Surface Segmentation and Modeling of 3–D Polygonal Objects from Stereoscopic Image Pairs

Reinhard Koch
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Abteilung "Automatische Bildinterpretation", Leitung C.–E. Liedtke
Universität Hannover, Appelstr. 9a, D–30167 Hannover, Germany,
E–Mail: koch@tnt.uni–hannover.de

Abstract

An approach to automatically generate 3–D polygonal models from stereoscopic image pairs of piecewise planar objects is presented. Dense disparity maps are computed by constrained epipolar block matching. Local surface orientations are computed from the quantized disparity map using a spline approximation under explicit consideration of quantization noise. The local surface orientation is then clustered into regions of similar surface orientation to find the dominant object planes. A photo realistic 3–D polygonal model of the object is constructed by fitting planar polygons to the surfaces and by mapping original image texture to the model.

1 Introduction

The recent advances in multimedia technology and virtual reality applications show that there is a wide range of applications where computer generated 3–D environments are desirable, like in architecture visualization [1], virtual television studios [2], virtual presence for video communications [3] and general "virtual reality" applications. In all these applications real objects have to be scanned and 3–D models must be generated automatically with low cost.

There have been some developments towards systems that automatically compute such models in a controlled environment like 3D scanners [4][5]. We concentrate on a system that analyses 3–D scenes from stereoscopic image sequences and operates either data driven [10] or by interpreting the scene based on explicit knowledge about the objects [7]. This system consists of three main modules: Stereoscopic image analysis, 3–D reconstruction, and Scene interpretation using an explicit knowledge base. Image analysis extracts 3–D features from the image sequence while the 3–D reconstruction solves the problem of view point estimation and data fusion from multiple view points. The interpretation module controls the modeling and maintains the modeling and scene knowledge.

This contribution focuses on the image analysis module, namely the reliable extraction of 3–D surface descriptions from a dense disparity map. The other modules of the scene analysis system are described in other publications [6][9].

Man–made objects like buildings are mostly composed of planar surfaces and can be approximated by 3–D polygonal models. The presented approach extracts such polygonal models from stereoscopic image pairs by segmenting stereoscopic depth maps into planar regions and extracting the bounding contours of these. Depth map segmentation can be viewed in the context of range segmentation and quite some work has been done in this field. Segmentation from local differential surface geometry was investigated[11],[12] as well as fitting parametric models [13]. Simultaneous regularization and crease detection is another way of surface segmentation [14]. According to [13], the main drawback of the differential methods is that data smoothing in the presence of sensor and quantization noise is difficult. Our approach exploits local surface properties but the data smoothing is adapted to the sensor noise (especially the quantization noise) and is able to compute local surface differentials while maintaining tight data fit and preserving discontinuities in surface orientation.

The generation of a generic 3D surface model from stereoscopic image pairs by computing image correspondence, surface interpolation and triangular mesh approximation is described in [9]. Based on the assumption that the object consists of locally planar surfaces, as is the case with buildings, the object can be segmented into planes by searching for the regions of similar surface orientations. The process of modeling 3–D polygonal objects will be described in this contribution.

Section 2 describes the computation of local surface orientation and segmentation from quantized disparity maps which are computed by stereoscopic correspondence analysis. Section 3 deals the generation of a 3–D polygonal model and Section 4 summarizes the approach.

2 Surface segmentation from a stereoscopic image pair

This section deals with the segmentation of an object into planar regions. The object is observed with a calibrated stereoscopic camera and a dense disparity map is estimated by image correspondence analysis. The disparity map is interpolated with a continuous spline surface to smooth the disparity measurements. From the spline surface the local surface orientation is computed by spatial differentiation and regions of similar orientation are clustered to find planar object surfaces.

In order to compute depth from the camera pair a sensor calibration is needed. In an off-line calibration process [16] 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 images are then 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 disparity estimation because epipolar search lines coincide with the image scan lines.

2.1 Disparity estimation

Disparity estimation exploits the fact that a surface point of the real object projects onto the images of the stereoscopic camera at different positions in each image. The projected image positions are constrained to lie on the epipolar lines because of the epipolar constraint in stereo vision. The position difference along the epipolar line is called disparity [17]. The algorithm used in this contribution was described in detail by Falkenhagen[18]. A disparity map is obtained by searching along the epipolar lines using correlation matching and dynamic programming. 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 [15]. 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. [19]. Further details on this procedure can be found in [18].

The disparity estimates are quantized to 1 pixel resolution due to the step size of the search procedure. Because the disparity in smooth regions changes slowly with low disparity variance the quantization error will effect the signal and must be taken into account. In addition to the disparity estimate d_i an estimate of the measurement quality c_i is computed by the local normalized cross correlation (NCC) and recorded in a disparity confidence map.

2.2 Disparity interpolation

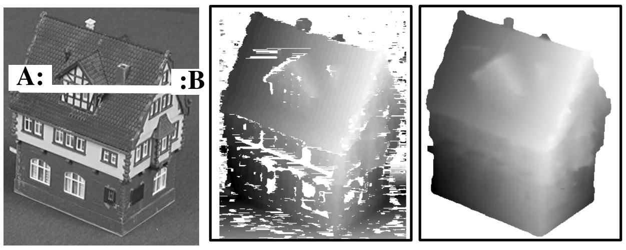
The estimated disparity map contains only discrete and quantized estimates of the real disparity from which local surface derivative can not be computed directly. Instead the disparity map is interpolated with a second order spline approximation that can be differentiated easily. A multi grid surface reconstruction algorithm described by Terzopoulos[21] was chosen to calculate the interpolation with a finite element approximation. It is assumed that the surface 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 white gaussian noise or no estimate is available. The physical model of a thin plate can be formulated as a variational functional of the Euler–Lagrange equation $\Delta^2 u_{(x,y)} = 0$ with additional constraints. The interpolation solves the problem of minimizing the total energy function **E** that consists of the inner energy of a thin plate $u_{(x,y)}$ which is deformed by external forces of the disparity measurements d_i at position (x_i, y_i).

$$\boldsymbol{E} = \sum_U \beta_i (u_i - d_i)^2 + \lambda \iint_U \frac{1}{2}\left(u^2_{xx} + u^2_{yy} + 2u^2_{xy}\right) dxdy \quad (1)$$

The indices (x,y) of u indicate the partial derivatives with respect to x and y. **E** consists of a measurement term and a regularization term. The measurement error $(u_i - d_i)$ is controlled by an individual weight β_i for each measurement d_i inside the object region **U**, while the plate energy term is weighted by the regularization factor λ. A high value for λ results in a smooth surface (strong regularization) while a small value of λ allows to fit the data more closely (weak regularization). The mechanical analogue to Eq. (1) is the model of a thin plate under external forces which are formed by springs with spring constant β_i and elongation $(u_i - d_i)$. The measurement weight β_i is derived from the disparity confidence c_i. It is switched off in undetermined areas with measurement confidence below a threshold c_{min}. Undetermined areas are merely interpolated.

$$\beta_i = \begin{cases} c_i & \text{for } (x_i, y_i) \ in\ \boldsymbol{U} \wedge (c_i \geq c_{\min}) \\ 0 & else \end{cases} \quad (2)$$

Fig. 1 displays some results of disparity estimation and interpolation. Fig. 1a shows the left original image of a stereoscopic image pair from the sequence VILLA, Fig. 1b the estimated disparity map, and Fig. 1c the interpolated disparity map. Disparity maps are printed color coded: light intensity values correspond to large disparities, dark values to small disparities, and transparent areas are undetermined.



a) left image (with measurement profile) b) quantized disparity map c) interpolated disparity map

Fig. 1: Disparity estimation and interpolation for the object VILLA.

2.3 Interpolation of quantized disparity maps

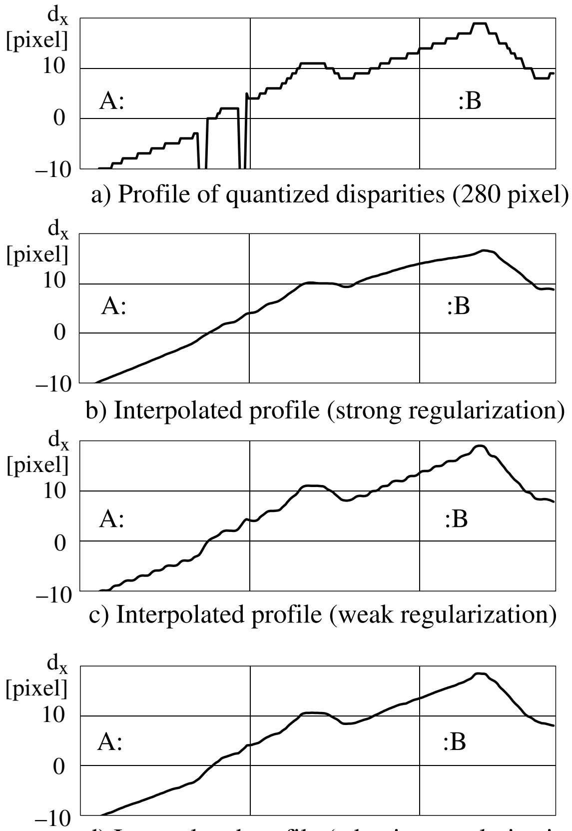
The interpolation function as defined in Eq. (1) is only valid for the model of uncorrelated gaussian measurement noise. The disparity map, however, is quantized due to the limited resolution of the search procedure. If the quantization is coarse and the measurement noise is small compared to the quantizer steps, then quantization noise is correlated with the signal and the minimization of **E** approximates the quantized signal rather then the true signal [20]. To circumvent this problem the quantization must be considered directly as part of the signal. From the assumption of locally planar surfaces we know that the surface has locally constant slope but changes rapidly at the region boundaries which are unknown in advance. This produces conflicting conditions for the regularization. Inside a planar region the data should be smoothed (strong regularization), at the boundaries the surface should fit to the data (weak regularization). We are therefore looking for a regularization that automatically adapts to region boundaries. Such a signal adaptive regularization can be found by introducing a quantizer threshold to the weight β_i.

All disparity measurements with a measurement error below the quantizer error do not contribute to the surface minimization and have to be discarded. Only measurement errors above the quantizer threshold are considered. This condition can be satisfied by non–linear switching of the measurement energy term in Eq. (1). The measurement weight in Eq. (2) is modified to incorporate the quantization noise. The mechanical analogue of this is to have a non linear spring constant with zero force below an elongation of $|(u_i-d_i)| < q/2$.

$$\beta_{qi} = \begin{cases} \beta_i & \text{for } \left(|u_i - d_i| > \frac{q}{2}\right) \\ 0 & \textit{else} \end{cases} \tag{3}$$

Minimizing Eq. (1) with weights according to Eq. (3) results in an adaptive interpolation. The disparity profile is smooth in planar regions but fits closely to the data at corners and boundaries.

The advantage of the proposed method is demonstrated in Fig. 2 by analyzing a disparity profile A:–:B that was taken from the roof area of the house extending over 280 pixel in x–direction (Fig. 1a). Parts of the front wall, the roof area and a small roof window area are covered by the profile. The disparity profile changes orientation at the corners of the object. Between the corners a smooth and almost linear disparity profile is expected. In Fig. 2a) the disparity profile from disparity estimation is displayed. The disparity range is about 30 pixel. The quantization effects can be seen as steps in the estimated disparity profile. Interpolation of the quantized profile is displayed in Profiles b= strong regularization, c=week regularization and d=adaptive regularization with the proposed method. In Fig. 2b the profile is smooth but deviates from the input data at orientation discontinuities. Weak regularization in Fig 2c fits closely to the data but the quantization error is visible. Interpolation according to Eq. (1) and (3) demonstrates the advantage of the proposed method. The quantizer error is eliminated and the disparity profile 2d is smooth while maintaining the surface creases.



a) Profile of quantized disparities (280 pixel)

b) Interpolated profile (strong regularization)

c) Interpolated profile (weak regularization)

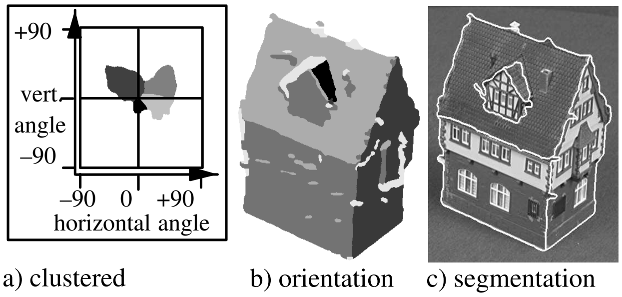
d) Interpolated profile (adaptive regularization)

Fig. 2: Comparison of interpolation for disparity profile A:–:B from Fig. 1a

2.4 Segmentation into planar regions

This segmentation exploits the knowledge that the scene to be modeled mostly consists of smooth or planar surfaces that may have some creases and breaks. The approach for the segmentation is therefore to extract regions of similar surface orientation and then to group these regions to form surfaces. Local surface orientation computed from the spatial surface derivative of the interpolated disparity map for each pixel and a two–dimensional histogram of surface orientations is accumulated. For each surface normal the corresponding horizontal and vertical angle is computed and added to the histogram. The local peaks in the histogram correspond to the most likely surface orientations while the valleys correspond to the boundaries between surfaces [22].

The surface orientation histogram is clustered by searching for the valleys and assigning each peak a surface label. Clustering takes into account that only major peaks of the histogram contain reliable region information and minor or local peaks are merged. Only the dominant directions receive an orientation label. Projection of the orientation labels into the image reveals the clustered surface regions. Fig. 3a shows the result of orientation clustering in the orientation histogram. Each color conforms to a certain surface orientation. The projection of the orientation labels onto the image plane in Fig. 3b reveals that all major orientations were segmented properly.



a) clustered orientation histogram b) orientation segments c) segmentation superimposed on image

Fig. 3: Segmentation from histogram clustering

Some post processing is performed to ensure spatial homogeneity of the surface regions and to cope with outliers due to noise. Small regions are merged with neighboring regions if the following conditions are met:

– the neighboring region is much larger than the observed region,

– the orientation angle between the regions is smaller than a threshhold,
– the average surface distance between the regions is smaller than a threshhold.

The selection of the conditions and the thresholds depend on the desired level of detail. For polygonal modeling only the dominant planes are kept and all the smaller regions (like the chimney and the balcony in the example VILLA) are discarded. For a very detailed surface model all regions are kept which may result in a complicated surface structure. In this case it may be appropriate to construct a generic wire–frame that covers the details [9]. The level of detail is chosen interactively or may be generated by an automatic scene interpretation system as discussed in [7]. In Fig. 3c the segmentation for the five largest planar regions is shown. Superimposing the region boundaries onto the image in Fig. 3c displays a good fit to the true surface boundaries.

3 Modeling of a 3–D polygonal surface model

Each segmented region corresponds to a smooth 3D–surface with boundaries at orientation discontinuities. For polygonal objects each region describes a planar surface. Each object consists of K surfaces and each surface is described by surface normal **n**, surface distance l, and the surface region boundary with arbitrary shape. For efficient surface representation a polygonal description is needed that approximates the true object shape with only a few points.

3.1 Plane fitting and plane intersection

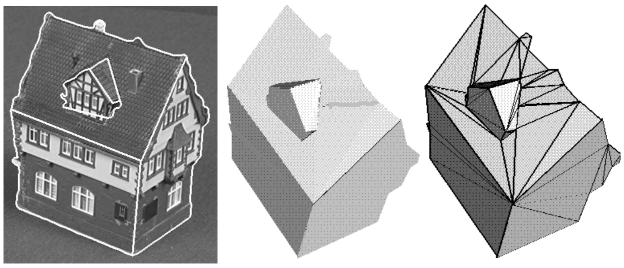
The object is composed of K surfaces. If the surface k contains I surface points, a weighted least squares regression is computed to estimate the surface parameters $(\mathbf{n}_k, l_k)$. The surface point $\mathbf{P}_i$ is computed from the disparity value d_i and the calibrated sensor parameters [10]. The weight β_i of each surface point is computed from the image uncertainty in Eq. (2).

$$\sum_{i=1}^{I} \beta_i (\boldsymbol{P}_i \cdot \boldsymbol{n}_k - l_k)^2 \rightarrow \min \qquad (4)$$

Each surface is bounded either by the boundary between object and background with an arbitrary shape or by the intersection with neighboring planes. The plane intersection is a 3D–line which corresponds to the edge between two object surfaces. It projects as a straight line segment onto the image. The intersection of three surfaces conforms to a point at an object corner. The intersection lines between two surfaces are of infinite length. The lines are limited by intersecting the projected line segments with the neighboring region boundaries in the image. Only intersections that project into the segmented surface regions and that are not occluded by other surfaces are considered.

3.2 Boundary approximation and polygonal surface modeling

Surface boundaries between object and background have an arbitrary shape that is approximated by straight line segments based on a distance threshold that measures the distance between the line segment and the boundary in the image plane. This threshold can be varied to control to the level of detail that should be preserved. The approximating line segments form a 3–D polygon for each region. Each polygon is tessellated into a triangular mesh. The original image texture is projected onto the triangles which allows to synthesize photo realistic views of the model from arbitrary view points [8].

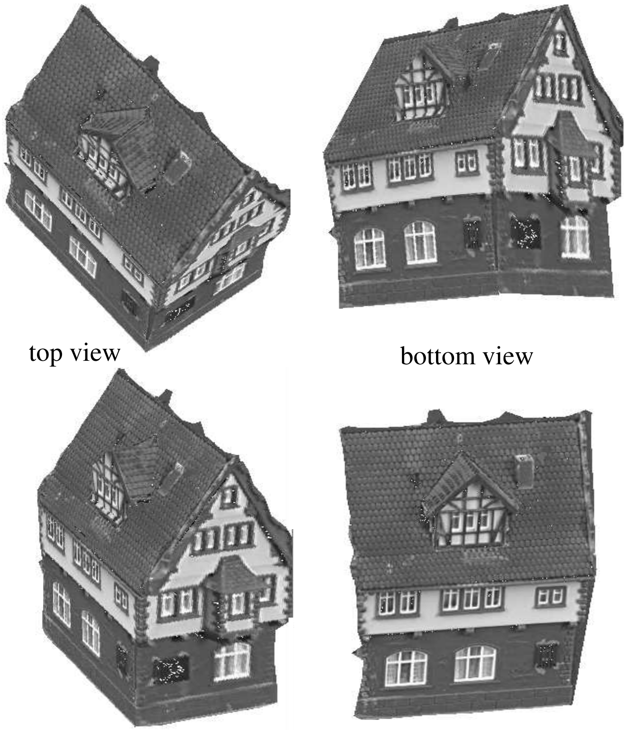


b) plane intersection boundaries super–imposed on image

b) 3–D polygonal model and c) wire–frame approximation (top view, shaded)

Fig. 4: 3–D polygonal surface modeling

Fig. 4 displays results of polygonal modeling for the scene VILLA. The intersections of the surface planes Fig. 4a shows a good fit with the true object boundaries when superimposed onto the image. A maximal boundary deviation of 5 pixel was allowed during boundary approximation. The polygonal model consists of six planar surfaces (Fig. 4b) that are approximated by 46 triangles (Fig. 4c). The occluded regions at the roof window were closed and form a good approximation of the true surface. The fine details, however, like the chimney and the balcony are suppressed.



top view

bottom view

right view

left view

Fig. 5: Synthesized views of textured model VILLA from arbitrary positions

The model was textured from two original images of the sequence and synthetic views of the model are displayed in Fig. 5. Four different views with angles about +/– 45° around the original viewing position were synthesized from the textured 3–D model and show good reconstruction results. Note, however, that due to the planar approximation the chimney and the balcony are not modeled and appear at wrong positions in the images. This problem can only be solved when finer details for the model are included.

4 Conclusions

An approach to automatically generate 3–D polygonal models from stereoscopic image pairs of piecewise planar objects was presented. From a quantized disparity map a smooth spline approximation was computed under explicit consideration of quantization noise. Regions of similar surface orientation were then clustered to find the dominant object planes. A photo realistic 3–D polygonal model of the object was constructed by fitting planar polygons to the surfaces and by mapping original image texture to the model.

The approach is embedded in a knowledge–based system to interpret 3–D scenes and to model objects from multiple stereoscopic image pairs [6][7]. At present some interaction of the user is still needed to adapt the segmentation thresholds and to classify which surfaces are relevant for the modeling and which can be discarded during the approximation. We are presently working to further automate this process during scene interpretation.

Acknowledgements

This research has been supported by a grant of the Research Institute of the German Postal Service TELEKOM.

References

[1] P. Durisch, ”Photogrammetry and Computer Graphics for Visual Impact Analysis in Architectur”, Proceedings of ISPRS Conference 1992, Vol. 29, B5, pp. 434–445.

[2] L. Blonde, ”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.

[3] 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.

[4] W. Niem, R. Buschmann, ”Automatic Modeling of 3D Natural Objects from Multiple Views”, 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] 3–Development Newsletter, Cyberware 3–D Scanners, Monterey, CA, 1995.

[6] C.–E. Liedtke, O. Grau, S. Growe, ”Use of Explicit Knowledge for the Reconstruction of 3–D Object Geometry”, 6th Int. Conf. Computer Analysis of Images and Patterns CAIP’95. 6.–8. Sept. 1995 Prague.

[7] O. Grau, ”A Knowledge Based Scene Analysis System for the Generation of 3–D Models”, 5th International Conference on Intelligent Systems. June 19–21, Reno, Nevada, USA, 1996.

[8] R. Koch, ”Automatic Modelling of Natural Scenes for Generating Synthetic Movies”. In: Vandoni, C.E. and Duce, D.A. (ed.) Eurographics Association 1990. Elsevier Science Publishers B.V. (North–Holland).

[9] R. Koch, ”Automatic Reconstruction of Buildings from Stereoscopic Image Sequences”, *Eurographics ’93*, Barcelona,Spain, 1993.

[10] R. Koch, ”3–D Surface Reconstruction from Stereoscopic Image Sequences”, Proceedings of the ICCV Conference 95, Cambridge, MA, USA, June 1995.

[11] P.J. Besl, R.C. Jain, ”Invariant Surface Characteristics for three–dimensional object recognition in range images”, CVGIP, 33(1), pp.33–80, 1986.

[12] P.J. Besl, R.C. Jain,”Segmentation through variable–order surface fitting”,IEEE PAMI 19(2), pp. 167–192, 1988.

[13] A. Leonardis, A. Gupta, R. Bajcsy,” Segmentation of Range Images as a Search for Geometric Parametric Models”, Intern. Journal Computer Vision Vol. 14, pp.253–277, 1995.

[14] A. Blake, A. Zissermann, ”Visual Reconstruction,” *MIT–Press*, Cambridge, 1987.

[15] D. Marr, ”Vision – A Computational Investigation into the Human Representation and Processing of Visual Information”, W.H. Freeman & Co., New York, USA, 1982.

[16] 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.

[17] O. Faugeras, ”Three–Dimensional Computer Vision – A Geometric Viewpoint”, MIT Press Cambridge, USA, 1993.

[18] L. Falkenhagen, ”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.

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

[20] N.S. Jayant, P. Noll,”Digital Coding of Waveforms, Principles and Applications to Speech and Video”, Prentice Hall Signal Processing Series, 1984.

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

[22] K. Pakzad,”Segmentierung von Tiefenkarten aufgrund ähnlicher Flächenorientierung”, Thesis, University of Hannover, June 1994.