

Highlight Substitution in Light Fields

Florian Vogt*, Dietrich Paulus†, Heinrich Niemann

Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg
Martensstr. 3, 91058 Erlangen, Germany
{vogt,paulus,niemann}@informatik.uni-erlangen.de

ABSTRACT

Highlights occur especially when recording medical (color) images during micro-invasive operations. They disturb the physicians who can sometimes only guess the tissue at the position of the highlights. In this contribution we present a new technique of highlight removal. A so-called *light field* is generated from the recorded image sequence. Then a binary highlight mask is computed for each image and used as *confidence map* for the light field pixels. The result is a light field in which pixels at highlight positions are interpolated by pixels which were not over-imposed by highlights. This leads to light fields with better images. We demonstrate and evaluate the technique on medical and synthetic image sequences.

1. INTRODUCTION

Highlights due to specular reflections may considerably disturb observers of images or image sequences, respectively. When regarding medical images, especially endoscopic images, the problem is even increased because light source and viewing direction are almost identical; thereby, wet tissue surfaces perpendicular to the viewing direction show highlights and the physicians can only guess the tissue at that position.

In this contribution we show how highlights can be substituted in image sequences when a light field [1] is created first, that is subsequently used to enhance image quality at locations, where the input images show defects. The main advantage of our technique is the substitution of highlight pixels by *real* intensity values which are obtained by interpolations of pixels which were not over-imposed by highlights. Note that all detectable image degradations could be substituted by our method, provided a robust detection method for the degradation is known for individual images.

*This work was funded by Deutsche Forschungsgemeinschaft (DFG) under grants SFB 603/TP B6 and SFB 603/TP C2. Only the authors are responsible for the content.

†Author's new address: Universität Koblenz-Landau, Institut für Computervisualistik, Rammsweg 1, 56016 Koblenz

2. HIGHLIGHT DETECTION

For di-electric inhomogeneous material, a model for separating specular reflectance from diffuse reflection exists [2], the so-called di-chromatic reflectance model. Algorithms based on this model have been applied, e.g., to remove highlights for stereo vision [3]. In [4], color gradients are used to detect highlights. Although human tissue does not fit the model of di-electric inhomogeneous material, these methods have been applied to detect highlights for biological material in several articles (e.g. [5]). We compared these methods to simple thresholds in *HSV* color space (Figure 1). As the differences are small we used simple *HSV* thresholds.

3. LIGHT FIELD GENERATION AND VISUALIZATION

Light fields [6, 7] in general describe a sampled set of the plenoptic function [8] which is suitable for the generation of novel views of a scene. The plenoptic function measures the outgoing radiance of a specific wavelength at every point in space in any direction at any point in time. By assuming a rigid scene and only measuring radiance along rays, the seven dimensions of the plenoptic function can be reduced to four.

A major challenge is the construction of a light field from a real (rigid) scene from an arbitrary captured image sequence of the scene [9]. The estimation of the intrinsic and extrinsic camera parameters for each frame of the sequence is called calibration. Most approaches are based on the knowledge of point correspondences between neighboring views in the image sequence. Particularly when light fields are generated from endoscopic sequences with low resolution and low signal-to-noise ratio, a highly robust approach is required to solve the correspondence problem. As we have continuous image streams, a corresponding point in one frame appears close to the corresponding location in the previous (and next) frame. This property is exploited by differential tracking approaches, e.g. [10] which is applied here. Knowing corresponding point features, a lot of mathematical methods exist for computing camera motion



Fig. 1. Comparison of highlight detection methods (image of a gall): original image (left), *HSV* thresholds (center), regions computed from color gradients with subsequent filling of closed contours (right).

from projections assuming a rigid scene [11]. To improve the light field quality and increase robustness, the intrinsic camera parameters of the endoscope are determined in advance using a calibration pattern [12] and have therefore not to be estimated. We reverse radial distortions which are considerable for the wide angle lenses used in endoscopic imaging. The extrinsic camera parameters (translation and rotation) are estimated by the weak-perspective version [13] of the originally orthographic factorization method [14]. This method is very robust with respect to outliers. A non-linear optimization step which is based on the perspective projection model is applied afterwards. The result is a sparse geometrical representation of scene surface which can be interpolated yielding approximative dense depth maps.

Light field rendering uses the calibration information and the sampled data set to reconstruct new views of the scene by interpolating the stored samples. To achieve visually convincing results at interactive rates, two-plane parameterized light fields and hardware-acceleration for rendering have to be used. As the cameras do not lie on the required two-plane setup an additional *warping* step is required where new images from virtual cameras lying on a plane are generated (interpolated). The approximative dense depth maps, resulting from the factorization step, provide additional information, so that the reconstructed views can be enhanced drastically [6].

4. HIGHLIGHT SUBSTITUTION

A *confidence map* for each image is calculated using the depth map in order to distinguish real depth information from interpolated or not available depth information. The confidence map has the same size as the input image and is set to zero if no depth information is available, e.g. at the border of the image. Currently, the confidence values for interpolated and real depth information are in both cases set to 1.

During warping, only pixels with confidence values > 0 are used to interpolate the images of the new virtual cam-

eras. The number of original cameras contributing to interpolation has to be chosen large enough to avoid black pixels (i.e. no information for interpolation is available).

It is seen easily that the confidence map can be used to substitute highlight pixels by setting their confidence value to zero. This means that highlight pixels are not used for interpolation during the warping step, i.e. no highlight pixels occur in the interpolated images of the new virtual cameras. If the object point of a highlight pixel was not overimposed by highlights in the contributing views, the interpolated value will be very close or even identical to the *real* value at that position.

The resulting light field uses images for rendering in which highlight pixels are substituted (interpolated).

5. EXPERIMENTS AND EVALUATION

The algorithm was evaluated on synthetic and real image sequences.

Signal-to-noise ratios could only be calculated for the synthetic sequences as only there ground truth data was available. For a synthetic sequence of a sphere and a cube with color texture, the value calculated from 50 images was increased by 6.6% (8.30 vs. 8.85) for the whole image and by 17.3% (5.21 vs. 6.11) at highlight positions (also see Figure 2). Note that the generally low signal-to-noise ratios are due to the black background of the synthetic images.

Two medical light fields were generated, one from a sequence of a gall (61 frames), the other from a sequence of the thoracic cavity (97 frames). The non-rigidity of the scenes during recording has been ignored exploiting the fact that the chosen calibration approach is robust against small changes of detected point features.

Highlights were detected by *HSV* thresholds where $H \in [0, 359]$, $S \in [0, 255]$ and $V \in [0, 255]$. We used the following thresholds for the gall sequence: $0 \leq H \leq 359$, $0 \leq S \leq 20$ and $0 \leq V \leq 200$. Afterwards the binary highlight mask was dilated three times. The thresholds for the thoracic cavity were: $0 \leq H \leq 359$, $0 \leq S \leq 40$ and

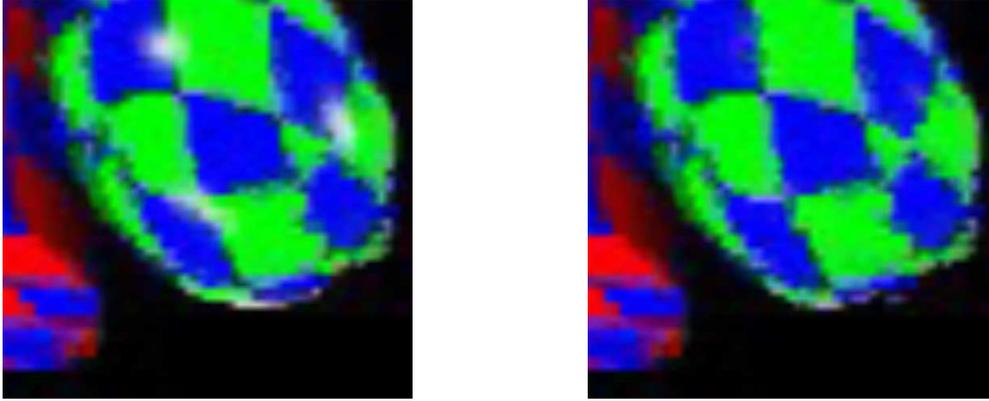


Fig. 2. Example images of the synthetic sequence: in contrast to image left, the image right was generated by using the highlight mask as confidence map.

$0 \leq V \leq 200$ and the binary highlight mask was dilated only two times.

Figure 3 shows two selected images of the gall light field (and their difference image). The images were rendered from the light fields and show the reduction of highlights.

In a double blind setup, physicians evaluated rendered images from the gall and the thoracic cavity light field. For each light field 50 images were rendered with the normal confidence map and another 50 images were rendered (using the same camera parameters) setting the detected highlight pixels in the confidence map to zero (XOR-operation between confidence map and highlight mask). These two times 50 images were compared pairwise. The physicians selected almost *always* the image where the highlights were reduced to be the image with the higher or better quality: 45 of 50 images at the gall sequence and 50 of 50 images at the thoracic cavity sequence. This shows very clearly the gained enhancement by substituting highlights.

6. CONCLUSION

We presented a new technique for highlight substitution. After the generation of a light field from a color image sequence of the scene, a highlight mask for each frame of the sequence is calculated by simple *HSV* thresholds. To substitute highlights, the confidence value of highlight pixels is set to zero. Thereby these pixels are not used at the interpolation during the warping step. The result is a light field in which pixels at highlight positions are interpolated by pixels which were not over-imposed by highlights.

The algorithm fails to substitute highlights if they do not move over the surface of the regarded object because then no real information is available. The algorithm works nicely as long as the highlights move and the objects and the lighting conditions remain almost unchanged during recording of the sequence.

For synthetic data, the calculated signal-to-noise ratios show that the image quality could be increased substantially, especially at highlight regions. The evaluation results of the physicians demonstrate that the image quality of endoscopic images rendered from light fields was enhanced by highlight substitution.

Up to now, highlights can only be substituted in images rendered from the light field which do not look exactly like the original images. In the future, we want to apply our algorithm to the original images directly, i.e. we want to substitute only the highlight regions in the original images by using the geometric information of the light field to obtain the real information for these pixels.

Acknowledgement

The authors express their thanks to a large number of people who cooperate in this project. In particular, Benno Heigl implemented most of the computer vision aspects of light field generation; the idea of subsequent non-linear optimization is due to him. Christian Vogelgsang provided the graphics and rendering part and jointly with Benno Heigl brought up the idea of confidence maps. The medical data was provided by the Surgical Clinic of the University of Erlangen-Nuremberg under guidance of Prof. Hohenberger who is also responsible for the medical aspects of the project. During surgeries of Christoph Schick the image sequences were recorded. He and Sophie Krüger did the medical evaluation. Prof. Greiner coordinates graphics. All of them jointly work in the project supported by DFG under grant SFB 603.

7. REFERENCES

- [1] R. Koch, B. Heigl, M. Pollefeys, L. Van Gool, and H. Niemann, "A geometric approach to lightfield cal-



Fig. 3. Rendered images of the gall light field at the same camera position with the same camera parameters with and without using the highlight mask as confidence map. Left: normal light field, center with substitution, right differences of the two methods

- ibration,” in *Computer Analysis of Images and Patterns — CAIP '99*, F. Solina and A. Leonardis, Eds., Heidelberg, 1999, number 1689 in Lecture Notes in Computer Science, pp. 596–603, Springer.
- [2] S. A. Shafer, “Using color to separate reflection components,” *COLOR research and application*, vol. 10, no. 4, pp. 210–218, 1985.
- [3] A. Koschan, “Analyse von Glanzlichtern in Farbbildern,” in *Dritter Workshop Farbbildverarbeitung*, D. Paulus and Th. Wagner, Eds., Stuttgart, 1997, pp. 121–127 & 95, IRB-Verlag.
- [4] Th. Gevers and H. M. G. Stokman, “Classifying color transitions into shadow-geometry, illumination highlight or material edges,” in *Proceedings of the International Conference on Image Processing (ICIP)*, Vancouver, BC, September 2000, pp. I:521–524, IEEE Computer Society Press.
- [5] C. Palm, T. Lehmann, and K. Spitzer, “Bestimmung der Lichtquellenfarbe bei der Endoskopie makrotexurierter Oberflächen des Kehlkopfs,” in *5. Workshop Farbbildverarbeitung*, K.-H. Franke, Ed., Ilmenau, 1999, pp. 3–10, Schriftenreihe des Zentrums für Bild- und Signalverarbeitung e.V. Ilmenau.
- [6] S. J. Gortler, R. Grzeszczuk, R. Szelinski, and M. F. Cohen, “The lumigraph,” *Computer Graphics (SIGGRAPH '96 Proceedings)*, pp. 43–54, August 1996.
- [7] M. Levoy and P. Hanrahan, “Light field rendering,” in *Computer Graphics Proceedings, Annual Conference Series (Proc. SIGGRAPH '96)*, 1996, pp. 31–42.
- [8] E. H. Adelson and J. R. Bergen, *Computational Models of Visual Processing*, chapter 1 (The Plenoptic Function and the Elements of Early Vision), MIT Press, Cambridge, MA, 1991.
- [9] C. Vogelgsang, B. Heigl, G. Greiner, and H. Niemann, “Automatic image-based scene model acquisition and visualization,” in *Workshop Vision, Modeling and Visualization*, Saarbrücken, Germany, Nov. 2000, pp. 189–198.
- [10] J. Shi and C. Tomasi, “Good features to track,” in *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, June 1994, pp. 593–600, IEEE Computer Society Press.
- [11] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.
- [12] R. Y. Tsai and T. S. Huang, “Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 13–27, 1984.
- [13] C. J. Poelman and T. Kanade, “A paraperspective factorization method for shape and motion recovery,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 206–218, March 1997.
- [14] C. Tomasi and T. Kanade, “Shape and motion from image streams under orthography: A factorization method,” *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, November 1992.