A Markov Chain Monte Carlo Approach to Stereovision

Julien Sénégas[1,2]

[1] Centre de Géostatistique, Ecole des Mines de Paris, 35 rue Saint-Honoré, 77305 Fontainebleau Cedex, France
[2] Istar, 2600 route des Crêtes, BP282 06905 Sophia-Antipolis Cedex, France
senegas@cg.ensmp.fr

Abstract. We propose Markov chain Monte Carlo sampling methods to address uncertainty estimation in disparity computation. We consider this problem at a postprocessing stage, i.e. once the disparity map has been computed, and suppose that the only information available is the stereoscopic pair. The method, which consists of sampling from the posterior distribution given the stereoscopic pair, allows the prediction of large errors which occur with low probability, and accounts for spatial correlations. The model we use is oriented towards an application to mid-resolution stereo systems, but we give insights on how it can be extended. Moreover, we propose a new sampling algorithm relying on Markov chain theory and the use of importance sampling to speed up the computation. The efficiency of the algorithm is demonstrated, and we illustrate our method with the computation of confidence intervals and probability maps of large errors, which may be applied to optimize a trajectory in a three dimensional environment.

Keywords: stereoscopic vision, digital terrain models, disparity, uncertainty estimation, sampling algorithms, Bayesian computation, inverse problems

1 Introduction

In this paper, we address the problem of assessing the uncertainty of depth estimates computed from a stereo system. We exclude the possibility of using any external source of information on the observed scene, i.e. only the radiometric information provided by the stereoscopic pair will be used. The application we have in mind is the certification of digital terrain models for the aircraft industry, but the methods we propose here are quite general and can be applied to any stereo system with some adaptations of the model used.

1.1 Errors in Stereovision

A stereo system consists of two cameras observing the same scene from two different points. With this set-up, a left and right image are obtained. A physical

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2352, pp. 97–111, 2002.

point M projects onto the image plane of each camera. These projections form a pair of homologous points, and their coordinates in each image plane can be expressed as a function of the stereo system parameters (mainly the baseline b and the focal length f) and the coordinates of M. The principle of depth estimation from stereovision relies on the inversion of this relation. An extensive introduction to three dimensional computer vision and depth from stereo can be found in [10].

Many different techniques are used for depth computation, depending on the type of sensors and the nature of the scene. However, they generally consist of the following steps. First, the stereo system parameters are computed, using calibration methods (see [8] for an example). It is then possible to reproject the images in the epipolar geometry, in which two homologous points have the same vertical coordinate. The difference between their horizontal coordinates is called the disparity. Hence, the rectification step reduces the search of homologous points to a mono-dimensional one (along the rows of the images). The stereo problem itself amounts then to the computation of the disparity map, which is done by matching pixels (or features) in the left image with the corresponding pixels in the right image [9]. Lastly, the computation of the real world coordinates, the triangulation, follows from the disparity map and the geometry of the stereo system.

Uncertainty in depth estimates arises from errors or approximations in each step described above, however it is important to distinguish between these errors. Calibration, rectification and triangulation rely on geometric relations verified by the stereo system. Even if these are only approximations of the true geometric constraints, the errors associated with them can be controlled and their range computed. For instance, quantization errors, which come from the knowledge of the image points coordinates at fixed positions only, with a precision equal to half the sampling interval δ [20], depend only on the stereo system parameters. Possibly, the precision can be improved by modifying the stereo system: for example, an increase of the product bf enables to obtain more accurate depth estimates [20].

In opposition, matching errors are strongly dependent on the nature of the radiometric information, and cannot be entirely controlled by the design of the stereo system. The reason for this lies in the implicit matching assumption that pairs of homologous points can be identified. In many situations, violations of this assumption are encountered: the geometric deformations induced by the projections may be strong, the images are corrupted by sensor noise or the radiometric information itself may be ambiguous (existence of repetitive features, lack of texture in the images,...). As a result, matching errors are likely to be dependent on the specific features in the images and to vary strongly over the disparity map.

1.2 A Stochastic Approach

A convenient framework to handle uncertainty in disparity computation is to consider the disparity D and the stereoscopic pair $Y = (I_1, I_2)$, as random

functions. This approach has been successfully applied in a Bayesian context to compute disparity estimates [1].
However, to our knowledge, the problem of uncertainty assessment has been only partly handled, mostly through the computation of standard deviation [23,16]. Standard deviation estimates give only a rough idea of the range of possible errors, and are not appropriate to describe large errors, which appear with very low probability and are spatially correlated. Therefore, specific methods have to be proposed. Typically, if we consider an object in motion in a 3D environment, one must find the path to a given target which minimizes the risk of collisions. Such problems require the computation of the whole conditional probability distribution $\pi_D(d|Y = y)$ of the disparity D *given* the stereoscopic pair Y. Note that once $\pi_D(d|Y = y)$ is known, we can easily compute the corresponding probability distributions for the real world coordinates, using the geometric transformation of the stereo system. This motivates the choice of the disparity as variable of interest.
The main contribution of this paper is to propose a theoretically founded approach to the problem of uncertainty assessment in disparity computation. Especially, we show in Sect. 2 that this problem amounts to sample from the posterior distribution π_D and to use Monte Carlo integrations. We also explicit the model π_D used for computation, but point out that this one is very dependent on the application. We give insight on how this model can be adapted to specific applications. On the opposite, the sampling algorithms we propose in Sect. 3 are quite general and can be used for different forms of probability models π_D. The only restriction is a gaussian assumption for the prior model. Lastly, in Sect. 4, we illustrate our methods with an application to the study of a stereoscopic pair of SPOT images.

In the sequel, we will assume that an estimate $\hat{d}$ of the disparity has been already computed, for example with a correlation based algorithm. This means that the algorithms we propose in the sequel are not used for disparity computation, but only for uncertainty assessment. One reason is that Monte Carlo algorithms are computationally cumbersome. Another reason is that we are interested in the uncertainty which is intrinsic to the radiometric information. Indeed, we consider the uncertainty estimation mostly at a postprocessing stage, and would like to propose a method that is widely independent of the disparity computation itself.

2 Stochastic Framework: Stereovision as an Inverse Problem

Let $Y = (I_1, I_2)$ be a stereoscopic pair in the epipolar geometry. We consider a physical point M, and its projections m_1 with coordinates (u_1, v_1) in I_1 and m_2, with coordinates (u_2, v_2) in I_2. The epipolar constraint imposes $v_1 = v_2$ and we can define the disparity in the reference system of I_1 as the difference:

$$d(u_1, v_1) = u_2 - u_1 \tag{1}$$

Therefore, a pair of homologous points is fully defined by the coordinates (u, v) of m_1 in I_1 and the disparity $d(u, v)$. Note that a symmetric definition of disparity in the cyclopean image is possible, see [1]. Although we proceed with definition 1, the approach we propose is readily applicable to this alternative definition. This purely geometric presentation however corresponds to the direct problem, where the coordinates of the homologous points are known. In practice, these are to be found, and generally a loss function $C_{u,v}(d)$, which expresses the similarity of the images in the neighborhood of $I_1(u, v)$ and $I_2(u + d, v)$, is used [9]. The estimate $\hat{d}$ of the "true" disparity is then taken to be the value which maximizes $C_{u,v}(d)$, possibly under some constraints, i.e.:

$$\hat{d}(u, v) = \arg\max_{d} C_{u,v}(d) \tag{2}$$

Therefore, when computing the disparity, and whatever method is used, one must move from a purely geometric definition 1 to an optimization criterion 2. In other words, disparity computation in stereovision is an inverse problem, and therefore, the result is only an *estimate* of the underlying disparity. This estimate is sensible to the amount of noise in the stereoscopic pair, but, even if perfect signal is assumed, uncertainty remains, because of the ambiguity of the information, due to occlusions, repetitive features, lack of texture in the images,... The Bayesian framework [3] is well adapted to inverse problems. Bayesian inference has been widely used in image analysis [2], and efficient algorithms have been proposed to solve the restoration problem, as in [13]. More recently, this formalism has been applied to the stereo-reconstitution problem [23,1]. Let $\pi_D(d|Y = y)$ be the posterior (or conditional) probability of the disparity D given the stereoscopic pair $Y = y$. Using Bayes relation, this one can be expressed by means of the conditional distribution of Y given $D = d$, $\pi_Y(y|D = d)$, and the marginal distributions $\pi_D(d)$ and $\pi_Y(y)$. Since here Y is known and remains equal to y, $\pi_Y(y)$ is a constant, and we can introduce the likelihood $\mathcal{L}(y|D = d) = \frac{\pi_Y(y|D=d)}{\pi_Y(y)}$, which can be computed up to a multiplicative constant, and for clarity, we denote $g(d) = \pi_D(d)$ the prior distribution. With these notations, the posterior distribution writes:

$$\pi_D(d|Y = y) = \mathcal{L}(y|D = d)g(d) \tag{3}$$

In stereovision, this formalism has been mostly applied to compute an estimate of the disparity, through maximum *a posteriori* (MAP) estimation [1], where the maximization of $\pi_D(d|Y = y)$ is used as criterion. However, here, we are interested in the computation of the whole posterior distribution $\pi_D(d|Y = y)$. Especially, for any π_D-measurable function h, we would like to compute the integral:

$$\mathbf{E}_{\pi_D}(h) = \int h(z)\pi_D(z|Y = y)dz \tag{4}$$

With an appropriate choice of h, every probability concerning D can be computed. For example, with $h(d) = 1_{|d-\hat{d}|\geq s}$ where $\hat{d}$ is an estimate of the disparity,

we obtain the probability that the estimation error $D-\hat{d}$ is greater than a threshold s:

$$P(|D-\hat{d}| \geq s) = \int_{|z-\hat{d}|\geq s} \pi_D(z|Y=y)dz \tag{5}$$

In the following, we explicit the computation of 4 and derive an expression for π_D.

2.1 MCMC Methods

In practice, the numerical computation of 4 is not possible since π_D is a function of n variables, where n is the size of the image. Another reason is that π_D is known only up to a multiplicative factor. A solution is provided by Monte Carlo integration which evaluates the integral in 4 by drawing samples $\{z_i, i = 1, ..., m\}$ from π_D and then approximating

$$\mathbf{E}_{\pi_D}(h) \approx \frac{1}{m}\sum_{i=1}^{m} h(z_i) \tag{6}$$

The laws of large numbers ensure that the approximation can be made as accurate as desired by increasing the sample size m.
Monte Carlo methods were used in similar problems [11,7], but mostly to compute *map* estimates. We show in this paper that more information can be extracted from the posterior distribution.

A method to draw samples from π_D consists of building a Markov chain having π_D as its stationary distribution. In practice, the problem amounts to determine a transition probability kernel which ensures the convergence to π_D. This is the principle of the well known Metropolis-Hastings algorithm [15] or the popular Gibbs sampler [13]. For a detailed introduction to Markov chain theory, the interested reader is referred to [18] and [24]. Practical applications of Markov chain Monte Carlo can be found in [14].

As a conclusion, assessment of disparity uncertainty and disparity estimation are two very different problems, and the former requires the development of specific methods. Uncertainty can be assessed under a Bayesian framework by drawing samples from the posterior distribution of the disparity given the stereoscopic pair. Two points remain to be solved: the specification of the model, which is treated in Sect. 2.2, and the choice of a sampling algorithm, which is the subject of Sect. 3.

2.2 Model Derivation

We need to specify the likelihood term $\mathcal{L}(y|D=d)$, which describes the relation between the images of the pair at fixed disparity, and the prior model g, which reflects the spatial structure of the disparity. Let us point out that the choice of the probability model π_D is specific to the application, i.e. to the type of images recorded and the nature of the scene observed. In our case, we use

stereoscopic pairs of SPOT images for the computation of mid-resolution digital terrain models (with a ground resolution approximately equal to 10 m). We first derive an expression for the image model (likelihood) and then for the disparity model (prior).

A common modeling for the likelihood [1] consists of assuming that the recorded intensity I for a specific feature is the same for both images I_1 and I_2 and sensor noise is additive. Therefore, for fixed disparity d, the relation between I_1 and I_2 writes:

$$I_1(u,v) - I_2(u + d(u,v), v) = \eta(u,v) \tag{7}$$

where $\eta(u,v)$ can be interpreted as the residual between the intensities after a correction on the coordinates with the disparity.
The choice of the image model amounts hence to the modeling of η, and especially its spatial correlation. If we assume the residual η to be uncorrelated gaussian noise with mean μ and standard deviation κ, as in [4,6,1], we end up with the following expression of the likelihood:

$$\mathcal{L}(y|D=d) \propto \exp\left(-\sum_{i,j} \frac{(\eta_{ij}-\mu)^2}{2\kappa^2}\right) \tag{8}$$

The assumption of uncorrelated noise η may be questioned. Especially, one may argue that if the deformations induced by the stereo system are relatively important, the residual η will show strong patterns. Therefore, this point is very dependent on the setting of the stereo system and the nature of the scene. For satellite images at mid-resolution, these deformations may be neglected in a first approach, and the assumption of uncorrelated gaussian noise is reasonable. However, this is definitely a challenging aspect of the problem.
Note that, as we have an estimate $\hat{d}$ of the disparity, we can compute from I_1 and I_2 an estimate of the residual $\hat{\eta}(u,v) = I_1(u,v) - I_2(u + \hat{d}(u,v), v)$. Hence, the model parameters μ and κ can be estimated by maximum likelihood.

For the disparity model, we adopt a gaussian random function framework, which is relatively common [23,4,1]. Under the gaussian assumption, the prior distribution is entirely specified by the mean m and the spatial covariance C. These parameters can be estimated from the disparity map $\hat{d}$, for example under an assumption of spatial stationarity [5]. Finally, the prior distribution g takes the following form:

$$g(d) \propto \exp\left(-\frac{1}{2}(d-m)^\top \mathrm{C}^{-1}(d-m)\right) \tag{9}$$

Here again, let us point out that the choice of the prior model is dependent on the application. Especially, the behavior of the covariance at the origin will reflect the smoothness of the disparity: for example, a model with a linear behavior (such as Brownian motion [1]) is liable to reproduce discontinuities. Our experience is that covariances which are differentiable at the origin (such as the

cubic covariance) are well adapted in the case of mid-resolution stereo systems. However, for high-resolution applications, such as urban or indoor scenes, other models may be chosen. For example, in [1], Belhumeur proposes to model directly the discontinuities at the edges of objects through Poisson processes. More generally, one may argue that for such applications, stochastic geometric models, which directly take into account the form of objects, may be better adapted. Such models were applied to building detection [12] and road extraction [22].

Let us point out that, although we made the gaussian assumption for the prior distribution g, the posterior distribution is not necessarily gaussian. The nature of the posterior distribution depends largely on the likelihood term $\mathcal{L}$, which is not a quadratic form of d. For example, the presence of repetitive features in the stereoscopic pair would result in a multimodal posterior distribution.

3 Sampling Algorithm

In this section we expose the details of the sampling algorithm. We recall that the probability distribution to be sampled from is of the form:

$$\pi(z|Y = y) = \mathcal{L}(y|Z = z)g(z) \tag{10}$$

where $\mathcal{L}(y|Z = z)$ is the likelihood function, which may be known up to multiplicative factor, and $g(z)$ is the prior distribution.

3.1 Markov Chain Sampling Algorithms

As stated previously, iterative Markov chain algorithms have to be used. In our case, the difficulty lies in the form of the likelihood term $\mathcal{L}(Y|D = d)$, which cannot be approximated efficiently and is expected to vary strongly with the coordinates (u, v). The Gibbs sampler, which in combination with Markov random fields models is very popular in image analysis, relies on the use of conditional densities of the form $\pi(z_{\mathcal{S}}|z_{-\mathcal{S}})$, where $\mathcal{S}$ is a subspace of the grid to be simulated and $-\mathcal{S}$ its complementary part [3]. Unfortunately, because of the expression of the distribution π_D, it is not possible to sample directly from the conditional distributions. Other alternatives are Metropolis-Hastings algorithms [14] and variants such as discretized Langevin diffusions [19].
The principle of the Metropolis-Hastings algorithm is to generate a transition from the current state z to the state z' from a proposal transition kernel $q(z, z')$, and to accept this transition with probability $a(z, z') = \min\left(1, \frac{\pi(z')q(z',z)}{\pi(z)q(z,z')}\right)$. The efficiency of this algorithm relies merely on the proposal transition kernel q, but when the target distribution π cannot be easily approximated, optimal choices of q are not possible, and very often an *ad hoc* transition kernel is used, like in the independence sampler or the random walk [14].

3.2 A New Sampling Algorithm

These difficulties have led us to propose a new sampling algorithm which makes use of the target distribution π to generate the transitions. Let us assume that the prior distribution $g(z)$ denotes now a gaussian distribution.
The algorithm is based on the following remark. Let W_1 and W_2 be two independent gaussian processes with zero mean and covariance C (which we denote $W_1, W_2 \sim g$). Then, for any $\theta \in [0, 2\pi[$, the linear combination $W(\theta) = W_1 \cos\theta + W_2 \sin\theta$ is also a gaussian process with zero mean and covariance C. We propose to use this relation, together with an appropriate choice of θ, to build the transitions of a Markov chain $Z = (Z_k, k \geq 0)$. The algorithm is the following:

1. sample $w \sim g$,
2. compute the proposal transitions:

$$z'_i = z_k \cos\left(i\frac{2\pi}{m}\right) + w \sin\left(i\frac{2\pi}{m}\right), \quad \forall i = 0, m-1 \tag{11}$$

   and the corresponding transition probabilities:

$$a(z, z'_i) = \frac{\mathcal{L}(y|Z = z'_i)}{\sum_{j=0}^{m-1} \mathcal{L}(y|Z = z'_j)}, \quad \forall i = 0, m-1 \tag{12}$$

3. choose $z'_i \sim a(z, z'_i)$ and set $z_{k+1} = z'_i$.

We prove in [21] that the Markov chain built according to this algorithm has π as stationary distribution.
A comparison study of this algorithm with the independence sampler and the Langevin diffusion can be found in [21]. The interesting feature of this algorithm is that at each iteration, a whole path of possible transitions is proposed, in opposition to a single one in the Metropolis-Hastings algorithm, and this is done at a low computational cost since it only requires the computation of linear combinations of the form 11. Therefore, transitions are likely to occur more often, which improves mixing within the chain and speeds up the convergence. An illustration of this feature is displayed on Fig. 1.

3.3 Size Reduction and Importance Sampling

For large grids, the computational burden of Markov chain algorithms can become very cumbersome. Gaussian samples can be generated very quickly (with algorithms such as those relying on the Fast Fourier Transform [5]), but the transition rate of the Markov chain tends to decrease drastically with the grid size, so that the sampling algorithm must be run longer to achieve convergence. To cope with this difficulty, we propose to split the sampling in two consecutive phases. Consider a subgrid $\mathcal{S}$ of the total sampling grid $\mathcal{G}$ and its complementary $-\mathcal{S}$ in

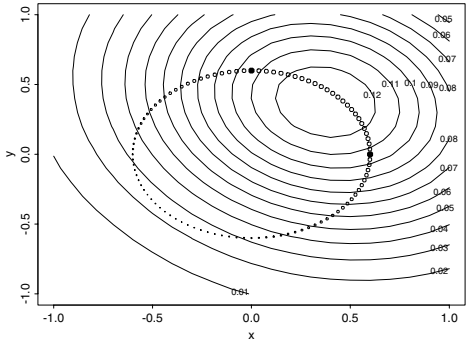


Fig. 1. Path of possible transitions generated from the current point and an independent simulation (black dots). The target distribution, displayed in the background with level lines, is a bigaussian distribution. The circles along the path, which correspond each to a different value of θ, are proportional to the transition probability $a(z, z_i')$.

$\mathcal{G}$. $z_{\mathcal{S}}$ denotes the restriction of the vector z to $\mathcal{S}$. The posterior distribution π can be decomposed as follows:

$$\pi(z|Y=y) = \frac{\mathcal{L}(y|Z=z)}{\mathcal{L}(y|Z_{\mathcal{S}}=z_{\mathcal{S}})} g(z_{-\mathcal{S}}|Z_{\mathcal{S}}=z_{\mathcal{S}}) \mathcal{L}(y|Z_{\mathcal{S}}=z_{\mathcal{S}}) g(z_{\mathcal{S}}) \qquad (13)$$

The term $\mathcal{L}(y|Z_{\mathcal{S}}=z_{\mathcal{S}})g(z_{\mathcal{S}})$ is the posterior distribution $\pi(z_{\mathcal{S}}|Y=y)$ on $\mathcal{S}$, whereas $g(z_{-\mathcal{S}}|Z_{\mathcal{S}}=z_{\mathcal{S}})$ is the conditional distribution of the gaussian vector $z_{-\mathcal{S}}$ given $z_{\mathcal{S}}$ (this is the conditional distribution which is used in the Gibbs sampler). Sampling from this conditional distribution can be done directly and very efficiently using an algorithm such as conditioning by kriging [5]. The first term $w = \frac{\mathcal{L}(y|Z=z)}{\mathcal{L}(y|Z_{\mathcal{S}}=z_{\mathcal{S}})}$ is a correction factor which shall be used to weight the samples in the Monte Carlo integration. We end up with the following algorithm:

1. run the Markov chain algorithm on the subgrid $\mathcal{S}$,
2. for each generated sample $z^i_{\mathcal{S}}$, generate $z^i_{-\mathcal{S}}$ using the conditional gaussian distribution $g(z_{-\mathcal{S}}|Z_{\mathcal{S}}=z_{\mathcal{S}})$, and construct $z_i = (z^i_{\mathcal{S}}, z^i_{-\mathcal{S}})$ the sample on the total grid $\mathcal{G}$,
3. compute the weight $w_i = \frac{\mathcal{L}(y|Z=z_i)}{\mathcal{L}(y|Z_{\mathcal{S}}=z^i_{\mathcal{S}})}$,
4. compute the integral

$$\mathbf{E}_\pi(h) \approx \frac{1}{\sum_{i=1}^m w_i} \sum_{i=1}^m w_i h(z_i) \ . \qquad (14)$$

The last step is an application of importance sampling theory [14]. One may run the risk that all weights w_i but a few become equal to 0, and only a few samples take the total weight. This could happen if the distribution used for sampling $\pi'(z|Y=y) = \pi(z_{\mathcal{S}}|Y=y)g(z_{-\mathcal{S}}|Z_{\mathcal{S}}=z_{\mathcal{S}})$ and the target distribution

$\pi(z|Y = y)$ are very different. In our case, this should be avoided due to the strong spatial correlations exhibited by the disparity D. The idea is that if we know the disparity on a subgrid $\mathcal{S}$, then we are able to compute accurate estimates of D on the total grid $\mathcal{G}$, and correct fluctuations can be reproduced by sampling from the conditional distribution $g(z_{-\mathcal{S}}|Z_{\mathcal{S}} = z_{\mathcal{S}})$. An overview of the sampling algorithm is shown on Fig. 2.

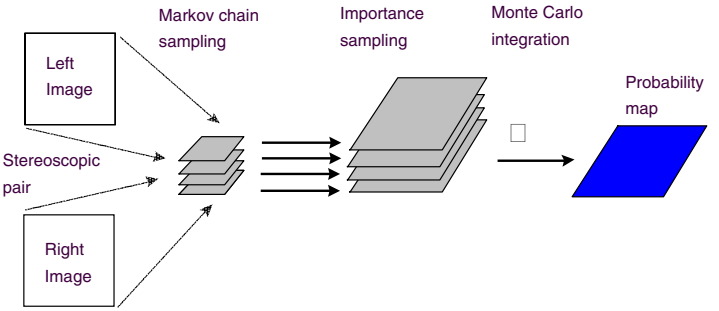
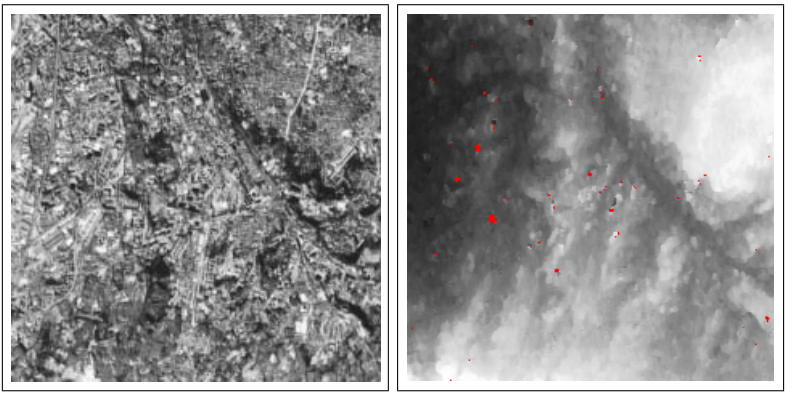


Fig. 2. Overview of the sampling algorithm. First, the Markov chain sampling algorithm is run on a sub-grid, conditional to the stereoscopic pair. Then, for each generated sample, the whole grid is generated. Lastly, the Monte Carlo integration is performed.

4 Experimental Results

We have applied the model of Sect. 2.2 and the sampling algorithms introduced in Sect. 3 to the study of a stereoscopic pair of rectified SPOT images of the Marseille area, of size 512×512 pixels (Fig. 3). A disparity map (Fig. 3) has been computed using a correlation based algorithm.

2000 samples have been generated. Using the importance sampling algorithm, the size of the sampling grid has been reduced to 64×64, yielding thus a size reduction factor equal to 64. To illustrate the efficiency of the sampling algorithms, we run a Metropolis-Hastings algorithm (denoted as MH) on the total grid, then the sampling algorithm of Sect. 3 (denoted TCP for transitions along continuous paths), first on the total grid, and then on the reduced grid with the importance sampling method (denoted further TCPIS). We compared the computation time required to generate 100 samples with at least 10 state transitions between two samples, assuming that these conditions ensure similar convergence properties. Computations were performed on a SUN Ultra 60 computer with a 400 Mhz ULTRA SPARC II processor. The results are shown on Tab. 1. The TCP algorithm allows to reduce the computation time by a factor 10 in comparison to the independence Metropolis-Hastings sampler. The size reduction coupled with



(a) One image of the stereoscopic pair

(b) Disparity

Fig. 3. One rectified image of the stereoscopic pair (Marseille area, France) and the corresponding disparity (right)

Table 1. Comparison of computation time for MH, TCP and TCPIS.

	MH	TCP	TCPIS
CPU time (sec.)	642571	55454	1326

the importance sampling algorithm (TCPIS) allows to reduce again the computation time, by a factor 40. The overall computation time is still relatively large (a few hours for our case study), but the TCPIS algorithm allows a tremendous gain in comparison with crude sampling algorithms.

The application we aim at is the detection of possible large errors; that is, given an error threshold s, which are the pixels for which disparity errors larger than s are likely to occur? This is typically the type of questions one may face in applications in the aircraft industry. This problem can be formulated as the computation of the probability $P(D-\hat{d} \geq s)$, and, for a given risk α, the location of the pixels (u, v) for which $P(D(u, v) - \hat{d}(u, v) \geq s) \leq \alpha$, which correspond to the safe area.

Figure 4 displays the probability maps obtained for $s = 2$ and $s = 3$. The probability of errors varies strongly with the location (u, v), that is why usual methods estimating error ranges over a large domain are not appropriate for such applications. It is then possible to threshold these maps to a given risk value, say 0.1%, to identify the safe areas.

In real applications however, it is not realistic to consider that such information is required for every pixel. For example, if one thinks of an object in motion over a domain, then one must guarantee that no collision occurs along

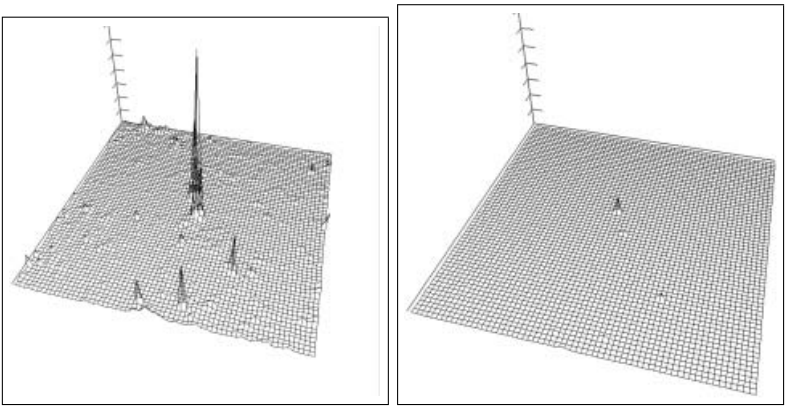


Fig. 4. Probability map of positive errors larger than 2 (left) and 3 (right) pixels. One scale unit on the vertical axis equals 10%.

its trajectory. Hence, we need to consider a whole domain, and compute the probability that no error greater than a given threshold s occurs in this domain, with a probability $1 - \alpha$. Since errors are spatially correlated, this is not equal to the product of the corresponding probabilities over the domain. We point out that this type of computation definitely requires the knowledge of the entire multivariate distribution π_D and not only the marginal distribution of a given pixel.

For this purpose, we divided the disparity map in 64 sub-domains of equal surface, and computed the probability that errors larger than 3 (respectively 4) pixels occur anywhere in the sub-domain (Fig. 5). Thus, as previously, it is possible to determine the safe areas as the sub-domains for which this probability is lower than $\alpha = 0.1\%$. The results obtained are rather pessimistic: the 3 pixel threshold map has only 2 safe domains, this contrasts with the situation we would have expected from the right map displayed on Fig. 4. This surprising result emphasizes the importance of spatial correlations and justifies the use of Monte Carlo methods, although they are computationally demanding.

We consider now the opposite situation, which is often encountered in practice: a risk value α is given, and one needs to compute a confidence interval $[Z_{\inf}, Z_{\sup}]$ for the variable Z such that $P(Z \in [Z_{\inf}, Z_{\sup}]) \geq 1 - \alpha$. An example of the use of confidence intervals in a different context of computer vision can be found in [17]. To derive a confidence interval, we propose to make use of the min and max of the disparity samples. Indeed, for $n + 1$ independent identically distributed random variables Z_i, we have the following result:

$$P(Z_{n+1} \in [\min_{i \leq n} Z_i, \max_{i \leq n} Z_i]) = \frac{n-1}{n+1} \tag{15}$$

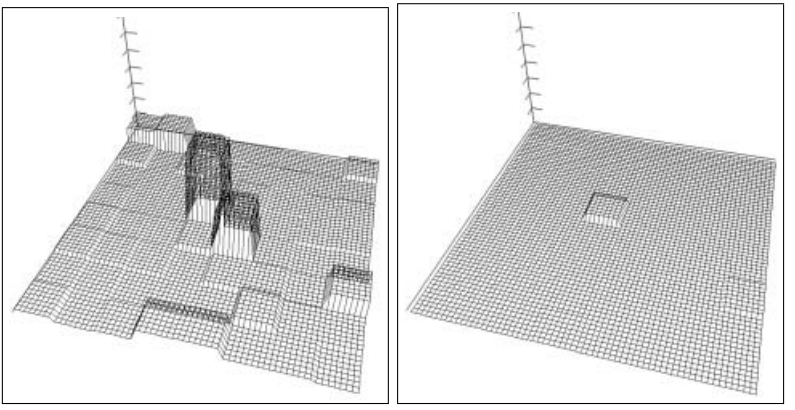


Fig. 5. Probability map of positive errors larger than 3 (left) and 4 (right) pixels in a subdomain. One scale unit on the vertical axis equals 10%.

Therefore, for any risk value α, we can choose the smallest integer n such that $\frac{n-1}{n+1} \geq 1-\alpha$ and compute for each pixel the min and the max of the n samples. The confidence interval has then the form of two embedding surfaces.
In order to validate our results, we have used a reference disparity map obtained from high resolution images, which has been degraded to the SPOT resolution. Using this reference map, it is possible to compute the percentage of points lying in the confidence interval computed previously and to compare this value to the theoretical value given by 15. The risk value was set equal to 0.1%, and therefore we used $n = 2000$ samples to construct the confidence interval. From the reference map, we found that 0.11% of the total points fell outside the confidence interval, which is in very good agreement with the theoretical value 0.1%. Moreover, the mean size over the grid of the computed disparity interval is 4.4 pixels: although the risk value is very small, the precision given by the confidence interval is reasonable. Therefore, the method proposed in this paper allows to compute accurate statistics.

5 Conclusion

In this paper, we have considered the problem of assessing the uncertainty of a disparity map with only the use of the stereoscopic pair as information. Since the usual standard deviation approach does not permit to handle the case of large errors which occur with very low probability and are spatially correlated, specific methods have to be proposed. A solution consists of computing the posterior distribution of the disparity given the stereoscopic pair. Hence, uncertainty assessment amounts to sample from this posterior distribution and to compute the desired probabilities using a Monte Carlo approach. From a

practical point of view, a model for the posterior distribution must be chosen, and an efficient sampling algorithm must be used to keep the computation time reasonable.
This method is very general and can be applied to different types of stereovision systems, but the choice of the stochastic model is specific to the application. The model explicited in this paper considers an application to the computation of mid-resolution digital terrain models, for which the disparity map can be assumed to vary relatively smoothly and the gaussian assumption is reasonable. For more complex applications, models which take explicitly into account the geometry of the scene may be more appropriate.
An important problem is the choice of the sampling algorithm. The algorithm we propose considers the case of gaussian priors, which is quite general. It is based on the construction of a continuous path of possible transitions, and compared to standard sampling algorithms such as the independence sampler proves very competitive. Moreover, we have shown that the computation can be tremendously speeded up by first running the Markov chain on a subgrid and then generating conditionally the samples on the whole grid.
This Markov chain Monte Carlo approach has been applied to the study of a stereoscopic pair of SPOT images. The results show that the disparity statistics are highly non-stationary. This is due to the nature of the radiometric information itself. These statistics can be used to determine the safe areas, for which large errors occur with probability inferior to a given risk value, or to compute precise confidence intervals. The range of applications is however brighter: for example, they may be used to optimize a functional over the domain under study. We think especially of the motion of an object in a three-dimensional environment, for which the path to a given target has to be optimized and collisions avoided.

References

[1] P.N. Belhumeur. A Bayesian approach to binocular stereopsis. *International Journal of Computer Vision*, 19(3):237–262, 1996.
[2] J. Besag. On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B*, 48:259–302, 1986.
[3] J. Besag, P. Green, D. Higdon, and K. Mengersen. Bayesian computation and stochastic systems. *Statistical Science*, 10(1):3–66, 1995.
[4] C. Chang and S. Chatterjee. Multiresolution stereo - A Bayesian approach. *IEEE International Conference on Pattern Recognition*, 1:908–912, 1990.
[5] J.P. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty.* Wiley and Sons, 1999.
[6] I.J. Cox. A maximum likelihood n-camera stereo algorithm. *International Conference on Computer Vision and Pattern Recognition*, pages 733–739, 1994.
[7] F. Dellaert, S. Seitz, S. Thrun, and C. Thorpe. Feature correspondance: A Markov chain Monte Carlo approach. In *Advances in Neural Information Processing Systems 13*, pages 852–858, 2000.

[8] F. Devernay and O. Faugeras. Automatic calibration and removal of distorsions from scenes of structured environments. In Leonid I. Rudin and Simon K. Bramble, editors, *Investigate and Trial Image Processing, Proc. SPIE*, volume 2567. SPIE, San Diego, CA, 1995.

[9] U.R. Dhond and J.K. Aggarwal. Structure from stereo - A review. *IEEE Transactions on systems, man and cybernetics*, 19(6):1489–1510, 1989.

[10] O. Faugeras. *Three-Dimensional Computer Vision*. Massachussets Institute of Technology, 1993.

[11] D. Forsyth, S. Ioffe, and J. Haddon. Bayesian structure from motion. In *International Conference on Computer Vision*, volume 1, pages 660–665, 1999.

[12] L. Garcin, X. Descombes, H. Le Men, and J. Zerubia. Building detection by markov object processes. In *Proceedings of ICIP'01, Thessalonik, Greece*, 2001.

[13] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[14] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1996.

[15] W.K. Hastings. Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[16] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9), september 1994.

[17] R. Mandelbaum, G. Kamberova, and M. Mintz. Stereo depth estimation: a confidence interval approach. In *International Conference on Computer Vision*, pages 503–509, 1998.

[18] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag London, 1993.

[19] G.O. Roberts and J.S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. Roy. Statist. Soc. Ser. B*, 60:255–268, 1998.

[20] J.J. Rodriguez and J.K. Aggarwal. Stochastic analysis of stereo quantization error. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):467–470, 1990.

[21] J. Sénégas, M. Schmitt, and P. Nonin. Conditional simulations applied to uncertainty assessment in DTMs. In G. Foody and P. Atkinson, editors, *Uncertainty in Remote Sensing and GIS*. Wiley and Sons, 2002. to appear.

[22] R. Stoica, X. Descombes, and J. Zerubia. Road extraction in remote sensed images using a stochastic geometry framework. In *Proceedings of MaxEnt'00, Gif Sur Yvette, July 8-13*, 2000.

[23] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-level Vision*. Kluwer Academic Publishers, 1989.

[24] R.L. Tweedie. Markov chains: Structure and applications. In B. N. Shanbhag and C.R. Rao, editors, *Handbook of Statistics 19*, pages 817–851. Elsevier Amsterdam, 1998.