OCCLUSION HANDLING FOR THE INTEGRATION OF VIRTUAL OBJECTS INTO VIDEO

Kai Cordes, Björn Scheuermann, Bodo Rosenhahn and Jörn Ostermann
Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover
Appelstr. 9, 30167 Hannover, Germany
{*cordes, scheuermann, rosenhahn, ostermann*}*@tnt.uni-hannover.de*

Keywords: Structure and Motion Recovery, Foreground Segmentation, Feature Tracking, Augmented Reality.

Abstract: This paper demonstrates how to effectively exploit occlusion and reappearance information of feature points in structure and motion recovery from video. Due to temporary occlusion with foreground objects, feature tracks discontinue. If these features reappear after their occlusion, they are connected to the correct previously discontinued trajectory during sequential camera and scene estimation. The combination of optical flow for features in consecutive frames and SIFT matching for the wide baseline feature connection provides accurate and stable feature tracking. The knowledge of occluded parts of a connected feature track is used to feed a segmentation algorithm which crops the foreground image regions automatically. The resulting segmentation provides an important step in scene understanding which eases integration of virtual objects into video significantly. The presented approach enables the automatic occlusion of integrated virtual objects with foreground regions of the video. Demonstrations show very realistic results in augmented reality.

1 INTRODUCTION

For the realistic integration of virtual objects into video, a highly accurate estimation of the camera path is crucial. State of the art techniques use a pinhole camera model and image features for the camera motion estimation. The camera motion estimation workflow consists of feature detection, correspondence analysis, outlier elimination, and bundle adjustment as demonstrated in (Pollefeys et al., 2004), for example.

Most techniques rely on feature correspondences in consecutive frames. Thus, temporarily occluded scene content causes broken trajectories. A reappearing feature induces a new 3D object point which adopts a different and therefore erroneous position. Recent approaches solve this problem by incorporating non-consecutive feature correspondences (Cornelis et al., 2004; Engels et al., 2008; Zhang et al., 2010; Cordes et al., 2011). The additional correspondences and their trajectories are used to stabilize the bundle adjustment and improve the reconstruction results. The reconstructed object points of these feature trajectories are not seen in several camera views. This information has not been used for further scene understanding so far. In many cases, the rejoined trajectories discontinue because of occlusion with foreground objects. We extract the reprojections of their object points and evaluate if they are occluded in each of the valid camera views.

We regard the occlusion and reappearance of scene parts as valuable scene information. It can be used to handle occlusions in video and result in a meaningful foreground segmentation of the images. This additional information eases the integration virtual objects into video significantly.

A typical input example is shown in Figure 1, top row. In this sequence, the background scene is temporarily occluded by a part of the swing rack and the swinging child. For the application of integrating virtual objects into the video, the foreground objects have to occlude the correct augmented image parts throughout the sequence. This is essential to provide realistic results. Otherwise the composed sequence does not look satisfactory as shown in the bottom row of Figure 1.

In literature, some approaches have been proposed for occlusion handling in video. A comparable objective is followed in (Apostoloff and Fitzgibbon, 2006). Occlusion edges are detected (Apostoloff and Fitzgibbon, 2005) and used for the video segmentation of foreground objects. However, no 3D information of the scene is incorporated and only edges of one foreground object are extracted which is not advantageous

Figure 1: Playground sequence (1280 × 720 pixels), top row: example frames $11, 33, 44, 76$ with temporarily occluded scene content resulting from static and moving foreground objects. Feature trajectories discontinue and their features reappear after being occluded. As shown in the bottom row, for integrating virtual objects, it is essential to handle foreground occlusions in the composition of virtual and real scenes.

for the following image based segmentation. In (Guan et al., 2007), the complete hull of occluded objects is reconstructed. For this approach, video streams from multiple, calibrated cameras are required in a shape from silhouette based 3D reconstruction. In (Brox and Malik, 2010), differently moving objects in the video are clustered by analyzing point trajectories for a long time. In this approach, a dense representation of the images is needed (Brox and Malik, 2011). In (Sheikh et al., 2009), a sparse image representation is used. The background trajectories span a subspace, in which foreground trajectories are classified as outliers. The idea is to distinguish between camera induced motion and object induced motion. These two classes are used to build background and foreground appearance models for the following image segmentation. However, many foreground trajectories are required to provide a reliable segmentation result.

Our approach is designed for the integration of virtual objects, and therefore can easily make use of the extracted 3D information of the reconstructed scene. It is not restricted to certain foreground object classes and allows for arbitrary camera movements. A very important step is the feature tracking. For the demanded accuracy, long and accurate trajectories are desired. In contrast to (Brox and Malik, 2010; Liu et al., 2011), our approach relies on a sparse representation of the images using reliable image feature correspondences as required for the structure and motion estimation. We propose a combination of wide-baseline feature matching for feature correspondences in non-consecutive frames and optical flow based tracking for frame to frame correspondences. The resulting trajectories are incorporated in an extended bundle adjustment optimization for the camera estimation. The additional constraints lead to an improved scene reconstruction (Zhang et al., 2010; Cordes et al., 2011). We identify foreground objects in the camera images as regions which occlude already reconstructed scene content. Resulting from the structure and motion recovery approach, reconstructed scene content is represented by 3D object points. In contrast to (Apostoloff and Fitzgibbon, 2006), this approach provides occlusion points inside the foreground objects, which is very desirable for the following segmentation procedure. The image segmentation is obtained by efficiently minimizing an energy function consisting of labeling and neighborhood costs using a contracted graph (Scheuermann and Rosenhahn, 2011). The algorithm is initialized with the automatically extracted information about foreground and background regions. The presented approach eases the integration of virtual objects into video significantly.

In the following Section 2, the structure and motion recovery approach is explained. Section 3 shows the automatic detection of foreground regions using correspondences in non-consecutive frames and their object points. In Section 4, the application of integrating virtual objects into video is demonstrated. Section 5 shows experimental results on natural image data. In Section 6, 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 (Pollefeys et al., 2004). The camera parameters of one camera are represented by the projection matrix A_k for each image $\mathtt{I}_k$, $k \in [1:K]$ for a sequence of K images. For the estimation, corresponding feature points are required. In case of video with small displacements between two frames, feature tracking methods like

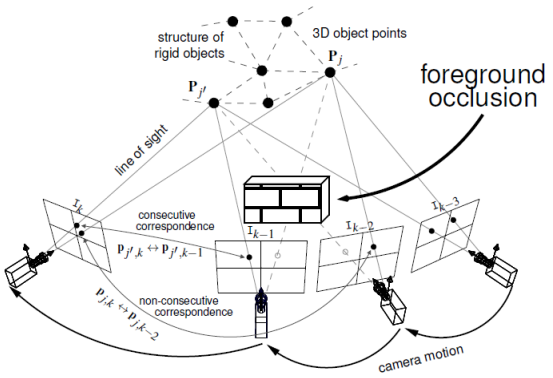

Figure 2: Common structure and motion estimation techniques use corresponding feature points in consecutive images only, for example $\mathbf{p}_{j',k} \leftrightarrow \mathbf{p}_{j',k-1}$. Due to foreground occlusion, trajectories discontinue and the corresponding scene content reappears in a later image. These trajectories are connected using a wide-baseline correspondence analysis, for example $\mathbf{p}_{j,k} \leftrightarrow \mathbf{p}_{j,k-2}$. A real world example is shown in Figure 1.

KLT (Lucas and Kanade, 1981) tend to produce less outliers and provide increased localization accuracy compared to feature matching methods (Thormählen et al., 2010).

Methods as presented in (Engels et al., 2008; Zhang et al., 2010; Cordes et al., 2011) additionally make use of feature correspondences in non-consecutive frames as shown in Figure 2 and therefore increase the reconstruction reliability. This is especially important if scene content disappears and reappears. This happens, if foreground objects temporarily occlude the observed scene. It follows, that non-consecutive correspondences induce occlusion information which is explicitly used in our approach for automatic foreground segmentation as explained in Section 3. The developed feature tracking scheme is presented in Section 2.1, and the bundle adjustment scheme is shown in Section 2.2.

2.1 Feature Detection and Tracking

The presented feature tracking scheme is designed for even large foreground occlusions while the camera is moving freely. Hence, a wide baseline analysis is required for establishing correspondences in non-consecutive frames. For a reliable feature matching, the SIFT descriptor (Lowe, 2004) is used for this task. Consequently, the feature selection uses the scale space for the detection of interest points. For a complete scene representation, the features in an image should be spatially well-distributed. For the results shown in this paper, the SIFT detector is used for newly appearing features and provides sufficiently distributed points. For sequences with very low texture content, a combination of different scale invariant feature detectors may be considered (Lowe, 2004; Matas et al., 2002; Dickscheid et al., 2010). For the tracking from frame to frame, the KLT tracker provides higher accuracy and less outliers than feature matching techniques.

The tracking workflow is shown in Figure 3. Newly detected SIFT features are tracked using KLT. The KLT tracked features are validated with RANSAC and the epipolar constraint. Inliers are used for the bundle adjustment leading to the estimation of the current camera $\mathtt{A}_k$ as well as to an update of the point cloud. Outliers and lost tracks which already have a valid object point are stored for a later match with the possibly reappearing feature. To represent newly appearing and reappearing scene structures, SIFT features are detected, which are at first compared to the stored discontinued trajectories. Validation with RANSAC and the epipolar constraint between $\mathtt{A}_k$ and $\mathtt{A}_{k-l}$, $l > 1$ result in non-consecutive correspondences of the current frame $\mathtt{I}_k$. They are used to stabilize the bundle adjustment and to extract occlusion information leading to the automatic foreground segmentation as explained in Section 3.

The combination of SIFT detection for newly appearing features, SIFT matching for non-consecutive frames, and KLT tracking for frame to frame tracking provides optimal performance for the presented occlusion handling and accurate scene reconstruction.

2.2 Bundle Adjustment

The main idea of bundle adjustment (Triggs et al., 2000) in structure and motion recovery approaches is that a reprojected 3D object point $\mathbf{P}_j$ should be located at the measured feature point $\mathbf{p}_{j,k}$ for each image $\mathtt{I}_k$, in which $\mathbf{P}_j$ is visible. The 3D-2D correspondence of object and feature point is related by:

$$\mathbf{p}_{j,k} \sim \mathtt{A}_k \mathbf{P}_j \tag{1}$$

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

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

The covariance of the positional error which is derived from the gradient images is incorporated in the estimation (Hartley and Zisserman, 2003) using the Mahalanobis distance for $d(\dots)$. The minimization of equation (2) results in the final camera parameters and object points.

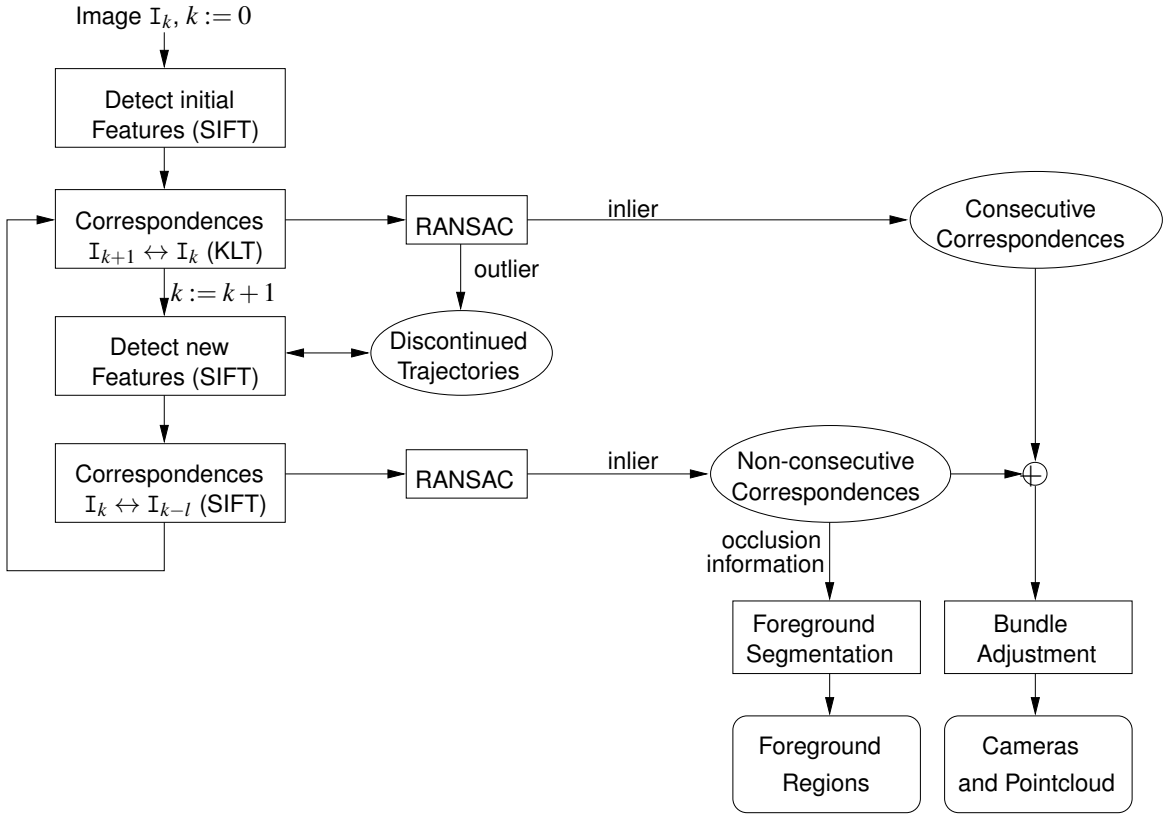

Figure 3: Workflow overview: features are tracked in consecutive frames by KLT while non-consecutive correspondences are established using the SIFT descriptor. Features of the current frame $\mathtt{I}_k$ are matched to features of previously discontinued trajectories in the images $\mathtt{I}_{k-l}$, $l = 2,\ldots,L$, $L \leq k$. For validation, RANSAC and the epipolar constraint between $\mathtt{I}_k$ and $\mathtt{I}_{k-l}$ is used. The bundle adjustment is based on consecutive and non-consecutive correspondences. The occlusion information is extracted from the non-consecutive correspondences and their trajectories. It is used to initialize the foreground segmentation algorithm.

3 AUTOMATIC FOREGROUND SEGMENTATION

The non-consecutive feature tracking connects discontinued trajectories to newly appearing features as shown in Figure 2. If the trajectory is discontinued because of an occlusion with foreground objects, the image coordinates of occluded scene content can be derived by reprojecting the corresponding reconstructed object points. These image locations are used to feed an interactive algorithm (Scheuermann and Rosenhahn, 2011), which is designed to segment an image into foreground and background regions with the help of initially known representative parts, called user strokes.

In (Scheuermann and Rosenhahn, 2011), the segmentation is initialized with manually drawn user strokes. In our work, the *strokes* are restricted to small discs and created automatically using the extracted occlusion information as explained in Section 3.1.

3.1 Occlusion Information

Let us assume, that foreground objects temporarily occlude the background scene. Thus, non-consecutive correspondences are established between the last occurrence of the tracked and the reappearing features after being occluded. By reprojecting their 3D object points onto the image planes, occluded locations of these points can be measured. A successfully established non-consecutive correspondence $\mathbf{p}_{j,k} \leftrightarrow \mathbf{p}_{j,k-l-1}$ in the current frame $\mathtt{I}_k$ is a part of a feature trajectory $\mathbf{t}_j^*$ as follows:

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

The object point $\mathbf{P}_j^*$ of $\mathbf{t}_j^*$ is occluded in l frames. It is visible in the current image $\mathtt{I}_k$ and in some previous images $\mathtt{I}_{j,k-l-1}, \mathtt{I}_{j,k-l-2}, \ldots$. It is occluded in the

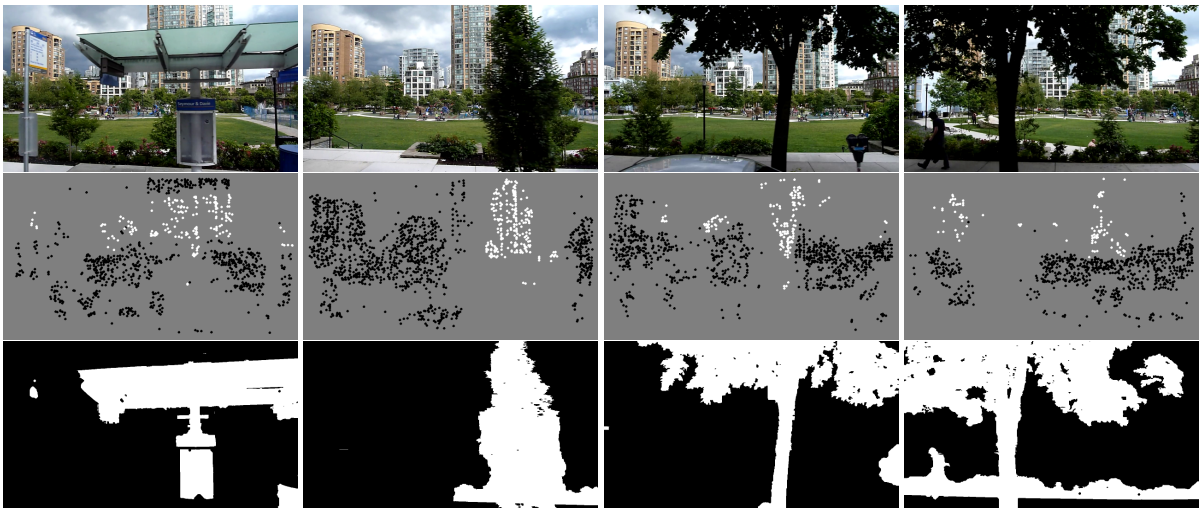

Figure 4: Foreground segmentation results of the *Bus* sequence (1280 × 720 pixels), top row: input sequence; center row: occluded (white) and not occluded (black) object points; bottom row: automatic segmentation of foreground objects using the occlusion information as initialization.

images $\mathtt{I}_{k-1},\ldots,\mathtt{I}_{k-l}$. The image location of the visible or not seen object point in each valid frame can be estimated with relation (1) after selecting the scale factor for the reconstruction.

To verify that the discontinued parts of a trajectory $\mathbf{t}_j^*$ are occluded by foreground objects, a similarity constraint between each point of the trajectory and the current feature point $\mathbf{p}_{j,k}^{visible}$ is evaluated. If the similarity constraint is fulfilled, the object point is visible in the camera view. Otherwise, the reprojection is an occluded image position. As similarity measure, the color histogram in a window around each reprojection $\mathtt{A}_{k-1}\mathbf{P}_j^*, \mathtt{A}_{k-2}\mathbf{P}_j^*, \ldots$ is evaluated. This step is important because non-consecutive feature correspondences may also arise from ambiguities in the image signal caused by repeated texture patterns or image noise. In this case, the assumption that the non-consecutive trajectory is discontinued because of a foreground occlusion would be wrong.

The result of this procedure is shown in Figure 4, center row and Figure 5, top row, respectively. The occluded image locations are visualized as white discs, the visible locations of the non-consecutive correspondences are black. The radius of a disc is set to 5 pel. These images provide the initialization for the segmentation procedure as explained in Section 3.2.

3.2 Foreground Segmentation

The images containing the occlusion information (Figure 4, center row) initialize an efficient image segmentation algorithm (Scheuermann and Rosenhahn, 2011). This algorithm provides the segmentation as the solution of the minimum cut / maximum flow problem on a contracted graph. The minimum cut of this graph minimizes the given energy function consisting of a regional and a boundary term. The regional term depends on regions that are assigned to either foreground or background. This information is usually given by the user marking foreground and background with strokes.

In this paper no user interaction is needed since the *strokes* are derived automatically as described in Section 3.1. The result is the desired foreground segmentation which is shown in Figure 4, bottom row and Figure 5, center row, respectively.

4 OCCLUSION OF VIRTUAL OBJECTS

An often used technique in movie production is the integration of virtual objects into a video. This technique allows the editor for including scene content that has not been there during image acquisition. The required data for this step are accurate camera parameters and a coarse reconstruction of the scene. This is the objective of structure and motion recovery approaches. If the integrated virtual object has to be occluded by real scene content, a segmentation is required, which is usually done manually (Hillman et al., 2010).

Our approach provides automatically segmented foreground regions. These regions have two properties: (1) their scene content temporarily occludes the background scene (see Section 3.1). (2) they are visually homogeneous (see Section 3.2). The resulting

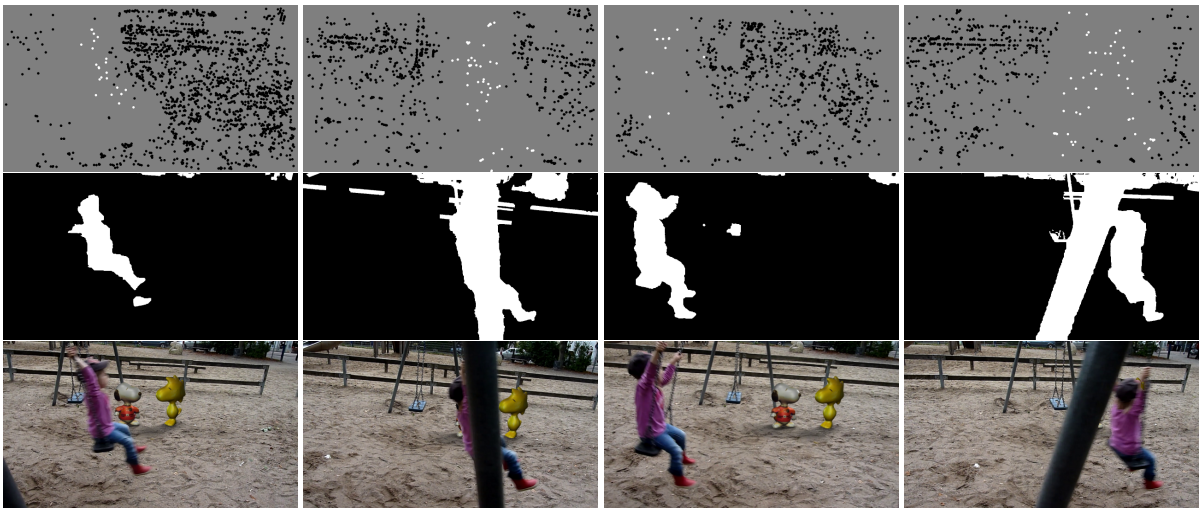

Figure 5: Result examples of *Playground* sequence from Figure 1: Top row: occluded (white) and not occluded (black) object points. center row: segmentation of foreground objects as described in Section 3.1 which is needed for the composition of real and virtual scenes; bottom row: final result of the integration of the virtual objects into the video sequence using the composition of the input sequence from Figure 1, top row and the augmented sequence from Figure 1, bottom row.

segmentation as shown in Figure 4, bottom row, and Figure 5, center row, is used in a compositing step for the occlusion of the augmented objects. The white regions are copied from the input, the black regions are copied from the augmented sequence.

5 EXPERIMENTAL RESULTS

The presented approach of foreground segmentation is tested using footage of a freely moving camera. Two example sequences are demonstrated.

The first sequence (270 frames) is recorded from a driving bus. Several foreground objects such as trees, bushes, signs, and a bus station occlude the background scene temporarily as shown in Figure 4, top row. The center row of Figure 4 shows the extracted occlusion information. The white discs depict foreground locations, the black ones are classified as background locations as described in Section 3.1. These images provide the initialization for the segmentation algorithm (Section 3.2). As shown in the bottom row, arbitrary and complex foreground object are segmented successfully, for example the structure of leaves of the trees.

The second sequence (120 frames) shows a playground scene with a child on a swing. The foreground objects are the swinging child and some parts of the swing rack as shown in Figure 1, top row. The resulting sequences are demonstrated in Figure 5. The occlusion information in the top row results from evaluating the non-consecutive correspondences. Again, the white discs belong to the foreground and the black discs belong to the background. These images initialize the segmentation algorithm, which leads to the foreground segmentation result shown in the center row. In the bottom row, the application of integrating virtual objects into the video sequence is demonstrated. This sequence is the composition of the rendered sequence from Figure 1, bottom row, and the input sequence. The composition is done using the foreground segmentation result. The pixels that are segmented as foreground regions (white pixels) are copied from the input sequence (Figure 1, top row). while the black labeled background regions are copied from the augmented sequence (Figure 1, bottom row).

The swinging child as well as the parts of the swing rack in the foreground are segmented reliably. The integration and the occlusion of the virtual objects is convincing and looks realistic. [1]

The computational expense for the evaluation of the occlusion information is marginal. It consists of reprojections of the object points $\mathbf{P}_j^*$, histogram calculations of their surrounding windows, and the image segmentation which is done in less than a second per image.

5.1 Limitations

Although the foreground is segmented reliably, some background regions are classified as foreground as

[1]The video can be downloaded at: http://www.tnt.uni-hannover.de/staff/cordes/

(a) (b)

Figure 6: Errors resulting from a misleading segmentation. (a): although there is no occlusion information in the fence, the segmentation classifies it to the foreground because of its appearance being similar to the swing rack; (b): although the point is correctly classified as foreground, it is isolated by the segmentation algorithm because of the strong motion blur of the foreground object.

well because of their visual similarity. Figure 6 shows two examples in detail. On the left, a small part of the fence which belongs to the background occlude the augmented objects because of a misleading segmentation. Here, the fence is visually very similar to the part of the swing rack which is a foreground region. On the right, the segmentation algorithm assigns a small part of the child to the background, although it has attached a correctly classified foreground disc. This is due to the strong motion blur. In these cases, the segmentation algorithm leads to suboptimal solutions. As these errors appear for an isolated frame only, we expect to solve this problem by incorporating temporal constraints into the segmentation algorithm which is left for future works. Even in the erroneous frames, the presented approach provides a meaningful initial solution within a few seconds which can easily be refined by adding a few user strokes and restarting the segmentation procedure. Note, that the results presented in this paper are fully automatic.

6 CONCLUSIONS

The presented approach provides the automatic handling of foreground occlusions designed for the application of integrating virtual objects into video. It is demonstrated that the required information for segmenting the foreground regions can be extracted from discontinued feature trajectories and their 3D object points. The information is extracted in a correspondence analysis step for non-consecutive frames. This technique is required for image sequences, in which foreground objects temporarily occlude large parts of the scene.

The presented feature tracking combines the highly accurate and reliable KLT tracker for correspondences in consecutive frames with wide-baseline SIFT correspondences for non-consecutive frames. The localization of occluded and not occluded scene content is gained from the reprojection of object points onto the camera planes. This data is successfully used as initialization of an efficient segmentation algorithm which results in visually homogeneous foreground regions. The resulting segmentation is used for the composition of virtual and real scenes.

The effectiveness of the approach is demonstrated in two challenging image sequences. Virtual object are accurately integrated and their occlusion with foreground objects is convincing. The proposed approach provides an additional step in scene understanding using feature based structure and motion recovery.

REFERENCES

Apostoloff, N. E. and Fitzgibbon, A. W. (2005). Learning spatiotemporal t-junctions for occlusion detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 553–559.

Apostoloff, N. E. and Fitzgibbon, A. W. (2006). Automatic video segmentation using spatiotemporal t-junctions. In *British Machine Vision Conference (BMVC)*.

Brox, T. and Malik, J. (2010). Object segmentation by long term analysis of point trajectories. In Daniilidis, K., Maragos, P., and Paragios, N., editors, *European Conference on Computer Vision (ECCV)*, volume 6315 of *Lecture Notes in Computer Science (LNCS)*, pages 282–295. Springer.

Brox, T. and Malik, J. (2011). Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(3):500 –513.

Cordes, K., Müller, O., Rosenhahn, B., and Ostermann, J. (2011). Feature trajectory retrieval with application to accurate structure and motion recovery. In Bebis, G., editor, *Advances in Visual Computing, 7th International Symposium (ISVC), Lecture Notes in Computer Science (LNCS)*, volume 6938, pages 156–167. Springer.

Cornelis, K., Verbiest, F., and Van Gool, L. (2004). Drift detection and removal for sequential structure from motion algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(10):1249–1259.

Dickscheid, T., Schindler, F., and Förstner, W. (2010). Coding images with local features. *International Journal of Computer Vision (IJCV)*, 94(2):1–21.

Engels, C., Fraundorfer, F., and Nistér, D. (2008). Integration of tracked and recognized features for locally and globally robust structure from motion. In *VISAPP (Workshop on Robot Perception)*, pages 13–22.

Guan, L., Franco, J.-S., and Pollefeys, M. (2007). 3d occlusion inference from silhouette cues. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1 –8.

Hartley, R. I. and Zisserman, A. (2003). *Multiple View Geometry*. Cambridge University Press, second edition.

Hillman, P., Lewis, J., Sylwan, S., and Winquist, E. (2010). Issues in adapting research algorithms to stereoscopic visual effects. In *IEEE International Conference on Image Processing (ICIP)*, pages 17 –20.

Liu, C., Yuen, J., and Torralba, A. (2011). Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(5):978 –994.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110.

Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 674–679.

Matas, J., Chum, O., Urban, M., and Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. In *British Machine Vision Conference (BMVC)*, volume 1, pages 384–393.

Pollefeys, M., Gool, L. V. V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., and Koch, R. (2004). Visual modeling with a hand-held camera. *International Journal of Computer Vision (IJCV)*, 59(3):207–232.

Scheuermann, B. and Rosenhahn, B. (2011). Slimcuts: Graphcuts for high resolution images using graph reduction. In Boykov, Y., Kahl, F., Lempitsky, V. S., and Schmidt, F. R., editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 6819 of *Lecture Notes in Computer Science (LNCS)*. Springer.

Sheikh, Y., Javed, O., and Kanade, T. (2009). Background subtraction for freely moving cameras. In *IEEE International Conference on Computer Vision and Pattern Recognition (ICCV)*, pages 1219–1225.

Thormählen, T., Hasler, N., Wand, M., and Seidel, H.-P. (2010). Registration of sub-sequence and multi-camera reconstructions for camera motion estimation. *Journal of Virtual Reality and Broadcasting*, 7(2).

Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (2000). Bundle adjustment - a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, IEEE International Conference on Computer Vision and Pattern Recognition (ICCV), pages 298–372. Springer.

Zhang, G., Dong, Z., Jia, J., Wong, T.-T., and Bao, H. (2010). Efficient non-consecutive feature tracking for structure-from-motion. In Daniilidis, K., Maragos, P., and Paragios, N., editors, *European Conference on Computer Vision (ECCV)*, volume 6315 of *Lecture Notes in Computer Science (LNCS)*, pages 422–435. Springer.