

# Learning Object Appearance from Occlusions Using Structure and Motion Recovery

Kai Cordes, Björn Scheuermann, Bodo Rosenhahn, and Jörn Ostermann

Institut für Informationsverarbeitung (TNT)

Leibniz Universität Hannover

{cordes,scheuermann,rosenhahn,ostermann}@tnt.uni-hannover.de

<http://www.tnt.uni-hannover.de/>

**Abstract.** Visual effect creation as used in movie production often require structure and motion recovery and video segmentation. Both techniques are essential to integrate virtual objects between scene elements. In this paper, a new method for video segmentation is presented. It incorporates 3D scene information from the structure and motion recovery. By connecting and evaluating discontinued feature tracks, occlusion and reappearance information is obtained during sequential camera and scene estimation.

The foreground is characterized as image regions which temporarily occlude the rigid scene structure. The scene structure is represented by reconstructed object points. Their projections onto the camera images provide the cues for regions classified as foreground or background. The knowledge of occluded parts of a connected feature track is used to feed the object segmentation which crops the foreground image regions automatically.

Two applications are presented: the occlusion of integrated virtual objects and the blurred background effect. Several demonstrations on official and self-made data show very realistic results in augmented reality.

## 1 Introduction

Structure and motion recovery consists of feature detection, correspondence analysis, outlier elimination, and bundle adjustment [1]. Recent approaches extend the correspondence analysis in consecutive frames by incorporating feature correspondences between arbitrary frames. This technique stabilizes the bundle adjustment and improves the reconstruction results [2,3]. In this paper, it is shown that non-consecutive correspondences induce valuable information about foreground and background. This information allows for a reasonable initialization of the foreground object segmentation. For applications in movie post production, the segmentation as well as the estimated camera parameters are required [4]. In many cases, the observed scene consists of a moving foreground object in front of a textured background. Examples are shown in Fig. 1. In Fig. 1(a), a child is throwing a stone in front of a static scene while the camera is moving freely. In Fig. 1(b), a similar setup is shown. This sequence is taken from an officially available dataset [5].



**Fig. 1.** Example sequences: moving object in front of textured background

Two application scenarios are presented in Fig. 2(a): the occlusion of virtual objects which are integrated into the video and in Fig. 2(b): the reduction of the depth of field in a video. This technique is used to focus the observers attention on the foreground objects. In movie production industry, the main tools to produce effects as shown in Fig. 2 are structure and motion recovery and object segmentation, so-called *matte painting*. While the camera parameters are usually obtained automatically, the *matte painting* is still done mainly manually [4].

In [6], the video segmentation of foreground objects is done by detecting occlusion edges. In [7], differently moving objects in video are clustered by analyzing point trajectories for a long time. Here, a dense representation of the images is needed. A sparse representation is used in [8]. The background trajectories span a subspace, in which foreground trajectories are classified as outliers. However, many foreground trajectories are required for the following segmentation of the image regions. The approach presented in [9] computes depth maps which are combined with a structure from motion technique to obtain stable results.

Current image segmentation techniques solve the minimum cut / maximum flow problem on a graph, which is initialized by the user marking foreground and background regions with strokes [10]. The locations of the strokes define the *hard constraints* while *soft constraints* are computed by a cost function combining regional and boundary properties of the segmentation. The regional properties are derived from the marked color values. The boundary properties guarantee the smoothness of the resulting segmentation. The problem of video segmentation is



(a) Automatic Occlusion of Integrated Virtual Objects (b) Background Blur Effect: left: ACCV advertisement in sequence from Fig. 1(b), right: integrated 3D objects in the *Column* sequence (b) Background Blur Effect: applied in post production to sequence from Fig. 1(a)

**Fig. 2.** Presented Visual Effect Applications

addressed by treating the image sequence as a single 3D volume. The necessary strokes are entered by the user with a 3D interface.

Our approach incorporates scene information from structure and motion recovery into the foreground segmentation approach. We automatically initialize the segmentation algorithm with reasonable foreground and background samples of the observed scene. Like in [10], the hard constraints are derived from the current image. As the appearance of the objects does not change significantly, the soft constraints are collected throughout the sequence during sequential camera tracking. The cues for foreground regions are identified as regions which occlude reconstructed background scene content. The background scene content is represented by 3D object points.

In contrast to [6], our approach provides points inside the foreground object regions, which is very desirable for the image segmentation procedure. It does not require trajectories on foreground objects like in [8]. In many cases, foreground objects are difficult to track because of few texture, motion blur, or non-rigid motions.

The image segmentation is obtained by efficiently minimizing an energy function consisting of regional and boundary costs. Temporal constraints of foreground and background are incorporated using information from structure and motion recovery. The automatically segmented images are used to ease several applications in movie post production.

In the following Sect. 2, the structure and motion recovery approach is explained. Section 3 shows the automatic detection of foreground regions using the extracted correspondences of discontinued trajectories. In Sect. 4, experimental results are demonstrated using publicly available as well as our data applying two proposed applications. In Sect. 5, the paper is concluded.

## 2 Structure and Motion Recovery

The objective of structure and motion recovery is the simultaneous estimation of the camera parameters and 3D object points of the observed scene [1]. The camera parameters of one camera are represented by the projection matrix  $A_k$  for each image  $I_k$ ,  $k \in [1 : K]$  for a sequence of  $K$  images. On the basis of corresponding feature points, camera matrices and object points are estimated using bundle adjustment [11]. Its idea is to minimize the distance between the reprojection of a 3D object point  $\mathbf{P}_j$  and the measured feature point  $\mathbf{p}_{j,k}$  for each image  $I_k$ , in which  $\mathbf{P}_j$  is visible. The 3D-2D correspondence of object and feature point is related by:

$$\mathbf{p}_{j,k} \sim \mathbf{A}_k \mathbf{P}_j \quad (1)$$

where  $\sim$  indicates that this is an equality up to scale. The bundle adjustment equation to be minimized is:

$$\epsilon = \sum_{j=1}^J \sum_{k=1}^K d(\mathbf{p}_{j,k}, \mathbf{A}_k \mathbf{P}_j)^2 \quad (2)$$

The covariance of the positional error which is derived from the gradient images is incorporated in the estimation [12] using the Mahalanobis distance for  $d(\dots)$  in equation (2). The minimization results in the final camera parameters and object points.

Feature tracking methods like KLT [13] are appropriate to measure correspondences in consecutive frames. In case of wide baselines between the camera positions, feature matching methods have proven superior performance [14]. Structure and motion recovery methods as presented in [2,3] make use of frame to frame feature tracking as well as feature matching in non-consecutive frames and therefore increase the reconstruction reliability and accuracy. Features in non-consecutive frames are of special importance if scene content disappears and reappears. This happens in case of foreground occlusion, if scene content temporarily leaves the field of view, or simply repeated texture and noise. Consequently, we use feature tracking (KLT) for frame to frame correspondences and wide baseline feature matching for the retrieval of discontinued trajectories. For the feature matching, SIFT [14] features are used.

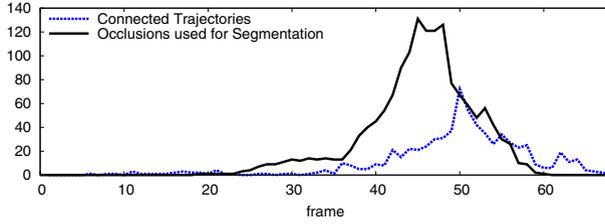
Outlier detection based on RANSAC [15] and the epipolar geometry is crucial in consecutive as well as in non-consecutive frames. Inliers are used for the bundle adjustment while outliers remain in memory to obtain matches with the possibly reappearing feature point. The resulting non-consecutive correspondences are used to stabilize the bundle adjustment and to extract occlusion information leading to the automatic foreground segmentation as explained in Sect. 3. The feature tracking using the combination of KLT tracking and SIFT matching is explained in detail in [16].

Non-consecutive correspondences are only established for discontinued feature tracks, which already have a reconstructed object point  $\mathbf{P}_j$ . The point  $\mathbf{P}_j$  is re-projected by the latest camera  $\mathbf{A}_{k-1}$  to limit the search range for the reappearing feature in the current image  $\mathbf{I}_k$ . This technique, called *guided matching*, avoids comparing a newly appearing feature to the very large data base of discontinued trajectories. Using guided matching is essential to guarantee enough correspondences. Otherwise, it is unlikely for a match to fulfill the uniqueness constraint using the second-closest neighbor [14] because the data base may consist of many similar features.

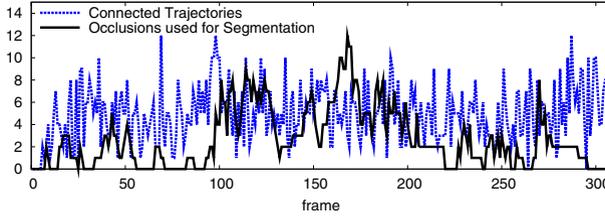
### 3 Learning Object Appearance from Occlusions

Our objective is to automatically segment foreground regions. These regions have two properties: (1) they temporarily occlude the background scene, (2) their appearance does not change significantly throughout the sequence.

While an object point in the background represents image content with nearly the same texture for each frame, its projections on occluding objects provide new foreground color information for every new frame. As the object appearance does not change throughout the sequence, the foreground color values from each available image can be collected to initialize the Gaussian Mixture Model (GMM) [17]. The GMM determines the soft constraints [10] while the hard



(a) *Person Sequence*, see Fig. 7



(b) *Hand Sequence*, see Fig. 5

**Fig. 3.** The number of connected trajectories in each frame and the number of reprojections used for the segmentation for each frame. If a connected trajectory results from occlusion, several reprojections of the corresponding 3D object point are usable for the segmentation.

constraints are derived from the current image only. Thus, a reliable segmentation can be obtained even if there is only few foreground information in the current image.

### 3.1 Occlusion Information

A successfully established non-consecutive correspondence  $\mathbf{p}_{j,k} \leftrightarrow \mathbf{p}_{j,k-l-1}$  in the current frame  $\mathbf{I}_k$  is a part of a feature trajectory  $\mathbf{t}_j^*$  as follows:

$$\mathbf{t}_j^* = (\mathbf{p}_{j,k}^{visible}, \mathbf{p}_{j,k-1}^{invisible}, \dots, \mathbf{p}_{j,k-l}^{invisible}, \mathbf{p}_{j,k-l-1}^{visible}, \dots)$$

The object point  $\mathbf{P}_j^*$  of  $\mathbf{t}_j^*$  is invisible in  $l$  frames  $\mathbf{I}_{k-1}, \dots, \mathbf{I}_{k-l}$ . It is visible in the current image  $\mathbf{I}_k$  and in some previous images  $\mathbf{I}_{j,k-l-1}, \dots$ . It may has been invisible several times before. The coordinates of each of the invisible image locations of  $\mathbf{t}_j^*$  can be estimated with relation (1) after selecting a scale factor for the reconstruction. These coordinates are used to extract occlusion information which provides the initialization of the automatic foreground segmentation.

If the object point  $\mathbf{P}_j^*$  is invisible in the current image  $\mathbf{I}_k$  because of occlusion, its reprojection  $\mathbf{A}_k \mathbf{P}_j^*$  belongs to the foreground in  $\mathbf{I}_k$ . However, experiments have shown, that many non-consecutive feature tracks are established without occluded scene content. To verify the occlusion property, a similarity constraint between each *invisible* point of  $\mathbf{t}_j^*$  and the current feature point  $\mathbf{p}_{j,k}^{visible} = \mathbf{A}_k \mathbf{P}_j^*$

is evaluated. If the similarity constraint is fulfilled, the object point is not occluded in the camera view. Otherwise, the reprojection is an occluded image position. As similarity measure, the color histogram in a  $d \times d$  window around each reprojection of  $\mathbf{A}_{k-1}\mathbf{P}_j^*, \mathbf{A}_{k-2}\mathbf{P}_j^*, \dots$  is computed. For the measurement, the Bhattacharyya histogram distance metric is chosen. This metric provides best results for comparing histograms [18]. Based on the size of the region used for a SIFT descriptor [14], the size  $d$  is chosen to  $d = 16$  pel. This step is important because the correspondence may be established a few frames after the feature reappears. Furthermore, non-consecutive feature correspondences may arise from ambiguities in the image signal (repeated texture patterns, noise) or if scene content leaves and re-enters the field of view.

In Fig. 3, the importance of this step is demonstrated for two publicly available example sequences from [5]. The *Person* sequence shows *simple* camera movements: the camera pans from the right to the left. Thus, many connected non-consecutive correspondences belong to scene content that is temporarily occluded by the person. Several reprojections of one object point are occluded and can be used for the segmentation as shown in Fig. 3(a). By contrast, the *Hand* sequence contains many discontinued trajectories that are connected because their object points leave and re-enter the field of view several times. Although many non-consecutive correspondences are established, only a few occlusions can be exploited and used for the segmentation as shown in Fig. 3(b).

The occlusion information is visualized as white discs, the visible locations of the non-consecutive correspondences are black (e.g. Fig. 4, second row). The diameter of a disc is set to  $d$ ,  $d = 16$  pel as described before. These images provide the initialization of the segmentation procedure as explained in Sect. 3.2.

### 3.2 Foreground Segmentation

Current methods for image segmentation minimize an energy term  $E(f)$  consisting of regional and boundary costs [10]. The idea is to determine the optimal segmentation as the minimum of the discrete energy function  $E : \mathcal{L}^n \rightarrow \mathbb{R}$ :

$$E(x) = \sum_{i \in \mathcal{V}} \varphi_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \varphi_{i,j}(x_i, x_j), \quad (3)$$

where  $\mathcal{V}$  corresponds to the set of all image pixels and  $\mathcal{E}$  is the set of all edges between neighboring pixels. For the problem of foreground segmentation the label set  $\mathcal{L}$  consists of a foreground and a background label. The unary term  $\varphi_i$  is given as the negative log likelihood using a Gaussian mixture model (GMM) model [17], defined by

$$\varphi_i(x_i) = -\log Pr(I_i | x_i = S), \quad (4)$$

where  $S$  is either foreground or background and  $I_i$  describes the feature vector of pixel  $i$ . The pairwise term  $\varphi_{i,j}$  of (3) takes the form of a contrast sensitive Ising model and is defined as

$$\varphi_{i,j}(x_i, x_j) = \gamma \cdot [x_i \neq x_j] \cdot \exp(-\beta \|I_i - I_j\|^2). \quad (5)$$

where  $[\cdot]$  denotes the indicator function. The parameter  $\gamma$  weights the impact of the pairwise term and  $\beta$  corresponds to the distribution of noise among all neighboring pixels. It has been shown that the energy function (3) is submodular and can be represented as a graph [10]. Represented as a graph, the minimum cut minimizes the given energy function.

The GMM's for foreground and background are estimated by image regions that are assigned to either fore- or background. Usually this information is given by the user marking foreground and background with strokes or bounding boxes. In this paper, the GMM's are estimated using the occlusion information that is derived automatically as described in Sect. 3.1. Hence, no user interaction is needed. The information about the foreground regions is given by reprojected 3D object points  $\mathbf{P}_i^*$ ,  $i = 1 \dots j$ . The occluded image region around the reprojected  $\mathbf{P}_i^*$  provides new texture information of the foreground object for each camera view. Thus, it is beneficial to collect the foreground information of the whole sequence to learn the object appearance. This benefit is shown in Fig. 4 and Fig. 5, respectively. For the segmentation results, a two-dimensional grid segmentation is used, which is explained in the next section.

### 3.3 Two-Dimensional Grid Segmentation

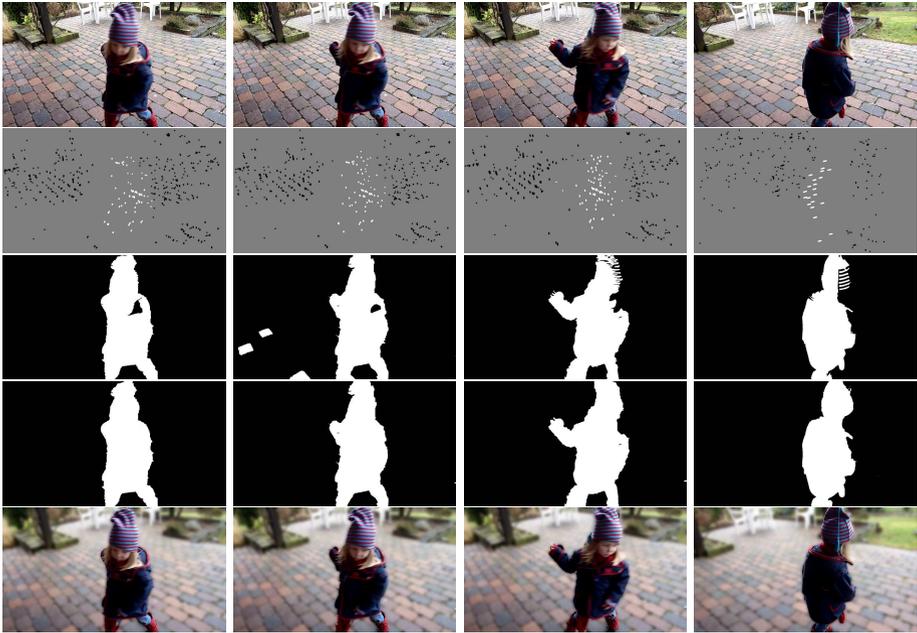
The two-dimensional grid segmentation uses the image grid for building the graph [10]. The segmentation is computed as the minimum cut of the graph.

The Figures 4 (third and fourth row) and 5 (third and fourth row) show examples of the two-dimensional grid segmentation using the second row as initialization. While the third row of Fig. 4, computes the GMM of the foreground from the corresponding single image only, the fourth row of Fig. 4 uses the learned object appearance from the whole sequence. It follows that the result improves in reliability and accuracy. For the *Hand* sequence in Fig. 5, only a few usable foreground occlusions are extracted (see Fig. 3(b)). Thus, the segmentation using only the current image fails (third row). Due to the proposed object learning scheme, the resulting segmentation is reliable and accurate (fourth row).

The computational expense for the two-dimensional grid segmentation is low. Since the graph is contracted to a SlimGraph [19], the result is obtained in less than a second for an image

### 3.4 Three-Dimensional Grid Segmentation

In more challenging scenes, a temporal consistent video segmentation is required to obtain high-quality results. The time domain is incorporated by using the three-dimensional grid to build the graph [10]. By using the additional constraints, a more consistent result with less flickering background objects is expected. On the other hand, the computational expense is large and the need for memory space increases drastically, because the complexity of the graph increases with the number of images. The cut for a sequence of 200 images (The



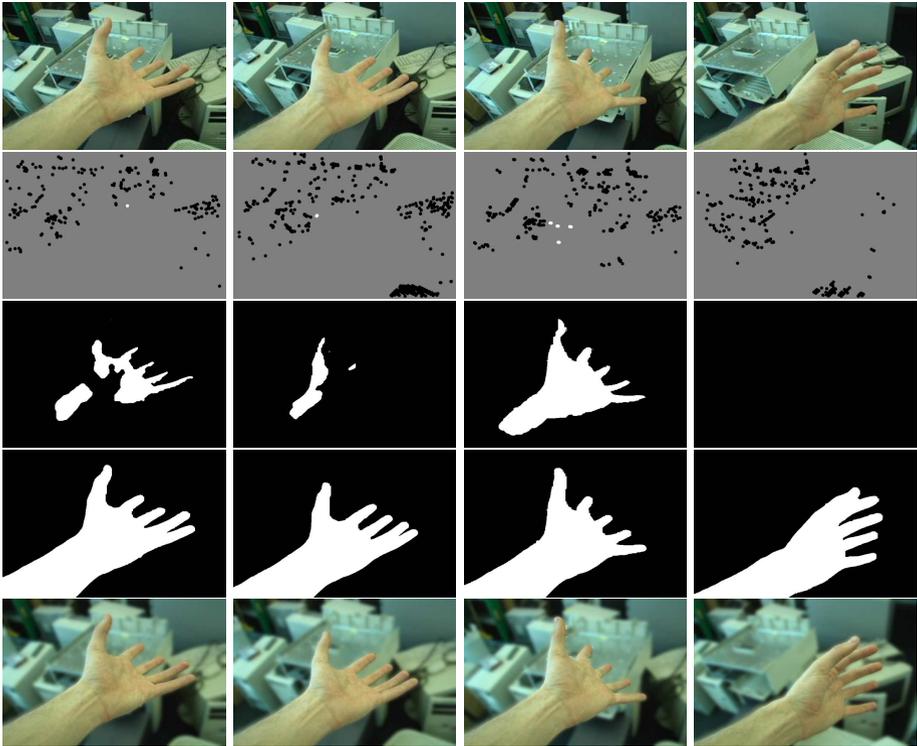
**Fig. 4.** The *Throw* sequence ( $1280 \times 720$ ), from top to bottom: input sequence; automatically generated strokes; segmentation result using only the current image and its corresponding strokes; result learning object appearance; blurred background effect.

*Column* sequence) is computed in 11.5 minutes using 30.8 GBytes of memory. The benefit of using the three-dimensional grid segmentation is demonstrated in Fig. 6 and Fig. 7, respectively.

## 4 Experimental Results

For the evaluation of the presented method of video segmentation and its applications, natural image sequences are used. Two sequences are taken from a publicly available data set <sup>1</sup>, which is used in [8] for the evaluation of video segmentation. Their resolution is  $720 \times 480$  Pixel. Two more sequences are generated using a standard consumer video camera with a resolution is  $1280 \times 720$  Pixel. In each of the sequences, a foreground object is temporarily occluding the background scene. The objective is to generate a reliable segmentation of this object. Then, the video segmentation is used for two different visual effects: (1) the automatic occlusion of integrated virtual objects and (2) the blurred background effect. For the integration of virtual objects, the accurate estimation of the camera parameters is crucial to guarantee the perspective correct relative position of the objects in each view.

<sup>1</sup> <http://rvsn.csail.mit.edu/pv/data/input-video/>



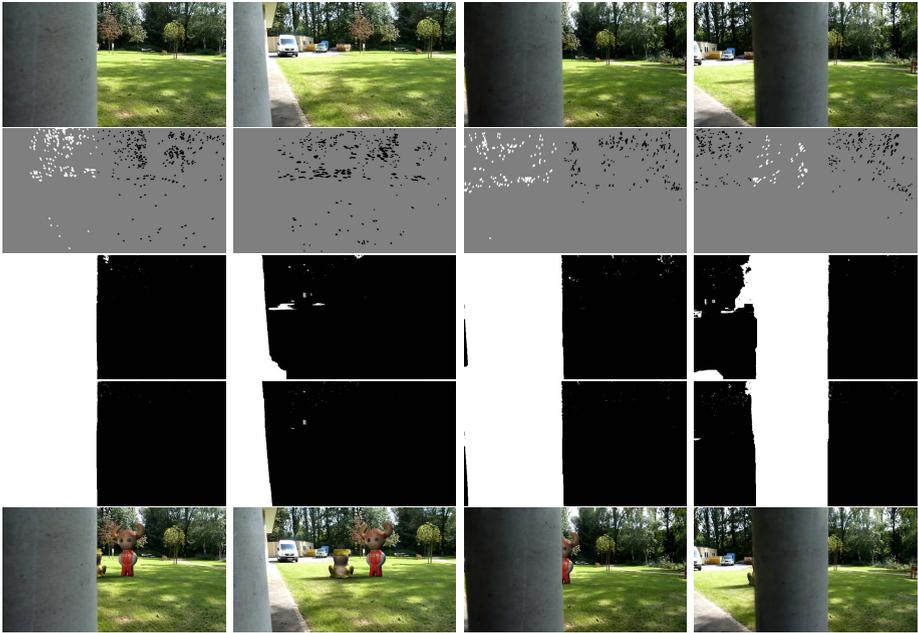
**Fig. 5.** The *Hand* sequence ( $720 \times 480$ ) from [5], from top to bottom: input sequence, automatically generates strokes, segmentation results using only the current image and its corresponding strokes, segmentation result using the proposed object appearance learning, and the blurred background effect.

The first two sequences demonstrate the benefit of the presented object appearance learning scheme (Sect. 3.1 and 3.2).

- *Throw* sequence (Fig. 4): This sequence shows a fast moving child throwing a stone. The extracted occlusion information is shown in the second row. The white regions depict foreground locations, the black regions are classified as background locations. The third row demonstrates the results obtained, if only information of the current image and the corresponding strokes is used for the segmentation of the foreground. By collecting the foreground information from each available frame of the sequence, a much better result is obtained (fourth row).

*Application:* By using the automatically generated segmentation, the background is blurred to focus the observer on the child (fifth row).

- In the *Hand* sequence (Fig. 5), scene content leaves and re-enters the field of view. Due to the small movement of the foreground object, only few foreground information is available as shown in Fig. 3(b). The visualization of



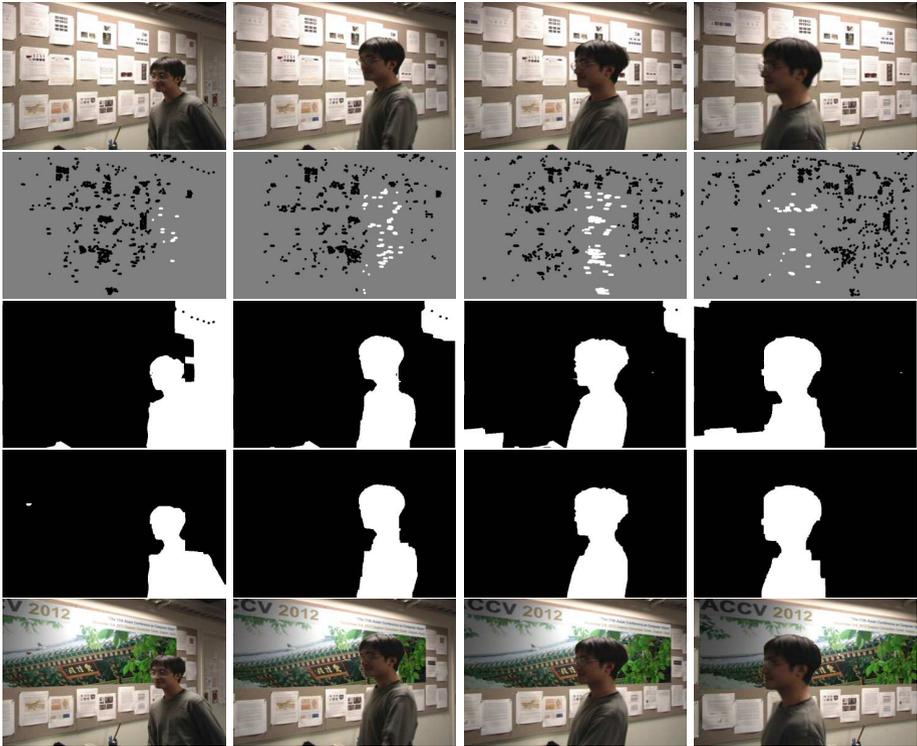
**Fig. 6.** The *Column* sequence ( $1280 \times 720$ ), from top to bottom: input sequence, automatically generates strokes, segmentation result with 2D grid (third row) and 3D grid (fourth row), and the integration of virtual objects which are occluded by foreground objects.

the occlusion information is shown in the second row of Fig. 5. This initialization is not sufficient for a segmentation using the current frame only (third row). Nevertheless, a reliable segmentation is obtained using the presented object appearance learning scheme. Although, no foreground information is available in the last image, the resulting segmentation is correct.

*Application:* In the last row, the application of the *Background blur* effect is demonstrated for this sequence.

The following two sequences compare the two-dimensional grid segmentation (Sec. 3.3) with the three-dimensional grid segmentation (Sec. 3.4). In both comparisons, the object appearance is learned from the whole sequence.

- *Column* sequence (Fig. 6): In this sequence, a column is passing the field of view twice (first row). The second row shows the extracted occlusion information. Initialized with these images, the two-dimensional grid (2D grid) segmentation results in the third row while the three-dimensional grid (3D grid) segmentation results in the fourth row. Several foreground artefacts that are present in the 2D grid approach are correctly assigned to the background with the 3D grid.



**Fig. 7.** The *Person* sequence ( $720 \times 480$ ) from [5], from top to bottom: input sequence, automatically generates strokes, segmentation result with 2D grid (third row) and 3D grid (fourth row), and the integration of the *ACCV* logo on the board. It is occluded by the foreground object.

*Application:* Two synthetic objects are integrated into the scene. Due to the accurate estimation of the camera parameters, the objects are perspective correct in each camera view. They show no drift. The occlusion of the objects using the segmentation results is convincing.

- *Person* sequence (Fig. 7): A person is walking from right to left while the camera is moving (top row). As demonstrated in Fig. 3(a), many points are usable for the segmentation of the foreground object, which is shown visually in the second row of Fig. 7. Again, these images are used to initialize the 2D grid segmentation (third row) and the 3D grid segmentation (fourth row), respectively. The usage of the 3D grid segmentation leads to significantly better results.

*Application:* The segmentation is used to occlude the integrated virtual object, the *ACCV* logo in a compositing step. Again, the integrated virtual object is occluded accurately as shown in the bottom row.

## 5 Conclusion

This paper presents a new approach for automatic video segmentation. It incorporates 3D scene information of sequential structure and motion recovery. Due to the occlusion of scene content with foreground objects, feature trajectories discontinue. They are connected using guided matching and the SIFT descriptor. The location of occluded and visible scene content is estimated from the re-projection of object points into the reconstructed camera planes. The appearance of the foreground objects is learned from occluded regions of the whole sequence. This information is used for the initialization of a segmentation method which minimizes an energy function consisting of regional and boundary costs. It is shown that the results improve by using the temporal constraints in a three-dimensional grid graph.

The resulting segmentation is used for two application in movie post production: the occlusion of perspectively correct integrated virtual objects, and the background blur effect which focuses the observers attention on the foreground objects.

The effectiveness of the approach is demonstrated in several natural image sequences. Virtual objects are accurately integrated between scene structures and the blurred background effect is convincing.

## References

1. Pollefeys, M., Gool, L.V.V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. *International Journal of Computer Vision (IJCV)* 59, 207–232 (2004)
2. Zhang, G., Dong, Z., Jia, J., Wong, T.-T., Bao, H.: Efficient Non-consecutive Feature Tracking for Structure-from-Motion. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part V. LNCS*, vol. 6315, pp. 422–435. Springer, Heidelberg (2010)
3. Cordes, K., Müller, O., Rosenhahn, B., Ostermann, J.: Feature Trajectory Retrieval with Application to Accurate Structure and Motion Recovery. In: Bebis, G. (ed.) *ISVC 2011, Part I. LNCS*, vol. 6938, pp. 156–167. Springer, Heidelberg (2011)
4. Hillman, P., Lewis, J., Sylwan, S., Winquist, E.: Issues in adapting research algorithms to stereoscopic visual effects. In: *IEEE International Conference on Image Processing (ICIP)*, pp. 17–20 (2010)
5. Sand, P., Teller, S.: Particle video: Long-range motion estimation using point trajectories. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 2195–2202 (2006)
6. Apostoloff, N.E., Fitzgibbon, A.W.: Automatic video segmentation using spatiotemporal t-junctions. In: *British Machine Vision Conference, BMVC* (2006)
7. Brox, T., Malik, J.: Object Segmentation by Long Term Analysis of Point Trajectories. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part V. LNCS*, vol. 6315, pp. 282–295. Springer, Heidelberg (2010)
8. Sheikh, Y., Javed, O., Kanade, T.: Background subtraction for freely moving cameras. In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 1219–1225 (2009)

9. Zhang, G., Jia, J., Hua, W., Bao, H.: Robust bilayer segmentation and motion/depth estimation with a handheld camera. *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)* 33, 603–617 (2011)
10. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In: *IEEE International Conference on Computer Vision (ICCV)*, vol. 1, pp. 105–112 (2001)
11. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment - a modern synthesis. In: *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, IEEE International Conference on Computer Vision (ICCV)*, pp. 298–372. Springer (2000)
12. Hartley, R.I., Zisserman, A.: *Multiple View Geometry*, 2nd edn. Cambridge University Press (2003)
13. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 674–679 (1981)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)* 60, 91–110 (2004)
15. Fischler, R.M.A., Bolles, C.: Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communications of the ACM* 24, 381–395 (1981)
16. Cordes, K., Scheuermann, B., Rosenhahn, B., Ostermann, J.: Occlusion handling for the integration of virtual objects into video. In: Csurka, G., Braz, J. (eds.) *International Conference on Computer Vision Theory and Applications (VISAPP)*, pp. 173–180. SciTePress (2012)
17. Rother, C., Kolmogorov, V., Blake, A.: Grabcut: interactive foreground extraction using iterated graph cuts. *ACM SIGGRAPH Papers* 23, 309–314 (2004)
18. Thormählen, T., Hasler, N., Wand, M., Seidel, H.P.: Registration of sub-sequence and multi-camera reconstructions for camera motion estimation. *Journal of Virtual Reality and Broadcasting* 7 (2010)
19. Scheuermann, B., Rosenhahn, B.: SlimCuts: GraphCuts for High Resolution Images Using Graph Reduction. In: Boykov, Y., Kahl, F., Lempitsky, V., Schmidt, F.R. (eds.) *EMMCVPR 2011. LNCS*, vol. 6819, pp. 219–232. Springer, Heidelberg (2011)