3–D Modeling of Human Heads from Stereoscopic Image Sequences

Reinhard Koch

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Abteilung "Automatische Bildinterpretation", Prof. Dr.–Ing. C.–E. Liedtke
Universität Hannover, Appelstr. 9A, 30167 Hannover
email: koch@tnt.uni–hannover.de

Abstract. The modeling of heads from image sequences is of great interest in the context of virtual reality, telecommunication and computerized animation systems. In this contribution a passive 3–D scanning system is described that automatically computes a complete 3–D surface model of a human head and shoulder part from a stereoscopic image sequence of a person rotating freely in front of the camera. No active element like a laser scanner is required for the modeling. 3–D shape, surface texture and the 3–D motion of the head is extracted solely from the image sequence. A dense map of 3–D surface points is obtained by image correspondence, object segmentation, interpolation, and triangulation. The surface geometry of the object is approximated by a triangular wire–frame which stores the surface texture in texture maps. Sequence processing serves to track head motion and to fuse surfaces from different view points into a consistent 3–D surface model. From the textured 3–D model highly realistic image sequences from arbitrary view points can be synthesized using computer graphics techniques.

Key Words. Stereoscopic Image Sequence Analysis, Scene Reconstruction, Shape and Object Representation, Motion Analysis, Close Range Photogrammetry

1 Introduction

A system to automatically compute 3–D surface models from image sequences is presented. 3–D surface models are needed for a wide range of applications where computer generated 3–D environments are desirable, like in architecture visualization [7], virtual television studios [4], virtual presence for video communications [9] and general "virtual reality" applications.

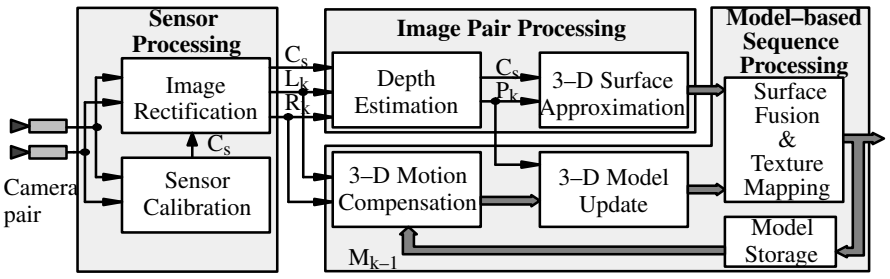
3–D modeling from image sequences is a recent topic in computer vision research. Great effort went into developing algorithms that estimate 3–D object shape from various sources, termed shape from motion, stereo, and others [10],[2]. On the other hand research was conducted to find solutions to the problem of rigid object motion [1],[18]. The problem of dynamic nonrigid bodies was addressed by [22],[19]. Some approaches are reported from monocular vision systems to compute dense depth maps and surface reconstruction for orthographic [24] and perspective projection [20] as well as stereoscopic analysis [17],[23].

In this contribution a system for automatic 3–D surface reconstruction from stereoscopic image sequences is discussed. The system is aimed on automatically computing 3–D models of any scene that is composed of smooth surface patches. Earlier contributions on this topic describe the general outline of the system with emphasis on the modeling of polyhedral objects like buildings and street scenes [13],[15]. The modeling system was extended to more general objects with arbitrary curved, smooth surfaces. A detailed analysis of the modeling system can be found in [16]. The application described here focuses on the automatic 3–D modeling of a human head and shoulder part from a person that rotates freely in front of the camera system. Geometric and photometric models as well as object and camera motion are extracted from the image sequence. The geometry is measured from the stereoscopic image pairs and the motion is tracked directly from the image sequence. All measurements obtained throughout the image sequence are integrated into a consistent 3–D scene model that contains not only the scene geometry but also texture maps of the object surface. In contrast to the currently available active laser scanning devices that need sophisticated and expensive mechanic alignment, this system is capable of modeling objects with high precision just from video images.

Chapter 2 sketches the general structure of the modeling system. In Chapter 3 the surface reconstruction from a single image pair is discussed. Chapter 4 deals with the sequence processing which consists of view point tracking and the fusion of surfaces from different view points.

2 System Overview

The goal of the surface reconstruction is to extract a complete 3–D surface model of the scene that contains the surface geometry and the surface colors. Input to the modeling is a stereoscopic image sequence recorded by a binocular stereoscopic camera. For the head modeling application, a person is asked to turn in front of the camera system and a stereoscopic image sequence is recorded. No constraints on the motion are required, only the head and shoulder part should be kept rigid throughout the recording. The structure of the analysis system is shown in Fig. 1. Three main modules can be identified.



L_k, R_k: stereoscopic image pair **P_k**: depth map **M_{k-1}**: stored 3–D model
C_s: sensor calibration coordinates index **$_k$**: sequence index of sensor view point

Fig.1: Structure of 3–D Scene Analysis from Sequences of Stereo Images

Sensor Processing: In order to compute depth from the camera pair a sensor calibration is needed. In an off-line calibration process before the actual scene recording, the internal camera parameters focal length and radial distortion are computed together with the relative position and orientation of both cameras in a sensor calibration coordinate system. The real cameras may have arbitrary relative external orientation like a convergence angle, and may have differing internal parameters. During scene recording those differences are compensated in a rectification process. The images are rectified with projective mapping onto a virtual camera target in such a way that the virtual camera system has coplanar image planes. This rectification greatly simplifies the image pair processing in the next stage.

Image Pair Processing: For each camera view point a stereoscopic image pair is recorded. This image pair is used to extract the surface depth information from this particular view point. The analysis consists of depth estimation and surface approximation. During depth estimation the correspondence between the stereoscopic image pair is evaluated to compute a dense depth map. The discrete depth map is then converted into a structured, parametric surface representation. As geometrical representation a triangular wire–frame is fitted to the surface. The surface color of the real object for this view point is stored as texture map on the wire–frame surface.

Model–based Sequence Processing: Image pair processing computes a 3–D surface reconstruction of a particular view point. The complete model is obtained by fusing the different view points into a consistent surface model. This fusion is performed in the sequence processing which consists of three main modules. The view–point dependency of the surfaces are removed by tracking the camera and object motion between the view points. The model is transformed into the current view point which allows to fuse the new surface with the existing model. If a particular model surface can be seen from several view points, then all the depth measurements of this surface are integrated in the model update phase using a kalman filter for each vertex of the surface wire–frame. When new object areas become visible the model wire-frame is extended to include the newly visible parts as well. Finally all surface areas are textured to allow realistic visualization of the model.

3 The Image Pair Analysis

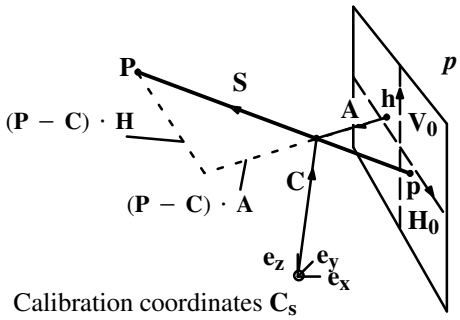
Image pair analysis computes a view point dependent 3–D surface model from the rectified stereoscopic image pair. It is assumed that the objects to be modeled consist of piecewise continuous, smooth surfaces. The steps to be performed are image correspondence analysis, surface segmentation, depth interpolation, and surface approximation. The modules of the analysis pipeline are described in detail in [13] and will be reviewed only briefly.

3.1 Camera Model

The imaging camera system is described as a pinhole camera with perspective projection as in Fig. 2. The camera is oriented arbitrarily in space with respect to a calibration coordinate system $\mathbf{C_S}$. The focal point of the camera is located at $\mathbf{C}$, while the camera orientation is described by a coordinate system ($\mathbf{A}$, $\mathbf{H_0}$, $\mathbf{V_0}$). $\mathbf{A}$ defines the optical axis and the imaging sensor is spanned by $\mathbf{H_0}$ for the image scan line and $\mathbf{V_0}$

for the image columns. This camera system corresponds to the camera model as introduced by Yakimowski and Cunningham [26]. The optical axis penetrates the image plane at center point $\mathbf{h} = (h_x, h_y)^T$ and the effective focal length c is scaled in pixel coordinates with the sensor pixel size $\mathbf{s} = (s_x, s_y)^T$. All camera parameters are calibrated using the calibration according to Tsai [25].

Perspective projection of **P** onto the image plane at **p** along the line of sight $\mathbf{S} = (\mathbf{P} - \mathbf{C})$ is computed in Eq. (1). The definition of four vectors for the camera is redundant because $(\mathbf{A}, \mathbf{H_0}, \mathbf{V_0})$ form a right–handed coordinate system. The orientation of the camera is uniquely described by a rotation matrix $\underline{R}$ with three independent rotation angles $\mathbf{R} = (R_x, R_y, R_z)^T$.



$$p = \frac{1}{(P - C) \cdot A} \cdot \begin{pmatrix} (P - C) \cdot H \\ (P - C) \cdot V \end{pmatrix} \quad (1)$$

$$\text{with: } H = \frac{c}{s_x} \cdot H_0 + h_x \cdot A \quad ,$$

$$V = \frac{c}{s_y} \cdot V_0 + h_y \cdot A$$

$$H_0 = \underline{R} \cdot e_x, V_0 = \underline{R} \cdot e_y,$$

$$A = \underline{R} \cdot e_z$$

Fig. 2: Camera projection model.

3.2 Correspondence analysis

Correspondence analysis exploits the fact that a surface point **P** of the real object projects onto both camera targets of the stereoscopic camera at $\mathbf{p_l}$ and $\mathbf{p_r}$. Because of perspective projection all points on the line $(\mathbf{P} - \mathbf{C_l})$ project onto the single point $\mathbf{p_l}$ in the left camera while in the right camera each point projects onto a different point $\mathbf{p_r}$ on the epipolar search line. Because of rectification the epipolar search line simply corresponds to the image scan line. From the rectified images a disparity map is obtained by searching along the epipolar lines using correlation matching and dynamic programming [8]. The cross correlation of a small image patch (typically 7*7 pixel) around each point on the epipolar line is computed as similarity measure.

The search for the best match between the points on the epipolar line is controlled by uniqueness and ordering constraints. These constraints are based on the fact that there can be no more than one match between left and right image points and that matches are in order for physical surfaces. All possible correspondences are evaluated in an optimum search procedure using dynamic programming that matches all correspondences between left and right image that lie on the same epipolar line. The dynamic programming algorithm was adapted from the work of Cox et al. [6]. The disparity value obtained for each candidate is recorded in a disparity map.

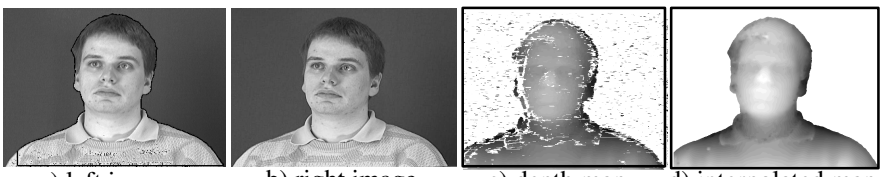
The disparity estimates computed with this approach give dense and reliable depth maps with a disparity resolution of 1 pixel due to quantization of the search space. This quantization leads to a discontinuous and rough surface that can not be tolerated

in surface reconstruction. Therefore in a second stage the quantized disparity estimates are taken as a starting value for a sub–pixel accuracy gradient matching with affine transformation. Assuming that the object mostly consists of smooth surfaces with arbitrary orientation, the effect of perspective distortions between both views is corrected by an affine transformation of a small image patch of about 11x11 pixel. The difference in grey level distribution between the corresponding patches is minimized by fitting the patch with a robust least squares estimation of the affine transformation [5]. With this approach the disparity is estimated to an average accuracy of 0.1 pixel, which is sufficient for most reconstruction tasks.

The disparity map is now converted into a depth map where each pixel $\mathbf{p_l}$ of the map contains the length $\tau = |\mathbf{S}|$ of the line of sight vector to the real surface point **P**. The 3D–Position **P** is computed from the corresponding points in sensor coordinates by space intersection between the corresponding lines of sight.

3.3 Interpolation

The estimated depth map still contains some undefined regions which have to be interpolated. A dense depth map of the object surface is therefore computed by a parametric surface. Only the head region is needed for the modeling and a simple color keying technique was employed to separate the person from the background. A multi–grid surface reconstruction algorithm as described by Terzopoulos[21] was chosen to calculate the interpolation with a finite element approximation. It is assumed that each segmented area contains a smooth coherent surface that can be modeled as a thin plate with a certain stiffness and that inside such a region the disparity measurement is either corrupted by noise or no estimate is available. The interpolation solves the problem of minimizing the potential energy function of the thin plate that is deformed by the disparity measurements. Fig. 3 displays the result of depth estimation of the image pair *Head* from left (a) and right (b) image. The estimated depth map 3c) is interpolated in the object region and the interpolation is displayed in Fig. 3d) color coded, with light values near to the camera and dark values further away.



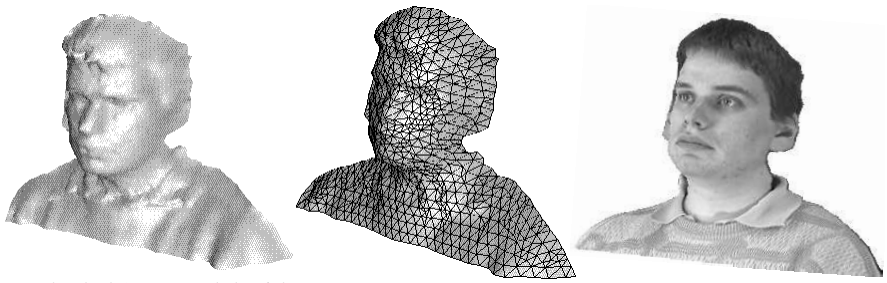
a) left image b) right image c) depth map d) interpolated map
depth color coded: light = near, dark=further away

Fig. 3: Depth estimation from one view point of scene *Head*.

3.4 3–D surface approximation

The amount of data for storing the finite element surface is very high, so an approximation of the surface is needed. The interpolated disparity map is converted into a parametric 3–D surface description by spanning a triangular wire–frame in space for each segmented object surface while preserving the discontinuities at the surface boundaries. The triangular mesh was chosen because it is capable of approximating

arbitrary surface geometries without singularities. On the surface of each triangle the object surface texture is stored in a texture map from which naturally looking views can be synthesized with texture mapping. Fig. 4 shows the surface reconstruction for a single view point of the scene *Head*. The model was computed from the front view and rotated 30° to show the 3–D shape as shaded model (a), the approximated wire–frame (b), and the textured surface model (c).



a) shaded 3–D model with 135,600 surface points

b) 3–D wireframe with 726 vertices

c) textured 3–D surface model

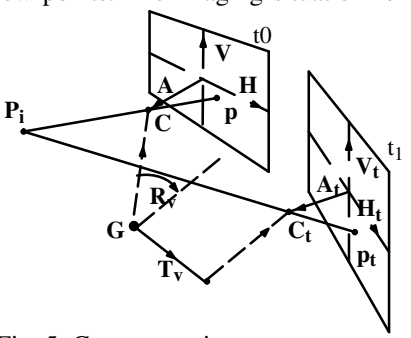
Fig. 4: 3–D surface reconstruction from front view point of the scene *Head*.

4 Model–based Sequence Processing

The modeling of complex scenes requires that more than one view is evaluated to account for occluded objects and to improve the object geometry. Surface models from different view points have to be registered into a consistent object coordinate system. This registration is achieved by tracking the camera and object motion directly from the image sequence. Newly visible surface areas are fused with the existing surfaces to complete the scene model.

4.1 Camera motion tracking

Intensity–based tracking of moving objects observed by a static camera was discussed in detail in [11],[12] and extended to a moving camera system in [14]. It will be reviewed briefly for the case of camera motion. Each object in the scene is treated as an arbitrarily shaped, rigid surface. The shape of the surface is modeled by the control points $\mathbf{P_i}$ of the triangular surface mesh. The object is observed from different view points. The imaging situation for the moving camera is depicted in Fig. 5.



$$
\begin{aligned}
\boldsymbol{C}_t &= \boldsymbol{C} + \underline{Rv} \cdot (\boldsymbol{C} - \boldsymbol{G}) + \boldsymbol{Tv} \\
\boldsymbol{A}_t &= \boldsymbol{A} + \underline{Rv} \cdot \boldsymbol{A} \\
\boldsymbol{H}_t &= \boldsymbol{H} + \underline{Rv} \cdot \boldsymbol{H} \\
\boldsymbol{V}_t &= \boldsymbol{V} + \underline{Rv} \cdot \boldsymbol{V}
\end{aligned}
\qquad (2)
$$

$$
\textit{with} \quad \boldsymbol{T_v} = \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}, \quad \underline{R_v} = \begin{pmatrix} 0 & -R_z & R_y \\ R_z & 0 & -R_x \\ -R_y & R_x & 0 \end{pmatrix}
$$

Fig. 5: Camera motion geometry.

Tracking of the camera motion can be computed directly from the spatial–temporal intensity gradients. The motion from frame t_0 to t_1 is described by a rotation $\mathbf{R_v}$ of the camera axes **A**, **H**, **V** and the focal point **C** around the rotational center **G** and by a translation of **G** with **Tv**. For small rotations the matrix $\underline{R_v}$ is linearized and the motion is expressed by Eq. (2) with unknown motion parameters $\underline{R_v}$ and $\mathbf{T_v}$.

The projection of **P** onto the moving image plane for the times (0, t) generates an optical flow $\mathbf{d} = (d_x, d_y)^T = \mathbf{p}_t - \mathbf{p}$. Insertion of the camera motion from Eq. (2) in Eq. (1) and solving with some small linearization delivers an optical flow as function of the camera motion parameters. The vector (**P**–**G**) is abbreviated as **PG**.

$$\boldsymbol{d} = \frac{1}{(\boldsymbol{P}-\boldsymbol{C})^T \cdot \boldsymbol{A} + \boldsymbol{PG}^T \cdot \underline{R_v} \cdot \boldsymbol{A} - \boldsymbol{A}^T \cdot \boldsymbol{Tv}} \begin{pmatrix} \boldsymbol{PG}^T \cdot \underline{R_v} \cdot (\boldsymbol{H} - p_x \cdot \boldsymbol{A}) - (\boldsymbol{H} - p_x \cdot \boldsymbol{A})^T \cdot \boldsymbol{Tv} \\ \boldsymbol{PG}^T \cdot \underline{R_v} \cdot (\boldsymbol{V} - p_y \cdot \boldsymbol{A}) - (\boldsymbol{V} - p_y \cdot \boldsymbol{A})^T \cdot \boldsymbol{Tv} \end{pmatrix} \quad (3)$$

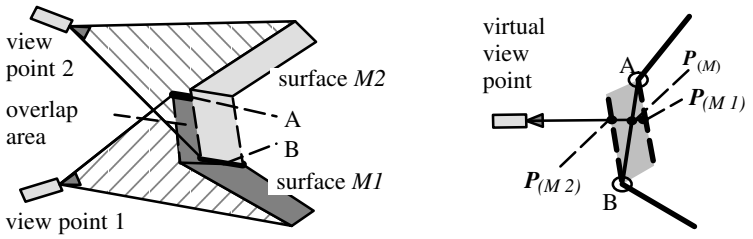
The optical flow **d** is derived directly from the spatial–temporal image gradients when the image model of locally linearized image intensities is assumed. This brightness–continuity constrained is widely used in flow computation. The relationship between the image gradients **g** and the camera motion parameters is found by substituting the flow equation (3) into the measurement equation (4):

$$\Delta I = \boldsymbol{g}(x,y)^T \boldsymbol{d}(x,y) = \frac{\left(\boldsymbol{PG}^T \cdot \underline{R_v} - \boldsymbol{T_v}^T\right)\left(g_x \cdot \boldsymbol{H} + g_y \cdot \boldsymbol{V} - (g_x p_x + g_y p_y + \Delta I) \cdot \boldsymbol{A}\right)}{(\boldsymbol{P}-\boldsymbol{C})^T \cdot \boldsymbol{A}} \quad (4)$$

The intensity–based measurement equation (4) can be computed for each point in the scene that holds image gradients above a noise threshold. From all points the motion vector $\mathbf{X_v} = (\mathbf{R_v}, \mathbf{T_v})^T$ is computed using a minimum variance estimator over all measurements [16].

4.2 Surface Fusion

While the object rotates in front of the camera, new surface regions appear that need to be modeled and appended incrementally to the existing surface model once a considerable amount of new scene contents appears. Fig. 6 explains the principle of surface fusion. Two partially overlapping surfaces are computed from the view points 1 and 2 (Fig. 6a). The overlap area is bounded by the lines A and B. Due to depth modeling uncertainties and misalignments in the motion tracking the surfaces do not overlap perfectly. Fusion must take care of the misalignment and provide a smooth transition between surface *M1* and *M2* in the overlap area. The gap between both surfaces is closed by weighted depth interpolation. A virtual camera is placed facing



a) Overlap of surface models

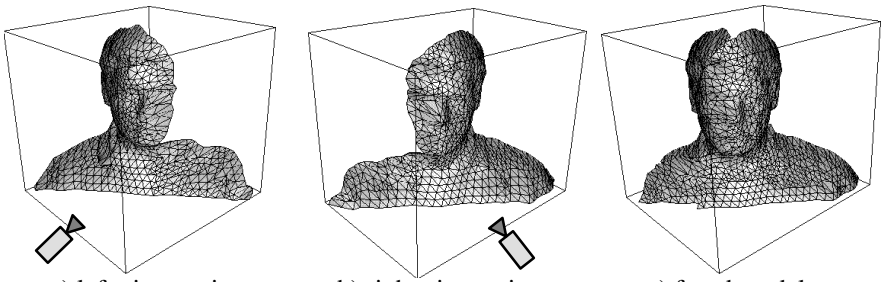
b) closing the overlap area with interpolation

Fig. 6: Fusion of overlapping surface models.

the overlap area and the depth of both surfaces to the camera is computed (Fig. 6b). The fused surface in the overlap area is then interpolated between *M1* and *M2* with a weighting function $I_{(P)}$ that ensures a continuous and smooth transition in the overlap area.

$$\boldsymbol{P}_{(M)} = I_{(P)} \cdot \boldsymbol{P}_{(M1)} + (1 - I_{(P)}) \cdot \boldsymbol{P}_{(M2)} \tag{5}$$

An example for the fusion of two views, rotated 60°, for the sequence *Head* is shown in Fig. 7 where the surfaces from a left (a) and right (b) view of the head are fused to the more complete surface model (c).



a) left view point b) right view point c) fused model

Fig. 7: Result of fusing surfaces from two view points of scene *Head.*

5 Results of Head Modeling

The overall performance was tested with the sequence *Head* where a person rotates on a swivel–chair in front of a stereoscopic camera and 160 stereoscopic image pairs were taken during one rotation of the person. Starting with a single view point, a first surface model was computed as described in Chapter 3. The object motion was tracked between the frames and a new surface model was computed for each 10th frame (rotation 25°). Fusion between the view points was performed each 40th frame (rotation 100°) until the surface model was completely closed. Only the top of the head is still open because it was never seen during recording. Surface texture was applied by selecting the best view point for each triangle texture map and smoothing the intensity transitions between different view points. Fig. 8 displays synthesized

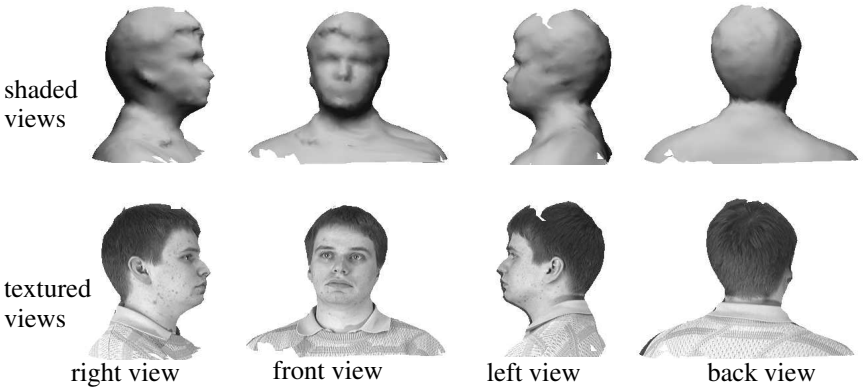


Fig. 8: Synthesized views of the complete 3–D head model.

views of the complete model from all sides. The geometry is modeled with high precision and the surface texture adds realism. The average rotation error is below 1% after a full rotation and the shape is measured to an accuracy of 3 mm [16]. The model consists of a wire–frame with 3250 vertices and a texture map with 1024x960 pixels. It is stored in Inventor file format which enables the user to visualize the model in computer graphics systems.

6 Conclusions

An automatic 3–D scene modeling system was discussed that is capable of building complete surface models from stereoscopic image sequences. The system computes dense depth maps and 3–D surface models as well as the object and camera motion from a stereoscopic image sequence and fuses the surfaces from different view points into a consistent object model. The system successfully models a variety of scenes assuming piecewise smooth surfaces. The application presented in this contribution enables the automatic modeling of the head–shoulder part of a person with high precision. No prior knowledge of the person's shape or motion is necessary for this task.

Acknowledgement: This work has been supported by a grant of the Research Institute of the German TELEKOM.

7 References

[1] J.K. Aggarwal, N. Nandhakumar, "On the Computation of Motion from Sequences of Images –A Review," *Proc. of the IEEE*, Vol. 76 (8), pp. 917–935, Aug. 1988.

[2] J. Aloimonos, D. Shulman, "Integration of Visual Modules," *Academic Press*, San Diego, U.S.A., 1989.

[3] H.H. Baker, T.O. Binford: "Depth from edge and intensity based stereo," *Proc. seventh Int. joint Conf. Artif. Intell*. pp. 632–636, 1981.

[4] Blonde, L.,"The MONA LISA Project: General Presentation", *Proceedings on the European Workshop in Combined Real and Synthetic Image Processing for Broadcast and Video Production*, VAP Media Centre, Hamburg, Germany, Nov. 1994.

[5] Brink, G.,"Subpixelgenaue Disparitätsschätzung aus Stereobildpaaren", *Master Thesis*, Universität Hannover, 1994.

[6] Cox, I., Hingorani, S., Maggs, B., Rao, S.," Stereo without Regularisation", *British Machine Vision Conference,* Leeds, UK, pp. 337–346, David Hogg & Roger Boyle (ed.), Springer Verlag, 1992.

[7] Durisch, P. "Photogrammetry and Computer Graphics for Visual Impact Analysis in Architecture", *Proceedings of ISPRS Conference 1992*, Vol. 29, B5, pp. 434–445.

[8] Falkenhagen, L.," Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continuos Surfaces", *European Workshop on Combined real and synthetic image processing for broadcast and video productions*, 23–24. 11. 1994, Hamburg, Germany.

[9] H. Harashima, F. Kishino, "Intelligent Image Coding and Communications with Realistic Sensations – Recent Trends,", *IEICE Transactions*, Vol. E 74 (6), pp. 1582–1592, June 1991.

[10] R. A. Jarvis, "A Perspective on Range Finding Techniques for Computer Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5 (2), pp. 122–139, March 1983.

[11] F. Kappei, "Modellierung und Rekonstruktion bewegter dreidimensionaler Objekte aus einer Fernsehbildfolge", *Ph.D. Thesis*, University of Hannover,1988.

[12] R. Koch,"Dynamic 3D Scene Analysis through Synthesis Feedback Control", *IEEE Trans. Patt. Anal. Mach. Intell., Special issue on analysis and synthesis,* Vol. 15:6, June 1993.

[13] R. Koch, "Automatic Reconstruction of Buildings from Stereoscopic Image Sequences", *Eurographics '93*, Barcelona,Spain, 1993.

[14] R. Koch, "3–D Surface Reconstruction from Stereoscopic Image Sequences", *Proceedings of 5th International Conference on Computer Vision*, pp. 112–114, IEEE Computer Society Press,Los Alamitos, CA., USA, 1995.

[15] R. Koch, "Surface Segmentation and Modeling of 3–D Polygonal Objects from Stereoscopic Image Pairs", *ICPR Conference '96*, Aug. 1996, Wien.

[16] R. Koch, "Automatische Oberflächenmodellierung starrer 3D–Objekte aus stereoskopischen Rundumansichten", Dissertation, Universität Hannover, 1996.

[17] Negahdaripour, S., Hayashi, B.Y., Aloimonos, Y., "Direct Motion Stereo for passive Navigation", *IEEE Trans. Robotics and Automation*, Vol. 11 (6), December 1995.

[18] A.N. Netravali, J. Salz, "Algorithms for Estimation of Three–Dimensional Motion," *AT&T Technical Journal*, Vol. 64 (2), 1985.

[19] A. Pentland, B. Horowitz, "Recovery of Nonrigid Motion and Structure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13 (7), pp. 730–742, July 1991.

[20] Robert, Ph, Ogor, F.,"Joint Estimation of depth maps and camera motion in the Construction of 3D Models from a Mobile Camera", *Proceedings on the European Workshop in Combined Real and Synthetic Image Processing for Broadcast and Video Production*, VAP Media Centre, Hamburg, Germany, Nov. 1994.

[21] Terzopoulos, D., "The computation of visible–surface representations", *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol 10, pp.417–438, 1988.

[22] D. Terzopoulos , D. Metaxas, "Dynamic 3D Models with Local and Global Deformations: Deformable superquadratics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13 (7), pp. 703–714, July 1991.

[23] Tirumalai, A.P., Schunck, B.G., Jain, R.C.,"Dynamic Stereo with Self–Calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14 (12), pp. 1184–1189, Dec. 1992.

[24] Tomasi, C., Kanade, T, "Shape and motion from image streams under orthography: A factorization method", *IJCV 9*, 1992, pp. 137–154.

[25] Tsai, R.Y.,"A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology using off–the–shelf TV Cameras and Lenses", *IBM Research Report RC 11413*, Sept. 1985, Yorktown Heights, NY, USA.

[26] Yakimowski,Y., Cunningham,R,"A System for Extracting 3D Measurement from a Stereo Pair of TV Cameras", *CVGIP, Vol 7*,pp 195–210, 1978.