INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Multiresolution Approach for Shape from Shading Coupling Deterministic and Stochastic Optimization

Alain Crouzil — Xavier Descombes — Jean-Denis Durou

N° 5006

Novembre 2003

THÈME 3

Rapport de recherche

ISSN 0249-6399

A Multiresolution Approach for Shape from Shading Coupling Deterministic and Stochastic Optimization

Alain Crouzil * , Xavier Descombes , Jean-Denis Durou †

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Ariana

Rapport de recherche n° 5006 — Novembre 2003 — 34 pages

Abstract: Shape from shading is an ill-posed inverse problem for which there is no completely satisfactory solution in the existing literature. In this technical report, we address shape from shading as an energy minimization problem. We first show that the deterministic approach provides efficient algorithms in terms of CPU time, but reaches its limits since the energy associated to shape from shading can contain multiple deep local minima. We derive an alternative stochastic approach using simulated annealing. The obtained results strongly outperform the results of the deterministic approach. The shortcoming is an extreme slowness of the optimization. Therefore, we propose an hybrid approach which combines the deterministic and stochastic approaches in a multiresolution framework.

Key-words: shape from shading, optimization, simulated annealing, multiresolution.

Ce rapport est également publié comme rapport de recherche de l'IRIT sous le numéro IRIT/2003–19–R

* IRIT, Université Paul Sabatier, 118 route de Narbonne, 31062 TOULOUSE Cedex 4
† IRIT, Université Paul Sabatier, 118 route de Narbonne, 31062 TOULOUSE Cedex 4

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Une approche multirésolution pour le *Shape from Shading* couplant des optimisations stochastique et déterministe

Résumé : Le *shape from shading* est un problème inverse mal posé pour lequel aucune méthode de résolution complètement satisfaisante n'a encore été proposée. Dans ce rapport technique, nous ramenons le *shape from shading* à un problème d'optimisation. Nous montrons d'abord que l'approche déterministe fournit des algorithmes efficaces en termes de temps de calcul, mais est d'un intérêt limité lorsque l'énergie comporte des minima locaux très profonds. Nous proposons comme alternative une approche stochastique utilisant le recuit simulé. Les résultats obtenus dépassent largement ceux de l'approche déterministe. La contrepartie est l'extrême lenteur du processus d'optimisation. Pour cette raison, nous proposons une approche hybride qui combine les approches déterministe et stochastique dans un cadre de multi-résolution.

Mots-clés : *shape from shading*, optimisation, recuit simulé, multi-résolution.

1 Introduction

Shape from shading is a well-known problem in computer vision [1, 2], that consists in recovering the 3D-shape of a scene through the analysis of the greylevels in a single image of this scene. Under certain circumstances, the greylevel directly gives the slope of the shape and the difficulty is then to find the direction of steepest descent. This problem is known to be ill-posed, that is, it cannot be solved without some assumptions. For that reason, shape from shading suffers of a bad reputation in the community, in comparison with other 3D-shape reconstruction techniques, such as shape from stereopsis, photometric stereo (see [3] for an up-to-date presentation of reflectance based shape recovery), shape from movement, etc., which require several images of the scene and are, for this reason, more often well-posed[1]. In the case of shape from shading, well-posedness is dependent on a certain number of assumptions, which can seem to be arbitrary.

1.1 Usual assumptions

The usual assumptions [2] which make the problem well-posed concern:

- The scene: it is supposed to have the same "reflectance map" at each point and the uniform reflectance map has to be known. Most of the time, the surface is supposed to be Lambertian with constant albedo.
- The light sources: it is supposed that there is a unique point light source, far from the scene, so that the incident beam can be approximated by a uniform parallel beam $\overrightarrow{S}$. Most of the time, the interreflections are neglected.
- The camera: it is supposed to produce an orthogonal projection of the scene on the image plane (up to a scale factor), that is, the perspective projection is neglected, as well as various distorsions due to the camera. Moreover, the receiver is supposed to be linear, so that the greylevel is proportional to the optical energy received.

1.2 The basic equations

Under these assumptions, and taking the optical axis as Oz (z increases when one goes closer to the camera) completed with Ox and Oy in the image plane, so that $Oxyz$ forms an orthogonal basis, the visible part of the scene can be described by the equation:

$$z = h(x, y) \tag{1}$$

where the unknown is the height function h, which can *a priori* be non derivable (if edges) and even non continuous (if occlusions). At each point (x, y) where h is derivable, the basic equation of shape from shading is the "image irradiance equation" [1]:

$$R(\overrightarrow{S}, p(x, y), q(x, y)) = E(x, y) \tag{2}$$

[1] In this technical report, the definition of well-posedness is that of Hadamard: a problem is well-posed if it admits a unique solution and if the solution is stable.

where E designates the greylevel, R the reflectance map, and p and q are the usual notations for $\partial h/\partial x$ and $\partial h/\partial y$. In this equation, only $\overrightarrow{S}$ and h, by way of its first derivatives, are unknown.

1.3 Shape from shading with part of the usual assumptions

In some works, some of the assumptions cited above are not made:

- The Lambertian reflectance map, which is a very simple model, cannot occur for real materials. More realistic models have been studied [4, 5].
- The lighting conditions are, most of the time, more complicated than a single infinite point light source [6]. Even in the case of a simple lighting, the effects of secondary reflections are important [7, 8].
- Some other works involve a cooperation between shape from shading and stereo, like in [9], or in [10], allowing to deal, for instance, with perspective projection.

Even if these works are of great interest and give directions for future works, it seems that they come too early, insofar as the existing shape from shading methods are not very satisfactory, even in the best conditions.

1.4 Resolution of the image irradiance equation

There are a lot of ways of dealing with the image irradiance equation, which make it very difficult to compare the different methods. In a recent work [11], Zhang *et al.* have implemented and tested six existing algorithms on a common set of four synthetic images (a vase and a Mozart's bust lighted from two different directions) and three real images (another vase, peppers and Lena). Their conclusion is clear-cut:

1. *All the SFS algorithms produce generally poor results when given synthetic data,*
2. *Results are even worse on real images, and*
3. *Results on synthetic data are not generally predictive of results on real data.*

The poorness of the results on real data can be explained, since the lighting vector $\overrightarrow{S}$ is preprocessed thanks to existing specific methods which make additional assumptions [12, 13], the reflectance map is supposed to be Lambertian, and this could be rather false, and, moreover, the albedo is not constant (especially in the case of Lena!). However, it is much more difficult to justify the poorness of the results for synthetic images, insofar as all the ideal assumptions cited before are perfectly verified. Would that mean that, even under these assumptions, the problem remains ill-posed?

1.5 Existence and uniqueness of solutions to the image irradiance equation

Under the previous assumptions, and if one supposes that $\overrightarrow{S}$ is known, the problem of existence and uniqueness for the solutions of (2) has been much studied [14, 15, 16, 17, 18, 19]. It depends, as ever with PDEs, on the set in which the solutions are searched. If h is not necessary continuous, there is always an infinity of solutions. Thus, it is usually admitted that h has to be continuous, but this is not sufficient to make the problem well-posed [20]. Even if we suppose that h is analytical, it has been proved that there can exist a family h_c of solutions, where c can vary continuously [21]. At this point, one has to add boundary conditions to make the problem well-posed, that is, one has to add a knowledge on h on the boundary $\delta\Omega$ of the reconstruction domain Ω. The boundary conditions can be Dirichlet conditions (knowledge of h on $\delta\Omega$) or Neumann conditions (knowledge of p and q on $\delta\Omega$). Moreover, such pieces of knowledge can come either as additional data (for instance, for an island observed from a satellite, the height is constant on the coast), or through two specific kind of points [14]:

- Points of maximal greylevel ("singular points"). For such points, if the surface is known to be Lambertian, the normal $\overrightarrow{N}$ to the surface is parallel to $\overrightarrow{S}$.
- Points of the silhouette (occluding contour), where $\overrightarrow{N}$ is contained in the image plane and is normal to the silhouette.

However, it has been proved [18, 20] that the silhouette is not as constraining as it could appear, towards the well-posedness of the problem. In fact, the best way of making the problem well-posed is firstly to add "constraints" on the solutions, so that usually the problem becomes over-constrained, and secondly to solve it in the way of finding the best solution, with respect to a given criterion which is, most of the time, the least squares criterion [22]. Let us note, however, that other criteria have been tested, which imply robust statistics [23].

1.6 Different methods of resolution

The resolution methods of SFS can be classified into three sets:

- The classical methods of resolution of PDEs: characteristic strips expansion [1], level sets methods [24], power series expansion [21] and viscosity solutions [16, 25, 26, 27]. While a power series expansion does not lend itself to discretization, characteristic strips expansion and level sets methods are propagation methods which are very sensitive to noise. The viscosity solutions seem to be the most promising. Moreover, these solutions are very well adapted to images containing edges.
- The second set contains the methods based on optimization. The problem is modeled by a functional which has to be minimized [22]. The functional is defined by the sum

of several terms. The first term reflects the data and is minimal when equation (2) is satisfied everywhere on the reconstruction domain Ω. To make the problem well-posed, some extra terms, reflecting some *a priori* knowledge on the solution, are added. Some local deterministic algorithms have been proposed to minimize the resulting functional [28, 29, 30, 31, 32, 33, 34, 23]. We have obtained more costly but global solutions with a stochastic optimization technique: the simulated annealing [35].

- The third set groups all other methods. Among them, the method by Tsai and Shah [36] gives rather good results [37] and is surprisingly simple to implement. Furthermore, local methods [38], as well as linear methods [39], do very strong additional assumptions, and for this reason, are a little limited, regarding possible applications.

We can now explain the purpose of this technical report and describe its organization.

1.7 General organization of the technical report

In this technical report, our contribution is to present a new formulation of shape from shading in the Bayesian framework, and a new algorithm of shape from shading using simulated annealing. Moreover, we will show why boundary conditions are not necessary to make the problem well-posed (apart from the classical concave/convex ambiguity). In section 2, we describe how optimization can be applied to shape from shading. In section 3, we present a first new method, named **M1**, for which some results are exhibited. In section 4, we present an alternative method, based on the classical algorithm of simulated annealing, and this method, named **M2**, is shown to work better than **M1** on a complex image. In section 5, we show that the main problem with **M2**, *i.e.* slowness, can be diminished by using a multiresolution version of **M1**, after having applied **M2** to the image of the smallest resolution. This leads to a third new method named **M3**. Conclusion and future work are covered in section 6.

2 Shape from shading as an optimization problem

In their survey [11], Zhang *et al.* concluded that the best results, in terms of accuracy, were given by the optimization methods (we will often say "minimization" in the following), although they are also the slowest. We address optimization for our new algorithms, not only because of the good reputation it has in the shape from shading community, but also because it offers a general framework which is very adaptable to various situations. As it would be too long, we are not going to comment here all the existing methods of shape from shading which use optimization. This could constitute a paper apart. We will only discuss some points, which will highlight the choices that we did for our new methods **M1**, **M2** and **M3**.

2.1 Choice of the unknowns

From now on, and as most authors, we will suppose that $\overrightarrow{S}$ is known, thanks to preprocessing by a specific method (some authors deal simultaneously with the recovery of the shape and that of $\overrightarrow{S}$ [40, 12]). Thus, the unknown is the function h which appears in the image irradiance equation only through its first derivatives p and q, which are two non independent functions, since, if h is of class C^2:

$$\partial p/\partial y = \partial q/\partial x \tag{3}$$

As equation (3) is a constraint on p and q, the most natural functional associated with equations (2) and (3) is:

$$\begin{aligned} \mathcal{F}_1(p,q,\mu) = & \iint_{(x,y)\in\Omega} \left[R(p(x,y),q(x,y)) - E(x,y)\right]^2 \, dx\, dy \\ & + \iint_{(x,y)\in\Omega} \mu(x,y) \left[\partial p/\partial y(x,y) - \partial q/\partial x(x,y)\right] \, dx\, dy \end{aligned} \tag{4}$$

where μ is a Lagrange multiplier. Horn and Brooks [22] have proved that the three Euler equations associated with $\mathcal{F}_1$ can be reduced, for h being of class C^2, to the Euler equation associated with the following functional:

$$\mathcal{F}_2(h) = \iint_{(x,y)\in\Omega} \left[R(\partial h/\partial x(x,y), \partial h/\partial y(x,y)) - E(x,y)\right]^2 \, dx\, dy \tag{5}$$

which implies directly the height function. Dealing with $\mathcal{F}_2$ seems to be more tractable than with $\mathcal{F}_1$ as, on the one hand, it contains only one unknown h which is the real unknown of the problem and, on the other hand, $\mathcal{F}_1$ is the sum of two terms which are not homogeneous one with the other. However, no author deals with $\mathcal{F}_2$. It seems that the reason for this apparent paradox is that it leads to very slow algorithms (in [37], it is claimed that several hours of CPU are required for an image of size 256×256!). As $\mathcal{F}_1$ and $\mathcal{F}_2$ have in fact the same minima, the most evident alternative to $\mathcal{F}_1$ is the following functional [22]:

$$\begin{aligned} \mathcal{F}_3(p,q) = & \iint_{(x,y)\in\Omega} \left[R(p(x,y),q(x,y)) - E(x,y)\right]^2 \, dx\, dy \\ & + \lambda_{\text{int}} \iint_{(x,y)\in\Omega} \left[\partial p/\partial y(x,y) - \partial q/\partial x(x,y)\right]^2 \, dx\, dy \end{aligned} \tag{6}$$

where λ_{int} is a positive constant named "integrability factor". As we will see later, this factor poses a problem, in the sense that the choice of its value is difficult to justify. This functional leads to quite satisfactory results [33]. However, a serious problem with $\mathcal{F}_3$ is that it is generally non convex, because of its first term. A way of partly solving this drawback is to add to $\mathcal{F}_3$ a third term [22]:

$$\begin{aligned}\mathcal{F}_4(p,q) = {} & \mathcal{F}_3(p,q) \\ & + \lambda_{\mathrm{smo}} \iint_{(x,y)\in\Omega} \left[|\nabla p(x,y)|^2 + |\nabla q(x,y)|^2\right] dx\,dy\end{aligned} \tag{7}$$

where λ_{smo} is a positive constant named "smoothing factor". The presence of a "smoothing term", in addition to the "integrability" and "data" terms, tends indeed to decrease the number of local minima of the functional, as we will see in the next section.

Furthermore, it has been dealt simultaneously with the three unknowns (h,p,q) [31]. Finally, other unknowns have also been used: among them, the stereographic coordinates (f,g) of the normal [28, 29, 41] present the great interest to be bounded on silhouettes, contrary to (p,q).

2.2 Recovery of the height from p and q

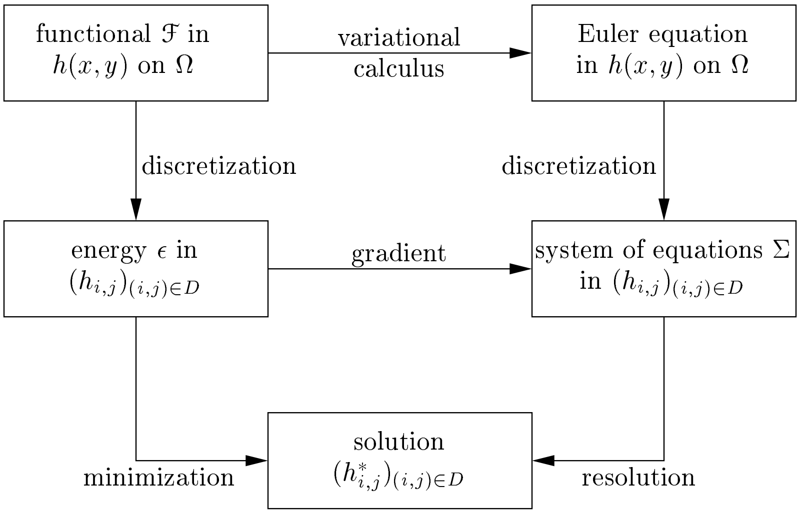
Besides the problem of selecting adequate values for the constants λ_{int} and λ_{smo}, the choice of (p,q) as unknowns poses another important problem, that of the computation of h from (p,q). The equations linking these functions being $\partial h/\partial x = p$ and $\partial h/\partial y = q$, it is natural to introduce the other following functional [22]:

$$\mathcal{F}_5(h) = \iint_{(x,y)\in\Omega} \left[(\partial h/\partial x(x,y) - p(x,y))^2 + (\partial h/\partial y(x,y) - q(x,y))^2\right] dx\,dy \tag{8}$$

Now, let us discuss the different strategies to find the minimum of a functional.

2.3 Minimization of a functional

The two main strategies allowing to numerically find the minimum of a functional $\mathcal{F}$ appear in the following figure, inspired by [33], where ϵ is the discrete approximation of $\mathcal{F}$, and D is the set of pixels contained in Ω:



- On the one hand, one can directly minimize the functional, by means of its discrete formulation ϵ, named an "energy".
- On the other hand, it is possible to produce one Euler equation for each unknown function, which characterizes the extrema of the functional. A discrete version of the Euler equations can be obtained, either by discretizing them, or by differentiating the energy, as shown in [22], and leads to a new system Σ.

It is surprising to find out that most of the algorithms of shape from shading use the second choice. As quoted by Szeliski [33], solving the system of equations Σ coming from the Euler equations is not strictly equivalent to minimize the energy, since all local minima and maxima (and inflexion points) of the energy are solutions of Σ. Moreover, when the equations of Σ are non linear (as in shape from shading), they have to be solved iteratively, and there exists no general proof of convergence for the produced schemes (proofs of divergence have even been given for two such methods in [42]). In [41], a proof of convergence for a method given in [28] is provided, but it has been shown in [37] that the admissible values for λ_{smo}, so that convergence is guaranteed, makes the method extremely slow. A much more serious problem with the Euler equations is that a well-posed problem can become ill-posed, if the "natural boundary conditions" [22] are not used. Let us focus on the crucial problem of well-posedness, in relation to boundary conditions.

2.4 Well-posedness and boundary conditions

We will deal with the functionals $\mathcal{F}_4$ and $\mathcal{F}_5$, for which the existence of one global minimum at least holds [43]. Is there a unique global minimum? This question is easier to handle with for the energies ϵ_4 and ϵ_5 associated with $\mathcal{F}_4$ and $\mathcal{F}_5$. As we will see later, the discrete approximation of a functional $\mathcal{F}$, *i.e.* its energy, is expressed through the use of forward finite differences. For instance, the approximation of $\partial h/\partial x$ at a pixel $(i,j) \in D$ involves the values $h_{i,j}$ and $h_{i+1,j}$, that is, the values of the unknown h at two neighbouring pixels (i,j) and $(i+1,j)$. If (i,j) is situated on the boundary of D, then $(i+1,j)$ may be out of D. If we denote by $\widetilde{D}$ the subset of D containing the pixels for which the forward finite differences used in the energy imply only pixels of D, and if $\widetilde{N} = \text{card}\,(\widetilde{D})$, then:

- The energy ϵ_4 is the least squares formulation of a set of $N + 5\widetilde{N}$ equations with the $2N$ unknowns $(p_{i,j}, q_{i,j})_{(i,j)\in D}$. This system of equations is over-constrained, which makes the minimization of ϵ_4 well-posed, in the absence of additional knowledge on the boundary.
- The energy ϵ_5 corresponds to a set of $2\widetilde{N}$ equations with the N unknowns $(h_{i,j})_{(i,j)\in D}$ so its minimization is a well-posed problem as well.

For these reasons, no knowledge on the boundary $\delta\Omega$ is required to make the shape from shading problem well-posed (neither Dirichlet nor Neumann conditions are required for this). In the algorithms that we will propose, indeed, we will not put any additional knowledge on h, p or q.

When dealing with the discretization of the Euler equations, the obtained system Σ is always under-constrained, because it consists in $2\widetilde{N}$ equations with $2N$ unknowns (or $\widetilde{N}$ equations with N unknowns), and $\widetilde{N} < N$. In fact, there exist, as mentioned by Horn and Brooks [22], "natural boundary conditions". To our knowledge, they have never been used in shape from shading. Therefore, this would be useful, as it would add the exact number of equations necessary to make Σ well-constrained.

2.5 Direct minimization of an energy

A lot of problems can be reformulated through the minimization of an energy. The two energies ϵ_4 and ϵ_5 that we have to minimize are fairly different, in the sense that ϵ_5 is issued from linear equations, and thus is convex, in the same time as ϵ_4 is generally non convex, because of the image irradiance equation, which is non linear. All the methods of minimization are iterative, that is, they construct a series of configurations $(\omega^k)_{k\in\mathbb{N}}$, beginning from an initial configuration ω^0, and hoping that $\lim\limits_{k\to+\infty} \omega^k = \omega^\star$, where $\omega^\star$ is the absolute minimiser of ϵ. Two kinds of methods of minimization can be distinguished:

- The first family of optimization methods are the deterministic methods. At each step k of the iteration, the computation of the new configuration ω^{k+1} uses the gradient of ϵ (and, sometimes, the Hessian matrix of ϵ) at the step k (and, sometimes, at the steps $0, \cdots, k-1$) and needs, most of the time, the search for a minimiser d^k of the function $\phi_k(d) = \epsilon(\omega^k - d\,v^k)$, where v^k is a "descent direction". The main problem is then to find d^k. This problem is called "line search" (see [44]). The conjugate gradient descent, used by Szeliski [33], is well adapted for energies derived from linear systems, since in such cases, it can converge towards $\omega^\star$ in a finite number of steps. Unfortunately, for an energy like ϵ_4, no general proof of convergence exists. Nevertheless the results obtained in [33] are good: the iteration seems to be stable and works faster than other methods associated with the same energy. A convenient way to ensure convergence is to stop the iteration if ϵ grows. The first method that we propose in this paper, named **M1**, uses, on the one hand, the very classical optimal gradient descent and, on the other hand, a method of line search based on parabolic interpolation: supposing that $\phi_k(d)$ can be approximated by a parabola, find a value $\widetilde{d} > 0$ for which $\phi_k(\widetilde{d}) > \phi_k(0)$, and take d^k at the "bottom" of the parabola passing through the three points $(0, \epsilon(\omega^k))$, $(\widetilde{d}, \phi_k(\widetilde{d}))$ and $(\widetilde{d}/2, \phi_k(\widetilde{d}/2))$. The advantage of **M1**, compared to the conjugate gradient descent, is that this method converges definitely.

- The second family of optimization methods are the stochastic methods, which allow to find a global minimum of an energy, contrary to the other methods, for which the possibility of finding a global minimum depends, for a non convex energy, on the initial configuration ω^0. In section 4, we will use the simulated annealing algorithm for the minimization of ϵ_4, which will produce the method **M2**.

3 Deterministic minimization for shape from shading

In this section, we will describe our first algorithm developped on the basis of the previous discussion.

3.1 General framework

From now on, we suppose that $\overrightarrow{S}$ is known and vertical ($\overrightarrow{S} = (0\ 0\ 1)^T$) and that the surface is Lambertian, so that the image irradiance equation (2) takes the particular form called the "eikonal equation" [14]:

$$\frac{E_{\max}}{\sqrt{1 + p(x,y)^2 + q(x,y)^2}} = E(x,y) \tag{9}$$

where $E_{\max}$ is the maximal value of E and is reached at points where $\nabla h = 0$. The discrete form of (9) is:

$$\frac{E_{\max}}{\sqrt{1 + {p_{i,j}}^2 + {q_{i,j}}^2}} = E_{i,j} \tag{10}$$

For the synthetic images which will be used for the tests, the greylevels will be computed from equation (10), with $E_{\max} = 255$.

3.2 Description of the used energies

Let us introduce the distance δ between two neighbouring pixels. We decide arbitrarily that, for each square scene, the side length is equal to 12.8, so that $\delta = 0.1$ for a 128×128 image,

$\delta = 0.2$ for a 64×64 image, and so on. Using forward finite differences, the discretization of equation (3) gives:

$$\frac{p_{i,j+1} - p_{i,j}}{\delta} = \frac{q_{i+1,j} - q_{i,j}}{\delta} \tag{11}$$

In the same way:

$$|\nabla p|^2_{i,j} \approx \left(\frac{p_{i+1,j} - p_{i,j}}{\delta}\right)^2 + \left(\frac{p_{i,j+1} - p_{i,j}}{\delta}\right)^2 \tag{12}$$

and idem for q, so that the energy ϵ_4 corresponding to the functional $\mathcal{F}_4$ is:

$$\begin{aligned}\epsilon_4(\omega_4) = &\,\delta^2 \sum_{(i,j)\in D} \left[\frac{E_{\max}}{\sqrt{1 + {p_{i,j}}^2 + {q_{i,j}}^2}} - E_{i,j}\right]^2 \\ &+ \lambda_{\text{int}} \sum_{(i,j)\in \widetilde{D}} \left[(p_{i,j+1} - p_{i,j}) - (q_{i+1,j} - q_{i,j})\right]^2 \\ &+ \lambda_{\text{smo}} \sum_{(i,j)\in \widetilde{D}} \left[(p_{i+1,j} - p_{i,j})^2 + (p_{i,j+1} - p_{i,j})^2 + (q_{i+1,j} - q_{i,j})^2 + (q_{i,j+1} - q_{i,j})^2\right]\end{aligned} \tag{13}$$

where $\omega_4 = (p_{i,j}, q_{i,j})_{(i,j)\in D}$ and where $\widetilde{D}$ is the subset of D containing the pixels (i,j) such that $(i+1,j)$ and $(i,j+1)$ are in D. It has been proved in [22] that this energy is essentially independent of the image resolution. This will be important when we will deal with multiresolution. In the same way, the energy ϵ_5 associated to $\mathcal{F}_5$ is:

$$\epsilon_5(\omega_5) = \sum_{(i,j)\in \widetilde{D}} \left[(h_{i+1,j} - h_{i,j} - \delta\, p_{i,j})^2 + (h_{i,j+1} - h_{i,j} - \delta\, q_{i,j})^2\right] \tag{14}$$

where $\omega_5 = (h_{i,j})_{(i,j)\in D}$.

3.3 First stage of the method M1

Given an image, a set of pixels D, one value for λ_{int} and another one for λ_{smo}, and given an initial configuration $\omega_4^0 = (p^0_{i,j}, q^0_{i,j})_{(i,j)\in D}$, the first stage of the method **M1** is an iterative process defined by the following steps:

- Compute $\nabla\epsilon_4$ at the current configuration ω_4^k.
- Find a local minimum d^k of the function $\phi_k(d) = \epsilon_4(\omega_4^k - d\,\nabla\epsilon_4(\omega_4^k))$.
- Compute the new configuration $\omega_4^{k+1} = \omega_4^k - d^k\,\nabla\epsilon_4(\omega_4^k)$.

The iteration is stopped at the first configuration $\omega_4^\star$ for which $|\nabla\epsilon_4(\omega_4^\star)| < \beta\sqrt{2N}$ (the threshold β is taken equal to 1.0, in our tests). The factor $\sqrt{2N}$ is necessary, in the previous inequality, as the Euclidean norm of $\nabla\epsilon_4$ is approximately proportional to its number of coordinates.

A configuration ω_4 can be represented through a "needle diagram", which consists in drawing the vectors $(p_{i,j}, q_{i,j})$ by little arrows (when $p_{i,j} = q_{i,j} = 0$, no dot appears in the needle diagrams processed by Matlab).

3.4 Second stage of the method M1

Given the result $\omega_4^\star$ of the first stage (which is now the data for the second stage) and given an initial configuration $\omega_5^0 = (h_{i,j}^0)_{(i,j)\in D}$, the second stage of **M1** is an iterative process which is, in all points, similar to the previous one. The final configuration $\omega_5^\star$, verifying $|\nabla\epsilon_5(\omega_5^\star)| < \beta\sqrt{N}$, is the final result of the method.

The energy ϵ_5 is invariant when a constant value is added to all $h_{i,j}$. This expresses the fact that shape from shading can only reconstruct a shape up to a constant. In our tests, the height h_{ref} of one pixel of D has to be fixed. Therefore, if we want to measure the precision of the reconstructed shape $\omega_5^\star$ (when the real shape $\widetilde{\omega_5}$ is known), the gap $|\omega_5^\star - \widetilde{\omega_5}|$, in the sense of a certain norm, for instance L_2, is in fact a function of h_{ref}. It would have no sense to fix arbitrarily h_{ref}. So, we choose the value of h_{ref} which gives the best fitting between $\omega_5^\star$ and $\widetilde{\omega_5}$.

A configuration ω_5 is a shape and can be represented in a more evident way than ω_4.

3.5 Remarks on the computation of $\nabla\epsilon$

When computing the gradients of ϵ_4 and ϵ_5, a special care must be taken for the pixels lying on the boundary of D. Let us illustrate this remark in the case of ϵ_5. For almost all pixels (i,j) of D:

$$\frac{\partial\epsilon_5}{\partial h_{i,j}} = 2\left[4h_{i,j} - (h_{i+1,j} + h_{i-1,j} + h_{i,j+1} + h_{i,j-1}) + \delta\,(p_{i,j} - p_{i-1,j} + q_{i,j} - q_{i,j-1})\right] \quad (15)$$

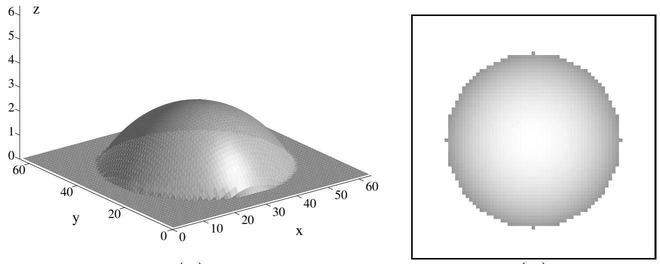
This expression is not valid, for instance, for a pixel (i_0,j_0) of D having (i_0-1,j_0) as unique neighbour contained in D. For such a pixel:

$$\frac{\partial\epsilon_5}{\partial h_{i_0,j_0}} = 2(h_{i_0,j_0} - h_{i_0-1,j_0} - \delta\,p_{i_0-1,j_0}) \quad (16)$$

There are in fact a certain number of special cases, which make the computation of these gradients a little tedious. However, experiments have shown that an exact computation of the gradient is absolutely necessary, in order to find the right solution. In fact, the interpretation of these special cases is the following: the normal case (15) corresponds to the Euler equation associated with ϵ_5, as all special cases like (16) correspond to the natural boundary conditions, which have already been quoted.

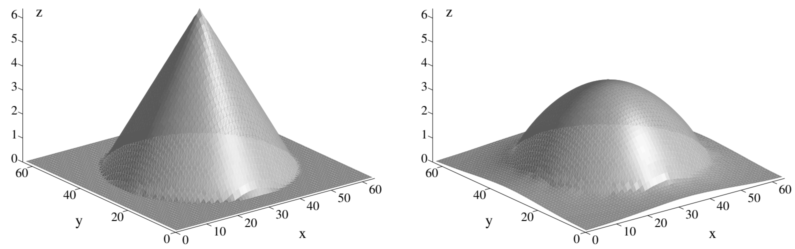
3.6 Experimental results

Before all, let us recall that no boundary condition is used in our tests. For the tests concerning the method **M1**, λ_{int} is equal to 10, λ_{smo} is equal to 50 and D contains all the pixels, except in one case which will be discussed. In figure 1(b), a first test image of size 64×64 has been computed, which corresponds to the simple shape represented in figure 1(a). Starting from the initial configurations ω_4^0 and ω_5^0 corresponding to the shape in figure 2(a), the method **M1** converges towards the shape represented in figure 2(b) in about 3 sec. of CPU time on a 733 MHz PC. The L_2-distance between 2(b) and the real shape 1(a), using the same scale as on the z axis, is equal to 0.015.



(a) (b)

Figure 1: (a) Spherical cap on a background and (b) 64×64 associated image.



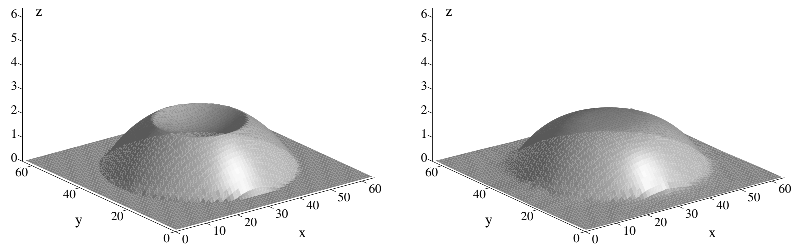
(a) (b)

Figure 2: (a) Starting shape and (b) associated result by **M1** ($\lambda_{\mathrm{int}} = 10$ and $\lambda_{\mathrm{smo}} = 50$).

In the same way, the result obtained for ω_4^0 and ω_5^0 corresponding to the shape in figure 3(a) is represented in figure 3(b).

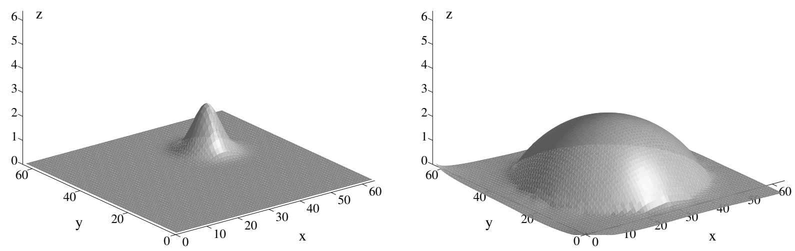
As these two initial shapes are fairly similar to the real shape, one could be doubtful about the robustness of the process. In figure 4(a), a third starting shape, which is not symmetrical with respect to the center of the image (in contrast with 2(a) or 3(a)), gives even so a result as good as previously, which is represented on figure 4(b).

As a proof for the necessity of using other terms than the "data term" in ϵ_4, let us set λ_{int} and λ_{smo} to 0 simultaneously. The problem becomes thus under-constrained. However, the method **M1** is able to deal with such a situation. Starting from the configuration which



(a) (b)

Figure 3: (a) Second starting shape and (b) associated result by **M1** ($\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 50$).



(a) (b)

Figure 4: (a) Third starting shape and (b) associated result by **M1** ($\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 50$).

corresponds to the shape 2(a) and which is represented, through a needle diagram, in figure 5(a), the first step of **M1** converges towards the configuration represented in figure 5(b), which corresponds to a good solution. On the other hand, starting from the configuration which corresponds to the shape 4(a) and which is represented in figure 6(a), the first step of **M1** converges towards the configuration represented in figure 6(b). This configuration is not satisfactory because it is not symmetrical with respect to the center of the image, unlike 5(b). One can observe that the orientations of the arrows in 6(b) are greatly induced by the initial configuration 6(a): the problem is definitely ill-posed, in this case.

Now, what is the effect of tuning λ_{int} and λ_{smo}? When setting λ_{int} to 10 and λ_{smo} to 0, starting from 3(a), the result is represented in figure 7(a), which seems to be a local minimum of the energy. So, it appears that the smoothing term of ϵ_4 reduces the number of local minima, since the shape 3(a) has given a good solution when $\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 50$ (figure 3(b)). On the other hand, when setting λ_{int} to 0 and λ_{smo} to 50, starting from 3(a), the solution is represented in figure 7(b). There is no local minimum anymore, but the ridge, at the bottom of the spherical cap, is smoothed, compared to the shape 7(a). Another effect of tuning λ_{int} and λ_{smo} can be seen when the method deals with a noisy image, as that of figure 8(a), obtained by adding a Gaussian noise with variance 10 to the image 1(b). With $\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 50$, starting from the shape 2(a), the solution is represented in figure 8(b). When $\lambda_{\text{int}} = 100$ and $\lambda_{\text{smo}} = 0$, the noise has obviously been interpreted as a rough texture of the surface (figure 9(a)). When $\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 200$, the solution is smoothed (figure 9(b)).

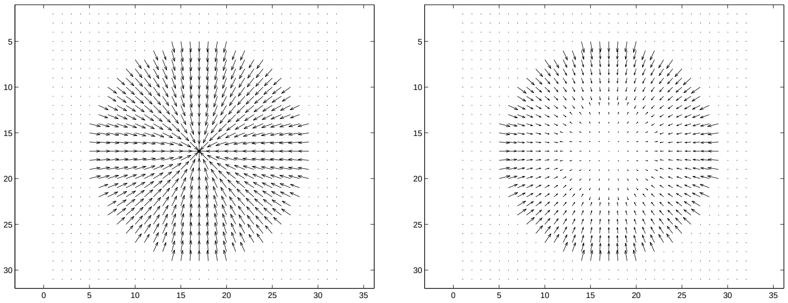
Let us use the more complex (but still synthetic) 64×64 image shown in figure 10(b), which corresponds to a vase on a background (figure 10(a)). Starting from the shape 1(a), with $\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 50$, the solution is represented in figure 11(a). It is not as satisfactory as the previous results. Therefore, compared to 1(a), the shape 10(a) is much more complex, because there are sharp edges at the top and at the bottom of the vase, as well as silhouettes on both sides. When a pixel (i, j) is situated on the background and $(i+1, j)$ is situated on the vase, it makes no sense to keep terms like $(p_{i+1,j} - p_{i,j})^2$ or $(h_{i+1,j} - h_{i,j} - \delta\, p_{i,j})^2$ in the energies ϵ_4 or ϵ_5, because h, p and q are not differentiable there. A way of overcoming this difficulty consists in withdrawing, in ϵ_4 and in ϵ_5, the terms which imply simultaneously pixels lying on the background and pixels lying on the vase. Making so, each new energy can be separated into two parts: $\epsilon(\omega) = \epsilon^{\text{back}}(\omega^{\text{back}}) + \epsilon^{\text{vase}}(\omega^{\text{vase}})$, where ω^{back} contains the unknowns for the background pixels and ω^{vase} the unknowns for the vase pixels, so that $\omega = (\omega^{\text{back}}, \omega^{\text{vase}})$. Therefore, the minimization of ϵ can be separated into two independent problems: the minimization of ϵ^{back}, and that of ϵ^{vase}. The only additional problem, besides the knowledge of the partition vase/background, with that way of processing is that ω_5^{back} and ω_5^{vase} will be known only up to two independent constants, so we will have to join the two reconstructed shapes. After joining ω_5^{back} and ω_5^{vase}, the final result is represented in figure 11(b). This result seems to be satisfactory. The CPU time is equal to 20 sec. (on a 733 MHz PC). The L_2-distance between 11(b) and the real shape 10(a) is equal to 0.761.

Now, let us study a much more complex shape, named DEM and represented in 12(a), associated to the 64×64 image shown in 12(b). The 3D interpretation of this image through human vision is quite impossible! Starting from a shape "similar" to the real shape, represented in 13(a), **M1** yields 13(b) as solution, which is not totally satisfactory. Starting from 14(a), the result obtained (see 14(b)) is not satisfactory at all, even from a qualitative point a view. On the other hand, when $\lambda_{\text{int}} = 500$ and $\lambda_{\text{smo}} = 20$, the results obtained from 13(a) and 14(a) are represented in figures 15(a) and 15(b). It is obvious that 15(a) is excellent, in the same time as 15(b) is very bad. We can thus conclude that **M1** has two residual disadvantages:

- The choice of λ_{int} and λ_{smo} is relatively arbitrary and has a great effect on the solution (compare 13(b) and 15(a)).

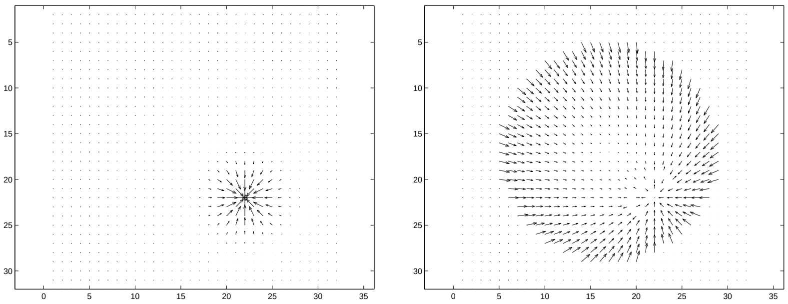
- The choice of the starting shape is of great importance (compare 15(a) and 15(b)), as ever with deterministic minimization.

In the next section, we address simulated annealing to overcome this second problem, and we let the first problem for future work.



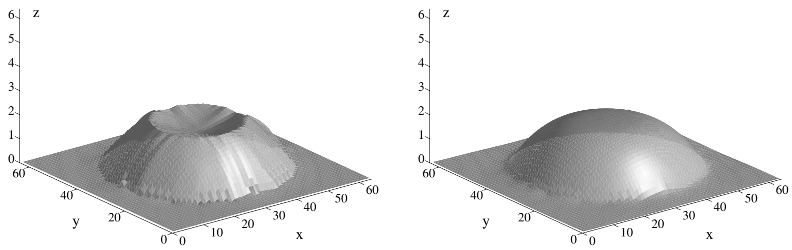
(a) (b)

Figure 5: (a) Needle diagram of 2(a) and (b) needle diagram of the associated result by **M1** ($\lambda_{\text{int}} = 0$ and $\lambda_{\text{smo}} = 0$).



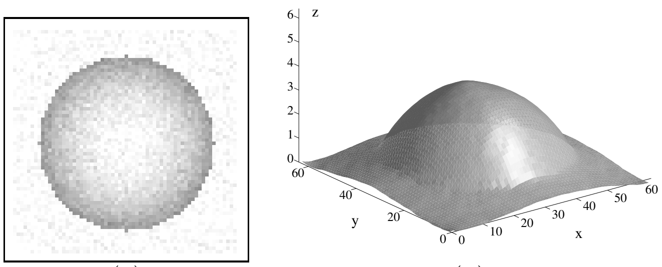
(a) (b)

Figure 6: (a) Needle diagram of 4(a) and (b) needle diagram of the associated result by **M1** ($\lambda_{\text{int}} = 0$ and $\lambda_{\text{smo}} = 0$).



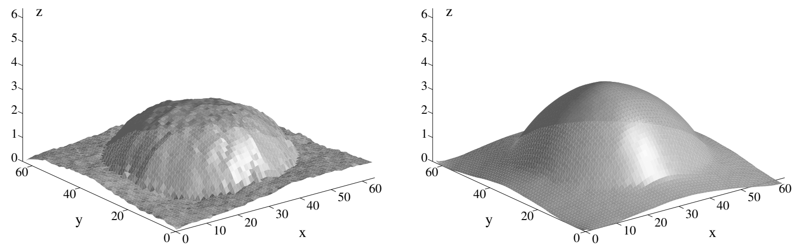
(a) (b)

Figure 7: Results obtained by **M1** starting from 3(a): (a) $\lambda_{\mathrm{int}} = 10$ and $\lambda_{\mathrm{smo}} = 0$; (b) $\lambda_{\mathrm{int}} = 0$ and $\lambda_{\mathrm{smo}} = 50$.



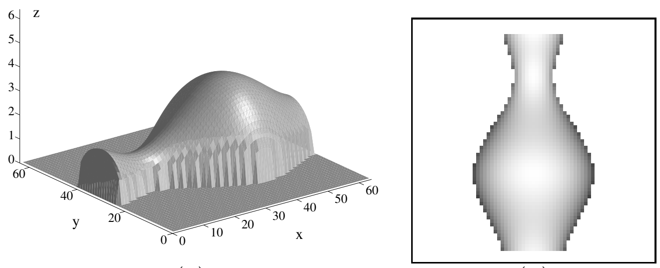
(a) (b)

Figure 8: (a) Noisy 64×64 image of 1(a) and (b) result obtained by **M1** starting from 2(a) ($\lambda_{\mathrm{int}} = 10$ and $\lambda_{\mathrm{smo}} = 50$).



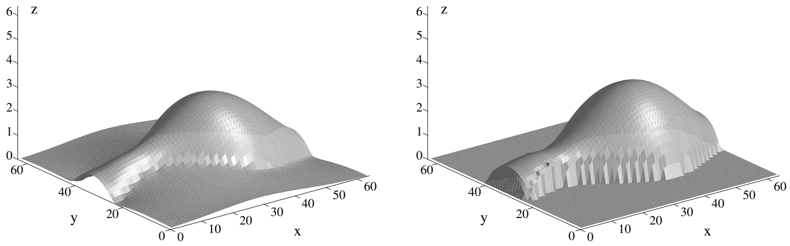
(a) (b)

Figure 9: Results obtained by **M1** starting from 2(a): (a) $\lambda_{\text{int}} = 100$ and $\lambda_{\text{smo}} = 0$; (b) $\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 200$.



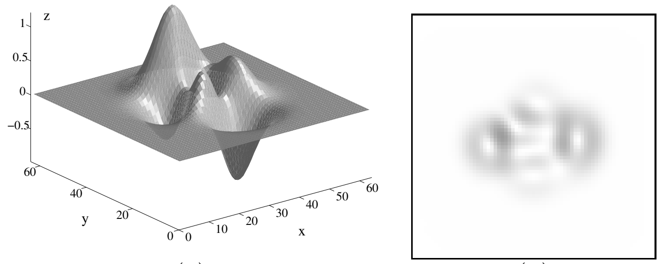
(a) (b)

Figure 10: (a) Vase on a background and (b) 64×64 associated image.



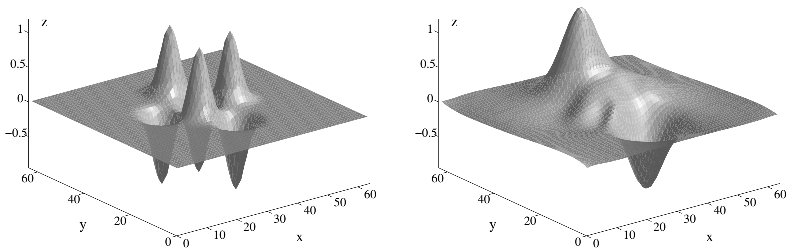
(a) (b)

Figure 11: **M1** applied (a) to the whole image 10(b) and (b) to two subparts joined afterwards ($\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 50$).



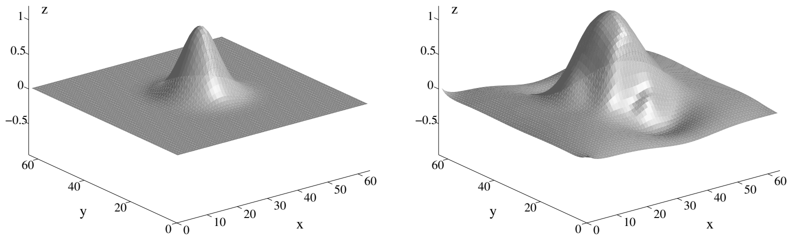
(a) (b)

Figure 12: (a) DEM and (b) 64×64 associated image.



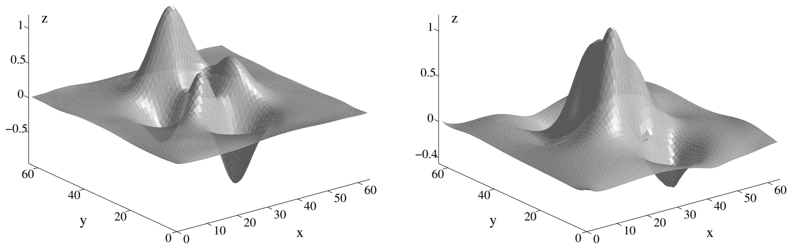
(a) (b)

Figure 13: (a) Starting shape and (b) associated result by **M1** ($\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 50$).



(a) (b)

Figure 14: (a) Second starting shape and (b) associated result by **M1** ($\lambda_{\text{int}} = 10$ and $\lambda_{\text{smo}} = 50$).



(a) (b)

Figure 15: When $\lambda_{\text{int}} = 500$ and $\lambda_{\text{smo}} = 20$: results obtained by **M1** starting (a) from 13(a) and (b) from 14(a).

4 A Bayesian approach for shape from shading

Shape from shading belongs to the class of ill-posed inverse problems which can be reformulated as an optimization problem in a Bayesian framework. In this section, we recall the Bayesian approach and rewrite the shape from shading problem into a stochastic framework. We then propose a simulated annealing algorithm to optimize the derived model.

4.1 Markov Random Fields in a Bayesian framework

Let E denote the data, $S \subset \mathbb{Z}^2$ the lattice (set of sites) and $\Lambda \subset \mathbb{R}^2$ the set of states. A configuration in $\Lambda^{\operatorname{card}(S)}$ is denoted by ω and the state of the site $s = (i, j)$ by $\omega_s = (p_s, q_s)$. We want to find the configuration which maximizes the *a posteriori* probability $P(\omega|E)$. Using the Bayes rule, we have:

$$P(\omega|E) = \frac{P(E|\omega)P(\omega)}{P(E)} \propto P(E|\omega)P(\omega) \tag{17}$$

The first term of the right hand side is the likelihood and represents the information provided by the data. The second term $P(\omega)$ is referred to as the prior. This probability distribution embeds some *a priori* properties, such as smoothness, which constrain the solution. The Markov Random Fields (MRF) are common and efficient models in this framework [45].

MRF are discrete stochastic processes whose global properties are controlled by means of local properties [46, 47, 48]. They are defined by local conditional probabilities which can model some interactions between neighbour pixels.

Definition: A random field (ω, P) is said to be Markovian if and only if:

$$\begin{cases} \forall\, \omega \in \Lambda^{\operatorname{card}(S)},\ P(\omega) > 0 \\ \forall\, s \in S,\ \forall\, \lambda \in \Lambda,\ P(\omega_s = \lambda | \omega_t, \forall\, t \in S \backslash \{s\}) = P(\omega_s = \lambda | \omega_t, \forall\, t \in \mathcal{V}_s) \end{cases} \tag{18}$$

where $\mathcal{V}_s$ is the neighbourhood of s.

Let us denote by c any finite subset of S and call it a clique. The Hammersley-Clifford theorem [46] allows us to write an MRF as a Gibbs field:

$$P(\omega) = \frac{1}{Z} \exp\left[-U(\omega)\right] = \frac{1}{Z} \exp\left[-\sum_{c \in \mathcal{C}} V_c(\omega_s, s \in c)\right] \tag{19}$$

where U is the energy, V_c is a function from $\Lambda^{\operatorname{card}(c)}$ onto $\mathbb{R}$ and Z is the partition function (normalization constant). Here, V_c refers to the potential associated with the clique c. The set of cliques $\mathcal{C}$, which is included in the set of finite subsets of S, induces the neighbourhood system of the associated MRF which is described by $\mathcal{V}_s = \{t \in S \backslash \{s\} : \exists c \in \mathcal{C}, \{s, t\} \subset c\}$.

4.2 The simulated annealing

The Maximum A Posteriori (MAP) is then obtained by minimizing the so called energy which, if we assume a conditional independency of the likelihood, is given by the following functional:

$$U = \sum_{c \in \mathcal{C}} V_c(\omega_s, s \in c) + \sum_{s \in S} -\log(\pi(E_s|\omega_s)) \tag{20}$$

where π denotes the local likelihood. The energy is usually not convex. To obtain the MAP criterion, we run a simulated annealing scheme using a Metropolis-Hastings dynamic [49]. This algorithm allows us to escape from the local minima of the energy function and has been proved to converge to the MAP solution. It is expressed as follows:

1. Initialize a random configuration $\omega^0 = (\omega_s^0, s \in S)$, set $T = T_0$ and $k = 1$.
2. For each site s (the sites being empirically numbered from 1 to $\operatorname{card}(S)$):
   - 2.a. Choose a random value new in Λ following a proposal distribution $Q(\omega_s^k = cur \rightarrow \omega_s^{k+1} = new)$ where cur is the current state of s and new is the proposed state.
   - 2.b. Compute the acceptation ratio:

     $$R = \left(\frac{P(\omega^{new})}{P(\omega^{cur})} \right)^{1/T_k} \frac{Q(\omega_s^k = new \rightarrow \omega_s^{k+1} = cur)}{Q(\omega_s^k = cur \rightarrow \omega_s^{k+1} = new)} \tag{21}$$

     where $\omega^{\text{value}} = (\omega_1^{k+1}, \cdots, \omega_{s-1}^{k+1}, \omega_s = \text{value}, \omega_{s+1}^k, \cdots, \omega^k_{\operatorname{card}(S)})$, for value $= cur, new$.
   - 2.c. Accept the proposition with probability $\min(1, R)$. If the proposition is accepted set ω_s^{k+1} to *new*, else set ω_s^{k+1} to *cur*.
3. If the stopping criterion is not reached, decrease the temperature $T_{k+1} = f_{\text{dec}}(T_0, k)$, increment k and go to 2.

4.3 A Markov model for shape from shading

We derive a Markov Random Field, adapted to the shape from shading problem, which contains four terms $P(E|\omega)$, $P_1(\omega)$, $P_2(\omega)$ and $P_3(\omega)$. The first term refers to the data and defines the likelihood. The three last terms define the prior and respectively define a smoothing constraint, an integrability constraint and a prior on the (p,q) distribution.

In the previous section, we have derived an energy ϵ_4 (see equation (13)) modelling the shape from shading problem (in fact, the first part of the problem). This energy embeds the eikonal equation, the integrability constraint and a smoothing constraint. Besides, this energy can be written as a sum of local functions which depend only on neighbouring pixels. Therefore, ϵ_4 can be considered as the energy associated with a Gibbs field. We thus propose to write the three first terms as follows:

$$P(E|\omega)P_1(\omega)P_2(\omega) = \frac{1}{Z} \exp\left[-\epsilon_4(\omega)\right] \tag{22}$$

To define the prior on the (p, q) distribution, we consider the sites to be independent with respect to this term, therefore we have the following:

$$P_3(\omega) = \prod_{s \in S} \pi_3(\omega_s) \tag{23}$$

To define the prior probability density π_3, we assume that the normal $\overrightarrow{N}$ of the surface is uniformly distributed on the Northern hemisphere of the Gaussian sphere. We use the spherical coordinates (see figure 16), so that the slope $\rho = \sqrt{p^2 + q^2}$ is equal to $\tan\theta$ and:

$$\begin{cases} p = \rho \cos\phi \\ q = \rho \sin\phi \end{cases} \tag{24}$$

On the Gaussian sphere, the surface of the crown $(\theta, \theta + d\theta)$ (see figure 16) is equal to $2\pi \sin\theta \, d\theta$. We therefore consider the distributions $\pi_\theta(\theta) \propto \sin\theta$ and $\pi_\phi(\phi)$ uniform on $[0, 2\pi[$ to define the prior on (p, q). Since $\rho = \tan\theta$, we have $\sin\theta \, d\theta = \frac{\rho}{(1+\rho^2)^{3/2}} \, d\rho$. The *a priori* distribution on ρ is then given by:

$$\pi_\rho(\rho) \propto \frac{\rho}{(1 + \rho^2)^{3/2}} \tag{25}$$

In practice, we bound the state space associated with ρ to $[0, 2\rho_{\max}]$, $\rho_{\max}$ being the maximal value of ρ. This value is computed by applying the eikonal equation (10) to the data ($\rho_{\max}$ corresponds to the lowest greylevel). We obtain:

$$\pi_\rho(\rho) = \frac{\dfrac{\rho}{(1 + \rho^2)^{3/2}}}{\displaystyle\int_0^{2\rho_{\max}} \frac{\rho}{(1 + \rho^2)^{3/2}} \, d\rho}, \quad \rho \in [0, 2\rho_{\max}] \tag{26}$$

4.4 Acceptation ratio

To reduce the CPU time of the optimization, we have to reduce the time necessary to simulate the proposal and the number of steps necessary to reach the convergence. To reduce the number of steps, we have to consider a proposal similar to the considered model. In case of a model with interactions, this leads to a proposal long to simulate. If we consider the data values in the proposal, we have to compute a different proposal at each site. Therefore, we propose to only consider the non interacting term of the prior model in the proposal distribution.

The proposal defined in the simulated annealing algorithm is as follows:

$$Q(\omega_s^k = cur \rightarrow \omega_s^{k+1} = new) = Q(\omega_s^{k+1} = new) = \pi_3(new) \tag{27}$$

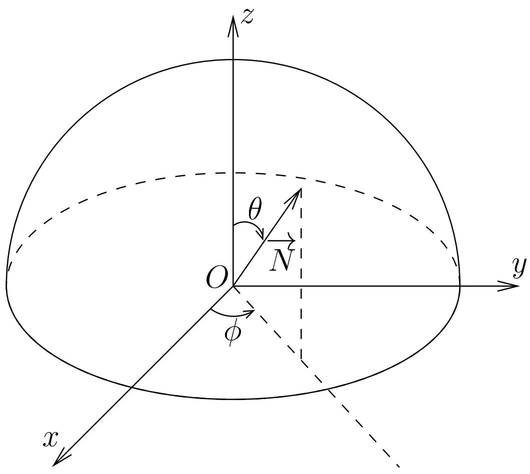


Figure 16: The Gaussian sphere.

Denote $\omega\backslash\{s\} = (\omega_1^{new}, ..., \omega_{s-1}^{new}, \omega_{s+1}^{new}, ..., \omega_{\mathrm{card}\,(S)}^{new}) = (\omega_1^{cur}, ..., \omega_{s-1}^{cur}, \omega_{s+1}^{cur}, ..., \omega_{\mathrm{card}\,(S)}^{cur})$. The acceptation ratio is then given by:

$$\begin{aligned} R &= \left(\frac{P(\omega^{new})}{P(\omega^{cur})}\right)^{1/T} \frac{Q(cur)}{Q(new)} \\ &= \left(\frac{\pi(E_s|\omega_s^{new})P_1(\omega_s^{new}|\omega\backslash\{s\})P_2(\omega_s^{new}|\omega\backslash\{s\})}{\pi(E_s|\omega_s^{cur})P_1(\omega_s^{cur}|\omega\backslash\{s\})P_2(\omega_s^{cur}|\omega\backslash\{s\})}\right)^{1/T} \left(\frac{\pi_3(new)}{\pi_3(cur)}\right)^{1/T-1} \end{aligned} \tag{28}$$

where $\pi(E_s|\omega_s) \propto \exp\left[-\delta^2\left(E_{\max}/\sqrt{1+{p_s}^2+{q_s}^2} - E_s\right)^2\right]$.

So, using the Markov property, we have to compute:

$$R = \exp\left[-\frac{F(p_s^{new}, q_s^{new}) - F(p_s^{cur}, q_s^{cur})}{T}\right] \left(\frac{\rho_s^{new}(1+(\rho_s^{cur})^2)^{3/2}}{\rho_s^{cur}(1+(\rho_s^{new})^2)^{3/2}}\right)^{1/T-1} \tag{29}$$

with:

$$\begin{aligned} F(p_s, q_s) = \delta^2 \left(\frac{E_{\max}}{\sqrt{1+{p_s}^2+{q_s}^2}} - E_s\right)^2 + \lambda_{\mathrm{smo}} \sum_{s' \in \{s_1, s_2, s_3, s_4\}} \left[(p_{s'} - p_s)^2 + (q_{s'} - q_s)^2\right] \\ +\lambda_{\mathrm{int}}\Big\{ \left[(p_{s_1} - p_s) - (q_{s_2} - q_s)\right]^2 + \left[(p_s - p_{s_3}) - (q_{s_6} - q_{s_3})\right]^2 + \left[(p_{s_5} - p_{s_4}) - (q_s - q_{s_4})\right]^2 \Big\} \end{aligned}$$

where $s_1 = s + (0,1)$, $s_2 = s + (1,0)$, $s_3 = s + (0,-1)$, $s_4 = s + (-1,0)$, $s_5 = s + (-1,1)$, $s_6 = s + (1,-1)$.

4.5 Results on the DEM

In this subsection, we focus on the DEM for which the method **M1** fails, and test it with the new method called **M2**. The results on the sphere and on the vase are similar to those obtained with **M1**. A crucial point to obtain the MAP estimator, using a simulated annealing scheme, lies in the definition of the cooling scheme $T_k = f_{\text{dec}}(T_0, k)$. From a theoretical point of view, the cooling scheme should be of the form $T_k = \frac{T_0}{\log(k+1)}$ (see [48]), where T_0 is proportional to the maximal difference between the energy of any two configurations. In practice, a geometrical scheme $T_k = \alpha^k T_0$ is often used to speed up the convergence. When applying this approach to image segmentation or image restoration, the values of α reported in the literature lie between 0.95 and 0.99. The tests that we performed have revealed that this cooling scheme is too fast for the shape from shading problem. Using this, the simulated annealing algorithm converges to a local minimum which is far from the actual surface (see the result obtained with $\alpha = 0.99$ in figure 17(a)). This is due to very deep local minima of the shape from shading functional. To actually converge to the global minimum we have to use $\alpha = 0.999998$ which leads to 6.10^6 iterations. The result is shown in figure 17(b). The initial configuration has no influence on the result. To obtain figure 17(b) we used a plane as initial configuration. We have obtained similar results using a random configuration as initialization. Because of the stochastic perturbations during the iterative scheme, the configuration has escaped from the local minima. The result is very close to the true surface[2]. It has been obtained on an image of the DEM of size 32×32 and has required about one hour of CPU time. The L_2-distance between 17(b) and the real shape 12(a) is equal to 0.353.

It is interesting to compare the needle diagrams of this result (figure 18(a)) and of the real shape (figure 18(b)). They are very similar but the arrows in the two diagrams are all oriented in opposite directions. In fact, a shape has exactly the same energy as its opposite, and we have observed indeed that **M2** gives sometimes the shape 17(b) and sometimes its opposite. The shape 17(b) corresponds to the needle diagram opposite to 18(a).

Apparently, we could consider the slopes as known since, from (10), ρ directly follows from E. Under this assumption, we probably could have cooled the temperature with a greater value for α, and then CPU time would have been shortened. However, since this remark does not hold anymore for noisy images, it did not seem appropriate to use such a simplification.

To handle with images having a more realistic size than 32×32, it is needed to speed up the convergence. The simulated annealing is helpful to escape from the local minima but, if the configuration is in the region of attraction of the global minimum, the method **M1** is much more efficient. This motivates the multiresolution approach **M3** proposed in the next section. The idea is to run a simulated annealing at a low resolution to obtain the global minimum at this resolution. We then compute **M1** at higher resolutions taking the previous result as initial configuration. Therefore, we assume that the global minimum at

[2] A movie corresponding to this processing can be found at the following address:
`http://www.irit.fr/~Jean-Denis.Durou/RECHERCHE/RECUIT/IMAGES/recuit.mpg`

low resolution belongs to the region of attraction of the global minimum when projected at higher resolutions.

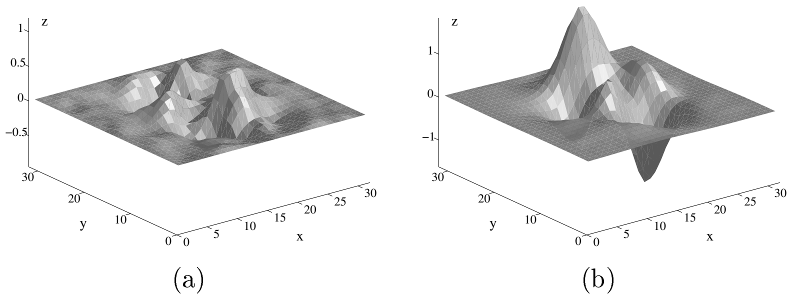


Figure 17: **M2** applied to image 12(b): (a) $\alpha = 0.99$ and (b) $\alpha = 0.999998$ ($\lambda_{\text{int}} = 500$ and $\lambda_{\text{smo}} = 20$).

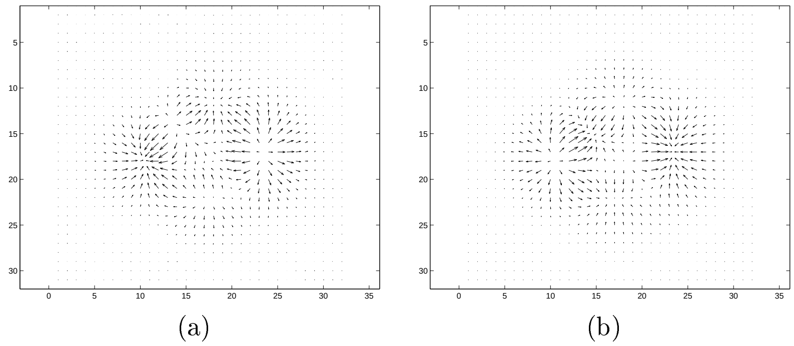


Figure 18: Needle diagrams (a) of the shape opposite to 17(b) and (b) of the DEM.

5 A multiresolution hybrid method

As already said in the previous section, the main drawback of the simulated annealing is the slowness of the computation. A solution consists in a hierarchical multiresolution approach that we are going to describe in this section.

5.1 Multiresolution for shape from shading

Multiresolution has been already used in the context of shape from shading [50, 33, 51] in order to accelerate the convergence of classical algorithms. Multiresolution techniques in image analysis are generally based on the construction of an image pyramid. For each level, the image is obtained by blurring then subsampling the image of the previous level. If the low pass filter used is the Gaussian filter, the pyramid is called a Gaussian pyramid.

In the case of shape from shading, the relation between the data (the image) and the unknown (the shape) is not linear. Consequently, reductions of the shape resolution and of the image resolution do not product the same effects. For this reason, we use the non linear multiresolution algorithm proposed in [51], which gives a better approximation of the shape resolution reduction. More precisely, the ideal situation would be to reduce the resolution of the shape and then to calculate the corresponding images. Unfortunately, the shape is the unknown of the problem, but the shape slopes $\rho_{i,j} = \sqrt{{p_{i,j}}^2 + {q_{i,j}}^2}$ can be computed from the greylevels using equation (10). Peleg and Ron [51] have shown that blurring the slopes is always a better approximation of blurring the shape than directly blurring the image.

If E^0 denotes the initial image (finest resolution) and E^l the image of size $2n \times 2n$ at level l, the coarser image E^{l+1} is computed using the following steps:

1. Calculate slopes $\rho^l_{i,j}$ from greylevels using eikonal equation (10) at level l:

$$\rho^l_{i,j} = \sqrt{\frac{{E_{\max}}^2}{{E^l_{i,j}}^2} - 1}, \quad (i,j) \in [0, 2n-1]^2 \tag{30}$$

2. Blur the slope array using a Gaussian-like convolution mask.

3. Subsample the slope array by discarding every other row and column to form ρ^{l+1} of size $n \times n$.

4. Calculate the corresponding image using eikonal equation (10) at level $l+1$:

$$E^{l+1}_{i,j} = \frac{E_{\max}}{\sqrt{1 + \left(\rho^{l+1}_{i,j}\right)^2}}, \quad (i,j) \in [0, n-1]^2 \tag{31}$$

After the pyramid construction, a first result is obtained on the coarser level and this result is then oversampled by interpolation and used as the starting configuration for the next level. This process is repeated level after level until the result with the original resolution is obtained.

5.2 The hybrid method M3

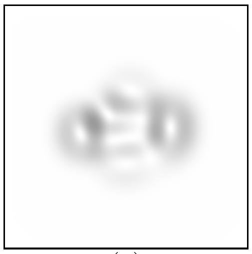
The remaining question is to decide what method to use at each level to provide the starting configuration for the next level. As said earlier, the method **M1** can fail leading to local minima. On the other hand, the method **M2** overcomes this problem, but is not completely applicable in the multiresolution scheme because its convergence speed does not depend on the starting configuration. Therefore, we chose to use the method **M2** at the coarser level of the pyramid in order to avoid local minima and the method **M1** at the other levels. The slowest method is then only used on a small image and, for the other levels, the number of iterations are reduced because the starting configuration is close to the final one.

5.3 Experimental results

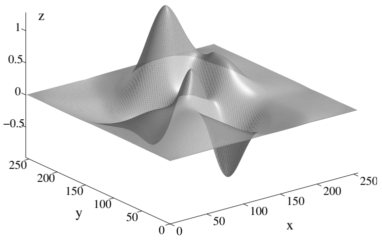
In our experiments we used a linear interpolation for the result propagation of (p, q) between levels and the following convolution mask (recommended in [51]) to blur the slope array during pyramid construction:

$$\begin{bmatrix} 0 & 0.125 & 0 \\ 0.125 & 0.5 & 0.125 \\ 0 & 0.125 & 0 \end{bmatrix}$$

We tested **M3** on the image 256×256 of the DEM represented in figure 19(a). The pyramid was thus created with four different resolutions (256×256, 128×128, 64×64, 32×32). The result is represented in figure 19(b). It is a little less satisfactory than 17(b) but the L_2-distance between 19(b) and the real shape 12(a) is equal to 0.471 only. Regarding the CPU time, it is almost totally due to the simulated annealing (about one hour, as already said). If we had applied **M2** to the image 19(a), the CPU time would have been equal to something like 4^3 hours!



(a)



(b)

Figure 19: (a) Image 256×256 of the DEM and (b) result by **M3** using 4 levels ($\lambda_{\text{int}} = 500$ and $\lambda_{\text{smo}} = 20$).

6 Conclusion

In this technical report, we have dealt with shape from shading expressed as a minimization problem. The non linearity of the shape from shading equation leads to the minimization of a highly non convex functional. In such a case, the deterministic approaches converge to a local minimum of the functional. They provide accurate results in simple cases where the surface itself is convex or concave, as we have shown on the spherical cap and on the vase, or when we can provide a good initial configuration, which means a configuration lying in the region of attraction of the global minimum. An alternative is to consider a stochastic algorithm which allows to escape from local minima. We have shown on a shape, referred to as DEM, which contains several convex and concave areas, that the simulated annealing algorithm provides accurate results, even in this difficult case. The local minima of the functional are very deep and we had to consider an unusually slow cooling scheme for the temperature. This fact may be one reason explaining that, up to our knowledge, the literature does not report on the association of shape from shading and simulated annealing. The counterpart of the good behavior of the simulated annealing for minimizing such complex functionals is the slowness of this algorithm. In order to combine the accuracy of the simulated annealing and the efficiency of the deterministic algorithm, we have proposed a multiresolution approach for which the simulated annealing is used at the lowest resolution and provide a good initial configuration for higher resolutions which are solved deterministically.

The functional we defined is written as the sum of local constraints and can be interpreted as the energy of a Markov Random Field. On each pixel, the acceptation ratio depends only on the four nearest neighbours. Therefore, we are currently studying a parallel version of the simulated annealing algorithm. In future work, we will also try to determine how the parameters, especially the factors λ_{int} and λ_{smo}, could be automatically estimated. Besides, we will address real images.

Acknowledgments

This work has been partially supported by the GdR ISIS within the "Young Researchers" program of the CNRS.

References

[1] B. K. P. Horn, "Obtaining Shape from Shading Information," in *The Psychology of Computer Vision*, P. H. Winston, Ed., chapter 4, pp. 115–155. McGraw-Hill, 1975.

[2] B. K. P. Horn and M. J. Brooks, Eds., *Shape from Shading*, MIT Press, 1989.

[3] R. Klette, R. Kozera, and K. Schlüns, "Shape from Shading and Photometric Stereo Methods," in *Signal Processing and Pattern Recognition*, B. Jähne, H. Haussecker, and

P. Geissler, Eds., vol. 2 of *Handbook of Computer Vision and Applications*, chapter 19, pp. 531–590. Academic Press, 1999.

[4] S. Bakshi, "Shape From Shading for Non-Lambertian Surfaces," M.S. thesis, University of Saskatchewan, Saskatoon, Saskatchewan, July 1994.

[5] M. Oren and S. K. Nayar, "Generalization of the Lambertian Model and Implications for Machine Vision," *International Journal of Computer Vision*, vol. 14, no. 3, pp. 227–251, Apr. 1995.

[6] Y.-L. Tian, H. T. Tsui, S. Y. Yeung, and S. Ma, "Shape from Shading for Multiple Light Sources," *Journal of the Optical Society of America A*, vol. 16, no. 1, pp. 36–52, Jan. 1999.

[7] A. J. Stewart and M. S. Langer, "Towards Accurate Recovery of Shape from Shading Under Diffuse Lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, pp. 1020–1025, Sept. 1997.

[8] D. A. Forsyth and A. Zisserman, "Reflections on Shading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 671–679, July 1991.

[9] P. Fua and Y. G. Leclerc, "Object-Centered Surface Reconstruction: Combining Multi-Image Stereo and Shading," *International Journal of Computer Vision*, vol. 16, pp. 35–56, Sept. 1995.

[10] J. E. Cryer, P.-S. Tsai, and M. Shah, "Integration of Shape from Shading and Stereo," *Pattern Recognition*, vol. 28, no. 7, pp. 1033–1044, Feb. 1995.

[11] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape from Shading: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 690–706, Aug. 1999.

[12] Q. Zheng and R. Chellappa, "Estimation of Illuminant Direction, Albedo, and Shape from Shading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 680–702, July 1991.

[13] A. P. Pentland, "Finding the Illuminant Direction," *Journal of the Optical Society of America A*, vol. 72, no. 4, pp. 448–455, Apr. 1982.

[14] A. Blake, A. Zisserman, and G. Knowles, "Surface Descriptions from Stereo and Shading," *Image and Vision Computing*, vol. 3, no. 4, pp. 183–191, 1985.

[15] J. Oliensis, "Uniqueness in Shape from Shading," *International Journal of Computer Vision*, vol. 6, no. 2, pp. 75–104, June 1991.

[16] P.-L. Lions, E. Rouy, and A. Tourin, "Shape-from-Shading, Viscosity Solutions and Edges," *Numerische Mathematik*, vol. 64, pp. 323–353, 1993.

[17] B. K. P. Horn, R. Szeliski, and A. L. Yuille, "Impossible Shaded Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 2, pp. 166–170, Feb. 1993.

[18] M. J. Brooks, W. Chojnacki, and R. Kozera, "Circularly Symmetric Eikonal Equations and Non-uniqueness in Computer Vision," *Journal on Mathematical Analysis and Applications*, vol. 165, no. 1, pp. 192–215, 1992.

[19] M. J. Brooks, W. Chojnacki, and R. Kozera, "Impossible and Ambiguous Shading Patterns," *International Journal of Computer Vision*, vol. 7, no. 2, pp. 119–126, 1992.

[20] J. Oliensis, "Shape from Shading as a Partially Well-Constrained Problem," *Computer Vision, Graphics, and Image Processing: Image Understanding*, vol. 54, no. 2, pp. 163–183, Sept. 1991.

[21] J.-D. Durou and D. Piau, "Ambiguous Shape from Shading with Critical Points," *Journal of Mathematical Imaging and Vision*, vol. 12, no. 2, pp. 99–108, Apr. 2000.

[22] B. K. P. Horn and M. J. Brooks, "The Variational Approach to Shape From Shading," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 2, pp. 174–208, Feb. 1986.

[23] P. L. Worthington and E. R. Hancock, "New Constraints on Data-Closeness and Needle Map Consistency for Shape-from-Shading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1250–1267, Dec. 1999.

[24] R. Kimmel and A. M. Bruckstein, "Tracking Level Sets by Level Sets: A Method for Solving the Shape from Shading Problem," *Computer Vision and Image Understanding*, vol. 62, no. 2, pp. 47–58, July 1995.

[25] F. Camilli and M. Falcone, "An Approximation Scheme for the Maximal Solution of the Shape-from-Shading Problem," in *Proceedings of the IEEE International Conference on Image Processing*, Lausanne, Switzerland, Sept. 1996, vol. I, pp. 49–52.

[26] P. Dupuis and J. Oliensis, "Direct Method for Reconstructing Shape from Shading," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, USA, June 1992, pp. 453–458.

[27] M. Falcone and M. Sagona, "An algorithm for the global solution of the Shape-from-Shading model," in *Proceedings of the 9th IEEE International Conference on Image Analysis and Processing*, Florence, Italy, Sept. 1997, vol. I, pp. 596–603.

[28] K. Ikeuchi and B. K. P. Horn, "Numerical Shape from Shading and Occluding Boundaries," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 141–194, Aug. 1981.

[29] R. T. Frankot and R. Chellappa, "A Method for Enforcing Integrability in Shape from Shading Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 439–451, July 1988.

[30] J. Malik and D. Maydan, "Recovering Three-Dimensional Shape from a Single Image of Curved Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 555–566, June 1989.

[31] B. K. P. Horn, "Height and Gradient from Shading," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 37–75, Aug. 1990.

[32] Y. G. Leclerc and A. F. Bobick, "The Direct Computation of Height from Shading," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, USA, June 1991, pp. 552–558.

[33] R. Szeliski, "Fast Shape from Shading," *Computer Vision, Graphics, and Image Processing: Image Understanding*, vol. 53, no. 2, pp. 129–153, Mar. 1991.

[34] K. M. Lee and C.-C. J. Kuo, "Shape from Shading with a Linear Triangular Element Surface Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 8, pp. 815–822, Aug. 1993.

[35] X. Descombes, J.-D. Durou, and D. Petit, "Recuit simulé pour le shape from shading," in *Actes du 18ème Colloque GRETSI sur le Traitement du Signal et des Images*, Toulouse, France, Sept. 2001, pp. 513–516, (in French).

[36] P.-S. Tsai and M. Shah, "Shape from Shading using Linear Approximation," *Image and Vision Computing*, vol. 12, no. 8, pp. 487–498, Oct. 1994.

[37] P. Daniel and J.-D. Durou, "From Deterministic to Stochastic Methods for Shape From Shading," in *Proceedings of the 4th Asian Conference on Computer Vision*, Taipei, Taiwan, Jan. 2000, pp. 187–192.

[38] A. P. Pentland, "Local Shading Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 2, pp. 170–187, Mar. 1984.

[39] R. Kozera and R. Klette, "Finite Difference Based Algorithms in Linear Shape from Shading," *Machine Graphics and Vision*, vol. 6, no. 2, pp. 157–201, 1997.

[40] M. J. Brooks and B. K. P. Horn, "Shape and Source from Shading," in *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, Los Angeles, California, USA, Aug. 1985, vol. II, pp. 932–936.

[41] D. Lee, "A Provably Convergent Algorithm for Shape from Shading," in *DARPA Image Understanding Workshop*, Miami Beach, Florida, USA, Dec. 1985, pp. 489–496.

[42] J.-D. Durou and H. Maître, "On Convergence in the Methods of Strat and of Smith for Shape from Shading," *International Journal of Computer Vision*, vol. 17, no. 3, pp. 273–289, Mar. 1996.

[43] T. Poggio, V. Torre, and C. Koch, "Computational Vision and Regularization Theory," *Nature*, vol. 317, no. 26, pp. 314–319, Sept. 1985.

[44] J.-F. Bonnans, J.-C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, Eds., *Numerical Optimization: Theoretical and Practical Aspects*, Springer-Verlag, 2002.

[45] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction*, Springer-Verlag, 1995.

[46] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society B*, vol. 36, no. 2, pp. 192–326, July 1974.

[47] G. R. Cross and A. K. Jain, "Markov Random Field Texture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 1, pp. 25–39, Jan. 1983.

[48] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[49] W. K. Hastings, "Monte Carlo Sampling Methods using Markov Chains and their Applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[50] D. Terzopoulos, "Efficient Multiresolution Algorithms for Computing Lightness, Shape from Shading, and Optical Flow," in *Proceedings of the 4th National Conference on Artificial Intelligence*, Austin, Texas, USA, Aug. 1984, pp. 314–317.

[51] S. Peleg and G. Ron, "Nonlinear Multiresolution: A Shape-from-Shading Example," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 12, pp. 1206–1210, Dec. 1990.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399