51. IWK

Internationales Wissenschaftliches Kolloquium

International Scientific Colloquium

PROCEEDINGS

11-15 September 2006

FACULTY OF ELECTRICAL ENGINEERING AND INFORMATION SCIENCE

INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING - DEVICES AND SYSTEMS, MATERIALS AND TECHNOLOGIES FOR THE FUTURE

Startseite / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=12391

TECHNISCHE UNIVERSITÄT
ILMENAU

Impressum

Herausgeber: Der Rektor der Technischen Universität Ilmenau
Univ.-Prof. Dr. rer. nat. habil. Peter Scharff

Redaktion: Referat Marketing und Studentische Angelegenheiten
Andrea Schneider

Fakultät für Elektrotechnik und Informationstechnik
Susanne Jakob
Dipl.-Ing. Helge Drumm

Redaktionsschluss: 07. Juli 2006

Technische Realisierung (CD-Rom-Ausgabe):
Institut für Medientechnik an der TU Ilmenau
Dipl.-Ing. Christian Weigel
Dipl.-Ing. Marco Albrecht
Dipl.-Ing. Helge Drumm

Technische Realisierung (Online-Ausgabe):
Universitätsbibliothek Ilmenau
ilmedia
Postfach 10 05 65
98684 Ilmenau

Verlag: isle
Verlag ISLE, Betriebsstätte des ISLE e.V.
Werner-von-Siemens-Str. 16
98693 Ilrnenau

ISBN (Druckausgabe): 3-938843-15-2
ISBN (CD-Rom-Ausgabe): 3-938843-16-0

Startseite / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=12391

51^{st} Internationales Wissenschaftliches Kolloquium
Technische Universität Ilmenau
September 11 – 15, 2006

A. Nowak, L. Hörchens, J. Röder, M. Erdmann

Farbbasierte Stanzmaskenerzeugung für die Fernsehproduktion

Colour-based Video Segmentation for TV Studio Applications

Introduction

The extraction of objects from video sequences is a standard problem in the context of television production. Commonly, techniques based on blue or green screens are employed in this task, despite of their drawbacks and the constraints they impose on the production process.

UIL student Lars Hörchens has investigated alternatives to the conventional chroma key technique in his diploma thesis. This thesis seeks to point out a possible approach on digital matting without a blue screen setup or additional sensor equipment. Based on a thorough review of classical matting techniques as well as recently developed methods for image segmentation, the structure of a suitable matting system is derived. This system can be used for conventional TV and film production as well as for multiview capturing and other applications in 3DTV.

Three main modules form the basic structure of this framework: an initial colour classification based on Gaussian mixture models, a Markov random field approach for the incorporation of spatial and temporal constraints and a natural matting procedure for the refinement of object boundaries. Possible realisations of the different modules of the proposed system are described and implemented. The developed methods have been applied to standard video sequences and their capabilities and limitations have been analysed and assessed.

System Design: A Model-Based Approach to Digital Matting

The seminal paper of Smith and Blinn [2] was one of the first publications that addressed the underlying problem of matting. It is fairly easy to see that this problem is generally underconstrained by taking the vectorial form of the compositing equation

$$\begin{pmatrix} C_R \\ C_B \\ C_G \end{pmatrix} = \alpha \begin{pmatrix} F_R \\ F_B \\ F_G \end{pmatrix} + (1-\alpha) \begin{pmatrix} B_R \\ B_B \\ B_G \end{pmatrix}.$$

As only C_R, C_G and C_B are given, there are three equations and seven unknown variables for the general matting problem. Even if we can assume that the background colour is known as this is the case for blue screen and difference matting, there are still four unknowns. Hence, the extraction of a foreground object is generally underconstrained.

A problem of this kind is usually termed an inverse or ill-posed problem, and most papers on digital matting state that there is no common solution to it, for instance [2], [4], [8] and [9]. The single observation of a pixel just does not provide enough information to calculate the matte.

The matting system is supposed to be used in the field of live television production and multiview capturing. This means that the system has to be able to work in real-time and it must be as easy to use as a conventional chroma keying system. A special colour for backdrop and floor should not be required, but a background that differs from the foreground objects to a certain extent would be appreciated. The proposed design of the system is shown in Figure 1.

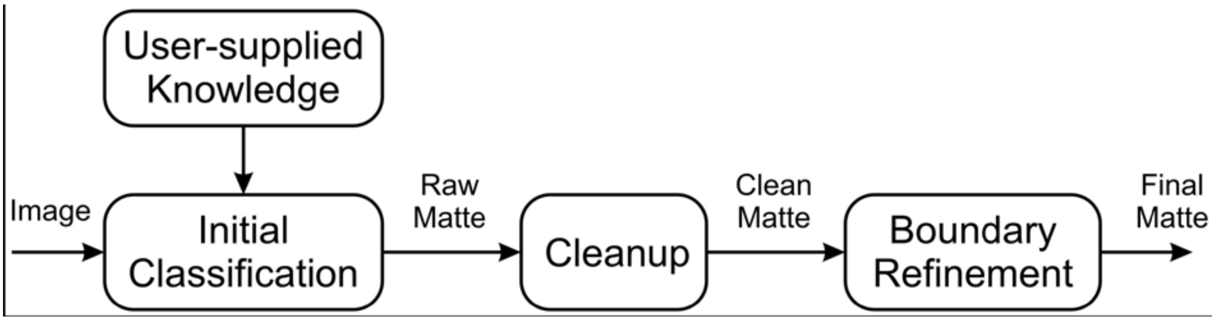

Figure 1: Proposed structure for the matting system

In the first step an estimation of region membership for each single pixel takes place. In the second step this rough foreground / background matte is cleaned up to close holes and smooth the matte. The borders of the matte and transparencies are refined in the last step.

Foreground / Background Segmentation

In our matting system a colour model is used in order to assign pixels to foreground, background or a mixture of both. Gaussian mixture models (GMMs) [3] [7] are used to accomplish this task. They provide a compact parametric description of colour distributions and they can be used for statistical inference. The given discrete distribution representing the colours of the input image is modelled by the sum of several Gaussian distributions. This is not only a simple approximation, but as well a proper modelling approach for colour distributions emanating from reflecting surfaces [11].

Assuming a standard three-dimensional colour space, the formula for the appropriate multivariate normal density *G* is given by

$$G(C,\mu,\Sigma)=\frac{1}{(2\pi)^{\frac{2}{3}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(C-\mu)^{T}\Sigma^{-1}(C-\mu)\right)$$

where *C* is a three-dimensional column vector representing colour, μ is the mean vector of the distribution, Σ is 3-by-3 covariance matrix. The mixture model with *k* components is then given by

$$M(C)=\sum_{i=1}^{k}w_{i}G(C,\mu_{i},\Sigma_{i})$$

with w_i as the weight for the respective mixture components. Expectation Maximisation (EM) is used to fit the Gaussian mixture components to a given image. A visualisation of the outcome is given by Figure 2. Eight mixtures were chosen for the foreground model $M_{FG}(C)$ (red) and the background model $M_{BG}(C)$ (blue), respectively. The left side of Figure 2 shows the input data, all voxels above a certain threshold are marked. On the right, 2σ-ellipsoids of the estimated Gaussian distributions are plotted for comparison.

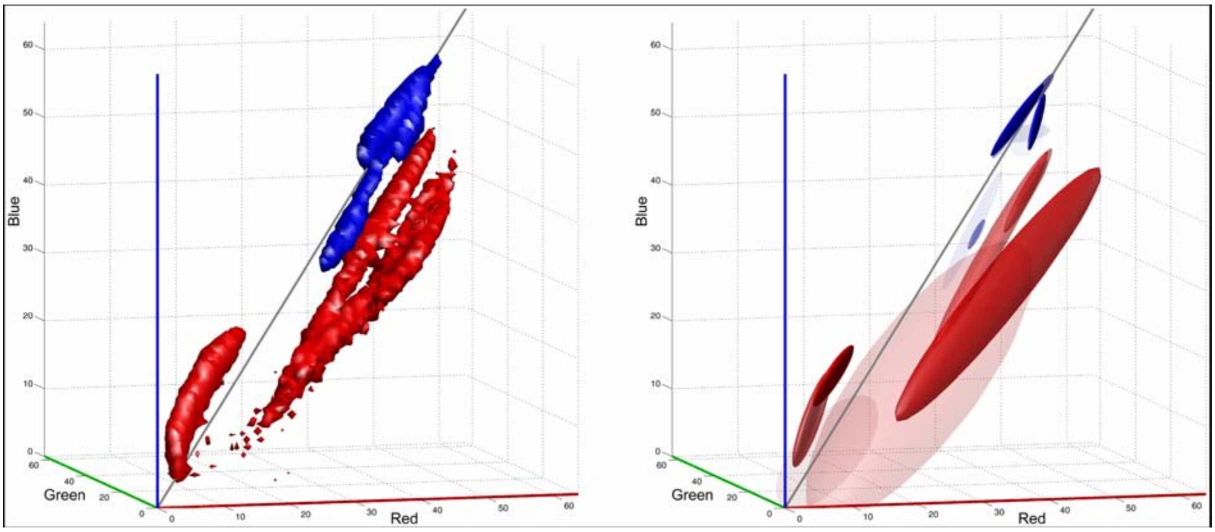

Figure 2: Colour distribution (left) and approximation by two Gaussian mixture models (right), 2σ-ellipsoids, opaqueness resembles density.

In the next step maximum a posteriori estimation (MAP) is used to perform the classification of pixels as foreground or background [13]. In this case only two classes are used. Figure 3 shows a training image and an input image for further processing.

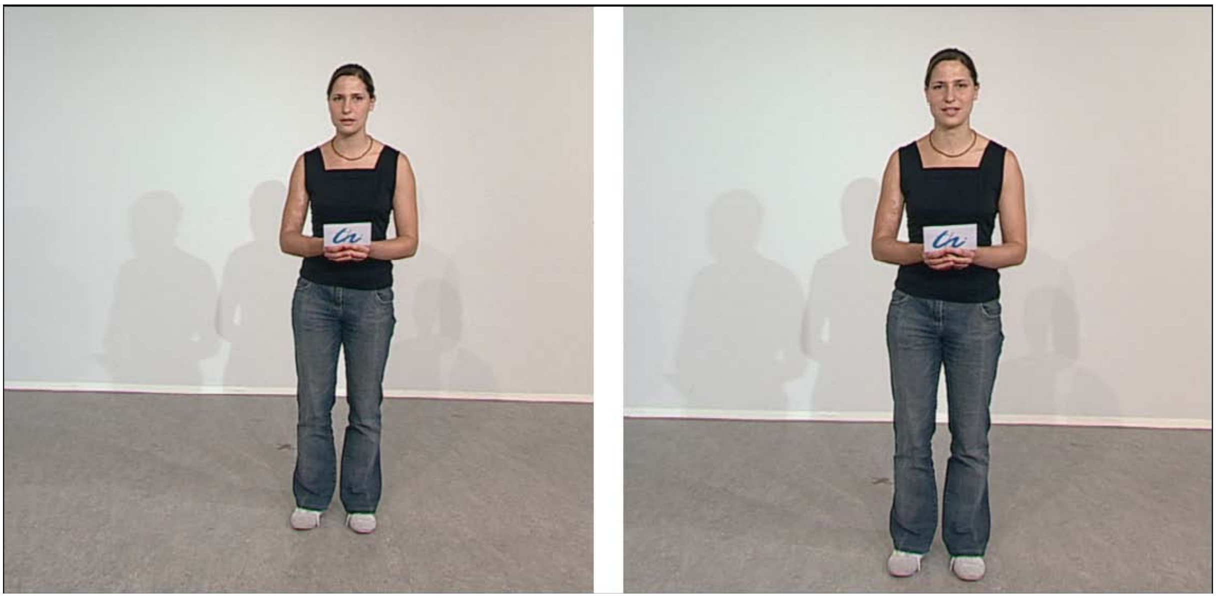

Figure 3: Training image (left) and input image (right)

A correct MAP estimate incorporates prior information in order to bias the outcome of the estimation towards a result that is more likely with respect to the general distribution of the variable to be estimated. A simple prior on the binary foreground/background *α* membership would be the usage of the relative frequencies of both classes from the training image. For our example image, only about 10% of the area are occupied by the foreground. A reasonable choice for a per-pixel estimation would thus be to set $p(\alpha)$=0.1 because in this image foreground pixels are in general by far less likely to occur than background pixels. Figure 4 shows the result of the MAP estimation after incorporating this information. As expected, less background pixels are incorrectly classified as belonging to the foreground.

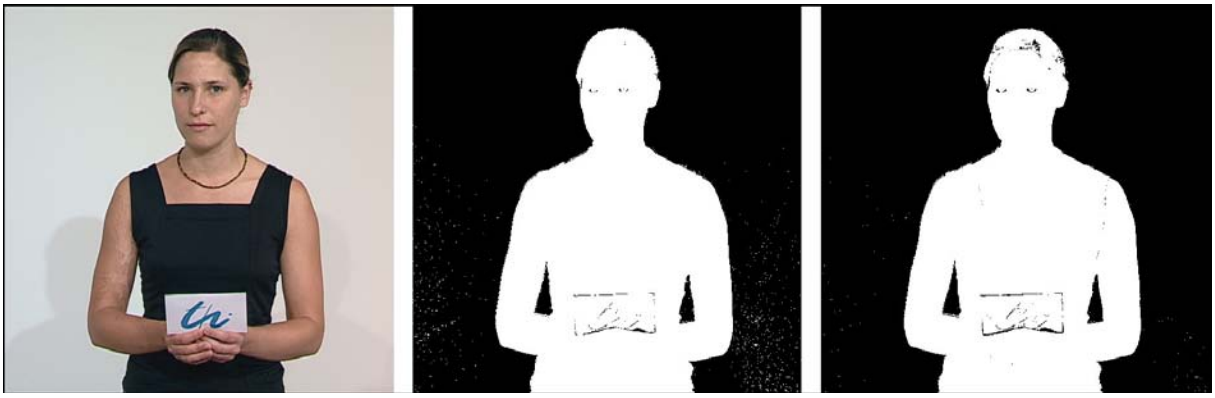

Figure 4: Input image (left), matte estimate without (left) and with prior (right)

Regularisation by Continuity

In order to regularise the under-determined matting problem, additional spatial as well as temporal constraints have to be imposed on the estimation process to obtain convincing results. Markov random fields (MRFs) represent a tool for modelling continuity between different entities interacting in a common context such as pixels in an image [10]. This is done to fill up small holes in the foreground objects. A combination called Gaussian Mixture Markov Random Field (GMMRF) [12] with spatial as well as temporal prior knowledge is used in our system to improve the quality of the binary mask. Figure 5 shows the detection results obtained with a Gaussian mixture model only along with the outcome of the GMMRF estimation when using spatial and temporal prior knowledge.

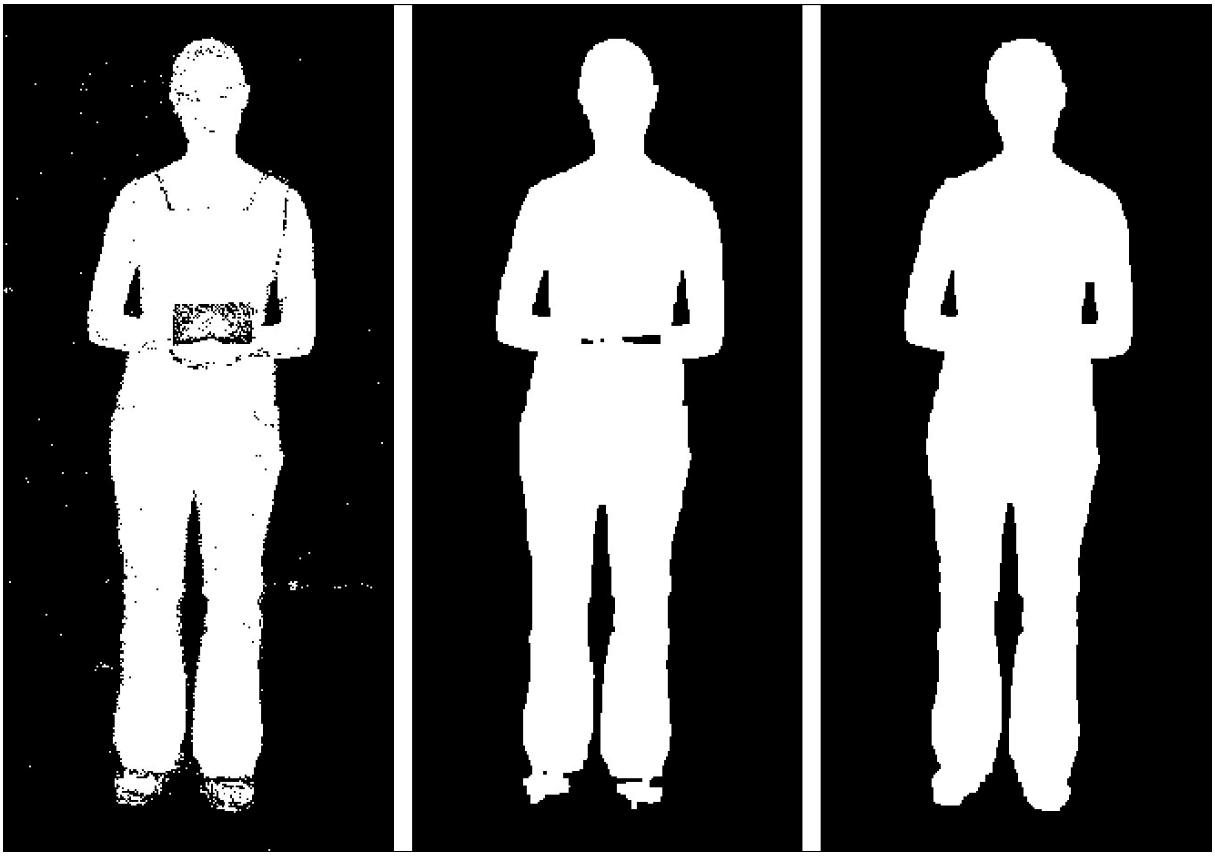

Figure 5: MAP-GMM detection result without prior (left), GMMRF with spatial constraints (middle), and GMMRF with both spatial and temporal constraints (right)

Border Matting

The remaining task is the conversion of the binary map obtained from the GMMRF estimation into an alpha matte of full range which then provides an approximation of the true foreground/background membership of each pixel that suffices for visually convincing composites. In order to correctly determine the alpha values along the object boundary, it is first necessary to derive a trimap from the image itself. Given this trimap and the image, one of the algorithms for natural matting can be applied to reconstruct the borders. The construction of a trimap from the binary segmentation is non-trivial. It is not feasible to use a ribbon of fixed width as the extent of a transition might vary between 2 and 20 pixels, depending on the camera aperture, the integration time and the speed of motion in the image sequence. Therefore, the boundary area must be estimated from the image itself. Nevertheless, the general idea of using snake contours or level sets in order to create a trimap seems to be an appropriate way. It is reasonable to start from the binary border which can be assumed to lie within or close to the unknown region of the trimap. This hard contour is then propagated outwards until it reaches the region of pure background colour. In the same way, it is propagated inwards until all points of the contour lie on pixels definitively belonging to the object. A kind of guidance field has to be incorporated in this process in order to stop the propagation of the boundary as soon as the final position is reached. We propose the comparison of the Mahalanobis distances *D* between the centres of the different mixtures.

$$D(C,\mu,\Sigma)=\sqrt{(C-\mu)^T\Sigma^{-1}(C-\mu)}$$

For reasons of simplicity, we compare the sum of Mahalanobis distances obtained for the foreground model to the sum of the distances from the background model. The grade of foreground/background membership for each pixel is then derived as

$$d(C)=\frac{\Sigma_{BG}D(C,\mu,\Sigma)}{\Sigma_{FG}D(C,\mu,\Sigma)+\Sigma_{BG}D(C,\mu,\Sigma)},$$

were Σ_{FG} and Σ_{BG} denote the sum of all foreground and background components, respectively. From this calculation, a Mahalanobis map can be derived, see Figure 6.

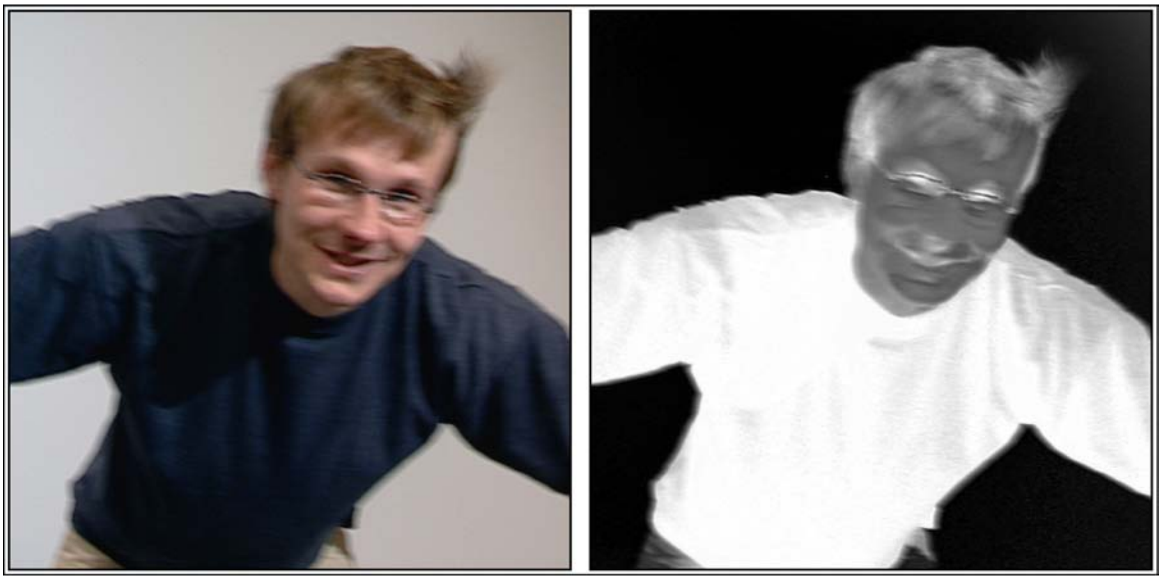

Figure 6: Input image (left) and Mahalanobis map (right)

The next step is the generation of a trimap given the binary segmentation result and the Mahalanobis map. The trimap is directly extended in a pixel-by-pixel manner, adding adjacent pixels to the border region of the trimap if they can be assumed to belong to the transition area. An example for an automatically extracted trimap is shown in Figure 7, along with a manually created one for comparison.

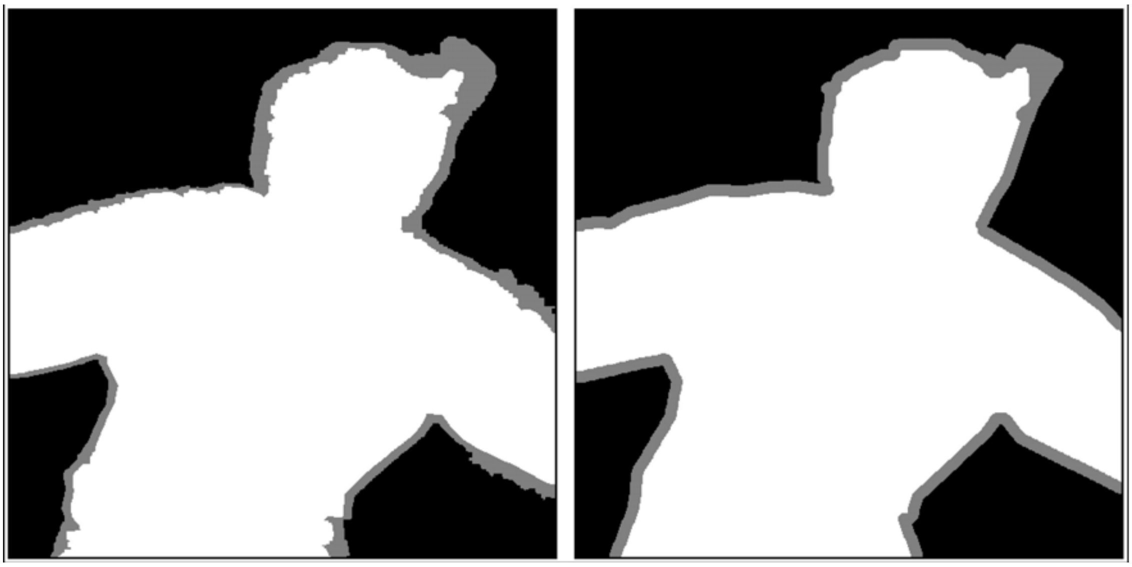

Figure 7: Computed trimap (left) and manually drawn version (right)

Re-implemented variants of Knockout [5] [6] and Poisson [9] matting were used for testing purposes. They deviate from the originally proposed algorithms due to some simplifications and a few improvements and should for this reason not directly been regarded as equivalent methods. In order to clearly indicate the difference, we shall call our implementations “Knockin” and “PDE matting”.

Both procedures require estimates of the pure foreground and background colours in the unknown area indicated by the trimap. In the current implementation, a simple scheme based on a local neighbourhood of 3 x 3 pixels is used, and weighting of the diagonal pixels is neglected. Figure 8 shows an example image and for visualisation purposes a full extrapolation of the foreground area. It is obvious that this method is too simple because it introduces hard edges in the extrapolated region. The averaging scheme is thus transformed into a two-step procedure. A first estimate of the pixels on the border of the known region is obtained as described above. In a second step, the newly gained estimates are incorporated in the estimation of adjacent pixels, too. The improved result is presented in Figure 9.

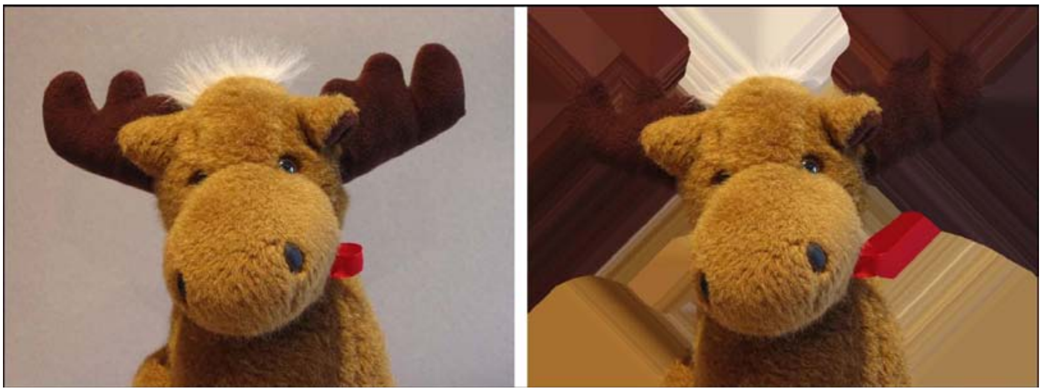

Figure 8: Example image (left) and extrapolated foreground (right)

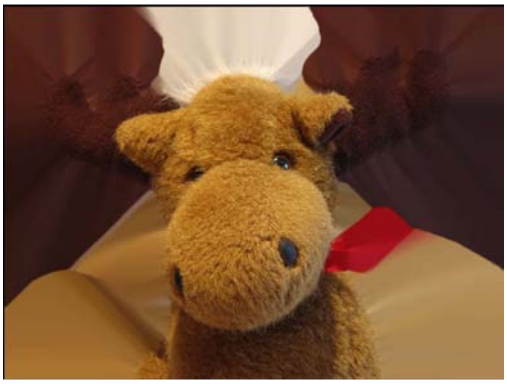

Figure 9: Two-step method for extrapolation

We calculate the alpha values according to

$$\alpha = \frac{C - B}{F - B},$$

assuming that the correct alpha value is obtained by projecting the observed colour *C* on the line connecting *F* and *B*. We update only the unknown region of the trimap. Note however that all important details of the procedure are protected by the respective patents quoted above. Figure 10 shows composites obtained with Knockin and PDE matting.

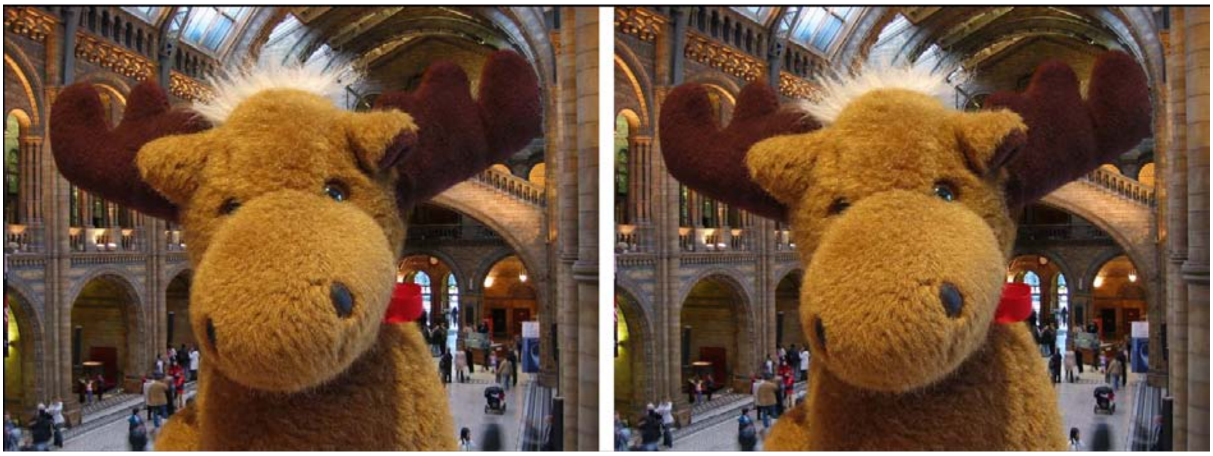

Figure 10: Composite obtained with Knockin (left) and PDE matting (right)

Experimental Results

We evaluated the capabilities of the matting system on three exemplary image sequences which we assume to represent typical applications and to contain standard problems a general matting system should be able to cope with.

The “TV Journal” sequence features a standard presentation scene. The camera zooms in on the actor who is quite static, see Figure 11.

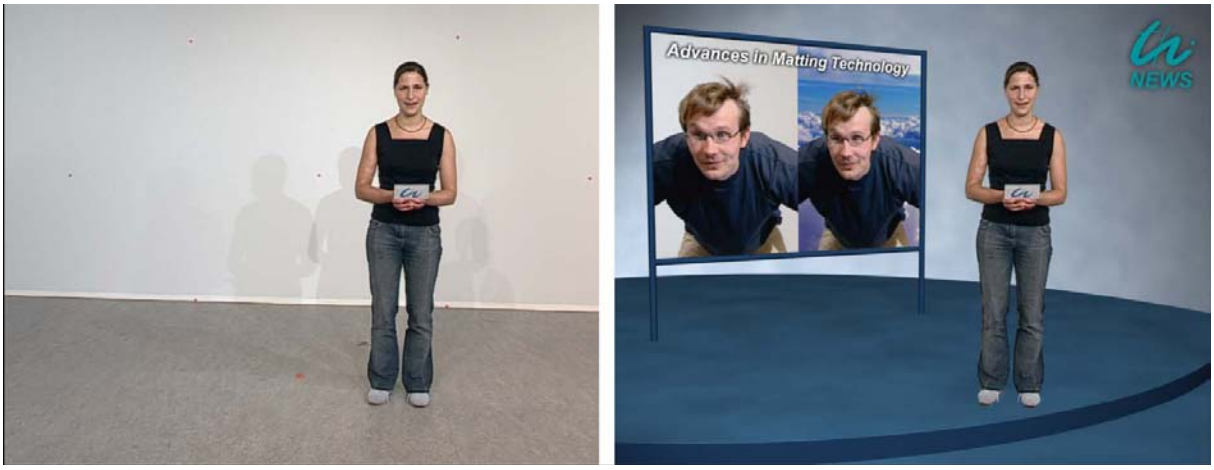

Figure 11: Input image (left) and result (right) from the “TV Journal” sequence

A major difficulty is the inclusion of the standard studio floor into the background, mainly because of its similarity to the presenter’s shoes. Even when incorporating spatial and temporal constraints, a consistent detection of the shoes remains difficult. The choice of a higher value for the temporal consistency parameter helps to keep the shape during the first frames, but when the zoom starts, the GMMRF looses track of the object boundary and does not adapt to the change in size. The temporal prior can thus only be exploited to a limited extent.

Several insights can be gained from the image sequences. The automatic trimap creation and the colour extrapolation form the problematic parts of the current implementation. Due to the serial arrangement of the modules, the performance of subsequent stages and thus the quality of the mattes and final composites is degraded. The weighting of the different constraints seems to depend heavily on the content of the scene. It is clear that colour is and always will be the most important information in the matting process. In addition, spatial coherence has proven to be another valuable tool which could successfully be applied in all examined cases.

Acknowledgements

The text above is a direct excerpt of the diploma thesis of Lars Hörchens [1]. This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV.

References

[1] Hörchens, Lars: Segmentation of Video Sequences for Compositing Applications in Television Production. Diploma Thesis. Technische Universität Ilmenau, 2004

[2] Smith, Alvy R.; Blinn, James F.: Blue Screen Matting. In: SIGGRAPH ’96: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1996, pages 259–268

[3] Raja, Yogesh; McKenna, Stephen J.; Gong, Shaogang: Segmentation and Tracking Using Color Mixture Models. In: Proceedings of the Asian Conference on Computer Vision, volume 1, 1998, pages 607–614

[4] Mitsunaga, Tomoo; Yokoyama, Taku; Totsuka, Takashi: AutoKey: Human Assisted Key Extraction. In: SIGGRAPH ’95: Proceedings of the 22nd Annual Conference on ComputerGraphics and Interactive Techniques. New York: ACM Press, 1995, pages 265–272

[5] Berman, Arie; Vlahos, Paul; Dadourian, Arpag: Comprehensive Method for Removing from an Image the Background Surrounding a Selected Subject. U.S. Patent 6,134,345; Assignee: Ultimatte Corporation, 2000

[6] Berman, Arie; Vlahos, Paul; Dadourian, Arpag: Method for Removing from an Image the Background Surrounding a Selected Object. U.S. Patent 6,134,346, 2000

[7] Delignon, Yves; Marzouki, Abdelwaheb; Pieczynski, Wojciech: Estimation of Generalized Mixture and Its Application in Image Segmentation. In: IEEE Transactions on Image Processing 6 (1997), number 10, pages 1364–1375

[8] Chuang, Yung-Yu; Curless, Brian; Salesin, David H. et al.: A Bayesian Approach to DigitalMatting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, volume 2, IEEE Computer Society, 2001, pages 264–271

[9] Sun, Jian; Jia, Jiaya; Tang, Chi-Keung et al.: Poisson Matting. In: ACM Transactions on Graphics: Proceedings of the 2004 SIGGRAPH Conference, volume 23, number 3. New York: ACM Press, 2004, pages 315–321

[10] Li, Stan Z.: Markov Random Field Modeling in Image Analysis. Tokyo: Springer, 2001

[11] Klinker, Gudrun J.: A Physical Approach to Color Image Understanding. Wellesley, Massachusetts: A K Peters, 1993

[12] Blake, Andrew; Rother, Carsten; Brown, M.: Interactive Image Segmentation Using an Adaptive GMMRF Model. In: Proceedings of the 8th European Conference on Computer Vision, volume 1. Piscataway, New Jersey: IEEE Press, 2004, pages 428–441

[13] Chalom, Edmond; V. Michael Bove, Jr.: Segmentation of an Image Sequence Using Multi-Dimensional Image Attributes. In: Proceedings of the IEEE Conference on Image Processing, volume 2. Piscataway, New Jersey: IEEE Press, 1996, pages 525–528

Authors:
Dipl.-Ing. Arne Nowak
Dipl.-Ing. Lars Hörchens
Dipl.-Ing. Jan Röder
Dipl.-Ing. Matthias Erdmann
TU Ilmenau, Institut für Medientechnik, PO 100 565
98684 Ilmenau
Phone: +49-3677-69-1577
Fax: +49-3677-69-1255
E-mail: arne.nowak@tu-ilmenau.de