STEREO-BASED IMAGE AND VIDEO ANALYSIS FOR MULTIMEDIA APPLICATIONS

M. Gelautz *, E. Stavrakis, M. Bleyer

Institute for Software Technology and Interactive Systems, Vienna University of Technology,
Favoritenstrasse 9-11/188/2, A-1040 Vienna, Austria - (gelautz, stavrakis, bleyer)@ims.tuwien.ac.at.
http://www.ims.tuwien.ac.at

KEY WORDS: Vision Sciences, Stereoscopic, Multimedia, Video, Reconstruction, Vision, Graphics, Rendering

ABSTRACT:

In this paper we utilize stereo images and video frames of real scenes to automatically generate artistic stereoscopic views with a hand-painted appearance ("painterly rendering"). The approach is motivated by contemporary artists who painted stereo image pairs on canvases. Although a variety of monoscopic painterly rendering algorithms have been proposed in the literature, the design of painterly rendering algorithms suited to stereo views of real scenes presents a largely unexplored field.
In our approach, we use stereo analysis to generate a disparity map which is then employed by the stereoscopic rendering algorithm to preserve coherence between the brush strokes of the two stereo views. The painting of the occluded regions is incorporated as a separate step into the rendering procedure. We utilize the disparity map to prevent paint spilling between surfaces located at different depths. Stereoscopic viewing of the stereo paintings produced by our algorithm demonstrates that the aesthetic impression resulting from the hand-painted appearance of the individual images has been enhanced with an additional perception of depth.

1. INTRODUCTION AND MOTIVATION

The 3D reconstruction of image and video scenes is an important requirement in many multimedia applications, including video editing, compositing, and the generation of artificial images based on real image content ("image-based rendering"). In this paper, we focus on a non-photorealistic rendering (NPR) application which uses an input stereo pair of images from a real scene to produce painterly rendering effects on stereo views, similar to artwork painted by hand. The computer-generated paintings can be viewed stereoscopically, which combines the artistic rendering effects of the individual images with the additional sense of depth.

1.1 Non-photorealistic Rendering

Over the past years, the development of techniques for non-photorealistic rendering (NPR) - as opposed to traditional photorealistic rendering - has received increasing attention in computer graphics research (Gooch and Gooch, 2001). Studies on computer-generated painterly effects include the work by (Litwinowicz, 1997), which deals with the automated generation of impressionist paintings. A comprehensive survey on stroke-based rendering techniques is presented by (Hertzmann, 2003). The combination of painterly rendering methods with computer vision techniques has been addressed by (Shiraishi and Yamaguchi, 2000) and (Gooch et al., 2002). The authors of the latter study compute the medial axis of segmented objects to guide the automated creation of brush strokes. As an extension, they suggest the incorporation of depth information into their stroke-based rendering system, which can, for example, assist the segmentation.

Most non-photorealistic rendering work that utilizes depth maps relies on synthetic depth maps computed from a 3D model. However, our approach makes use of stereo analysis to process images of real scenes, which usually provide richer and more visually interesting content than synthetic images rendered from 3D scene models.

1.2 Artistic Rendering of Stereo Views

Although not widely known, traditional stereoscopic paintings have been created by several well-known artists including René Magritte ("Man with Newspaper", 1928) and Salvador Dalí (Maur, 1989). (Ferragallo, 1972) promoted the idea of stereoscopic painting in his "Manifesto Directed to the New Aesthetics of Stereo Space in the Visual Arts and the Art of Painting".

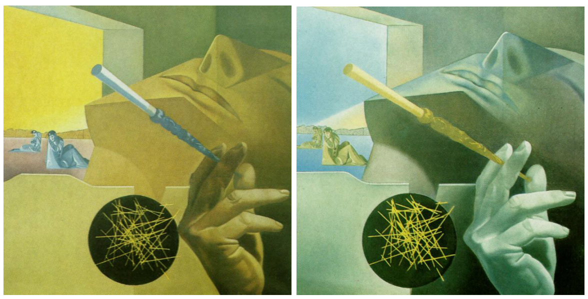
A stereo image pair painted by Dalí is shown in figure 1. When comparing the left and right view, one can clearly recognize the stereo parallaxes between the two images (see, e.g., the shift of the cigarette with respect to the head and window in the background). Interestingly, Dalí used different colors (warm and cool tones) for the left and right image, which fuse into a new color impression on the stereo view. For his 3D paintings, which combine real and surreal elements, he created the term "Metaphysical Hyperrealism" (Maur, 1989).

The manual creation of stereo paintings is a labor intensive task, since it requires the artist to reproduce the same composition twice from different viewpoints. Some painters used stereo photography to base their compositions on; others restricted themselves to geometric forms. We believe that the technical requirements and excessive effort associated with creating stereoscopic paintings were principal reasons why the technique has not become widespread. The potential of computer vision techniques to overcome these limitations provides a major motivation for our study.

When devising painterly rendering algorithms that operate on stereo images of real scenes, several requirements need to be taken into account. An obvious solution to computer-generated

* Corresponding author

stereo painting would be to apply the same single image rendering technique to both images individually. However, this will usually not produce satisfactory results. Depending on the paint style, individual painting of the two images may result in non-coherent stroke patterns that do not fuse well stereoscopically. Furthermore, brush strokes that cross the borders of an object can cause the color to spill to adjacent objects located at different depths. While intersurface paint spilling does not necessarily degrade the quality of a single painting, in stereo painting it becomes a more noticeable and mostly unpleasant effect. These and other related effects are still largely unexplored.



(a) left image (b) right image

Figure 1. Example of a stereoscopic painting by Salvador Dalí ("The sleeping smoker", 1972/73).

In this study, we propose a stereoscopic painterly rendering algorithm that uses a stereo-derived depth map to prevent paint spilling and to preserve coherence between the brush strokes of the two stereo views. We also tackle the related problem of occlusions. The algorithm is described in section 2. In section 3, we show experimental results obtained from a stereo benchmark pair and self-recorded video frames. Suggestions for future work are given in section 4.

2. ALGORITHM

2.1 Stereo Analysis

We developed a stereo matching algorithm that uses a color-based image segmentation for the precise location of depth discontinuities and the superior handling of large untextured regions. Whereas the majority of existing stereo matching algorithms were originally tailored to grey-value images, color images play an important role in many multimedia applications, especially computer-generated art. The need to deal with untextured regions of significant extent is illustrated by figure 1. Precise localization of depth discontinuities is required to prevent paint spilling, as described in the introduction.

The matching algorithm can be described as follows. We first apply color segmentation to the reference image and use a planar model to represent the disparity inside the derived segments. This approach is based on the assumption that for regions of homogeneous color the disparity varies smoothly and depth discontinuities coincide with the boundaries of those regions (Zhang and Kambhamettu, 2002), which holds true for most natural scenes. We compute an initial disparity map using window-based correlation with additional consistency checks to determine the reliability of the depth values computed for each segment. The initial disparity map is used to derive a planar model of each segment. The segments are grouped into depth layers by a mean-shift clustering algorithm. The layer assignment is refined iteratively using a global cost function. The quality of the current assignment is measured by projecting the reference image into the geometry of the second view and evaluating the difference between the projected and original pixel values in the second image. Besides pixel dissimilarity, the cost function also penalizes discontinuities and occlusions. Since the problem of finding an exact solution to the minimization problem can be shown to be np-complete, we employ an efficient greedy search strategy to find a local optimal solution. More details on the algorithm along with results obtained in benchmark tests are given in (Bleyer and Gelautz, 2004).

2.2 Stereoscopic Painterly Rendering

The proposed stereo painting algorithm is an extension of Hertzmann's single image painting algorithm (Hertzmann, 1998) that was designed for the image-based generation of hand-painted effects. The authors use a multiresolution approach, in which layers of successively finer strokes are placed on top of each other. The brush strokes are computed as B-splines that follow the isocontours of the original image. Each stroke is rendered by dragging a circular brush mask along the spline. The color of the stroke is derived from a blurred source image, which also serves as underpaint for the hierarchical refinement of the computer-generated strokes. Finer strokes are applied to regions of high-frequency content, whereas in relatively uniform areas the coarser strokes of previous layers of paint are preserved. For a more detailed description of the spline planning and rendering procedure, the reader is referred to the original paper by (Hertzmann, 1998).

Our modifications to the algorithm make use of the stereo-derived depth map to handle depth discontinuities and occlusions during the painting procedure as follows.

Depth discontinuities: When painting the reference image we use the depth map to prevent brush strokes from painting across depth discontinuities in order to avoid paint spilling between different layers of depth. As we calculate the control points for the B-splines, we check for each control point whether a significant depth discontinuity, according to a pre-defined threshold, is located in its vicinity, which is defined by the current size of the circular brush. If such a discontinuity is encountered, the stroke is terminated. This adds a fourth constraint to the three termination criteria (brush length, color, and magnitude of image gradient) that are already part of the original algorithm. Our tests in section 4 will demonstrate that the object boundaries are well preserved using this depth constraint.

Occlusions: We employ the depth map to project the rendered strokes from the reference view into the geometry of the second view in order to preserve coherence between the two images of the stereo pair. Those areas in the second view which are not visible in the reference view require special consideration. For the sake of simplicity, we refer to these regions as "occlusions", although the gaps can also be caused by other conditions (e.g., missing information along image borders due to the spatial displacement between the two cameras). A straightforward solution to filling the occlusion gaps would be to interpolate the missing information from the projected neighboring pixels or to employ the color values of the underpaint (i.e. the blurred original image) at those locations. While this may work for minor gaps, the resulting artifacts are clearly visible on larger regions and disturb the viewer. In our approach, we propose initially painting the occluded regions of the second view using the same painting algorithm as for the reference view, and then superimposing the projected pixel values of the non-occluded regions onto the paint applied to the occlusion gaps. When painting the occluded regions, we render those strokes whose starting points lie within the occluded regions. We allow the strokes to extend beyond the occlusion boundaries, since confining them to the occlusion gaps can change the stroke characteristics perceived by the user. We also found that a morphological dilation of the occlusion map prior to painting could improve the results.

In Hertzmann's original algorithm, the strokes are rendered in a random order, to prevent an undesirable appearance of regularity on the final painting. In our implementation, we use randomization within individual depth layers, and order the layers according to decreasing depth. This would support a future implementation of different levels of detail, rendered with coarser and finer stroke sizes, depending on the distance of an object from the viewer. Generally, regions rendered with many small brush strokes tend to attract the attention of the viewer. Whereas our current implementation focuses attention on image regions containing high-frequency information (i.e., fine image details), the depth layer implementation might be used to emphasize foreground elements.

3. TESTS AND RESULTS

We performed tests on both benchmark images obtained from the Middlebury Stereo Vision website (Middlebury, 2004) and self-recorded data. We captured our video frames in 24 bit RGB format using two Dragonfly IEEE-1394 video cameras (Pointgrey, 2004) in a stereo configuration. The images were calibrated according to (Zhang, 2000) and projected into epipolar geometry.

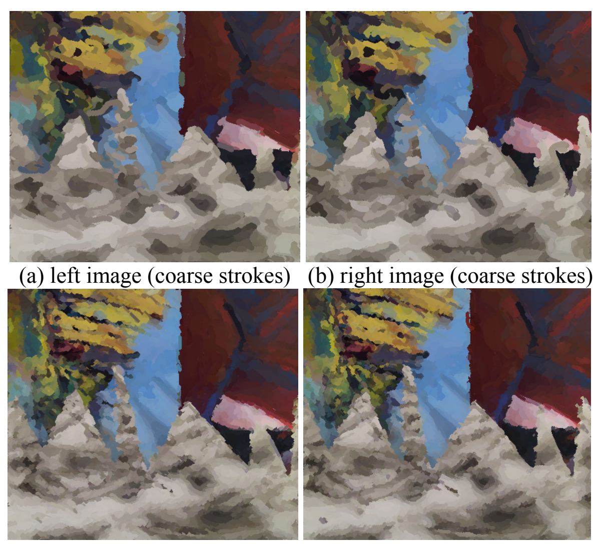
Figures 2 through 5 demonstrate the application of the proposed stereoscopic painting algorithm to the *Sawtooth* test set. The initial stereo image pair is given in figure 2. We started our stereoscopic rendering tests by first applying an implementation of Hertzmann's original algorithm to the left and right image individually. We chose a painting style that imitates work by impressionist artists. Figures 3 (a) and (b) show the results obtained after painting two consecutive layers with a circular brush of 8 and 4 pixels radius, respectively. Since the depth discontinuities are not taken into account in Hertzmann's original implementation, the object contours are not well preserved. Figures 3 (c) and (d) show the same original images rendered with finer brush strokes. In this case, the effect of paint spilling is less pronounced, but still present. In figure 3, one can also recognize the loss of coherence between the brush strokes computed for the left and right stereo view. This effect becomes more pronounced with larger brush sizes and increasing geometric dissimilarity between the images of the original stereo pair.

We employed our segmentation-based stereo matching algorithm described in section 2.1 to compute the corresponding disparity map, which is shown in figure 4 (a) in the geometry of the left image. For comparison, the ground truth for this data set can be seen in figure 4 (c). Visual comparison of (a) and (c) indicates a very good quality of the stereo-derived disparities in (a). This observation was confirmed by a quantitative analysis. The evaluation testbed provided by (Scharstein and Szeliski, 2002) found for our disparity map a percentage of 0.2 % "bad" pixels (i.e., unoccluded pixels whose absolute disparity error is greater than 1). At the time of writing this paper, this error rate resulted in a first rank of our matching algorithm for the *Sawtooth* image pair among 30 algorithms listed on the stereo evaluation website (Middlebury, 2004). The high percentage of correctly matched pixels is reflected by the almost perfect reconstruction of the depth layer boundaries in figure 4 (a).

We projected the computed depth map into the geometry of the right view and determined those pixels in the right view that are not visible in the left view. Figure 4 (b) shows the location of the occluded regions marked in white. These areas are painted in a separate step in our modified version of Hertzmann's algorithm, since paint cannot be propagated from the left view. The painting of the occlusion gaps from figure 4 (b) is illustrated in figure 4 (d).

(a) left image (b) right image

Figure 2. *Sawtooth* stereo pair (size 436 x 380 pixels) from the Middlebury Stereo Vision website.

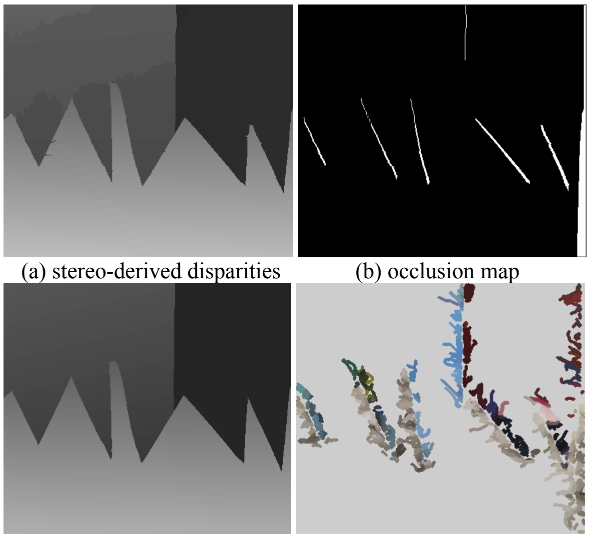


(a) left image (coarse strokes) (b) right image (coarse strokes)

(c) left image (fine strokes) (d) right image (fine strokes)

Figure 3. Results obtained by painting the stereo images from figure 2 separately using Hertzmann's original algorithm. The images in the top row were rendered using two layers of paint with stroke radii of 8 and 4 pixels. An additional layer of paint with a brush stroke radius of 2 pixels was applied to generate the refined version shown in the bottom row.

The results of the proposed stereo painting algorithm can be seen in figure 5. In contrast to figure 3, depth discontinuities and coherence between the two images are now well preserved. One can recognize that the painting of the occluded regions in figures 5 (b) and (d), which was computed in a separate step, blends well into the surrounding regions that were projected from the left view. The pink region in the right central part of the image is not a separate object, as indicated by the depth map. Its edges were therefore correctly overpainted by the rendering algorithm.



(a) stereo-derived disparities (b) occlusion map

(c) ground truth disparities (d) strokes related to occlusions

Figure 4. Disparity maps, occlusion map, and strokes emanating from the occluded regions for the *Sawtooth* set from figure 2.

Similar results were obtained from tests with other stereo pairs, as shown in figures 6 through 9. In figures 6 and 7, the top row gives the original stereo video frames (size 640 x 480 pixels) that we captured in our lab. The corresponding disparity maps produced by stereo matching can be seen in subfigure (c). The occluded regions are marked in subfigure (d). For these two data sets, the occluded regions are larger than in the previous *Sawtooth* example. Nevertheless, no obvious artifacts are visible along the occlusion boundaries in the final results (see figures 8 and 9), which demonstrates the usefulness of the proposed approach.

We displayed the painted images stereoscopically on both conventional computer monitors and a large Baron BARCO stereo table (Barco, 2004), and viewed them using active shutter glasses. The images were presented to several users. While most of the users were not aware of stereoscopic painting, they found the resulting paintings appealing and reported a better immersion into the artwork due to the new sense of depth. We varied the brush parameters to produce images with finer and coarser strokes, the latter making the painterly effect more noticeable. We found that the users tended to prefer coarser brush strokes on the stereoscopic presentation than when viewing the same image monoscopically. More tests would be required to examine whether this effect can be related to a reduced sensitivity of the eye due to stereoscopic fusion.

For comparison, we also showed the stereo views that were generated individually using Hertzmann's original algorithm (see figure 3). Several of the test persons reported difficulties in fusing the non-coherent strokes, especially on coarser paintings. As regards computation time, our algorithm avoids re-calculating the strokes for large parts of the second image and thus achieves a speed-up over the straightforward creation of two separate paintings, which becomes more significant as the image size increases. For an image of size 1600 x 1200 pixels, we measured a performance improvement of 13 %.

4. SUMMARY AND OUTLOOK

We have proposed a method for rendering artistic views of stereo images from real scenes. The stereo paintings produced by our algorithm give the impression of a painted world, in which the brush strokes are attached to objects.

As the next step, we plan to render stereo views in different painting styles and present them to a larger group of test persons, in order to learn more about the special requirements of stereoscopic painterly rendering. Furthermore, we will investigate the effect of lower-quality disparity maps on the painting result.

Our current implementation processes stereo videos on a frame-by-frame basis. For stereoscopic painterly animation, we plan to extend the algorithm using feature tracking in order to preserve both spatial and temporal coherence between brush strokes in consecutive stereo video frames.

ACKNOWLEDGEMENTS

This study has been supported by the Austrian Science Fund (FWF) under project P15663. We wish to thank "Verwertungsgesellschaft Bildender Künstler (VBK)", Austria, for permission to reproduce "The sleeping smoker".

REFERENCES

Barco, 2004. http://www.barco.com (accessed 14 Apr 2004).

Bleyer, M. and M. Gelautz, 2004. A layered stereo algorithm using image segmentation and global visibility constraints. Submitted to *ICIP'04*, Oct 2004, Singapore.

Ferragallo, R., 1974. On stereoscopic painting. *Leonardo*, 7(2), pp. 97-104.

Gooch, B. and A. Gooch, 2001. *Non-photorealistic rendering*. A. K. Peters Ltd, Natick, MA, United States, 250 pp., 2001.

Gooch, B., G. Coombe, and P. Shirley, 2002. Artistic vision: painterly rendering using computer vision techniques. In: *Proc. NPAR'02*, ACM Press, pp. 83-90.

Hertzmann, A., 1998. Painterly rendering with curved brush strokes of multiple sizes. In: *Proc. Siggraph'98*, ACM Press, pp. 453-460.

Hertzmann, A., 2003. A survey of stroke-based rendering. *IEEE Computer Graphics and Applications*, 23(4), pp. 70-81.

Litwinowicz, P., 1997. Processing images and videos for an impressionist effect. In: *Proc. Siggraph'97*, pp. 407-414.

Maur, K., 1989. *Salvador Dalí*. Verlag Gerd Hatje, Stuttgart, Germany, pp. 398-405.

Middlebury Stereo Vision Website, 2004. http://www.middlebury.edu/stereo (accessed 14 Apr 2004).

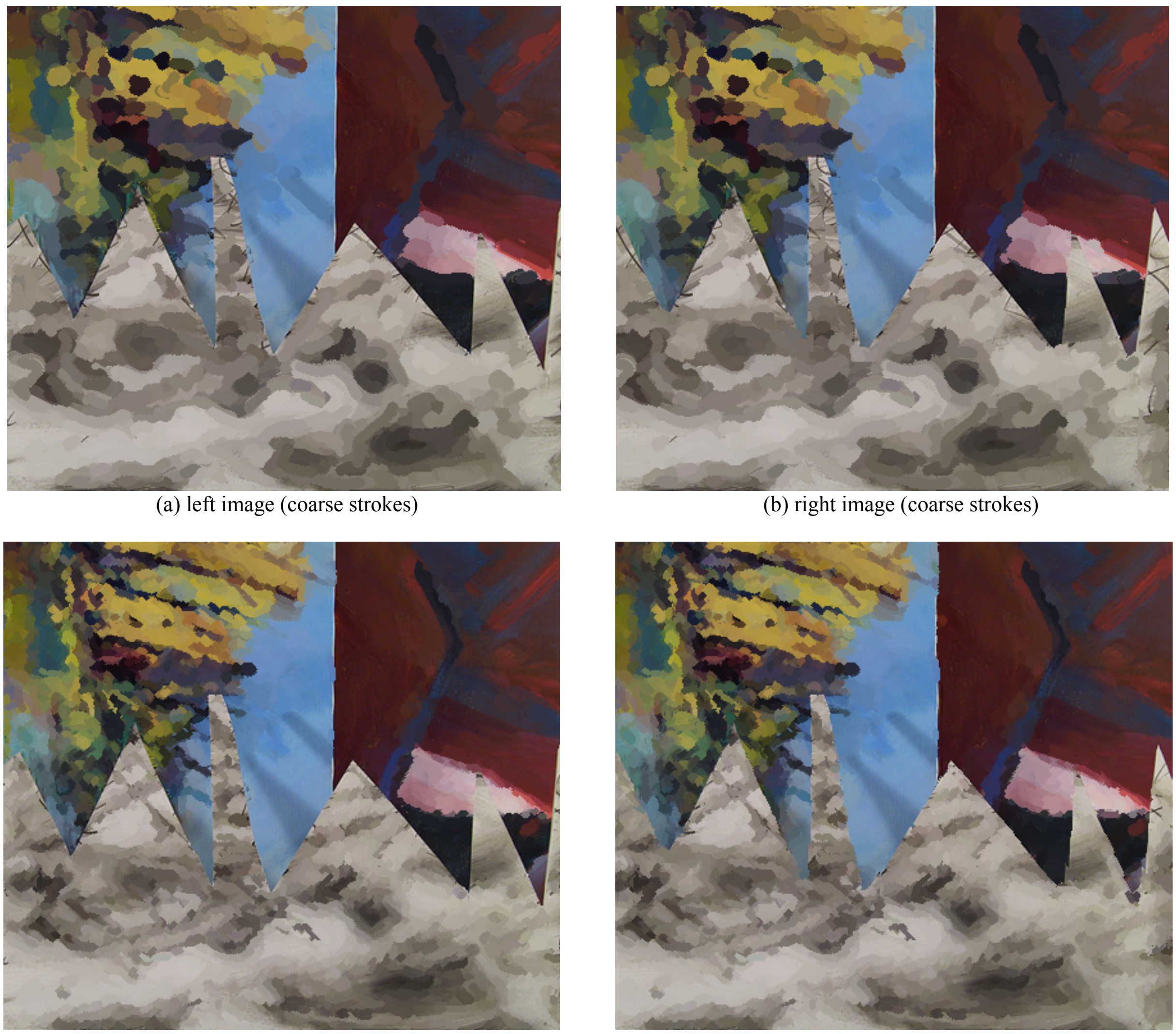
Pointgrey Research Inc., 2004. http://www.ptgrey.com (accessed 14 Apr 2004).

Scharstein, D. and R. Szeliski, 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 22(1/2/3), pp. 7-47.

Shiraishi, M. and Y. Yamaguchi, 2000. An algorithm for automatic painterly rendering based on local source image approximation. In: *Proc. NPAR'00*, ACM Press, pp. 53-58.

Zhang, Y. and C. Kambhamettu, 2002. Stereo-based matching with segmentation-based cooperation. In: *Proc. ECCV 2002*, pp. 556-571.

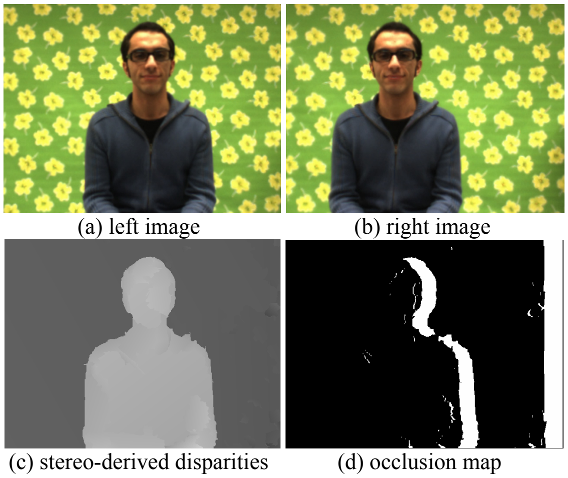
Zhang, Z., 2000. A flexible new techniques for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), pp. 1330-1334.



(a) left image (coarse strokes) (b) right image (coarse strokes)

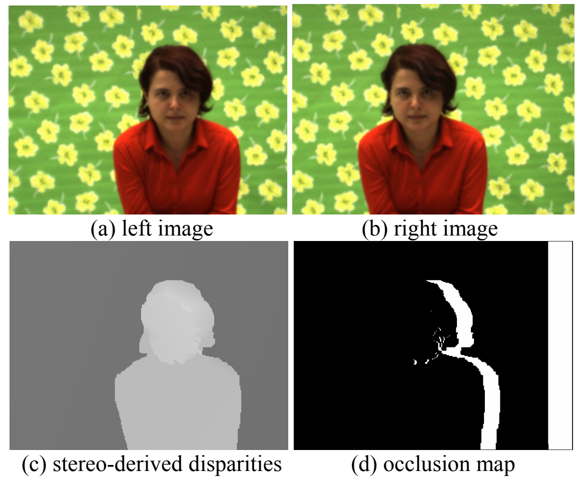
(c) left image (fine strokes) (d) right image (fine strokes)

Figure 5. Results obtained by applying the stereoscopic painting algorithm to figure 2. For the images in the top row, two layers of paint with brush stroke radii of 8 and 4 pixels were applied. The finer strokes in the bottom row were generated using three layers of paint with brush stroke radii of 8, 4, and 2 pixels.



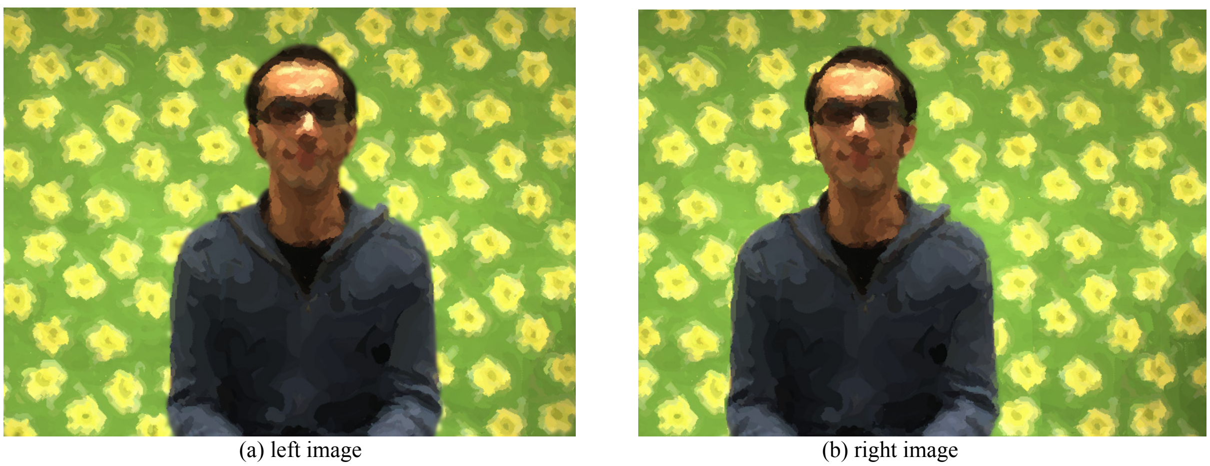
(a) left image (b) right image
(c) stereo-derived disparities (d) occlusion map

Figure 6. Stereo frame pair *Stathis* with stereo-derived disparities in the geometry of the left image and occlusion map in the geometry of the right image.



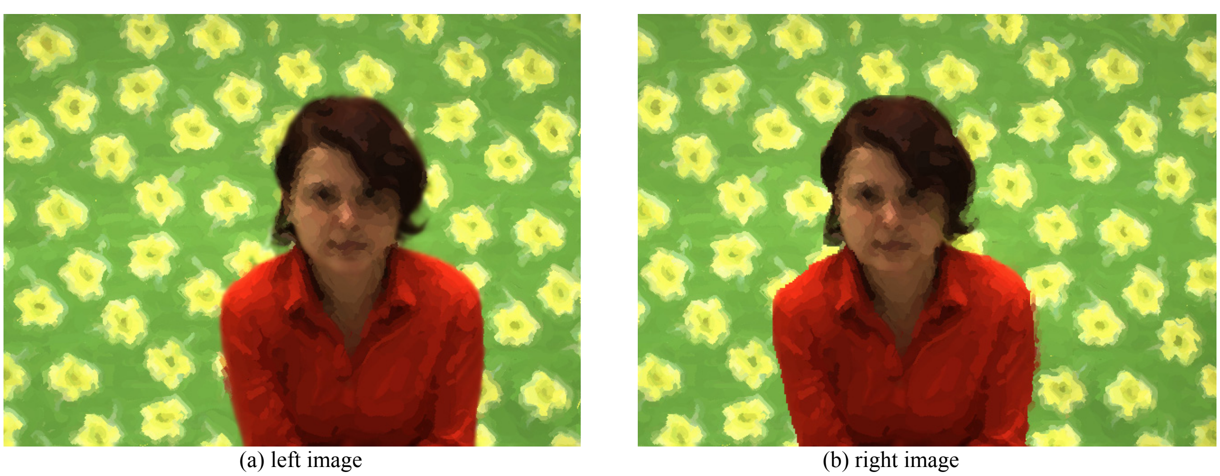
(a) left image (b) right image
(c) stereo-derived disparities (d) occlusion map

Figure 7. Stereo frame pair *Danijela* with stereo-derived disparities in the geometry of the left image and occlusion map in the geometry of the right image.



(a) left image (b) right image

Figure 8. Final stereo painting corresponding to figure 6. Three consecutive layers of paint with brush sizes of 8, 4, and 2 pixels were applied.



(a) left image (b) right image

Figure 9. Final stereo painting corresponding to figure 7. Three consecutive layers of paint with brush sizes of 8, 4, and 2 pixels were applied.