

Tracking Non-Rigid, Moving Objects Based on Color Cluster Flow

B. Heisele U. Kreßel W. Ritter

Daimler-Benz AG, Research and Technology

Pattern Understanding, F3M

P.O. Box 2360, 89013 Ulm, Germany

{Heisele,Kressel,Ritter}@DBAG.Ulm.DaimlerBenz.Com

Abstract

In this contribution we present an algorithm for tracking non-rigid, moving objects in a sequence of colored images, which were recorded by a non-stationary camera. The application background is vision-based driving assistance in the inner city. In an initial step, object parts are determined by a divisive clustering algorithm, which is applied to all pixels in the first image of the sequence. The feature space is defined by the color and position of a pixel. For each new image the clusters of the previous image are adapted iteratively by a parallel k-means clustering algorithm. Instead of tracking single points, edges, or areas over a sequence of images, only the centroids of the clusters are tracked. The proposed method remarkably simplifies the correspondence problem and also ensures a robust tracking behavior.

1. Introduction

Detection and recognition of moving objects are key functions in driving assistance. In extending the application area from highway to inner city scenarios, non-rigid objects such as pedestrians and bicyclists must be considered. If object detection is based on motion segmentation, the first essential step is tracking object parts or object features. Basically there are two different problems to be addressed in feature tracking: Which image features can be tracked successfully in the image sequence and how well do these features describe the objects being tracked. With an eye on these problems we evaluated a few common techniques. One often used approach is to track points with prominent features [1, 2] which are likely to be detected in consecutive images. However, the density of the tracked points is usually not sufficient for an object description. A natural extension is to track edges instead of points [3]. Although successful for artificial objects, it is hard to describe natural objects with a few edges. The next logical step in the search for im-

age features to be tracked, is to use regions instead of edges. We have shown in [4], *Color Blob Flow (CBF)*, that colored regions, determined by color segmentation, yield very promising results in detecting and tracking moving vehicles on highways. There are two main advantages to this method compared to the techniques mentioned above: good object descriptions achieved by color segments, and early reduction of data in the segmentation step, which speeds up the tracking process. However, the shape-dependent features used for matching color segments in consecutive images are not suited for non-rigid objects.

The approach represented here encompasses the advantages of the *CBF* while also solving the matching problem for varying-shaped color segments. The initial color image is divided into clusters containing pixels of similar color and position. The clusters can be seen as ‘fuzzy’ color blobs. In analogy to the *CBF* this approach is therefore called *Color Cluster Flow (CCF)*. Each cluster is represented by a prototype, which is the centroid of the cluster in the color/position feature space. In the consecutive images the prototypes, and thus the clusters, are shifted in the feature space by parallel k-means clustering to fit the new image data. This yields another advantage compared to the *CBF*: No explicit matching of corresponding clusters is required. A similar clustering technique was proposed by [5]. This approach also included spatio-temporal intensity gradients in the clustering procedure. However, in the case of non-rigid, moving objects, these motion-dependent features proved to be more detrimental than helpful. They also require much more computation.

In all algorithms described above, image features are determined and tracked without a priori knowledge about the objects in the image. For this reason these procedures may be used not only to track objects but also to detect moving objects via segmentation of the flow of image features. On the other hand there is a group of algorithms based on active contours [6], which are only suited for tracking objects. Before the tracking procedure is initiated, the contours of the objects must be known.

The plan of the paper is as follows: In Sec. 2 we describe the initial color/position clustering. Adaptive image sequence clustering is introduced in Sec. 3 and tracking objects is described in Sec. 4. Our results are summarized in Sec. 5.

2. Initial Color and Position Clustering

For clustering, each pixel in the image is described by a feature vector \mathbf{f} , containing its color in the RGB-space and its weighted position in the image: $\mathbf{f}_{ij} = (R_{ij}, G_{ij}, B_{ij}, w \cdot i, w \cdot j)$, where i is the row and j the column of the pixel. Important for the clustering procedure is the relation between the color and position space defined by the weighting factor w . For optimal segmentation results w must be chosen as a function of the current color distribution in the image and the size of the objects. In our investigations of traffic scene images (360 columns x 288 rows) we achieved good results for a w of about 1.



Figure 1. Results of the color/position clustering with 64 prototypes for a highway scene. The borderlines of the clusters are highlighted.

The task of clustering is to find a given number of K prototypes \mathbf{p}_k , which minimize the sum of quantization errors: $\sum_{ij} (\mathbf{f}_{ij} - \mathbf{p}_{min})^2$, where \mathbf{p}_{min} is the prototype closest to the feature vector \mathbf{f}_{ij} in the color/position feature space. We have chosen a fast, divisive vector quantization [7, 8], which starts with one initial cluster covering the whole data space. In each iteration the cluster with the highest variance is split in two by a hyperplane. The hyperplane is determined so that it is perpendicular to the direction of the highest variance and runs through the centroid of the original cluster.

The objective of this heuristic is to minimize the variances of the two new clusters. After the partitioning has been performed, the prototypes are calculated as the centroids of the new clusters. Instead of preselecting a maximum number of clusters, one can alternatively define an upper limit for the sum of quantization errors as the stop-criteria for the divisive clustering.

The result of the clustering process (64 clusters) for a highway scene is shown in Figure 1. The original color of each pixel is replaced by the color of the nearest prototype in the color/position feature space. As expected, large regions of similar color (highway and sky) are partitioned into tile-like structures. However, objects of interest with a sufficient color contrast (car, van) are clustered in a way that the shape of the object is reflected by the shape of the respective cluster.

3. Adaptive Image Sequence Clustering

For the very first image a set of prototypes is determined by the divisive clustering described above. For each following image the prototypes of the preceding image serve as seeds for the parallel k-means clustering [9, 10] in the current image. This clustering produces a new set of prototypes for the next image. Parallel k-means clustering, which is based on alternating optimization, consists of two steps per iteration:

$$\begin{aligned} \text{Clusters: } C_k(t+1) &= & (1) \\ \left\{ \text{pixel } ij \mid \left\| \mathbf{f}_{ij} - \mathbf{p}_k(t) \right\|^2 \leq \left\| \mathbf{f}_{ij} - \mathbf{p}_l(t) \right\|^2 \forall l \right\}, & \end{aligned}$$

$$\begin{aligned} \text{Prototypes: } \mathbf{p}_k(t+1) &= & (2) \\ \frac{1}{\text{size}[C_k(t+1)]} \cdot \sum_{ij \in C_k(t+1)} \mathbf{f}_{ij}. & \end{aligned}$$

In the partitioning step each pixel ij , characterized by its feature vector \mathbf{f}_{ij} , is assigned to a cluster $C_k(t+1)$ with the closest prototype $\mathbf{p}_k(t)$. After that, the prototypes $\mathbf{p}_k(t+1)$ are recomputed as the average of the data in their clusters. Each of these two alternating steps reduces the sum of quantization errors until no further changes occur. At this stage a local minimum of the sum of quantization errors is obtained.

Our investigations of different scenes of moving pedestrians have shown that one iteration is sufficient to adapt the prototypes to temporal changes in the image sequence, caused by motion of the objects, egomotion of the camera and new entering objects.

4. Tracking Objects

In the previous chapters we have shown how to determine and track a set of clusters which resulted in a CCF.

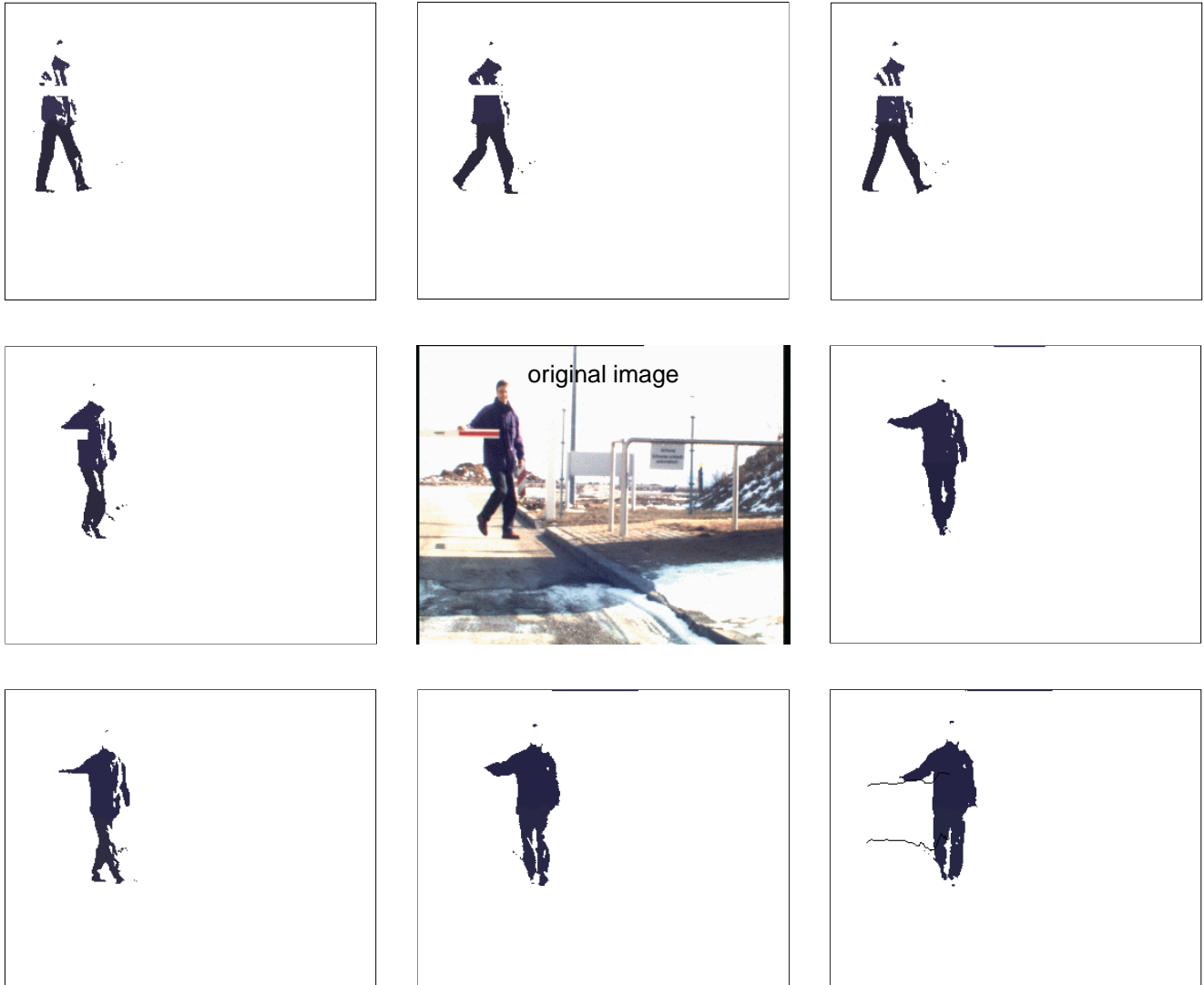


Figure 2. Example for tracking a pedestrian.

As explained in the introduction, our future goal is object detection based on segmentation of the *CCF*. Up to now the initialization of the objects in the first image is done manually by selecting an appropriate set of clusters for each object to be tracked.

In Figure 2 an example of a tracked pedestrian is shown. The pedestrian is represented by two clusters, one for the lilac jacket and one for the dark blue trousers. Despite strong shape variations and partial occlusions by a barrier, the two object parts are easily tracked over a sequence of about fifty images. The resulting trajectories of the prototypes (centroids of clusters) are marked in the last image of Figure 2. Despite the complexity of leg and arm movements, the trajectories are smooth and quite similar to each other. This is caused by the contrarotating motion of the two arms and

two legs. Thus only the lateral motion is reflected in the trajectories. This may play an important role if the *CCF* is used for motion segmentation. Furthermore, the example shows that additional motion-dependent features, e.g. spatio-temporal intensity gradients, in the clustering process are not suited for tracking pedestrians. Rotational movement of the arms, for example, causes permanent changes of motion-dependent features over time. As there is no consistency in these features of the arms they can hardly be described by one cluster over the whole sequence. This leads to non-trackable small clusters.

On the next page, Figure 3 shows that the algorithm can also deal with difficult situations, such as changes in illumination (top) and partial occlusions (bottom) of the moving object.

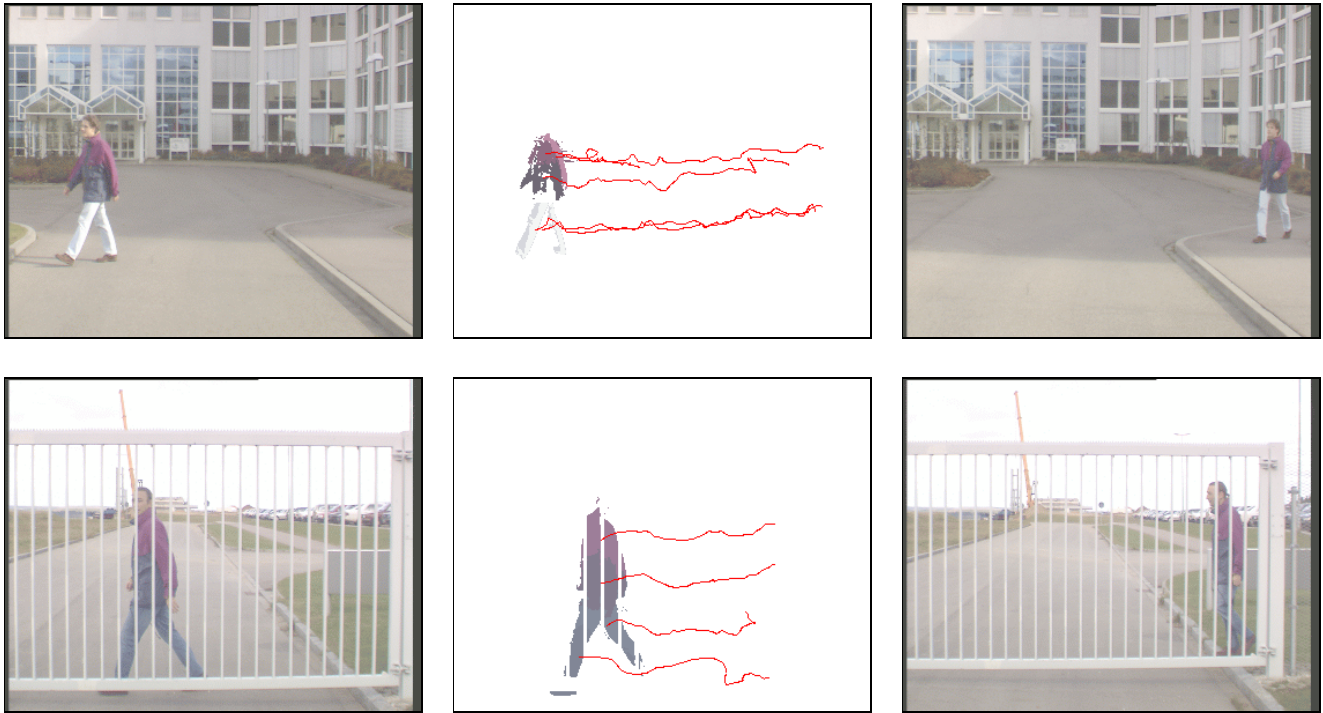


Figure 3. Tracking pedestrians under difficult conditions.

5. Summary and Outlook

In this paper we proposed a new algorithm for tracking non-rigid objects, which is based on clustering in the color/position feature space. With an initial divisive clustering we determine a set of clusters in the first image. These clusters are used as seeds for a parallel k-means clustering, which implicitly performs the tracking of clusters over a sequence of images. The clustering algorithms are conceptually clear and purely data driven. No heuristics, which must be carefully tuned and often do not work outside the lab, are necessary. The tracking algorithm proved to be robust with respect to shape variations and partial occlusions of the objects. Moreover the smoothness of the trajectories of the cluster centroids, even for objects with large shape variations, suggests that object detection based on segmentation of the *CCF* will be possible. The algorithm can also be used in stereoscopy, where matching of image features in pairs of images is a similar problem to tracking features over two consecutive images in a temporal sequence.

References

[1] S. Ullman: *The Interpretation of Visual Motion*. MIT Press, Cambridge, 1979.

- [2] S. T. Barnard and W. B. Thompson: *Disparity analysis of images*. IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-2, no. 4. pp. 333–340, 1980.
- [3] F. Thomanek, E. Dickmanns, and D. Dickmanns: *Multiple object recognition and scene interpretation for autonomous road vehicle guidance*. Proceedings Intelligent Vehicles Symposium '94, pp. 231–236, Paris, 1994.
- [4] B. Heisele and W. Ritter: *Obstacle detection based on color blob flow*. Proceedings Intelligent Vehicles Symposium '95, pp. 282–286, Detroit, 1995.
- [5] M. Etoh and Y. Shirai: *Segmentation and 2D motion estimation by region fragments*. Proceedings International Conference on Computer Vision 4, pp. 192–199, Berlin, 1993.
- [6] M. Kass, A. Witkin, and D. Terzopoulos: *Snakes: Active contour models*. Proceedings First International Conf. Comp. Vision, pp. 259–268, 1987.
- [7] Y. Linde, A. Buzo, and R. Gray: *An algorithm for vector quantizer design*. IEEE Transactions on Communications, COM-28/1, pp. 84–95, 1980.
- [8] N. Bartneck and W. Ritter: *Color segmentation by classification with supervised and unsupervised learning*. Proceedings of the 5th advanced technology conference of the United Postal Service, Washington, 1992.
- [9] U. Kreßel and J. Schürmann: *Pattern classification techniques based on function approximation*. Handbook on optical character recognition and document image analysis, P. Wang and H. Bunke (eds.), World Scientific Publishing Company, 1997.
- [10] R. O. Duda and P. E. Hart: *Pattern Classification and Scene Analysis*. Wiley-Interscience, New York, 1973.