Adaptive Regularization: Towards Self-Calibrated Reconstruction

Mads Nielsen

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Adaptive Regularization: Towards Self-Calibrated Reconstruction

Mads Nielsen

N° 2351

Septembre 1994

PROGRAMME 4

Robotique,

image

et vision

Rapport de recherche

1994

ISSN 0249-6399

Adaptive Regularization: Towards Self-Calibrated Reconstruction

Mads Nielsen

Programme 4 — Robotique, image et vision
Projet ROBOTVIS

Rapport de recherche n° 2351 — Septembre 1994 — 28 pages

Abstract: Regularization is often applied to the ill-posed problem of surface reconstruction. This implies the incorporation of *a priori* knowledge in the solution. The reconstruction depends strongly on this *a priori* information. Typically, qualitative *a priori* information is used, leaving it to the user to estimate parameters, (eg. the weak string [4]). Other methods depend strongly on assumptions of the viewing geometry and/or statistics of the scene [3], [21]. Neither the viewing geometry nor the scene statistics are in general known for an active observer. In dynamic vision, however, the *a priori* knowledge can be extracted from the reconstruction of the previous scenes. This leads to an adaptive regularization scheme capable of capturing the resulting scene statistics in the camera coordinate system. Consequently, calibration is not needed. The *a priori* knowledge required is the amount of noise in the data, and that the statistics of the scene is only varying slowly. It is shown that the adaptive regularization yields results which are comparable to those of the weak string if the input is piecewise planar. Inputs having only a few different surface orientations are reconstructed robustly. The adaptive system is shown to have 1 stable fixpoint.

Key-words: Learning in computer vision, Integration of modules and cues, Surface reconstruction, Bayesian estimation.

(Résumé : tsvp)

This work was supported by the Danish Council of Technical Sciences, and was partly carried out at Depertment of Computer Science, University of Copenhagen, Denmark

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Régularisation adaptative: Vers une Reconstruction Autocalibrée

Résumé : Un des domaines principaux d'application de la régularisation est la reconstruction de surfaces. Cela implique l'incorporation de l'information *a priori* dans la solution. La reconstruction est extrêmement dépendante de cette dernière. Habituellement, une information qualitative est utilisée *a priori*, l'estimation des paramètres étant laissée à l'utilisateur (comme par exemple dans le cas de la membrane faible [4]). D'autres méthodes dépendent fortement des suppositions faites sur la géométrie et/ou sur les distributions statistiques caractérisant la scène [3], [21]. En général, ni les paramètres géométriques de l'observation ni les distributions ne sont connus. Dans le cadre de la vision dynamique, l'information *a priori* peut être extraite des reconstructions précédentes de la scène. Cela conduit à un système de régularisation adaptatif, qui est capable d'inférer les distributions statistiques de la scène mesurées par rapport au repère image, rendant ainsi toute calibration superflue. Les seules connaissances *a priori* requises sont la valeur du rapport signal sur bruit et le fait que les distributions statistiques de la scène varient lentement dans le repère image. Nous montrons dans ce rapport que la régularisation adaptative donne des résultats comparables avec les résultats de le membrane faible si les données sont constantes par morceaux. Des expériences conduites avec seulement quelques orientations de surface montrent une reconstruction robuste. Nous démontrons aussi que le système a une point fixe stable.

Mots-clé : Vision dynamique, Régularisation adaptative, Reconstruction de surfaces, Estimation bayesienne

1 Introduction

The problems of depth-extraction or surface reconstruction are ill-posed in the sense of Hadamard [1]. This is mainly due to noise corruption, but also often due to lack of information (Eg. in a 2D image about the 3D structure of the scene). 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 measurements are created by addition of noise N to the data.

$$M(x) = D(x) + N(x) \tag{1}$$

To overcome the ill-posedness, regularization is applied. This regularization might be formulated in many ways:

- Reformulation of the ill-posed problem as well-posed by adding a stabilizing term [24].
- Markov Random Fields (MRF) [9]
- Bayesian Maximum A Posteriori (MAP) estimation [8]
- Minimum Description Length (MDL)[13]

These techniques might be used separately or together. No matter which formulation is used, the equations turn out to be identical, and the choice of technique is a question of convenience [13], [8]. The result is that a stabilizing term is added to the original problem.

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 [24] used quadratic sums of the derivatives of the solution to gain well-posedness. This implies the minimization of an energy term, which in the case of use of the first derivative in the smoothness yields:

$$E(R) = E_{\mathrm{Data}}(M, R) + \sum \lambda R_x^2$$

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. The solution can be found by solving the Euler-Lagrange equation if the data term is simple.

Geman and Geman [9] introduced a line process which is used as an alternative to the smoothness term in a Markov Random Field. 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: Minimize

$$E(R) = E_{\mathrm{Data}}(M,R) + \sum \eta\lambda R_x^2 + (1-\eta)P \tag{2}$$

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 found one of the solutions of low energy by simulated annealing. Blake and Zisserman [4] reformulated the same strategy as a thresholding of the smoothness term. This implies the minimization of

$$E(R) = E_{\mathrm{Data}}(M,R) + \sum \lambda f(R_x)$$

where

$$f(t) = \begin{cases} t^2 & \text{if } t^2 < P \\ P & \text{otherwise} \end{cases}$$

Blake and Zisserman found the solution by the graduated non-convexity algorithm (GNC), which does not guarantee an optimal solution [18], but is deterministic and faster than the simulated annealing [5]. The GNC was generalized to other regularization schemes by Nielsen [19].

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 have 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. [16]. Each of these methods depends on a number of parameters which have to be estimated in advance. Actually, implying a model (such as the weak string) is a limitation from an infinite number of parameters (a function on the real axis) to a few parameters. Techniques such as cross-validation has been used to determine these few parameters, when the "shape" of the smoothness term has been chosen [23].

A wide range of adaptive estimation methods have been developed in the field of system identification [14]. These methods are able to find parameters of a system, but are dependent on pairs of input and output of the system which is to be identified. In general, we are not given information of the input, why standard system identification techniques cannot be used. In terms of learning, we will try to build a system, which given the measurement are able to tell us the input. In general we cannot feedback to a learning system the answer whether the prediction was correct or not. This means that we are in the case of unsupervised learning. In Neural Networks a class of unsupervised learning methods exists [12]. They do, however, depend on *a priori* information (knowledge of the smoothness term). One example is the Boltzmann Machine, which leads to the formulation on Markov Random Fields, which is identical to our approach, when the smoothness term is known.

The reconstruction depends strongly on the smoothness term. Actually, in mathematical sense, **the smoothness term and the data term are equally important, why the smoothness term should be picked very carefully**. 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". In general, the smoothness term should reflect the knowledge of the scene (eg. in terms of statistics of surface normals, curvatures etc.), and of the viewing geometry (how the statistics are transformed into the reconstruction domain). The scene information could be that some surface orientations are more probable than others, as in *Legoland.* It might even be that no surface orientations are more probable than others, which leads to discontinuous regularization [21], when the viewing geometry is utilized.

The techniques of MAP-estimation and MDL can be used to find the smoothness term from a statistic model of the world [3], [21]. MAP and MDL techniques results in the same equations letting the choice of technique to be of convenience and taste. We have chosen to use the formulation from MAP estimation, which is the topic of Section 2. In this paper, the observer in mind is a consecutively grabbing stereo reconstructor, but the technique can be applied in any system working over time. To calculate the smoothness term for a stereo reconstruction system (using MAP or MDL techniques), one needs to know the statistics of the scene and the viewing geometry. Belhumeur [3] is using a polygonal scene model to infer the *a priori* information, but is not using the viewing geometry. Nielsen [21] is using the viewing geometry, but also a simple model of isotropy of surface normals in the scene. Subjected to other conditions, these system would fail. **To make a regularization scheme, which will work robustly under varying types of scenes and for an observer of unknown or partly unknown geometry, the regularization scheme must be able to adapt to the current statistics of the scene.**

If the surface reconstruction is performed consecutively in many frames, the *a priori* knowledge to the ith frame might be extracted from the previous frames instead of being a general model known in advance. In this way, the regularization scheme might adapt to the characteristics of the scene. Because the statistics are measured directly in the camera coordinate system (the coordinate system of the reconstruction), the calibration of the camera system is not needed.

Before looking into the adaptive regularization (Section 3) and the experimental results (Section 4), we need to pay attention to Bayesian estimation theory, to see how the scene statistics can be used to find the smoothness term of a regularization scheme.

2 Bayesian estimation

Bayesian estimation is a technique using the Bayesian rules from probabilistic theory to compute the expectation value of a cost function. When the expectation value of the cost function is minimized an optimal estimate is found.

In advance one has to know: A measurement M, the probabilistic dependency $p(M|D)$ of the measurement on the true 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 true value. Bayes theorem yields

$$p(D|M) = \frac{p(M|D)p(D)}{p(M)} \tag{3}$$

where $p(M)$ is a constant, when the measurement M is known. Thus we can perceive $p(M)$ as a normalizing constant found by integration. Furthermore, we can denote the prior (the *a priori* distribution) by $p(D_x)$ if it is only dependent on the derivative of the solution. The mean value $C(R|M)$ of the cost function when M has been measured, can be found by integration.

$$C(R|M) = \int c(R, D)p(D|M)dD \tag{4}$$

The optimal Bayes estimate of D is the R which minimizes $C(R|M)$. How to find the optimal estimate depends on the structure of the cost function and of the density function. Note, that we in Equation 4 are performing an integration over a function space. The quantity dD is in general not well defined, but a discretization into N variables x_i can make the integral well-defined:

$$dD(x_1, x_2, \ldots, x_N) = f(x_1, x_2, \ldots, x_N)dx_1dx_2 \cdots dx_N$$

Normally, and in this report, the function f is chosen to unity. In general other functions might be chosen. When concerning depth measures, a logarithmic function might be expedient, because it corresponds to a constant relative uncertanty in depth [20].

2.1 Deviation as cost and Maximum A Posteriori estimate

When the cost of choosing the reconstruction R when D is the correct solution can be expressed as a function c_{dev} of the difference of R and D Equation 4 yields:

$$C(R|M) = \int c_{\text{dev}}(R - D)P(D|M)dD = c_{\text{dev}}(R) * P(R|M) \tag{5}$$

In the special case where all wrong decisions are equally bad, and only the exactly correctly estimated parameter R is considered right, a cost function can be constructed as:

$$c_{\text{dev}}(R - D) = -\delta(R - D)$$

where δ is the Dirac delta function. It yields a constant punishment for all wrong estimations and a significantly lower punishment for a correct estimate. It is easily seen that

$$C(R|M) = -\delta(R-D) * P(R|M) = -P(R|M)$$

and that the R minimizing Equation 4 is the one maximizing the *a posteriori* probability:

$$\arg\sup_R P(R|M)$$

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. For a discussion of the implications of different cost functions see Nielsen [21].

2.2 Bayesian regularization

Regularization can be formulated as Bayesian estimation. If the noise model and the *a priori* distribution of the derivative are independent of the image position and are mutually independent in all image positions we find (using 3).

$$\begin{aligned} p(D|M) &= \frac{1}{Z}\Pi_i p(M^i|D^i)p(D_x^i) \\ &= \frac{1}{Z}e^{\sum_i \log p(M^i|D^i)+\log p(D_x^i)} \end{aligned} \tag{6}$$

where superscript denotes the discrete position $x=i$ and Z is a normalizing constant. Equation 6 is only valid if the measurement are only dependent on the data locally. The assumption of mutually independence and independence of the image position i of the prior

$$p(D_x) = \Pi_i p(D_x^i)$$

will in the following be called the assumption of stationarity of the prior. We now find the Bayesian estimate using the MAP cost function (if the data dependency and the *a priori* distribution of the derivative is given) as the one minimizing the exponent of Equation 6.

When the noise is uncorrelated Gaussian noise and has the same standard deviation σ in each point (ie. stationarity of the data term), the measurement dependendcy on the real data will be:

$$p(M|D) = \frac{1}{Z} e^{-\sum_i \frac{(M_i - D_i)^2}{2\sigma^2}} \tag{7}$$

This stationary data dependency will be used in the following.

3 Adaptive regularization

Regularization can be performed as MAP estimation using the distribution of Equation 6 where the data dependency is quadratic in the exponent due to Gaussian distributed noise as in Equation 7. We need only to establish the *a priori* distribution of the derivative and the amount of noise in the measurements. In estimation theory, techniques for handling situations of unknown *a priori* distribution has been developed:

3.1 Unknown prior

In the following we will assume, that the *a priori* distribution is stationary (we have the same distribution of the derivative at any position on the surface). This is not always an obvious assumption. In many applications, we might expect to find the ground plane in the lower part of the image, while the upper part might be expected more or less vertical. For the purpose of studying adaptive regularization, we will make the stationarity-assumption. In future studies, the non-stationarity might be taken into account.

In estimation techniques a so called MINIMAX estimation is proposed in the case of unknown *a priori* distribution [10]. The MINIMAX estimation correspond to finding the *least favorable a priori* distribution. This implies that we find the solution as a saddle point in the solution space spanned by all possible solutions and all possible *a priori* distributions. This is defined as: If R^* is the MINIMAX estimate and $p^*(R)$ is the least favorable *a priori* distribution we have

$$\forall p(R)\ \forall R: \ C(R^*|p(R)) \leq C(R^*|p^*(R)) \leq C(R|p^*(R))$$

where C is the mean penalty from Equation 4. Unfortunately, the MINIMAX estimation does not make any sense in the case of a MAP penalty function. When all misses are equally punished, we can easily find a p yielding a higher expected penalty if we just zero-out the chosen R in the distribution. Afterwards an estimation yielding a lower expected penalty can be found by changing the estimation into a solution not having zero probability. This means that the saddle point does not exist (in the interior of the solution space) in case of MAP estimation

If we instead find the *a priori* distribution as the *most favorable*, the optimum exists. By solving the Euler-Lagrange equation using Lagrange multipliers to condition the probability distribution to have integral 1 and be positive, we find that the most favorable distribution is the current distribution of R. This means that the intuitively understandable assumption that the *a priori* distribution should be identical to the *a posteriori* distribution corresponds to finding the most favorable distribution.

The criterion of the process to be stationary (identical in all image positions) is very important at this point. If we let the distribution vary from point to point, the distribution is able to adapt totally to the data, and the smoothness term would in practice disappear. The reconstruction will be the data, and the resulting prior will be the one having the data with probability 1. The method of most favorable prior needs an assumption of smoothness in the statistics in order to be non-trivial. The situation which is considered in this paper is constant prior (stationarity assumption). We could also make the abstraction of demanding a smoothly varying prior. This will assemble the scheme of Marroquin [15] where the reconstruction itself is a smoothly varying probability distribution.

3.2 Sequential reconstruction

When an observer is observing over time another degree of freedom is introduced. A more stable reconstruction can be made if the solution is integrated over time [17]. In earlier work, the assumption, that the scene is not changing much over time has been explored (eg. by Bergen et. Al. [2]). In this work, we will not use the assumption of the scene changing slowly. The reconstruction will be restarted from scratch in every frame. However, the scene statistic (the