determination of a motion trajectory based on a trisectional structure. I.e., firstly it distinguishes between real world objects, secondly extracts image features like Motion Blobs and colour-Patches and thirdly abstracts objects like Meta-Objects that shall denote real world objects. The efficiency of the suggested technique for determination of motion trajectory of moving objects will be demonstrated in this paper on the basis of analysis of strongly disturbed real image sequences.

## 1 Introduction

Video object segmentation is required by numerous applications ranging from high-level computer vision tasks [1], traffic monitoring [2] to second-generation video coding [3]. The suggested technique pursues the target of segmentation of each moving object automatically and furthermore the determination of the motion trajectory of these moving objects making use of a multi feature correlation, MFC. Our approach differs from the other methods of the motion analysis (*blockmatching*, BM [4], *optical flow* [5], *feature-based methods* [6,7] and *deformable model-based method* [8]) due to the fact that on one hand object-adapted regions will be applied but not fixed block and on the other hand the colour information is evaluated. Another significant difference of the proposed work as compared to the work reported by other researchers, is that instead of image-primitives (*e.g. edges and corners*) the *hierarchical features extracted* from the moving image regions for the solution of the corresponding problem in image sequences. This is because of the fact that the image primitives occur in a large number and generally they don't have good characteristics for the removal of ambiguities. In fact, a complete search is impossible for primitive-oriented concepts in the contemplated image. In contrast, our method will allow the extraction for the solution of the corresponding problem in image sequences in such a way that only moving regions are extracted. This procedure facilitates the reduction of image data and also allows robust determination of motion information.

## 2 Tracking Paradigm

In image processing the question of “*what is an object or a car*” is of a philosophical nature. One can try to find objects by searching for them with pre-defined model-patterns that requires previous knowledge. Naturally, no clear set of models are available that can cover the whole spectrum of vehicles. Therefore the proposed method goes in a different way.

In this paper, since the real time processing and robustness by the segmentation of moving object is essential for the video surveillance and tracking analysis, hence a **modified difference-image-based (MDI)** approach is used for segmentation of moving objects. The whole procedure allows extraction of arbitrary objects automatically. After extracting regions of motion, there are always some residual errors in general, which are normally eliminated via morphological and the separation algorithm. Thereby a *robust* segmentation of moving image regions is reached despite the change of lighting condition that occurs often in the real environment. The resulting image regions present the object candidates that are used for further tracking analysis. Here the tracking **paradigm** can be defined as “the *Motion Blobs (MB)* and *colour patches (colour-structure code, CSC)* representing two *feature-levels* in a tracking hierarchy” (Fig. 1). Basically, these features describe the noticeable motion in an image pair of two consecutive images. To find and track objects in a longer image sequence an abstraction level is introduced. In this level a set of so-called *Meta-Objects (MO)* are suggested to denote real world objects. MOs are specified by a set of CSC-Patches. Now, the task to track MO splits into two parts. The first to do is a correlation at the feature level, i.e. a correlation of MO and CSC-Patches in image pairs. The second part is a correlation at the *abstraction level*, i.e. the assignment of the correlated CSC-Patches to the right MO. Generally the paradigm uses a trisection, i.e. it distinguishes between firstly the *real world* objects, secondly the image features like MB and CSC-Patches and thirdly the abstract objects like Meta-Objects that shall denote real world objects. The *relationship* between *feature* and *abstraction level* will be elucidated in detail after the introduction of the MB and CSC-Patch correlation.

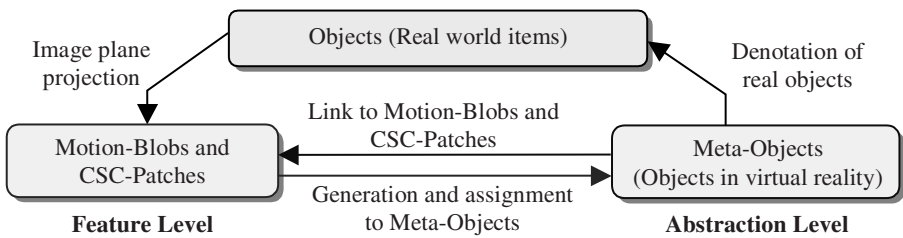


Fig. 1. The suggested technique for the automatic detection and tracking analysis

### 2.1 Automatic Segmentation of the Moving Objects

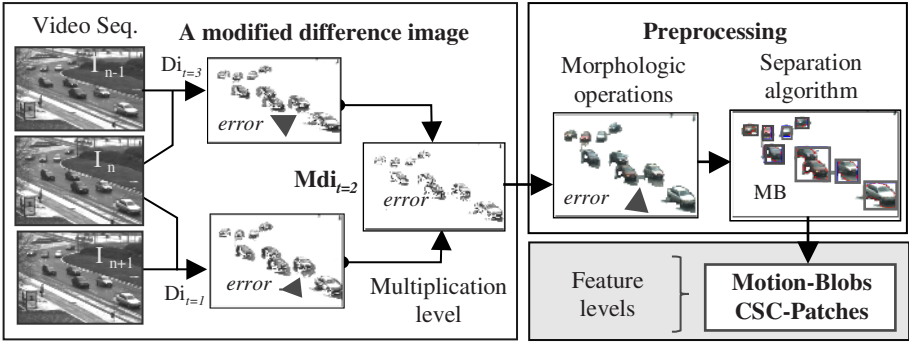
Compared with the motion-field-based methods [4] for segmenting moving objects, a difference image (DI) scheme represents a simple way to detect the moving objects in

a scene. This is because a difference image is produced quickly by simple subtractions. Here it is to be emphasised that the transition borders between regions (*i.e. discontinuity*) will not be cleaned up by this approach. This is because some parts of the stationary background is detected as a moving region (Fig. 2  $DI_{i=1}$ ). The cause is attributed to the fact that the background is *changed* behind the moving object and it is *detected* in the resulting  $DI_{i=1}$ -difference image as movement. A consequence of that is that the moving image regions will merge by this background. In extreme cases the stationary background will be detected as a moving region. This problem is represented in here as a motivation for a *modified approach* for segmentation of the moving objects. The principle for this **modified difference image-based (MDI)** approach is based simply on the fact that the resulting difference image will be received only if the extracted moving regions from two consecutive DIs are *multiplied* (Fig. 2). This operation takes place in accordance with Eq. 1. Thus zero crossings will appear in the DI image, however, an accurate position of the moving objects (middle image) at the time  $t$  is

$$Mdi(x, y, t) = DI_{t-\delta, t}(x, y) \cdot DI_{t, t+\delta, t}(x, y) \quad (1)$$

In general, a MDI approach has two good advantages. The first one is that motion regions on the multiplication level keeps the shape of the moving object at a time ' $t$ '. Regions on a normal DI do not express the shape of the object because it is a mixture of the object shape on the image plane at time ' $t-1$ ' and that at time ' $t$ '. The other advantage is that it is easy to detect whether the current frame contains motion information or not. If motion regions on a MDI approach are small or do not exist, it indicates that the moving object stands still and there is no need to estimate the pose in that frame. It should be recognised from Fig. 2 that the MDI approach, a substantial problem isn't solved yet. If one pixel of a moving object has coincidentally the same grey value as the stationary background or if these pixels do not have a texture foreground at this position, then the change of the intensity value will not be detected in the difference image, although the object moved there. These pixels are missing in the difference image, so that the resulting moving regions are occupied with holes as a result of this step. **The goal** of the following *pre-processing* algorithm is the suppression of the remaining errors (*holes, outliers or fusion of regions*) by the segmentation. The holes are, therefore, filled up first via morphologic operations. A suitable structural element must be defined by the application of morphologic operations. The optimal size of the structural element is reached through the following steps: (1) *Closing-Dilation* with a larger square structural element, in order to interconnect ranges with much movement (2) *Closing-Erosion* with a small structural element, in order to separate closer objects. Subsequently, a separation algorithm is applied for the separation of merged image regions (*Removal of thin background connection between objects*). The separation happens, if two peripheral points have an Euclidean distance  $d$  to each other, which is below a threshold value  $d_{\min}$  and thereto a minimum number exists between the two points of  $n$  peripheral points. This algorithm is used further for the *smoothing* of the contour-region, then outliers will be *removed* simply and quickly from the previously segmented regions (Fig. 2). The resulting image is binary only after execution of the multiplication operation and not before. Thereby the weak

edges, which move fast or the strong edges, which move slowly, are considered. The generation of the motion mask can be accomplished, by setting these regions with a threshold value. After the moving image regions are extracted, the motion parameters for each moving region will be determined. This happens in the next step of the suggested technique.

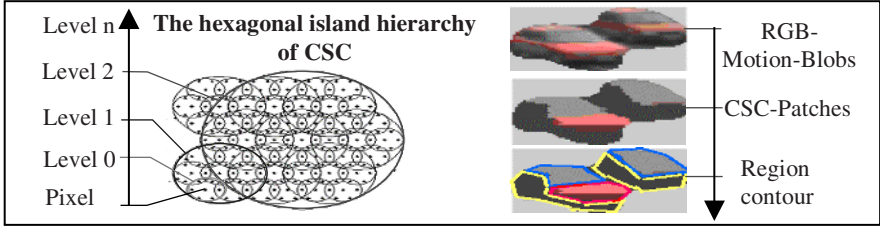


**Fig. 2.** The *segmentation* of moving objects making use of the MDI approach and pre-processing with morphological and separation operators

### 2.2 Generation of Motion Blobs and CSC-Patches (Feature Level)

In the next step object candidates and their corresponding motion trajectories will be determined. This will be based on the analysis of all previously found regions of Motion-Blobs (MB). For this task it becomes necessary to do a further segmentation of the MB. There are several algorithms, which are potentially more suitable for this purpose. The algorithm which combines most of the features required is the Colour-Structure-Code (CSC). It produces a stable real-time capable for segmentations and ease of control. The CSC is an advanced region growing approach that combines the pros of fast local region growing algorithms and the robustness of global methods [7,9]. In a region growing approach usually small segments are built first, which grow bottom-up in a second step of the procedure. That means the initial segments are merged with other segments that fulfils a similarity criterion. Common problems in region growing approaches are *transitive errors*, which emerge from local similarities. Through successive local similarities, distant pixels with different colours may fall into the same segment. Also the segmentation result depends on the order of all merging steps. The CSC avoids both disadvantages through the hierarchical hexagonal topology. It assumes colour images that have all pixels arranged in a hexagonal structure. This can easily be simulated for conventional orthogonal images. In a hexagonal structure each pixel has six equidistant neighbours. Each second pixel per row and each second pixel per column represent the central pixel of a so-called “*island of level 0*”, which consists of exactly this central pixel and six neighbours. Also, all pixels but the central pixels belong to exactly two islands of level 0. In general, each fourth island of level  $i$  represents the centre of an island of level  $i+1$ , which itself consists of seven islands of level  $i$ . The CSC algorithm takes advantage of the numerous overlap-

ping islands. In the *initial* stage, a colour segmentation is conducted for each island of level 0, resulting in segments of at least two pixels size. During the iterative process segments are merged or split, depending on a colour similarity criterion. The parameter to control the similarity is the Euclidean distance between two colours in an arbitrary colour-space [7,9].

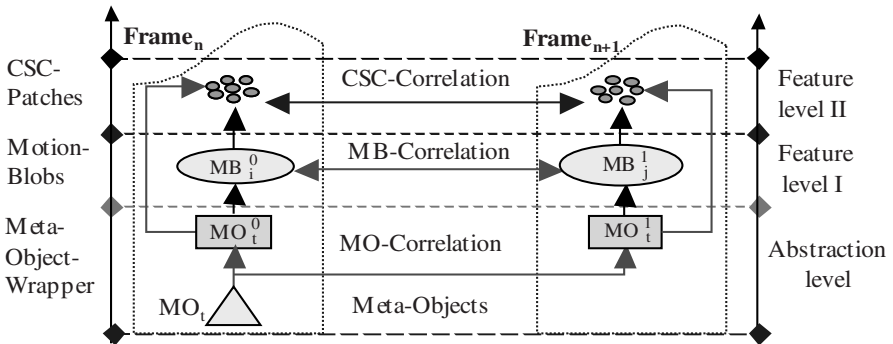


**Fig. 3.** The principle of the colour feature extraction of the motion region using CSC

The real-time capable CSC approach gives superior results in the quality of segmentation and computing time in comparison to other approaches like the top-down Split-and-Merge or Recursive-Histogram-Splitting algorithm [10]. Thus, the CSC is an adequate method to further subdivide a motion region. The search for the correspondences between subdivided motion regions in subsequent images takes place in the next section.

**2.2.1 Tracking Analysis**

To find and track objects in a longer image sequence an abstraction level is introduced. In this level the so-called *Meta-Objects* (MOs) are suggested to denote real world objects. MOs are specified by a set of CSC-Patches. Now, the task to track MO splits into two parts (Fig. 4). The first to do is a correlation at the feature level, i.e. a correlation of MO and CSC-Patches in image pairs. The second part is a correlation at the *abstraction level*, i.e. the assignment of correlated CSC-Patches to the right MO. The relationship between feature and abstraction level will be explained in detail after the introduction of the MB and CSC-Patch correlation.



**Fig. 4.** Relationship of the different correlation levels

**2.2.1.1 Motion-Blob Correlation**

The MB correlation (MBC) is the basis for a subsequent CSC-Patch correlation, which itself is the basis for the MO correlation. The MBC is founded on the assumption of real world objects, which are the reason for any MB can only cover a limited distance between two frames. The MBC is manifold, the simplest case is a 1:1 assignment when one MB in the image  $i$  is assigned to another one in image  $i+1$ . In general, an arbitrary  $m : n$  assignment is possible, where many MB are simultaneously split and merged. The following section will describe the algorithm for the automatic assignment of Motion-Blobs between two frames. In the initial situation there is a MB-set  $S_0$  of frame  $i$  and another set  $S_1$  of frame  $i+1$ . Task of the algorithm is to entirely partition  $S_0 = \{ b_0, b_1, b_2, \dots, b_n \}$  and  $S_1 = \{ b'_0, b'_1, b'_2, \dots, b'_m \}$ , additionally each subset  $U_{0i} \subseteq S_0$  shall be assigned to another subset  $U_{1j} \subseteq S_1$ . The assignment is based on spatial vicinity and a similarity measure for the area covered by  $U_{0i}$  and  $U_{1j}$ . To exactly characterize the relation between  $U_{0i}$  and  $U_{1j}$  a definition by cases becomes necessary. The different cases (Fig. 5) are the following:

- {(a)  $0 : 1$  – emerge; (b)  $1 : 0$  – vanish; (c)  $1 : 1$  – simple movement; (d)  $1 : n$  – split; (e)  $n : 1$  – merge; (f)  $2 : 2$  – simultaneous split and merge of 2 blobs; (g)  $n : m$  – simultaneous split and merge of  $n$  blobs at a time}.

Case (a) eventuates for any MB  $b_{1i}$  of  $S_1$  that has not been correlated in the algorithms main loop. The knowledge of the present case of the assignment is of great help to afterwards realize the Meta-Object correlation. This will be the point when actual objects are located inside of a MB.

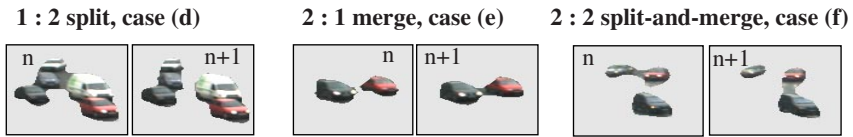


Fig. 5. Examples MB-Correlation Cases

**2.2.1.2 CSC-Patch-Correlation**

Once all MB are correlated, the second feature level can be processed. This task is performed for each correlated MB-Set-Pair, i.e. all related CSC-Patches will be correlated themselves on the basis of different matching criteria. The matching process is realized through a combination of *four separately weighted correlation* tables that achieves a high accuracy at a low computational expense. The matching is done between two CSC-Patch-Sets  $M_0$  and  $M_1$  that belong to previously correlated Motion-Blobs. Matching criteria of CSC-Patches within the MB are (1) *relative location*, (2) *Inter-frame-distance*, (3) *colour value* and (4) *size*.

Each correlation value is computed separately, assessed with respect to reliability and combined to a single overall similarity measure. When the similarity table is filled, those CSC-Patches of  $M_0$  and  $M_1$  that have the highest correlation value  $V$ , which is at least as high as a minimal similarity threshold value  $V_{min}$ , becomes mutually linked. All four similarity measures are weighted and combined to a single correlation table, which is used to a set of finally assigned CSC-Patches of the successive

frames. In evaluation of the table, each CSC-Patch  $p0_i$ , will be linked to exactly one CSC-Patch  $p1_j$  if there is a correlation value  $V$  that is maximal for the respective pair. Additionally  $V$  needs to be as least as high as a minimal threshold  $V_{min}$ . However, there may occur cases of *equivocation* when the table is not of quadratic format and several patches  $p0_i$  become linked to one patch  $p1_j$ . To resolve that problem patches are redistributed to their counterparts with the next best correlation value.

### 2.2.1.3 Correlation in the Abstraction Level Assignment of Meta-objects

After accomplishing the correlation in the feature level, i.e. the correlation in the last pair of blob frames, which resulted in the inter-frame assignment of MB and CSC-Patches, now the assignment of Meta-Objects can be conducted. In contrast to the prior correlation, which is always regarded as only two different frames, this step considers a longer history that is related to the Meta-Objects. So-called Meta-Object-Wrapper (MOW) are used to “wrap” the MO specific information of a certain frame pair. The Fig. 4 presents a general overview about the connection of the different correlation levels. As shown in Fig. 4, the abstraction level is connected to the feature level via MOW, which has access to Motion-Blobs as well as to CSC-Patches that are specific for a particular MO. Meta-Objects are “kept alive” from frame to frame with the help of their MOW. To initialise this process, we must specify how a MO can initially be generated. By definition, in the initial stage of the tracking system each MB shall be one MO. This is a very unlikely assumption, but in the course of time it is approximated to best fit the actual case.

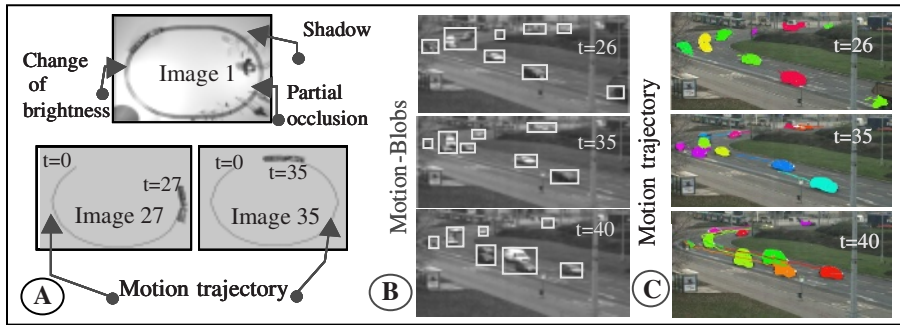
## 3 Results

We want to *briefly* demonstrate the analysis of a real image sequence, which is overlaid by image-specific disturbances (*brightness change, shadow and small partial occlusion*). There, the results of segmentation of moving objects is represented by means of the suggested technique as motion trajectory (Fig. 6). It is to be recognized that during the evaluation, the segmentation of moving objects and the determination of motion trajectory is reliable. A *robust* segmentation of the moving MB objects was reached by means of applying the suggested technique.

In Fig. 6 the resulting movement vectors in form of motion trajectory (Fig. 6) and a tracking contour (MB) are represented. When using the conventional procedure the contour no longer describes the actual object position due to the shadow of that object. Whereas the suggested technique pursues the possible regarded image region over the entire image sequence, because the object regions are stable due to the saturated colours of moving regions.

## 4 Summary and Conclusion

A robust algorithm has been developed for automatic segmentation of moving objects and robust tracking under the influence of disturbed image situations. The robustness is reached by the use of a MDI approach for extracting moving objects. The solution



**Fig. 6.** The analysis for moving objects in real video sequence with the suggested technique. Part **A** presents the analysis under the influence of brightness change, shadow and small partial occlusion. Part **B** shows the results of the motion segmentation (MB) in the first feature level for a sequence. In **C**, the analyses in the second level are presented.

of the correspondence problem in the tracking has taken place in the next stage of the algorithm via *hierarchy feature correlation* from moving image regions between consecutive images. The matching process is realized through the combination of four separately weighted correlation tables that achieve a high accuracy at lower computational expenses. Each correlation value is computed separately, assessed with respect to the reliability and combined to a single overall similarity measure. By the suggested technique, reliable results are achieved despite overlaid by image-specific disturbances (*brightness change, shadow and small partial occlusion*). Further reached improvement is the invariant with object enlargement, object miniaturisation or object rotation.

**Acknowledgements.** This work was supported by LSA/ Germany grant (FKZ: 3133A/0089B).

## References

1. Ullman S.: High-Level Vision: Object Recognition and Visual Cognition, MIT Press, Cambridge, MA, 1996 Torres.
2. Badenas J.; Sanchitz J.M. and Pla F.: Using Temporal Integration for Tracking Regions in Traffic Monitoring Sequences; 15<sup>th</sup> ICPR; Barcelona 2000; Sept; Vol. III; pp.1137–1140.
3. Torres L.; Delp E.J.: New trends in image and video compression, in X European Signal Processing Conference, Tampere, Finland, September 4–8, 2000.
4. Al-Hamadi, A.; Michaelis, B.: Intensity-based method for tracking f objects in colour video sequences under the influence of non-cooperative situations; IASTED; SPPRA, Crete, Greece, June 25–28,, 2002; pp. 62–67 [ISBN:88986-338-5].
5. Klette, R.; Koschan A.; Schluens K.: Computer Vision; Vieweg T.; [ISBN 3-528-06625-3]
6. Deriche R. and Faugeras O.: Tracking line segments. Image and Vision Computing; 8(4): 261–270,1990.



7. Rehrmann V.: Object-oriented Motion Estimation in colour image sequences; in Proc. Of the 5<sup>th</sup> European conference on computer vision; Springer-Verlag, June-1998. Freiburg.
8. Blake A., Curwen R. and Zisserman A: A framework for spatio-temporal control in the tracking of visual contours; International Journal of CV; 11(2):127–145; 1993.
9. Priese L.; Rehrmann V.: On hierarchical colour segmentation and applications; in proceedings of the computer vision and pattern recognition; pp. 633-634,IEEE computer society press, June 1993, New York city.
10. Lutz Priese: Vergleich von Farbsegmentierungstechniken, Labor fuer Bilderkennung Universitaet Koblenz, [www.uni-koblenz.de/~lb](http://www.uni-koblenz.de/~lb).
11. Kim c.; Hwang J.N.:: A fast and robust moving object segmentation in video sequences; IEEE International Conference on Image Processing; Kobe Japan;1999, pp. 131–134.