Detection and Tracking of Moving Objects for Mosaic Image Generation

Birgit Möller[1] and Stefan Posch[2]

[1] Technische Fakultät, AG Angewandte Informatik, University of Bielefeld
P.O. Box 100 131 , 33501 Bielefeld
bimoelle@Techfak.uni-bielefeld.de
[2] University of Halle, posch@informatik.uni-halle.de

Abstract. Mosaic images provide an efficient representation of image sequences and simplify scene exploration and analysis. However, the application of conventional methods to generate mosaics of scenes with moving objects causes integration errors and a loss of dynamic information. In this paper a method to compute mosaics of dynamic scenes is presented addressing the above mentioned problems. *Moving pixels* are detected in the images and not integrated in the mosaic yielding a consistent representation of the static scene background. Furthermore, dynamic object information is extracted by tracking *moving regions*. To account for unavoidable variances in region segmentation topologically neighboring regions are grouped into sets before tracking. The regions' and objects' motion characteristics are described by trajectories. Along with the background mosaic they provide a complete representation of the underlying scene which is idealy suited for further analysis.

1 Introduction

An important topic in computer based scene exploration is the analysis of image sequences, since motion within the scene cannot be extracted using single images only. However, the resulting amount of data to be processed usually limits the application area of image sequences. One possibility to reduce the amount of data is to create mosaic images. In doing so a sequence is integrated in one single mosaic image thus removing redundancies within the sequence. Applying conventional methods to generate mosaics to sequences with moving objects yields integration errors and a loss of dynamic information (see e.g. [1]). In the works of Mégret [6] and Davis [3] moving areas within a sequence are therefore detected and omitted for integration. Thus integration errors can be avoided but dynamic information still is lost. Cohen [2] suggests tracking of moving regions based on dynamic templates, however, if the shapes of objects change significantly templates are not sufficient. In Irani [4] tracking is realized by temporal integration, but no explicit data extraction is suggested.

* This work has been supported by the German Research Foundation (DFG) within SFB 360.

In B. Radig and S. Florczyk, editors, Pattern Recognition, Proc. of 23rd DAGM Symposium, Lecture Notes in Computer Science 2191, pages 208-215, Munich, Germany, September 2001. Springer.

The method presented in this paper is based on a two-step strategy for each image to process resulting in a mosaic of the static scene background and trajectories describing the object motion. In the first step pixels belonging to projections of moving objects, in the following referred to as *moving pixels*, are detected resulting in a motion map and omitted during subsequent integration of greylevel information to generate the background mosaic. In the second step regions, referred to as *moving regions*, are first extracted from the motion maps. They are subsequently grouped into connected components and matched against the components of the previous image. Thus temporal correspondences can be established and trajectories are derived.

The paper is organized as follows. In section 2 a brief introduction to image alignment and detection of moving pixels is given. Section 3 outlines the temporal correspondence analysis based on moving regions by which dynamic scene information is extracted. Results of applying the algorithms to various image sequences are presented in section 4, and finally a conclusion is given.

2 Motion Detection

To detect moving objects in an image sequence many methods have been proposed in the literature. Most algorithms rely on the analysis of pixel based intensity differences between the images after an alignment step in case of a non-static camera. In our approach the images are aligned using *perspective flow* developed by [5] and implemented in [8], where the current background mosaic serves as reference. The alignment is based on an underlying motion model describing the global motion between both images induced by the active camera. We chose a projective transformation, which yields correct global transformation e.g. for camera rotation around the optical centers (no translation) and arbitrary static scenes, while projections of moving objects result in violations of this model. These errors are subsequently detected either computing the average intensity difference or the mean magnitude of the local normal flow $\boldsymbol{n}(x, y)$ for each pixel (x, y) within a neighborhood μ, as illustrated in equation 1 for the mean normal flow $N(x, y)$. Taking neighboring pixels into account yields more robust classification results as the influence of image noise is reduced.

$$N(x, y) = \frac{1}{|\mu|} \cdot \sum_{(x', y') \in \mu} \| \boldsymbol{n}(x', y') \| \tag{1}$$

The classification of moving pixels itself is acomplished thresholding the resulting pixelwise motion measure. Thus detection of motion is achieved except for image regions where motion does not cause any changes in intensity. However, the resulting motion maps are often fragmented and are therefore smoothed applying a dilatation operator of size 7 x 7 to the thresholded motion maps. Hence moving areas become more compact and small gaps in between are closed.

The resulting motion information is used to integrate the current image into the mosaic where only static pixels are taken into account yielding a consistent representation of the static scene background. Further temporal correspondence analysis is based on this data as presented in the next section.

3 Temporal Correspondence Analysis

The resulting motion information is sufficient to generate mosaics of the static scene background. As a next step we aim at explicitly representing the dynamic information of the image sequence contained in the already calculated motion maps. Therefore in our approach moving regions resulting from region labelling the thresholded and dilated motion maps are tracked and trajectories describing their motions are generated to represent the dynamic information.

3.1 Tracking of Moving Regions

Tracking is based on moving regions. The matching criterion for tracking will now be developed for these regions, but subsequently applied for sets of regions (section 3.2).

Each moving region needs to be characterized by several features which allow us to match corresponding regions of consecutive images. When selecting appropriate features it has to be taken into account that due to variances within the segmentation process and because of scene events like object decompositions or collisions, the moving regions' shape and size may vary significantly even for consecutive images. Furthermore regions often do not show homogeneous intensity values since generally moving regions contain projections from different objects or surfaces. Of course, also regions resulting from one object may be of inhomogenous gray values. Due to these limitations features need to be chosen which are mainly invariant against scaling and changes in shape and which preserve an adequate description of the regions' pixel values. This is true for the histogram of the pixel values and the centroid. Based on these features robust detection of correspondences between regions of consecutive images is possible: Two regions are considered as corresponding if the distance between their centroids is smaller than a given threshold θ_d, usually set to 60 pixel, and if the distributions of their intensity values are similar. This is checked by computing the overlapping area F of the two normalized histograms, whereas the minima of the entries a_i or b_i in all cells i are summed up:

$$F = \sum_i \min(a_i, b_i) \tag{2}$$

The resulting intersection area F is required to exceed a given percentage θ_p for establishing a match. θ_p is usually chosen between 0.75 and 0.85. For robust region tracking the sizes A and B of both regions given by the number of pixels are compared in addition. However, only if large differences between A and B occure a match is rejected due to the region expansion induced by the dilatation operator which has to be taken into account explicitly. The difference in size between to areas is regarded too large if the following condition holds:

$$\frac{|A-B|}{A+B} \geq 0.2 \tag{3}$$

3.2 Tracking Components

Matching all pairs of moving regions from two consecutive images is not very efficient due to the combinatorics. Additionally regions frequently decompose or merge in the course of the image sequence induced by variances of the segmentation results (see e.g. motion maps in fig. 2) or events within the scene. In such cases correspondences cannot be established due to significant differences in the regions' histograms or size or too large distances between their centroids. Therefore we propose to track connected components instead, which are referred to as *components* subsequently, similarly as in [7] for color segmentation. In [7], regions are considered as neighboring, if they are spatially adjacent and similar with respect to their color features. In our case regions are assumed to be neighboring, if their minimum point distance is smaller than a given threshold. As mentioned, matching of connected components is based on the same criteria as developed to match moving regions (section 3.1) where the features can be derived directly from the features of the underlying regions. Searching for correspondences between these components reduces complexity, and variances within the segmentation can be handled since it is not required that components contain the same number of regions. Rather each component is treated as a single region. Using this strategy objects or parts of objects can be tracked robustly in most cases. However, in some cases correspondences cannot be established due to a changing arrangement of the regions within the components during the tracking process. To cope with these situations for each unmatched component all subsets of constituting moving regions are considered in a final step. Between all subsets of all components from consecutive images the best match is searched iteratively where the same match criteria as for components are applied. Matched subsets are removed from the sets and search continues until no more matches can be established or all given subsets have been matched.

3.3 Trajectories

As a result of tracking, correspondences between components of consecutive images have been established. To extract the dynamic information, in the present context given by the motion direction, trajectories for all tracked components are generated. The position of each component in an image is given by its centroid. A concatenation of the centroids of matched components, and matched subcomponents as well, yields a trajectory describing the objects' motions within the image sequence. In the case of translational motion this is sufficient to characterize the object motion. Future work will focus on a more detailed analysis of the trajectories, which may serve as a starting point to apply more sophisticated motion models. Furthermore they can be used to detect discontinuities of object motion and components incorrectly classified as moving in the later case. Trajectory points of these components show little variance which should make it possible to distinguish them from moving ones (see figure 4). Finally, based on the trajectories a reconstruction of homogenous objects which cannot be detected as a whole is possible. Different moving components belonging to one object show similar trajectories and could be grouped to reconstruct the object.

4 Results

The proposed algorithms for detection and tracking of moving objects have been tested on various image sequences. All sequences were aquired using an active camera which scanned the scene by pan and/or tilt movements. For the first three examples presented in this article the mosaics of the static scene background are shown along with the reconstructed trajectories. On the left side of each figure several images of the underlying sequence are depicted while in the lower half motion maps for different points of time within the sequence are shown. The last example illustrates the possibilities for false alarm detection and elimination of initial object positions (see below) based on trajectory analysis.

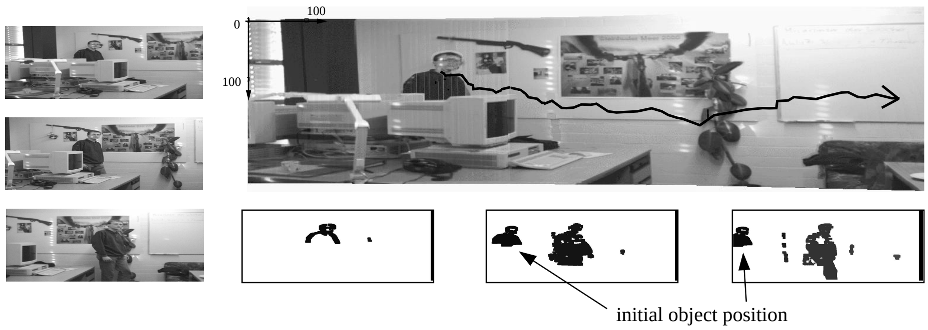

Fig. 1. Results of tracking a person: although the shape of the regions changes (maps below the mosaic) the motion trajectory (black line) can be reconstructed correctly

In figure 1 the person's motion is reconstructed correctly (using intensity difference for detection) despite the fact that the shape of the detected regions and their sizes vary significantly over time (as can be seen in the motion maps). The position of the person can be reconstructed almost completely and no parts of it are integrated so that a consistent representation of the static scene background results. However, it needs to be mentioned that the initial position of the person remains in the mosaic. It is integrated at the beginning of the sequence because it initially belongs to the static background. After the person has left his initial position, the beforehand occluded background becomes visible and results in large intensity differences at this image region, which consequently is detected as a "moving" region. Therefore at this position new intensity information can never be integrated. However, the centroid of this virtually moving region is nearly constant and the analysis of trajectory data is a promising starting point to correct this error in the future.

The second example (fig. 2) shows a mosaic image of a scene containing several pens. A pencil enters the scene from the top left corner and stops moving after hitting one of the other pens in the scene. Its trajectory (reconstructed using normal flow detection) is drawn black whereas the white arrow indicates the real motion direction. Especially at the beginning of the tracking process the motion given by the trajectory differs significantly from the true one. This originates from the fact that initially only parts of the pencil are visible. As new parts become visible a displacement of the centroid results and the reconstructed translation direction is distorted accordingly. As soon as the pencil is completely visible this divergences disappear and the centroids' displacements are caused by real object motions only, allowing a correct reconstruction of direction.

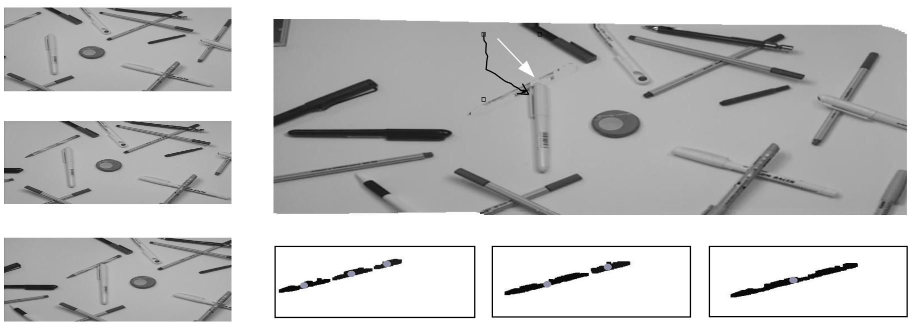

Fig. 2. Mosaic and reconstructed trajectory (black) of an object within a desk scenario. The motion maps show great variances (centroids of moving regions marked grey).

Figure 3 illustrates the mosaic and extracted trajectories of another desk scenario. The match box in the upper half of the four images on the left is opened and its two parts are moving in opposite directions afterwards. As in the first example, the initial box position remains part of the static mosaic whereas the following positions are omitted within the integration process. However, the object parts cannot be detected completely computing the intensity difference due to lack of contrast between the image background and the object parts. Especially the box top moving to the left causes integration errors due to incomplete object detection. Still the reconstructed trajectories describe the object motion almost correctly. Even the decomposition of the box (which forces the moving regions to be split up multiple times, see motion maps in figure 3) is identified correctly. Due to the fact that the regions resulting from the decomposition are grouped into one component they can be matched to the single region of the previous image. With increasing distance, however, they are eventually arranged into different components (points of time indicated by white circles within the figure). This causes a significant change within their centroids' positions (white arrows). However, the existing correspondences can be detected correctly and the scene events represented exactly.

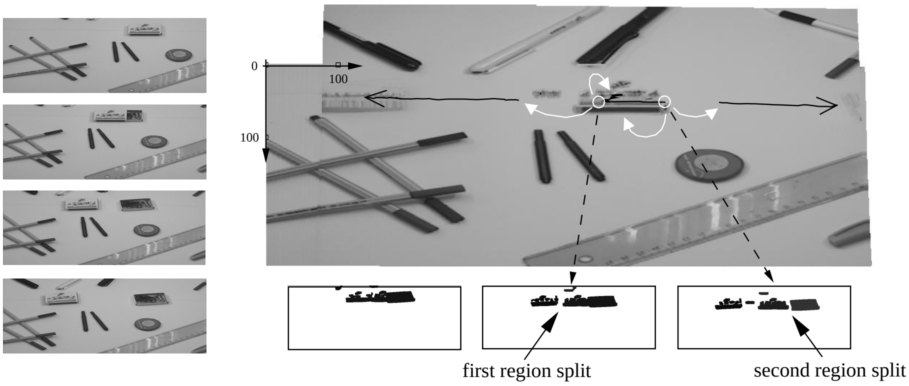

Fig. 3. Detection and tracking of an object decomposing in two parts

Concluding, the last example in figure 4 illustrates the former mentioned possibilities to detect false classified regions by trajectory analysis. The images of the sequence and the related mosaic depict several objects of the construction scenario of the SFB 360 where the dark ring in the center of the images moves to the right. Its initial position remains part of the mosaic. However, the plot of trajectory points at the bottom shows, that variance within these points is quite small and should be sufficient to identify this region as false classified. Hence the mosaic image can be corrected lateron by integrating local information from the current image although the region had been classified as moving beforehand.

5 Conclusion

In this paper an approach to generate mosaics of image sequences containing moving objects has been presented. A mosaic of the static scene background is generated and in parallel trajectories representing the dynamic information are extracted. To this end moving regions within the sequence are segmented based on pixelwise normal flow or intensity difference between two images. Subsequently the regions are grouped into sets of topologically neighboring regions

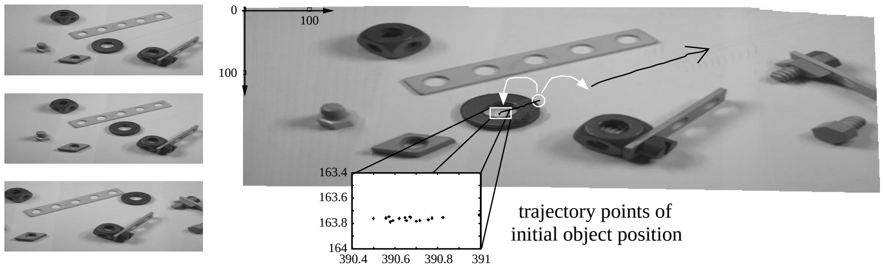

Fig. 4. The static initial position of the moving ring remains part of the mosaic. However, analysing the trajectory points indicates low variance suitable for identification.

tracking is based on. In this manner variances within the segmentation process and object decompositions can be handled. The components and if necessary subsets of them are robustly tracked over time by comparing their intensity histograms and centroid positions. Future work will focus on removing initial object positions from the mosaics and on detecting false alarms. As pointed out trajectory data as computed yield an ideal starting point to solve these problems.

References

1. Bergen, J., Anandan, P., Irani, M.: Efficient Representation of Video Sequences and Their Applications. Signal Processing : Image Communication (1996) (8):327–351
2. Cohen, I., Medioni, G.: Detection and Tracking of Objects in Airborne Video Imagery. CVPR Workshop on Interpretation of Visual Motion (1998)
3. Davis, J.: Mosaics of Scenes with Moving Objects. CVPR **1** (1998) 97-100
4. Irani, M., Rousso, B., Peleg, S.: Computing Occluding and Transparent Motions. International Journal of Computer Vision (IJCV) **12:1** (1994) 5-16
5. Mann, S., Picard, R. W.: Video Orbits of the Projective Group: A New Perspective on Image Mosaicing. M.I.T. Media Laboratory Perceptual Computing Section Technical Reports, No. 338 (1996)
6. Mégret, R., Saraceno, C.: Building the background mosaic of an image sequence. Tech. Report PRIP-TR-060, TU Wien (1999)
7. Rehrmann, V.: Object Oriented Motion Estimation in Color Image Sequences. Proc. of 5th ECCV (1998) 704-719
8. Sypli, D., Tappe, H.: Konstruktion von Mosaikbildern für die Bildanalyse. Masterthesis, University of Bielefeld (1999)