XVI IMEKO World Congress
Measurement - Supports Science - Improves Technology - Protects Environment ... and Provides Employment - Now and in the Future
Vienna, AUSTRIA, 2000, September 25-28

OBJECT PARAMETER ESTIMATION: A SYNERGY OF CV AND VR

Z. Houkes, F. van der Heijden and P.P.L. Regtien
Laboratory for Measurement and Instrumentation
Department of Electrical Engineering
University of Twente, NL-7500 AE, The Netherlands

Abstract: A model-based approach, used to estimate object parameters from grey-level images, is described. Position and orientation of a cube are estimated from mono- and stereo-images. The Digital Elevation Map (DEM) of a wedge is computed from an animated sequence of 2 successive images acquired by a moving camera.
Keywords: computer vision, model-based, parameter estimation, virtual reality

1 INTRODUCTION

Computer Vision (CV) [5] is "the science that develops the theoretical and algorithmic basis by which useful information about the world can be automatically extracted and analysed from an observed image, image set, or image sequence from computations made by special-purpose or general-purpose computers. Such information can be related to the recognition of a generic object, the three-dimensional description of an unknown object, the position and orientation of the observed object, or the measurement of any spatial property of an object, such as the distance between two of its distinguished points or the diameter of a circular section." Much progress has been made in this area since the 60's, which is the period the interest of scientists [4,11] for this area has been increased rapidly.

Virtual Reality (VR) is the inverse trajectory of CV, comprising the generation of 'realistic' images of 'virtual' scenes. Figure 1 shows an example of a 256x256 grey-level image generated by PARSCENE[1]. The ocular point spread function (psf) is a δ-function. Many programs for animation and rendering are already available for Windows, Mac OS and many other platforms. Most of these programs for computer animation are used for fun, but the use of VR, e.g. in scientific visualisation, is an upcoming application [9]. Serious research on modelling of virtual actors, including their dynamics to make them moving through a virtual world as realistic as possible, is being done at the Computer Graphics Lab (LIG) [13]. The dynamics are of great importance when image sequences are used to reconstruct the observed world.

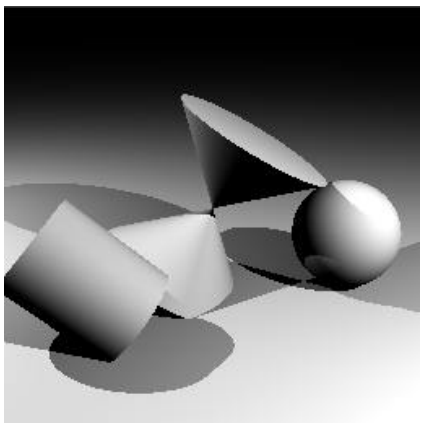

Figure 1. VR image of simple objects illuminated by 2 point sources

The synergy of CV and VR opens the possibility to improve the modelled world by comparing the observed world and the modelled world by their respective images. This can be done by human interaction, or automatically, as is in some sense the subject of this paper. In this work the *simultaneous* estimation of parameters of 3D-objects from *grey level images* using *a model-based approach* is presented.

2 IMAGED-BASED MEASUREMENT

Images ([1], [5]) are spatial representations of a 2-or 3-dimensional scene containing objects. In computer vision we usually deal with *digital* images, represented by m-vector discrete valued image functions $\boldsymbol{f}(\mathbf{x})$. Usually, m=1, and the domain and range of $f(\mathbf{x})$ are discrete. The domain of f is finite, usually a rectangle, and the range of f is positive and bounded: $0 \le f(\mathbf{x}) < N$, with N some integer. For binary images N=1, while for grey-level images N=255 is quite usual. The domain is determined by the sampling rate. Usual values for the size of the rectangle are 512, 256, etc.

The motivation to choose for using grey-level images instead of binary images as the measurements is determined by 4 points of consideration:

[1] PARSCENE is a software product developed by the Laboratory for Measurement and Instrumentation

1. Images contain information about *geometric* and *radiometric* parameters of 3D-objects. So the use of images is expected to allow the simultaneous measurement of quantities from at least 2 different physical domains.
2. The use of grey-level images is inspired by the idea to use all the information from the image(s), instead of throwing away most of the grey-level information, which is being done in 'early processing' [1] to recover intrinsic structure [2]. Every image element contains information about the quantities to be estimated. The suppression of the noise will take advantage of many observations. The variance of the noise will decrease under certain conditions inversely with *n*, the number of the observations.
3. The use of all the pixels allows geometric parameters determined at sub-pixel accuracy.
4. The noise occurring in grey-level images acquired by video cameras, although caused by different kinds of mechanisms, can be modelled approximately by Gaussian noise.

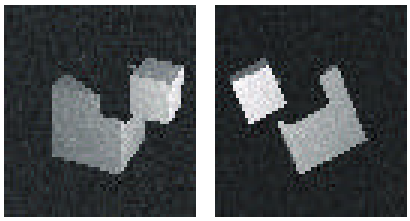

Figure 2. Noisy reference image pair (64x64) of an occluded cube for σ_n=10

On the other hand, the choice for grey-level images, is in some sense also a restriction. We could have used colour images, which are usually of lower spatial resolution, but which can be of interest for estimating radiometric parameters. We leave this as a point for further research.

The higher complexity of the modelling is a disadvantage. Errors in the *structural* descriptions of the scene or the image formation process will cause errors in the estimated parameters.

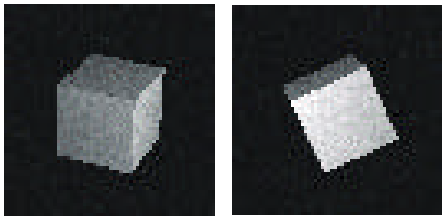

Figure 3. Noisy reference image pair (64x64 pixels) of a cube with σ_n=10

3 MODEL-BASED PARAMETER ESTIMATION

The grey-levels of the pixels are not directly related to the geometric quantities that we want to know. As a consequence a model will be required to reconstruct the information about geometric parameters like length, position, etc. A model-based approach also opens the use of the approach for more types of applications such as those mentioned in section 4.

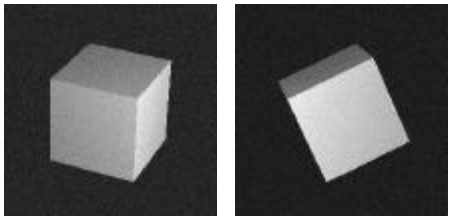

Figure 4. Noisy reference image pair (256x256 pixels) of a cube with σ_n=10

A problem encountered in computer vision problems is the occlusion (figure 2) problem. The model-based approach is able to cope in a 'natural' way with this problem, as long as the information available in the images guarantees the 'observability' of the geometric and/or radiometric parameters. The clue to cope with the occlusion problem is to use 3D-models to describe the scene, and to use a (non-linear) measurement function that models the image formation and acquisition system. Occlusion should be interpreted in its most extensive sense. It comprises self-occlusion as well as occlusion by other objects (figure 2). The model-based approach uses geometric and radiometric models to describe the objects, including the light sources and cameras constituting the scene. Figure 3 and 4 show examples of stereo images of a scene consisting of a single cube. The size of the ROI[2] is 64x64 and 256x256 respectively. The modelling of the dynamics of objects in the scene, e.g. a robot moving around or a camera mounted in an aeroplane observing the earth, is not part of the work described. This point is planned as part of further research.

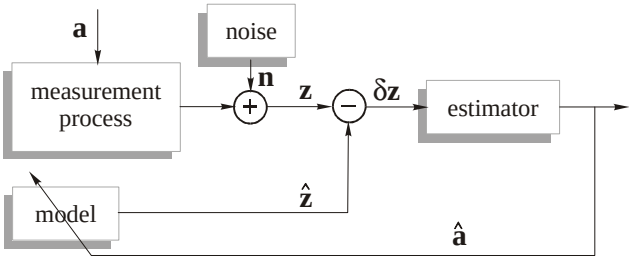

Figure 5. Iterative estimation of model parameters

3.1 Non-linear parameter estimation

For a set of geo- and radiometric parameters **a**, the physical imaging process, indicated in figure 5 as the measurement process, produces a set of images (e.g. the set in figure 3). The image pixel grey

[2] Region Of Interest

levels, which are put row by row in a vector **z**, constitute the set of measurements, which may be corrupted by (additive) noise $\mathbf{n}_0$.

$$\mathbf{z} = \mathbf{h}(\mathbf{a}) + \mathbf{n}_0 \tag{1}$$

The functional relation **h**(**a**), describing the image formation, is usually non-linear. In accordance with Liebelt [8], the estimator in figure 5 determines iteratively the unknown parameters using a least squares estimator. This estimator estimates the error

$$\delta\mathbf{a}_i = \mathbf{a} - \hat{\mathbf{a}}_i \tag{2}$$

still existing between the real parameters **a** and their estimate $\hat{\mathbf{a}}_i$, available after the *i*-th iteration, by using the difference

$$\delta\mathbf{z}_i = \mathbf{z} - \hat{\mathbf{z}}_i \tag{3}$$

between the measurements **z** and their prediction $\hat{\mathbf{z}}_i$ obtained with $\hat{\mathbf{a}}_i$ from the model described by $\hat{\mathbf{z}}_i = \mathbf{h}(\hat{\mathbf{a}}_i)$. This estimated error $\delta\hat{\mathbf{a}}_i$ is used to compute a new guess $\hat{\mathbf{a}}_{i+1}$ using $\hat{\mathbf{a}}_{i+1} = \hat{\mathbf{a}}_i + \delta\hat{\mathbf{a}}_i$. Under certain conditions, this process will converge to an optimal estimate of **a**. The measuring function **h(a)** is approximated by a Taylor series expansion about $\hat{\mathbf{a}}_i$. Using the first order approximation of **h**(**a**) and joining the higher order terms and the observation noise in a general noise term **n,** yields:

$$\delta\mathbf{z}_i = \mathbf{H}_i\delta\mathbf{a}_i + \mathbf{n} \tag{4}$$

with $\mathbf{H}_i$ being the Jacobian of **h**(**a**) with respect to the parameters at $\hat{\mathbf{a}}_i$. If there is no a priori information about $\delta\hat{\mathbf{a}}_i$, and assuming that **n** is (nearly) an equal-variance uncorrelated noise term, the correction $\delta\hat{\mathbf{a}}_i$ can be computed from

$$\delta\hat{\mathbf{a}}_i = (\mathbf{H}_i^T\mathbf{H}_i)^{-1}\mathbf{H}_i^T\delta\mathbf{z}_i \tag{5}$$

which is the least squares estimation of $\delta\mathbf{a}_i$. If these conditions are not satisfied, the use of a linear minimum variance unbiased estimator (see Liebelt [8]) should be considered.

3.2 Identifiability of the parameters

To be sure that the parameters can be estimated uniquely, the matrix $\mathbf{H}^T\mathbf{H}$ should be positive definite. This requires the determinant of $\mathbf{H}^T\mathbf{H}$ and all its sub-determinants to be > 0 (Beck [3]). This condition is called the 'identifiability condition'. If this condition is not satisfied, there will be no unique point at which the expectation of the squared sum $Q=E[(\mathbf{z}-\mathbf{h}(\hat{\mathbf{a}}))^T(\mathbf{z}-\mathbf{h}(\hat{\mathbf{a}}))]$ has a minimum. This condition yields under certain conditions the same solution as obtained when the *Trace*[$\mathbf{C_e}$] is minimised. The matrix $\mathbf{C_e}=E[\mathbf{e}\mathbf{e}^T]$ is the so-called error matrix, with the error defined as **e = â-a =** -ä**a** (Houkes [6]). The absence of a unique minimum might be caused by 'sensitivity coefficients' $h_{ij}=\partial h_i/\partial a_j$ of the Jacobian **H** being zero for at least one parameter, or by 'near dependency' of at least 2 columns of the Jacobian **H** containing the sensitivity coefficients $h_{ij}=\partial h_i/\partial a_j$, within the range of the measurements. In the first case the sensitivity coefficients for at least one parameter are zero, which is expressed in the relation for the j^{th} parameter:

$$\sum_i h_{ij}^2 = 0 \tag{6}$$

In the latter case we would like to check the 'degree of dependency' of the columns of the Jacobian **H** regardless of the magnitude of the sensitivity of the measurements to an individual parameter. The numerical computation of the sensitivity coefficients requires the variation of a parameter. The magnitude of this variation determines the magnitude of the sensitivity coefficients and consequently the eigenvalues. The sum of the eigenvalues equals the sum of the diagonal elements of the identifiability matrix! To be able to check the dependency regardless of the magnitude of the sensitivity coefficients, at first a scaled (Houkes [6]) parameter vector $\delta\hat{\mathbf{a}}'_i$ is estimated. From the scaled set of parameters the *j*-th element of the correction vector $\delta\hat{\mathbf{a}}_i$ is computed using

$$(\delta\hat{\mathbf{a}}_i)_j = \frac{1}{s_j}(\delta\hat{\mathbf{a}}'_i)_j \tag{7}$$

with s_j the square root of the j^{th} diagonal element of $\mathbf{H}^T\mathbf{H}$. The sum of the eigenvalues of this matrix equals the number of parameters. The identifiability condition mentioned before is satisfied when the eigenvalues of the normalised matrix differ significantly from zero. When one of the eigenvalues comes close to zero (e.g. <0.05), two or more of the columns of the Jacobian **H** containing the sensitivity coefficients are nearly dependent. See [6] for a detailed discussion about the normalisation of $\mathbf{H}^T\mathbf{H}$ and the eigenvalues.

3.3 Stereo vision

A common problem in computer vision is the extraction of depth parameters of an object from single images. It is not possible to estimate both its size and distance from a single image. Combinations of these parameters don't provide a parsimonious (Beck [3]) set of parameters. In terms of identifiability, it means that size and depth parameters can not be estimated uniquely, because the columns of the Jacobian **H** containing the sensitivity coefficients concerning these parameters are (nearly) dependent. An additional camera might solve this problem. In the case of two cameras observing the scene from different points of view, as in stereo vision, two sets of measurements can be used to estimate the scene parameters. The measurement vector **z** is now composed of the two sets $\mathbf{z}_1$ and $\mathbf{z}_2$ of the individual cameras. In the same way the prediction $\hat{\mathbf{z}}$ is composed of the individual predictions $\hat{\mathbf{z}}_1$ and $\hat{\mathbf{z}}_2$ for both camera images.

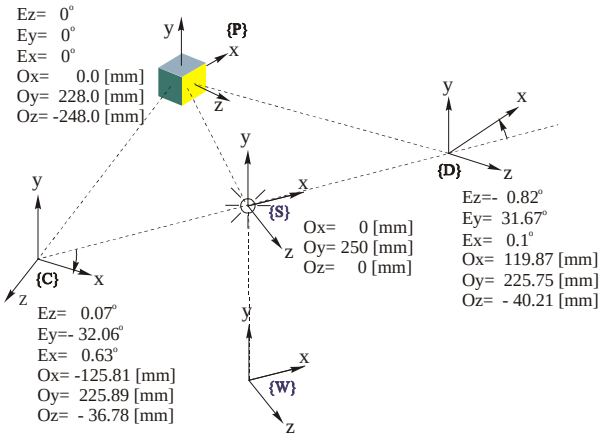

Figure 6. Set-up and parameters for real experiments

$$\mathbf{z}^T = (\mathbf{z}_1^T, \mathbf{z}_2^T); \quad \hat{\mathbf{z}}^T = (\hat{\mathbf{z}}_1^T, \hat{\mathbf{z}}_2^T) \tag{8}$$

The eigenvalues of the normalised identifiability matrix for a binocular camera set-up are expected to differ significantly from zero compared with those for a monocular set-up. This will realise the simultaneous estimation of depth and size parameters from sets of images.

The model-based approach, using grey-level images as the measurements in the described estimator, has the advantage that it can use all the available measurements in the same straightforward manner. The requirement is adequate modelling.

4 APPLICATIONS OF THE MODEL-BASED APPROACH

The model-based approach, as described before, opens the use for different kinds of applications:

1. parameter estimation of simple objects like a cube, a cylinder, etc., eventually extended to deformable objects composed of a set of solids or even to 3D-reconstruction;
2. alignment problems, e.g. occurring in remote sensing when 2 airborne images should be aligned;
3. motion estimation, using sequences of images for object reconstruction;
4. DEM-generation, e.g. by using a sequence of video images to determine terrain elevation;
5. camera-calibration;

The estimation of position and orientation of a cube is used to demonstrate the approach depicted in figure 5. The iterative model-based estimator, based on a least squares criterion, is extended with a Gaussian filter. The stereo images of the figures 2, 3 and 4 are used as the input of the estimator. The results are discussed in the next section.

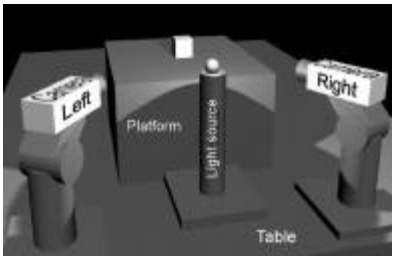

Figure 7. The set-up used for the experiments with real images.

Figure 7 shows an animated picture of the set-up used for experiments with images of a real scene. The (real) reference images of size 256x256, shown in figure 8, are produced with this set-up. The external parameters, comprising the position (O_x,O_y,O_z) and the orientation in Euler angles (E_x,E_y,E_z) of the co-ordinate systems of the cube ({P}), the light source {S} and the cameras {C} and {D}, are given in the experimental set-up depicted in figure 6. The principal distance *f* for both cameras is 12.5 mm. The internal parameters are based on the parameters of the CCD-chip used (see [12] for the details). The calibration has been carried out with the algorithm described in [7].

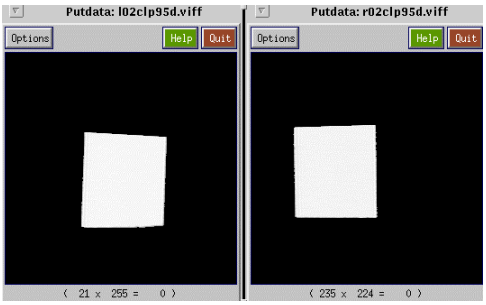

Figure 8. Real reference image pair (256x256) after clipping

The 2nd application concerns DEM-generation. To be able to better illustrate the capabilities of the approach, a wedge was modelled to generate the animated images with

texture shown in figure 9. The wedge has a maximum height of 29.7 [mm] and a width of 35.2 [mm]. The length of the wedge is 122.6 [mm]. The origin of the world co-ordinate system coincides with the upper right corner of the wedge, which is at the side where the height is 0. The pinhole is located at z=350 [mm] above the ground level. The xy-position of the camera, with a focal length of 6 [mm], is (-10,70) and (-10,50) after a displacement of 20 [mm] in the negative y-direction. The pixel period is 6.5 [µm] in x-direction and 6.25 [µm] in y-direction.

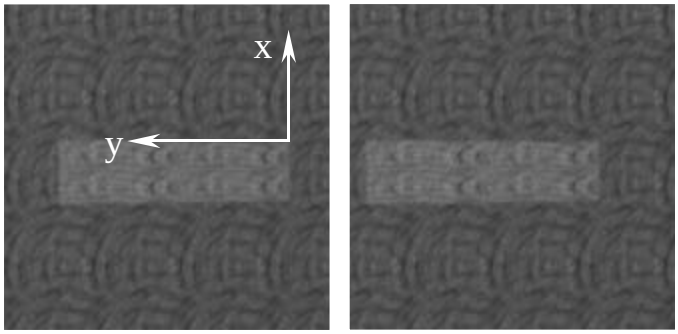

Figure 9. Two animated 512x512 images of a wedge. The 2nd image is obtained by moving the camera over a distance of 20 [mm] in -y-direction.

The model used for DEM-generation approximates the terrain (the wedge in this example) locally by a block of constant height. The 'radiometric' model that will be used in this problem is based on the assumption that the appearance of the object does not change essentially from one image to the next one in the sequence. It is used in combination with the geometric model and the motion of the camera to predict the appearance of the set of corresponding pixels in frame n+1 from a selected set of pixels in frame n. Because the movement of the camera is known, the only unknown parameter is the terrain height h_{ter}. An initial guess of this parameter is used to initialise the iterative estimation process. This procedure enables the computation of the sets of pixels in two consecutive images corresponding with the surface of the blocks given their height. The procedure starts with a set of pixels $\{I_n\}$ in the first image (frame n) to compute the corresponding area on the earth surface, given an initial estimate of the local terrain height. This area can be used to compute the corresponding set of pixels $\{I_{n+1}\}$ in the second image (frame n+1). Repeating the last part for a slightly increased estimated height of the terrain enables the computation of the Jacobian. The initial estimate of h_{ter} should be within the region of convergence of the estimator.

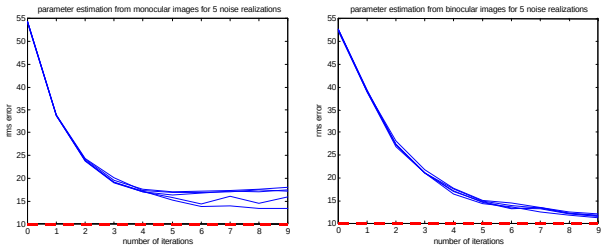

Figure 10. Propagation of the rms error

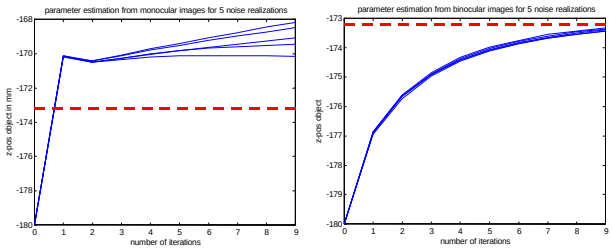

Figure 11. Cube distance parameter O_z

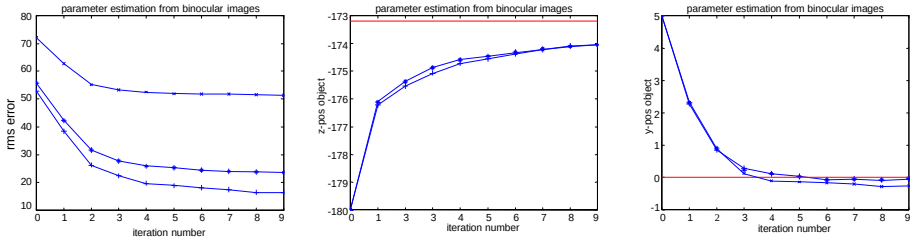

Figure 12. rms error, z-position (depth) and y-position of an **occluded** cube estimated from **binocular** images of **64x64** pixels disturbed with additive noise (σ_n=10, 20, 50) and Gaussian filtering (σ_{psf} =2.0).

5 SOME RESULTS AND CONCLUSIONS

Figures 10 and 11 show for 5 different realisations of the noise (σ_n=10) and for a filter parameter σ_{psf}=1, typical results of the correct convergence of rms error and the cube distance parameter *Oz* (depth parameter) for binocular (right) images of 64x64 pixels (figure 3) and the convergence to an incorrect value for monocular images (left). The results for 6 other parameters show a corresponding behaviour. Only for the case of binocular images (right) the rms error decreases to the sd of the noise. For monocular images a clear divergence occurs in the rms error after 4 iterations, indicating the occurrence of the same effect in the individual parameters. The keyword to determine the kind of behaviour is 'observability' or 'identifiability'.

The 1st results showed errors in the estimated parameters indicating that the calibration of some internal parameters could not be correct. The best responses that could be obtained with real images

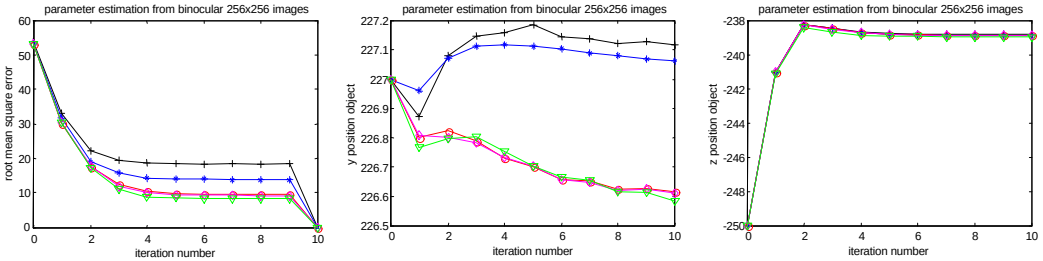

Figure 13. rms error, y-position and z-position (depth) of a real cube estimated from **binocular** images of **256x256** pixels shown in figure 8 and Gaussian filtering (σ_{psf} =8.0).

are depicted in figure 13. To obtain these results, some of the internal parameters were adjusted by hand using information from the error images to minimise the squared error. The reference images contain some 'noise' in the background, which is caused by the foundation of the 20 mm cube and by the reflectance of the background. This 'noise' does not fulfil the zero-mean condition. It has been removed from the reference images by a clipping operation to avoid effects on the estimated parameters. The camera model does not contain a model of the lens. This model discrepancy, which might cause errors (e.g. aliasing), has been left for further investigation.

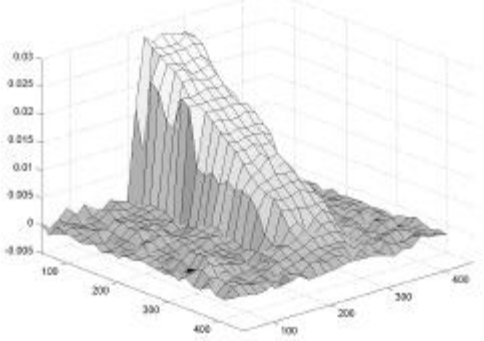

Figure 14. The reconstructed wedge.

The wedge reconstructed from these 2 images is shown in figure 14. The reconstruction used the estimated DEM of a previous reconstruction (started at an initial height estimate of 0 for every point) as the initial estimate. This improved the height estimates of the wedge at those points where, because of the maximum number of iterations was 3, the final value was not reached. The height in [m] is estimated at pixel positions in the 2nd image, started at (row,col)=[50,50] and increasing step by step with 15 pixels in row and column to (row,col)=[425,425]. The size of the region is 4 [mm] in world co-ordinates. The standard deviation computed varies between 1 and 2 [mm] except for a few points lying at the edge of wedge. The standard deviation varies from 2 [mm] at the low end to 8 [mm] at the high end of the wedge.

REFERENCES

[1] Ballard, D.H. and Brown, C.M., *Computer Vision*, Prentice-Hall, Inc., 1982
[2] Barrow, H.G. and J.M.Tenenbaum, "Recovering intrinsic scene characteristics from images." Technical Note 157, AI Center, SRI International, April 1978.
[3] Beck, J.V., K.J.Arnold, Par. estimation in engineering and science, J. Wiley & Sons, NY, 1977
[4] Binford, T.O., "Visual perception by computer", *Proc.*, IEEE Conf. on Systems and Control, Miami, December 1971 (B&B, p. 112)
[5] Haralick, R.M. and Shapiro, L.G.; *Computer and Robot Vision I*, Addison-Wesley, Inc., 1992
[6] Houkes, Z., Estimating 3D Object Parameters from 2D Grey-Level Images, PhD-thesis, UT, 2000, Enschede, The Netherlands, ISBN 90-365-14053
[7] Jaspers, G., Calibration of a set-up for stereo vision, MSc thesis, UT, 1991, rep. nr 91M042
[8] Liebelt, P.B., An introduction to optimal estimation, Addison-Wesly, Reading, MA, USA, 1967
[9] Liere, R. van, "Virtual Reality in Scientific Visualization": http://www.cwi.nl/~robertl/
[10] Lengyel, J., "The convergence of Graphics and Vision", *Computer*, July 1998, pp. 46 - 53.
[11] Roberts, L.G., "Machine perception of three-dimensional solids." In *Optical and Electro-optical Information Processing*, J.P.Tripett et al. (Eds.), Cambridge, MA: MIT Press, 1965. (B&B, p. 113)
[12] Schrijen, G.-J., "Research of the convergence of model-based estimation of object parameters from camera-images", internal report of an experimental investigation instruction.
[13] Thallmann, D., Swiss Federal Institute of Technology (EPFL) in Lausanne: http://ligwww.epfl.ch/

AUTHORS: Associate Prof.Dr.Ir Zweitze HOUKES, Dr.Ir. Ferdinand van der HEIJDEN, Prof.Dr.Ir Paul P.L. REGTIEN, Laboratory for Measurement and Instrumentation, Faculty of Electrical Engineering, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands, Phone Int. +31 53 4892796, Fax Int. ++53 4891067, E-mail: Z.Houkes@el.utwente.nl