Automatic Foreground Propagation in Image Sequences for 3D Reconstruction

Mario Sormann[1], Christopher Zach[1], Joachim Bauer[1], Konrad Karner[1], and Horst Bischof[2]

[1] VRVis Research Center, Inffeldgasse 16/II, A–8010 Graz, Austria
   `sormann@vrvis.at`
[2] Institute for Computer Graphics and Vision, Graz University of Technology
   `bischof@icg.tu-graz.ac.at`

Abstract: In this paper we introduce a novel method for automatic propagation of foreground objects in image sequences. Our method is based on a combination of the mean-shift operator with the well known intelligent scissors technique. It is effective due to the fact that the images are captured with high overlap, resulting in highly redundant scene information. The algorithm requires an initial segmentation of one image of the sequence as an input. In each consecutive image the segmentation of the previous image is taken as an initialization and the propagation procedure proceeds along four major steps. Each step refines the segmentation of the foreground object and the algorithm converges until all images of the sequence are processed. We demonstrate the effectiveness of our approach on several datasets.

1 Introduction

Efficient and interactive foreground/background separation of images have become a fundamental part of many applications in computer vision and 3D reconstruction [10]. Obviously manual segmentation is a tedious and time consuming process, especially when applied on a large number of images, as usual needed for a dense 3D reconstruction of complex objects. Therefore many segmentation algorithms have been developed recently [1, 7].

This paper addresses the problem of automatic propagation of a foreground object in a complex environment for 3D reconstruction, whose background can not be removed in a simple way. The key idea of our approach is to take advantage of the high overlap of the images. Essentially we utilize redundant scene information to automate the segmentation procedure and propagate an initial segmentation through the image sequence. The goal of our approach is to achieve a fast, automatic and robust foreground segmentation. Moreover, our method minimizes the expenditure of time to achieve an accurate foreground

segmentation. Its power can be derived from the fact that labelled image sequences simplify the correspondence problem dramatically and therefore, dense 3D reconstruction results of complex objects can be clearly improved.

The main methods to accomplish the propagation procedure are the well known mean-shift technique and the intelligent scissors approach. Intelligent scissors, introduced by Mortenson and Barrett [7], also known as Live Wire or Magnetic Lasso, allows the user to define a precise contour with minimized human interaction, by roughly tracing the objects contour with the mouse. A user selects interactively optimal contour segments by immediately displaying the minimum cost path between the so called current seed point and the previous one, where the current seed point is represented as the position of the mouse cursor. The optimal path is computed via dynammic programming by applying Dijkstra's graph search algorithm [4] to find the optimal spanning tree. The second related technique is mean-shift analysis, which was originally invented by Fukunaga and Hostetler [5] and recently successfully applied to image segmentation and tracking by Comaniciu and Meer [2, 3]. The mean-shift analysis approach is essentially defined as a gradient ascent search for maxima in a density function defined over a high dimensional feature space. The feature space include a combination of the spatial coordinates and all its associated attributes that are considered during the analysis. The main advantage of the mean-shift approach is based on the fact that it considers geometric coordinates and the associated attributes together at the same time.

The remainder of the paper is composed as follows. After a brief overview of our method we discuss the novel parts of the automatic foreground propagation algorithm in section 2. Experimental results and concluding remarks are presented in section 3 and 4.

2 The Automatic Foreground Propagation Algorithm

Our method is a multistage approach to separate a foreground object, for example a statue, from the background in all images of an image sequence. The algorithm requires as input an initial segmentation of one image, which can be obtained by utilizing intelligent scissors [7], GrabCut [9] or other interactive segmentation techniques. In this paper we focus on the propagation of this initial segmentation through all images of the sequence. The propagation task itself is mainly based on a region based matching algorithm. Therefore we segment the image into a certain number of regions. All these regions are classified into three different sets (foreground, background and uncertain regions), illustrated in Figure 2. The final contour can be extracted from these three sets. For dividing the image into regions we employ a mean-shift image segmentation proposed by Comaniciu and Meer [3].

Additionally, to improve the robustness of the propagation procedure, our algorithm requires the relative orientation of the images to be known. The

orientation is determined based on methods described by Horn [6], and Nister [8] and provides both, an accurate orientation and a set of corresponding points.

The workflow of our proposed approach can be roughly seen as the composition of the following consecutive subtasks:

1. Extract an area of interest, which is called initial contour ring, with an inner boundary and an outer boundary.
2. Utilize information acquired from the contour ring and from the corresponding points to identify foreground and background regions.
3. Perform a region based matching algorithm based on mean-shift information to separate the remaining regions in the contour ring in foreground regions, background regions and uncertain regions.
4. Extract true contour segments from adjacent foreground and background regions and utilize intelligent scissors to close uncertain contour segments.

This procedure is repeated until all images of the sequence are processed.

2.1 Initial Contour Ring

The first step in our approach consists of extracting an initial contour ring, which represents an area of interest where we expect to find the true contour. Therefore, the initial contour of the first image is swept along the epipolar lines of the next image. For each position a support function can be formulated as

$$S_c = \sum_{i=1}^{n} g_i(x, y)$$

where $g_i(x, y)$ is the gradient of the contour i and n the number of sweep positions. The position with the highest support function is confirmed and represent the initial contour C_i in the next image. This initial contour C_i contains a set of continuous points and deviates in general slightly from the true contour of the current foreground object. Hence it is necessary to extract a contour ring where we expect to obtain the final contour of the object. Our next step includes an Euclidean distance transform on C_i to compute the contour ring. The scale of the distance transform, respectively the width of the contour ring can be directly derived from the relative orientation of the images, which guarantees that the true contour of the object is within the area of interest.

Such a contour ring, computed in the first step leads to several advantages. First, the following processing steps can concentrate on a smaller number of regions, which increases the performance of the algorithm dramatically. Moreover, the inner and outer boundary of the contour ring can be used to separate foreground regions and background regions with high confidence, which is described in more detail in the next section. Lastly, the reduction of the search area reduce the error propagation.

2.2 Prior Information

As previously outlined, different prior information is incorporated, to simplify the consecutive tasks and improve the robustness of our approach. We distinguish between two important types of information:

1. Information provided by the contour ring.
2. Information provided by the corresponding points.

This information is used to label foreground regions and background regions with high confidence. In the former case the separation can be directly derived from the inner and outer boundary of the contour ring, which is illustrated in Figure 1.
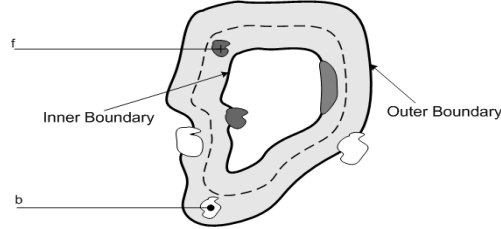

Fig. 1. Illustration of utilized information. Each region, which is adjacent to the inner boundary is labelled foreground (dark grey) and similar is applied for the outer boundary regions (white). Additionally **f** indicates a foreground correspondence, whereas **b** identifies a background correspondence, both acquired from the previous segmentation.

Consequently we label all mean-shift regions which are directly connected to the inner boundary of the ring to foreground and those connected to outer boundary as background. A similar procedure is performed for the information which is provided by the corresponding points. Here we take advantage of the direct relationship of correspondences and simple separate foreground regions from background regions by comparing their location in the previous image against the already segmented contour.

2.3 Extended Region Matching

As mentioned before, we use the well known mean-shift algorithm to segment the image into a set of regions. So far we have already classified some of the mean-shift regions in our area of interest. For the remaining regions we perform an region based matching algorithm against the previous segmented image.

Basically our region matching algorithm works as follows: A matching between two regions r_i of the previous image and r_j of the current image is assigned with a similarity measure $S_{i,j}$. The similarity measure $S_{i,j}$ is based

on the mean-shift parameters and the known relative orientation. Currently three different types of similarity measures are formulated. The first similarity measure S_{LUV} is represented by the LUV values of the mean-shift region, where L encodes luminance, and U and V encode color information. The other two similarity measures can be derived from the relative orientation. First, S_{Epi} encodes the distance of the epipolar line from region r_i to the center of gravity of region r_j. Second, the similarity measure S_{Corr} is composed from the distance of the nearest corresponding point to region r_i respectively to region r_j. The final distance function for two regions is formulated as:

$$d(r_i, r_j) = \omega_1 * S_{LUV} + \omega_2 * S_{Epi} + \omega_3 * S_{Corr}$$

where $\omega_1...\omega_3$ are weights to control the influence of the different similarity measures. We can distinguish between foreground regions, background regions and uncertain regions, by evaluating the introduced distance function for each remaining region against a user defined threshold. Uncertain regions are regions, which can be classified neither to foreground nor to background. In this case a further processing is necessary.

2.4 Foreground Extraction

The aim is to extract the final foreground object from previously labelled mean-shift regions. Obviously, the true contour lies between adjacent foreground regions and background regions or intersects an uncertain region. In the former case the final contour can be extracted with simple neighbourhood checks, whereas in the latter case the intelligent scissors algorithm is applied. Figure 2 illustrates the extraction of needed start and end points to initiate the intelligent scissors procedure.
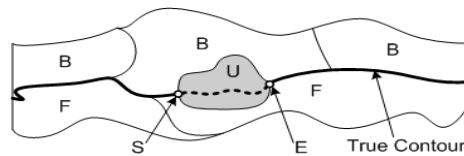

Fig. 2. Set of regions including foreground regions (F), background regions (B) and uncertain regions (U) and the highlighted true contour. Furthermore an illustration of start point (S) and end point (E) to automatically apply intelligent scissors.

Finally all obtained contour segments are combined to a closed continuous contour of the foreground object.

3 Experimental Results

All presented image sequences were taken with a calibrated high quality digital consumer camera with a 11.4 megapixels CMOS sensor. In a first evaluation

we used an image sequence consisting of 12 images of a garden gnome which is approximately 23cm tall with a diameter of 10cm. Figure 3 shows the garden gnome with the overlayed segmentation, whereas Figure 4 demonstrates all intermediate results of our method.

Fig. 3. Illustration of five images with overlayed segmentation of the garden gnome image sequence. The garden gnome is approximately 23cm tall with a diameter of 10cm. The last image illustrates the obtained 3D reconstruction.

Figure 5 illustrates a more complex dataset consisting of 12 images showing a statue of St. Barbara. The statue is 55cm tall with a diameter of 13cm at the pedestal. As shown in Figure 5, automatic approaches will sometimes lead to incorrect results. In our method, if the segmentation result is not satisfactory, a user has the possibility to correct a miss-segmentation, by manual assignment of the critical mean-shift regions or by an assisted intelligent scissors algorithm.

Finally, Figure 6 demonstrates the usability of our approach on a real-world dataset, which depicts a statue on the roof of the Austrian National Library.

4 Conclusion and Future Work

We have developed an automatic foreground propagation method that performs well in terms of accuracy, robustness and efficiency. Our approach takes advantage of the redundant scene information, which is typically provided from image sequences for 3D reconstruction. The primary purpose of our method is the improvement of our 3D reconstruction results. Moreover, the tedious process of an interactive segmentation of all images is dramatically reduced, thus our method simply requires one initial segmentation.

Though the results are very promising, there are several improvements that can be made to our approach. In order to achieve more accurate results we are currently working on extending the similarity measures introduced by extended region matching. Another consideration is to utilize active contour models to extract foreground objects with sub-pixel accuracy.
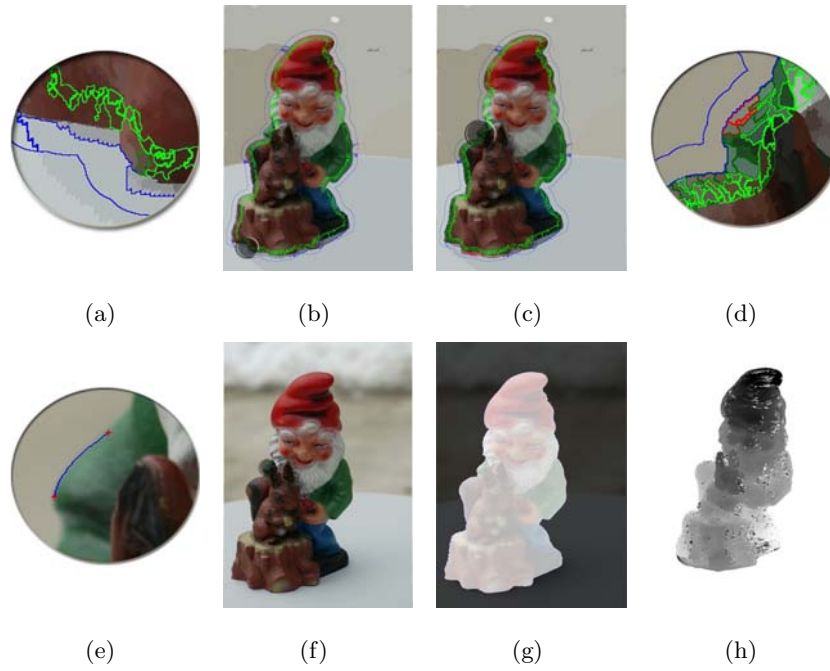
(a)　　　　　　(b)　　　　　　(c)　　　　　　(d)

(e)　　　　　　(f)　　　　　　(g)　　　　　　(h)

Fig. 4. Intermediate results of the automatic foreground propagation algorithm illustrating one image of the garden gnome image sequence. **(a)** Close-up from **(b)** showing labelled foreground (green) and background (blue) regions in the contour ring after incorporating prior information. **(d)** Close-up from **(c)** illustrating labelled foreground (green), background (blue) and uncertain regions (red) after applying extended region matching. **(e)** Close-up from **(f)** showing start and end point (red crosses) of intelligent scissors and the obtained true contour (blue) of a uncertain region. **(g)** Garden gnome image with the final segmentation overlayed. **(h)** Illustration of the achieved 3D reconstruction represented as depth map.

Fig. 5. Four images showing a statue of St. Barbara and achieved segmentation. The Barbara statue is approximately 55cm tall with a diameter of 13cm. One image of the sequence illustrates a small miss-segmentation, which can be corrected by human assisted intelligent scissors. The last images consists of our obtained 3D reconstruction result.

Fig. 6. Four images of a statue on the roof of the Austrian National Library and obtained propagation results. The first image illustrates the achieved mean-shift segmentation.

5 Acknowledgements

References

1. BOYKOV, Y., AND JOLLY, M. P. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *International Conference of Computer Vision* (Vancouver, Canada, July 2001), vol. 1, pp. 105–112.
2. COMANICIU, D., AND MEER, P. Mean shift analysis and applications. In *International Conference of Computer Vision* (Corfu, Greece, June 1999), vol. 2, pp. 1197–1203.
3. COMANICIU, D., AND MEER, P. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 5 (May 2002), 603–619.
4. DIJKSTRA, E. W. A note on two problems in connexion with graphs. In *Numerische Mathematik*, vol. 1. Mathematical Centre, Amsterdam, The Netherlands, 1959, pp. 269–271.
5. FUKUNAGA, K., AND HOSTETLER, L. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory 21*, 1 (January 1975), 32–40.
6. HORN, B. Relative orientation. *International Journal of Computer Vision 4*, 1 (January 1990), 59–78.
7. MORTENSON, E. N., AND BARRETT, W. Intelligent scissors for image composition. *Graphical Models and Image Processing 60*, 5 (September 1998), 349–384.
8. NISTER, D. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*, 6 (June 2004), 756–777.
9. ROTHER, C., KOLMOGOROV, V., AND BLAKE, A. Grabcut - interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics 23*, 3 (August 2004), 309–314.
10. ZIEGLER, R., MATUSIK, W., PFISTER, H., AND MCMILLAN, L. 3d reconstruction using labeled image regions. In *Eurographics/ACM SIGGRAPH symposium on Geometry processing* (Granada, Spain, September 2003), vol. 1, pp. 248–259.