

Isotropic Regularization

Mads Nielsen
DIKU,
University of Copenhagen,
Universitetsparken 1, DK-2100 Ø

Abstract

Regularization of ill-posed problems has been applied in various ways to surface reconstruction problems. Typically the smoothness term has been found in an *ad hoc* manner. In this work the *a priori* knowledge of the surface is used to construct the smoothness terms. If all surface normals are considered equally probable, different smoothness terms can be derived for different projections using Bayesian estimation. These smoothness terms implement discontinuous regularization by the Lorentzian estimator under orthographic and perspective projection, and imply a convex solution space, if the standard deviation of the noise is smaller than some quantity. Under stereo projection, occluded areas are punished dependent on the distance from the cameras, as they are more probable to exist on nearby objects than on distant objects. A general scheme of developing smoothness terms under assumptions of isotropy or anisotropy is outlined.

1 Introduction

The problems of depth-extraction or surface reconstruction are ill-posed in the sense of Hadamard [1]. A typical surface reconstruction problem is to find the reconstruction R of the data D from the measurements M , when it is known that the measurement are created by addition of noise N to the data.

$$M(x) = D(x) + N(x) \quad (1)$$

To overcome the ill-posedness, regularization is applied. The regularization implies a reformulation of the ill-posed problem as well-posed by adding a stabilizing term [2]. This stabilizing term is often called *smoothness term*, and incorporates some *a priori* knowledge of the solution. The solution is found by minimizing a weighing of the original problem against the smoothness term. Tikhonov [2] uses quadratic sums of the derivatives of the solution. This implies the minimization of an energy term, which in the case of the first derivative in the smoothness term yields: $E(R) = E_{\text{Data}}(M, R) + \sum_R \lambda R_x^2$ where R is the solution, subscript x denotes the derivative according to x , and λ is a weighing constant between the data term and the smoothness term.

Geman and Geman [4] introduces a line process which is used as an alternative to the smoothness term. The line process is a constant punishment of discontinuities in the solution. The punishment is used in those points, where it yields a lower energy than the smoothness term. In these points, there is no further punishment of a high derivative, and the data term will totally govern the solution, which in these points will have discontinuities. The mathematical formulation is:

$$E(R) = E_{\text{Data}}(M, R) + \sum_R \eta \lambda R_x^2 + (1 - \eta)P \quad (2)$$

where η has to be varied as well as R . P is the constant punishment of the line process. In Equation 2 the solution is not necessarily unique, neither is the solution space convex. Geman and Geman finds one of the solutions of low energy by simulated annealing. Blake and Zisserman [5] reformulates the same strategy as a thresholding of the smoothness term. This implies the minimization of

$$E(R) = E_{\text{Data}}(M, R) + \sum_R \lambda f(R_x) \quad \text{where} \quad f(t) = \begin{cases} t^2 & \text{if } t^2 < P \\ P & \text{otherwise} \end{cases}$$

Blake and Zisserman find the solution by the graduated non-convexity algorithm (GNC), which does not guarantee an optimal solution [6], but is deterministic and faster than the simulated annealing [9]. The GNC is generalized to other regularization schemes by Nielsen [7]. Jensen and Nielsen use genetic algorithms to perform the optimization of integer valued curve fitting using the weak string [11].

The choice of smoothness term is in the above mentioned case made by convenience, and not by any quantitative argumentation. In robust statistics many other shapes of smoothness terms has been introduced. By tradition a least square estimation has been used. To make this robust, outliers has to be detected and not weighed as much as the good data. A survey of methods applied to computer vision is given by Meer et. al. [10]. One typical method is to use a quadratic measure for small values of derivatives, a linear for medium values of derivatives and finally a constant punishment for high values of derivatives.

In neither of the above mentioned works, the smoothness term have been established from more exact assumptions than "The derivatives of the solutions are expected to be small, except in some discontinuity points". Nearly all methods use some threshold of discontinuities, which has to be known in advance. The solutions are very sensitive to this parameter, and to the weighing constant between data and smoothness terms. Furthermore, the implications of the parameter values are far from intuitive [6]. This work provides a method for the incorporation of a *priori* knowledge in depth estimation. If information, which might be of qualitative nature, about the scene is present, this can be used as a *priori* information in the regularization scheme. In an indoor environment, the information, that most surfaces in the scene often are perpendicular to each other, can be used. This information (that the indoor world can be modeled as *Legoland*) is used by Straforini et. al. in a rule based system, and yields very robust results [15]. In a forest environment, information about the structure of vertical stems of the trees and the horizontal ground plane, can be used. The above mentioned *a priori* knowledge are of very specialized nature. In general such knowledge is not present. A very weak assumption is, that the surface normals in the scene are equally distributed in all directions. This corresponds to the weak isotropy of texture introduced by Gårding [13].

2 Bayesian estimation

Bayesian estimation is a technique, where the Bayesian rules from probabilistic theory is used for minimizing the expectation value of a cost function yielding an optimal estimate. In advance one has to know: A measurement M , the probabilistic dependency $p(M|D)$ of the measurement on the real value D , an *a priori* density of the real values $p(D)$ and a cost function $c(R, D)$, which is

the cost of choosing R , when D is the real value. From Bayes theorem we get $p(D|M) = p(M|D)p(D)/p(M)$ where $p(M)$ can be perceived as a normalizing constant, when the measurement M is known. The mean value $C(R|M)$ of the cost function, when M has been measured, can be found by integration. The optimal Bayes estimate of D is the R which minimizes $C(R|M)$.

$$\inf_R C(R|M) = \inf_R \int c(R, D)p(D|M)dD \quad (3)$$

If all wrong decisions are equally bad, and only the exactly correctly estimated parameter R is considered right, a cost function can be constructed as: $c(R, D) = 1 - \delta(|R - D|)$ where δ is the Dirac delta function. The R which minimizes Equation 3 is the one which maximizes the *a posteriori* probability: $\sup_R p(R|M)$. In other words: The Maximum A Posteriori (MAP) estimate is the one which maximizes the probability of correct decision, but punishes a near miss as much as every other miss. The MAP estimate is often used in regularization.

Regularization can be formulated as Bayesian estimation. Let us assume that the measurements are created by addition of uncorrelated noise and an *a priori* distribution $p(D_x)$ of the first derivative of the data is given. From Bayes theorem we get

$$p(D|M) = \frac{p(M|D)p(D)}{p(M)} = \frac{p(M|D)p(D_x)}{p(M)}$$

as D_x just is a basis change of D in the discrete case [8]. For a set of measurements we find

$$p(D|M) = \frac{1}{Z} \prod_i p(M^i|D^i)p(D_x^i) = \frac{1}{Z} e^{\sum_i \log p(M^i|D^i) + \log p(D_x^i)}$$

if the measurement is only dependent on the data locally and superscript denotes the discretized position $x = i$ and Z is a normalizing constant. The Bayesian estimate can be found using the MAP cost function if the data dependency and the *a priori* distribution of the derivative is given. If the noise is uncorrelated Gaussian noise, having the same standard deviation σ in each point, the measurements will depend on the real data as: $\log p(M^i|D^i) = (M_i - D_i)^2/2\sigma^2 - \log Z^i$, where $\log Z^i$ can be omitted. This data dependency will be used in the following.

3 Orthographic projection

Orthographic projection is used as model, when the depth measurements are found as a measure perpendicular to some axis or plane. As argued above, a general assumption in depth extraction modules is isotropy of the scene. This means that no direction (no angle to the image plane) is preferable to others. Because a frontoparallel area has a larger projection onto the image plane than a tilted area, it is more likely and is not punished as much. This is formalized in the following.

The angle to the image plane can be measured by the first derivative of the depth function. Our goal is to establish an *a priori* distribution of the first derivative of the depth function D_x (which is the easily measurable quantity) from distribution of the angle θ (which is the quantity that initially has been argued

equally distributed). The orthographic projection of a line parameterized by the arc length s is

$$\begin{pmatrix} x(s) \\ D(s) \end{pmatrix} = \begin{pmatrix} a + s \cos \theta \\ b + s \sin \theta \end{pmatrix}$$

The equal distribution of θ in the scene results in an *a priori* distribution of θ observed in the projection, which is proportional to the length of a projection of a unit length.

$$f(X = \theta) = \frac{1}{n} \left. \frac{dx}{ds} \right|_{s=0} = \frac{1}{2} \cos \theta$$

where $n = 2$ is a normalizing constant. The derivative of the depth D according to x is constructed as

$$D_x = \left. \frac{ds}{dx} \frac{dD}{ds} \right|_{s=0} = \left. \frac{dD}{ds} / \frac{dx}{ds} \right|_{s=0} = \tan \theta$$

In general, if a new stochastic variable Y is created from X as $Y = \phi(X)$, where X is distributed as $f(X)$ and $\psi = \phi^{-1}$, then the distribution g of Y will be

$$g(Y) = f(\psi(Y)) \left| \frac{d\psi(Y)}{dY} \right| \quad (4)$$

In this case of orthographic projection we have:

$$\psi(D_x) = \arctan(\theta) \Rightarrow g(Y = D_x) = \frac{1}{2} \frac{1}{\sqrt{1 + D^2}} \frac{1}{1 + D^2} = \frac{1}{2} e^{-\frac{3}{2} \log(1 + D_x^2)}$$

This distribution has to be multiplied by the data constraint and the reconstruction R can be found as the MAP estimate. Because the exponential function is monotonously increasing, it has its maximum, where the following energy term $E(R)$ has its minimum:

$$E(R) = \sum_i ((M^i - R^i)^2 + 3\sigma^2 \log(1 + (R_x^i)^2)) \quad (5)$$

This means, that the smoothness term is $\log(1 + R_x^2)$. This is a special case of the well known Lorentzian estimator. It is a robust estimator yielding discontinuous regularization [8]. Furthermore, it yields a convex solution space for $\sigma^2 < \sqrt{2/3}h$, where h is the sampling distance [8]. It has earlier been used as smoothness constraint in flow analysis by Black and Anandan [12].

The isotropic regularization has been tested in the 2D case on simulated data. In Figure 3 it is tested on a sphere on a platform. It shows that the step edges are preserved, and that the reconstruction is not perfect, but has some small hills. On the sphere we see, that gradients, which are very high cause spurious discontinuities to be detected, as the weak string would do. The isotropic regularization yields results, which in general are as good as those of the weak string, when applied to step edges. In the example, the solution space is not convex as $\sigma = 25h$. The solutions are found by the Smoothness Focusing GNC algorithm [7].

4 Perspective Projection

Under perspective projection the isotropy constraint implies other smoothness constraints than under the orthographic projection. A given angle has a different probability of being seen, dependent on the position relative to the optical center. The perspective projection of a line parameterized by the direction θ and the arc length s is

$$\begin{pmatrix} x \\ D \end{pmatrix} = \begin{pmatrix} \frac{X}{D} \\ D \end{pmatrix} = \begin{pmatrix} \frac{a+s \cos \theta}{b+s \sin \theta} \\ b + \sin \theta \end{pmatrix}$$

where capital letters are world coordinates, D is the depth, and the focal length is normalized to 1. Calculations [8] similar to those of the orthographic projection, using the derivatives in $s = 0$, yields

$$g(Y = D_x) = \frac{D^2}{2(D_x^2 + (xD_x + D)^2)^{\frac{3}{2}} \sqrt{x^2 + 1}}$$

It should be noticed, that this is the distribution of the first derivative of the depth, when x and D are kept constant. To maximize the *a posteriori* of D_x and D in a given position of the retina x , we should maximize

$$p(D_x, D|x) = p(D_x|D, x)p(D|x)$$

This requires an *a priori* probability of the depths in a retina point. When this is unknown, we can overcome the problem by estimating a local mean (or expectation) value of the depth, and use this in the distribution. This implies, that the maximization of the *a posteriori* probability corresponds to minimization of the smoothness energy

$$E_S(R) = \frac{3}{2} \log(R_x^2 + (xR_x + \bar{R})^2) \approx \frac{3}{2} \log\left(1 + \left(\frac{R_x}{\bar{R}}\right)^2\right) + k$$

where terms which are constant in R_x has been removed, and R is denoted as mean value as it is expected to be locally constant. The approximation can be used when x is small (weak perspective projection). This correspond to the case of orthographic projection, where the derivative is scaled by the distance. This implies, that the perspective projection near the center of projection yields qualities similar to the qualities of the orthographic projection.

5 Stereo Projection

The stereo projection is more complicated than the above mentioned projections, but nevertheless it is possible to carry out the calculation. The stereo projection is more interesting from a computer vision point of view, as the monocular projections can only perform depth sensing if a kind of range finding equipment is used. Under stereo projection, we can find the depths measures directly from matches in the two images and knowledge of the camera geometry. Before the smoothness term is derived, we will have to look into the stereo projection, as it is used in this paper.

When the internal and external geometry of the cameras is known, the problem of matching features in the two images, is a one dimensional matching problem.

This is due to the epipolar constraint, which constrains all the scene-points placed in a plane coinciding with the two optical centers, to be located on the lines in the image planes, that coincide with the plane [19]. Because of this epipolar constraint, we will only look into the 2D-to-1D stereo projection.

A stereo projection can be constructed by drawing the curve(s) of matched points in 2D, and then transform it into a coordinate system in which left and right camera position are the two perpendicular axes (see Figure 1). In this coordinate system, the matching curve will be a single valued function m of the position on the diagonal axis u , if no change of order in matches occur (see Geiger and Giroi [16]). In this stereo configuration occlusion will be modelled as intervals of u , where

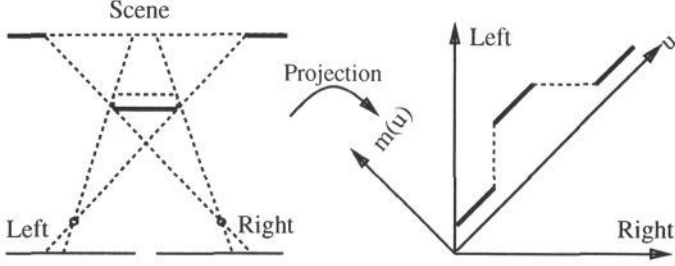


Figure 1: Stereo projection. The matched curve in the 2D world is transformed into a coordinate system of perpendicular axes. In this coordinate system will the matching curve be a single valued function m on the diagonal axis u , if the matches are not changing order from left to right camera.

$|\frac{\partial m}{\partial u}| = 1$. Where matches occur, $|\frac{\partial m}{\partial u}| < 1$. Like in the previous examples, we will have to parameterize the possible directions. For simplicity, we will assume, that the two cameras are identical and the two images planes coincide, thus the two optical axes are parallel. Using these simplification we can parameterize the stereo geometry and projected angles by the following parameters (see Figure 2): a is half the base line, X is the distance to the optical axis of the cyclopic coordinate system, and Z is the distance from the projection onto the optical axis to the cyclopic center. This projection can be written as a linear transformation of the positions x_L and x_R on the two retinas.

$$\begin{pmatrix} u \\ m(u) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_L \\ x_R \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \frac{X-a}{Z} \\ \frac{X+a}{Z} \end{pmatrix} \quad (6)$$

The surface, which is seen in the two cameras is a general curve but can be approximated to be locally linear and parameterized by a parameter s :

$$\begin{pmatrix} X(s) \\ Z(s) \end{pmatrix} = \begin{pmatrix} X_0 + s \cos(\theta) \\ Z_0 - s \sin(\theta) \end{pmatrix}$$

This leads after calculations similar to those of the previous examples to the following smoothness term:

$$E_S(m) = \frac{3}{2} \log(m_u^2 Z_0^2 + (a - X_0 m_u)^2) \quad (7)$$

This term can be used as smoothness term in these parts of the stereo images, where no occlusions are present. The first derivative of the matching function m_u

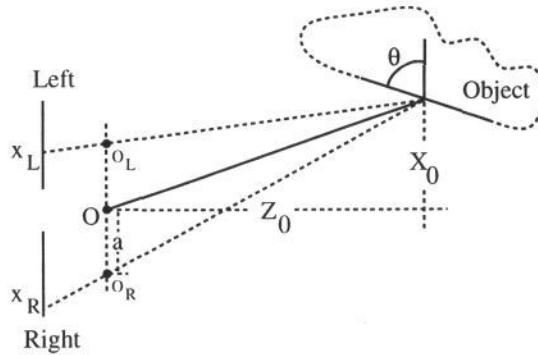


Figure 2: Stereo projection parameterization. Every point is parameterized by (Z, X, θ) , where θ is the angle to the image planes.

is limited to values in $[-1; 1]$, as larger absolute derivatives will lead to occlusions. In the smoothness term, the matching function m , and three more parameters are used. The half base line a is a constant, and yields no problems, except, that it should be measured in units of focal lengths, as the stereo projection geometry is scaled into focal length $f = 1$. The two other parameters, the depth Z_0 and the left-right position X_0 can be expressed directly from the matching parameter u and the matching function $m(u)$. By inversion of Equation 6 and substitution into Equation 7 we find:

$$E_S(m) = \frac{3}{2} \log\left(\left(\frac{2m_u}{m}\right)^2 + \left(1 - \frac{m_u u}{m}\right)^2\right) + 3 \log(a) \quad (8)$$

The last term is a constant and can be omitted without changing the solution. As in the case of perspective projection, this equation is only valid for locally constant parameters m and u .

The smoothness term is independent of the base line. It does, however, depend on the focal length, as u and m is measured in units of the focal length f , and the position of the principal point, as u is measured as the distance (with sign) from the principal point. The first term $\frac{2m_u}{m}$ does, in this way, depend on the focal length, but not on the position of the principal point. The second term $\frac{m_u u}{m}$ does not depend on focal length as $\frac{m}{u}$ and thereby $\frac{dm}{du}$ are independent of f . The second term does, however, depend on the principal point.

If the viewing field (the ratio of the size of the retina to the focal length) is small, the second term $\frac{m_u u}{m}$ is in practice much smaller than 1 and can be approximated to zero:

$$E_S(m) \approx \frac{3}{2} \log\left(1 + \left(\frac{2m_u}{m}\right)^2\right)$$

This smoothness term takes the same form as the term for the orthographic projection. This implies, that the properties from the orthographic projection can be transferred to the stereo projection, when the above approximation is used. The stereo projection will have a convex solution space if not too much noise is present, and the smoothness term will imply discontinuous regularization. The data term, under stereo projection, will rarely imply a convex solution space, but

often a solution space containing many local minima. This means that a gradient descend algorithm will not be guaranteed to find the global minimum, no matter that the smoothness term implies a convex solution space.

In the above theory of the stereo projection, no assumptions of the visibility of the object has been made. Some points in one image is not visible in the other image. Whether a point is visible in one, both or non images depends on the viewing angle θ and the position. In the total *a priori* probability distribution we will have to encounter the fact of occlusion. All the angles θ which are left-occluded will appear in the matching function as $m_u = 1$, and the right occluded will appear as $m_u = -1$. We find

$$\int_{-\infty}^{-1} g(Y = m_u) dm_u \approx \int_1^{\infty} g(Y = m_u) dm_u \approx \frac{Z_0 - \sqrt{a^2 + Z_0^2}}{2\sqrt{a^2 + Z_0^2}} \quad (9)$$

when $X_0 \approx 0$. The cases where X_0 takes other values are outlined in [8].

The constant from Equation 9 times a Dirac delta function has to be added in $Y = \pm 1$ (where the occlusions occur) and the probability of $|Y| > 1$ is set to zero. It is obvious, that these Dirac delta functions yields a problem when making a MAP estimation. In the points $m_u = \pm 1$, the energy will be minus infinity, which will totally dominate the solution space. Another cost function, without singularities has to be used. In Nielsen [8], it is shown, that a pseudo-MAP estimate, which takes not only the density, but also the local probability volume into account, can be constructed by a convolution of the *a priori* distribution by a Gaussian. In order to make a regularization, we will have to convolute the *a priori* distribution by a Gaussian of appropriate standard deviation, and then minimize the energy. This method might be applied to the occluded stereo problem. Furthermore, is it shown [7], that the convolution by a Gaussian, yields a Graduated Non-Convexity algorithm.

6 Anisotropic regularization

In many cases, the world is not isotropic. Some surface normals are more probable than others. If aerial photos of Denmark is the basis of a stereo analysis, the wideness might be several kilometers, while the depths are in a range of tens of meters. Under these circumstances the isotropic constraint is obviously not fulfilled. If an *a priori* distribution of the surface normals in the scene is well known, it can be used to construct smoothness terms corresponding to the MAP estimation. This scheme is identical to the scheme from the previous sections, we do only have to substitute the new $f(\theta)$ into Equation 4. If high density values, which only represents a small probability volume, are present, the distribution can be convolved by a Gaussian to obtain the estimate. Furthermore, the *a priori* distribution, do not have to be known analytically. It might as well be measured. If the density of surface normals do not change in time, the result from a previous reconstruction can be used as *a priori* in the following frame. In this way an adaptive regularization scheme is constructed.

7 Conclusion

A theory of using quantitative *a priori* knowledge is outlined and applied to the isotropy-constraint in the case of orthographic, perspective and stereo-projection.

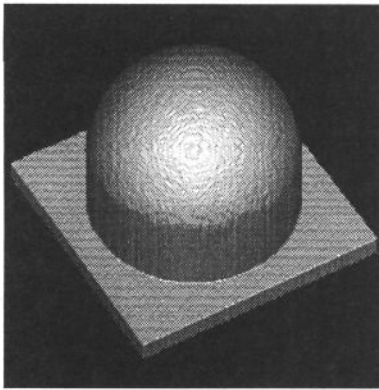
The isotropy constraint implies the well known Lorentzian robust estimator. It needs only one parameter: The amount of noise, which, if not known in advance, can be estimated by eg. the technique proposed by Olsen [18]. In the case of stereo projection occluded areas are punished by an increasing function of the distance, as they are less likely to occur on distant objects than on nearby objects. The *a priori* density function of gradients in the matching function contains singularities in the occluded case, why MAP estimation is not applicable. An estimation method and the application to the stereo projection under the isotropy constraint is to be developed.

The isotropy constraint implies discontinuous regularization, and has a convex solution space for small amounts of noise. For larger amounts of noise, the solution space is non-convex, so a more expensive optimization technique has to be applied. In this paper the GNC by Smoothness Focusing [7] has been used. In future work, also second order regularization might be included in the theory. The *a priori* assumption might be: Every sphere is equally probable, but large spheres have a larger projection and therefore low curvature is more probable than high curvature.

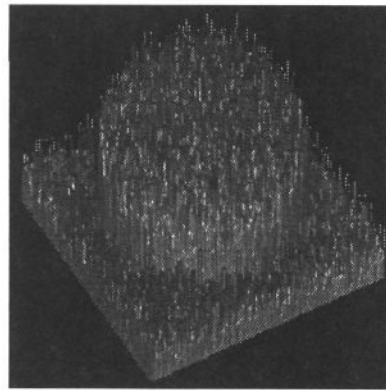
References

- [1] J. Aloimonos and D. Shulman, "*Integrating of visual modules*" Academic Press, 1989.
- [2] A. N. Tikhonov and V. Y. Arseninn "*Solution of ill-posed problems*" V.H Winston & Sons, John Wiley, 1977.
- [3] W. E. L. Grimson, "*From images to surfaces*", MIT Press, 1981.
- [4] Stuart Geman and Donald Geman, "*Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images*", IEEE PAMI Vol. 6, No. 6, pp. 721-741, April 1984.
- [5] A. Blake and A. Zisserman, "*Visual Reconstruction*" MIT Press, 1987.
- [6] M. Nielsen, "*Scale Space Extensions of the Weak String*", Proc. 8th SCIA, Tromsø, Norway, may 25-28, 1993.
- [7] M. Nielsen, "*Graduated Non-Convexity by Smoothness Focusing*", Proc. 4th BMVC, Guildford, England, 21-23/9, 1993.
- [8] M. Nielsen, "*Isotropic regularization*", Technical report, DIKU.
- [9] A. Blake, "*Comparison of the Efficiency of Deterministic and Stochastic Algorithms for Visual Reconstruction*" Proc. of PAMI Conference, 1989.
- [10] P. Meer, D. Mintz, A. Rosenfeld and D. Y. Kim, "*Robust Regression Methods for Computer Vision: A Review*", Int. Jour. of CV 6:1, pp. 59-70, 1991.
- [11] J. B. Jensen and M. Nielsen, "*A Simple Genetic Algorithm Applied to Discontinuous Regularization*", Proc. 2nd IEEE workshop on NNSP, Copenhagen, 29/8-2/9, 1992.
- [12] M. J. Black and P. Anandan, "*A Framework for the Robust Estimation of Optical Flow*", Proc. of 4th ICCV, Berlin, Germany, May, 1993.
- [13] J. Gårding, "*Shape from texture and contour by weak isotropy*", Jour. of Art. Int. 1993 (in press).
- [14] R. L. Stevenson and E. J. Delp, "*Fitting Curves with Discontinuities*", Proc. Int. Workshop on Robust CV, Seattle, WA, USA, 1-3 Oct, 1990.
- [15] M. Straforini, C. Coelho, M. Campani and V. Torre. "*Using Geometrical Rules and a Prior Knowledge for Understanding of Indoor Scenes*" Proc. Int. Workshop on Robust CV, Seattle, WA, USA, 1-3 Oct, 1990.

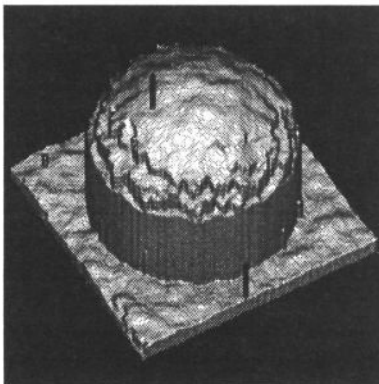
- [16] D. Geiger, B. Ladendorf and A. Yuille, "*Occlusions and Binocular Stereo*", Proc. 2nd ECCV, Santa Margherita Ligure, Italy, May 19-22, 1992.
- [17] N. M. Nasrabadi, S. P. Clifford and Y. Liu "Integration of stereo and optical flow by using an energy-minimizing approach", Jour. of Optical Society, Vol 6, No. 6, June 1989.
- [18] S. I. Olsen, "*Noise Variance Estimation in Images*", Proc. 8th SCIA, Tromsø, Norway, may 25-28, 1993.
- [19] S. I. Olsen, "*Epipolar Line Estimation*", Proc. 2nd ECCV 92, Santa Margherita Ligure, May 19-22, 1992.



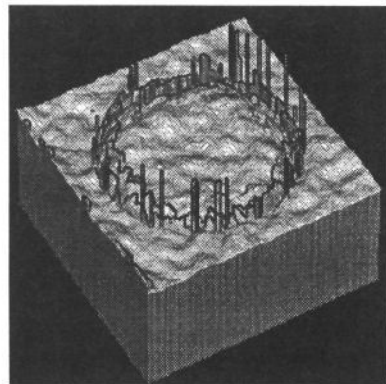
(a)



(b)



(c)



(d)

Figure 3: Orthographic projected depth data. (a) is the original data. (b) noise corrupted signal, width $SNR = \sqrt{2}$ on the step edges. (c) is the reconstruction, and (d) is the normalized residual.