Monte Carlo methods in stereovision

Julien Sénegas
Ecole des Mines de Paris
35 rue Saint Honoré
77305 Fontainebleau
France

Michel Schmitt
Ecole des Mines de Paris
35 rue Saint Honoré
77305 Fontainebleau
France

Philippe Nonin
Istar
2600 route des Crêtes
06905 Sophia-Antipolis
France

Abstract

The goal of this paper is to provide statistical tools for assessing the uncertainty of a disparity map computed from a stereoscopic pair. Under a Bayesian framework, it is possible to compute the posterior distribution of the disparity map given the stereoscopic pair. The analysis of uncertainty consists then of sampling the entire posterior distribution, which can be achieved through MCMC simulations. This method allows the prediction of large errors which occur with low probability.

The main contribution of this paper is the proposition of a new sampling algorithm for Gaussian priors, and the use of importance sampling to speed up the computation. The performance of this algorithm is demonstrated and an application to a stereoscopic pair of SPOT images is given.

1 Introduction

Bayesian inference has been widely used in image analysis [2, 3], and efficient algorithms have been proposed to solve the estimation problem, as in [8]. More recently, this formalism has been applied to the stereo-reconstruction problem [15, 1], where not a characteristic of the image itself has to be estimated, but an underlying variable deduced from the stereoscopic pair: the disparity. The main difficulty lies in the nature of the information contained in the initial images. Indeed, the signal is often perturbed by noise, but the correspondence problem is also altered by the presence of repetitive features, the lack of texture, the existence of occlusions or diachronism. In other words, the information is often ambiguous and does not permit to determine univocally the underlying disparity. As a consequence, the errors which occur in the disparity maps computed by maximum likelihood estimation or correlation, which are often auto-correlated and dependent on the initial information, cannot be automatically detected.

Especially, we consider that the only information available is the stereoscopic pair, and assume that a disparity map \(d\) has been already computed, with a correlation based algorithm for example. Thus, we consider the uncertainty analysis at a post-processing stage. So far, this problem has been mostly treated through the computation of standard deviations [15, 11], but this approach is not appropriate to describe large errors which occur with low probability. Typically, if one considers an object in motion in a 3D environment, one must find the path to a given target which minimizes the risk of collisions. Therefore, specific methods need to be proposed.

From a statistical point of view, a solution to this problem amounts to compute the posterior distribution \(\pi\) of the disparity given the stereoscopic pair.

\[ \pi(d|y) = \mathcal{L}(y|d)g(d) \]

where \(d\) denotes the disparity and \(y\) the stereoscopic pair. In this perspective, the stereoscopic pair is rather considered as a constraint to be fulfilled than as hard information. This constraint is expressed in the likelihood function, which attributes a cost to a given disparity map. The prior distribution \(g\) describes the spatial structure of disparity, and especially its regularity.

From a practical point of view, a solution is to sample the posterior distribution \(\pi\) and use a Monte Carlo algorithm, since analytical evaluation of \(\pi\) is not feasible. That is, the computation of any expectation of a \(\pi\)-measurable function \(f\) is replaced by the following sum:

\[ \mathbb{E}_\pi f \approx \frac{1}{n} \sum_{i \leq n} f(d_i) \]

where \(d_i, \ i \leq n\) is a sampling of \(\pi(d|y)\).

Markov chain algorithms provide a convenient framework to sample from posterior distributions. The principle is to generate a chain of consecutive states
which converges to the target distribution. The evolution of the chain is driven by a transition kernel, which describes the transition between two consecutive states. See [9] or [3] for a nice introduction to MCMC methods, with applications in image analysis.

However, MCMC methods can be quite computationally expensive, and there is a need for algorithms which keep the integration time reasonable. The computational burden becomes even more acute when dealing with large image files. The Markov chain algorithm introduced in section 3, together with the importance sampling methods, enables to handle such problems.

This paper is organized as follows. In section 2, we briefly introduce the bayesian model used to tackle the stereovision problem. The details of the simulation algorithm are described in section 3. Section 4 presents experimental results obtained with a stereoscopic pair of SPOT images.

2 A Bayesian model

In this section, we derive an expression for the distribution $\pi$. Let us point out that the choice of the stochastic model is dependent on the type of images and the nature of the scene. Here, we consider the case of mid-resolution stereo systems, but give insight on how the model can be extended to other applications.

Let $(I_l, I_r)$ be the stereoscopic pair in the epipolar geometry [7]. We consider a physical point M and its projections $m_l$ and $m_r$, with coordinates $(u_l, v_l)$ and $(u_r, v_r)$ in $I_l$ and $I_r$. The epipolar constraint imposes $v_l = v_r$, and we define disparity $d$ as:

$$d = u_r - u_l$$ (3)

A symmetric definition of disparity in the cyclopean image is possible, see [1]. We proceed with definition 3, but our approach is readily applicable to this alternative definition.

If we denote $I$ the initial intensity, $H$ the camera transformation, $\phi$ the coding transformation and $\eta$ an observation noise, and assume the recorded intensity is the same for the two cameras (lambertian assumption), the recorded intensity $I_l$ and $I_r$ write [8, 1]:

$$I_l(u_l, v_l) = \phi(I \star H)(u_l, v_l) + \eta_l(u_l, v_l)$$ (4)

$$I_r(u_r, v_r) = \phi(I \star H) + \eta_r(u_r, v_r)$$ (5)

Hence, using the definition of disparity 3, the residual $I_l(u, v) - I_r(u + d, v)$ becomes equal to a noise term

$$\eta(u, v) = \eta_l(u, v) - \eta_r(u + d, v).$$

The choice of the image model amounts therefore to the modeling of $\eta$. If we assume the image noise to be uncorrelated gaussian noise [4, 6, 1], the likelihood term $\mathcal{L}(I_l, I_r|d)$ takes the following expression:

$$\mathcal{L}(I_l, I_r|d) \propto \exp\left(-\sum_{i,j} (\eta_{ij} - \mu)^2 / 2\kappa^2\right)$$ (6)

Note that we can use the disparity estimate $\hat{d}$ to compute the residual $\eta$ and estimate the mean $\mu$ and standard deviation $\kappa$ by maximum likelihood. The assumption of uncorrelated gaussian noise may be questioned; especially, if the deformations induced by the stereo systems are important, the residual $\eta$ may exhibit strong patterns. For satellite images at mid-resolution, these can be neglected, but this is definitely a challenging aspect which may be taken into account in other applications.

For the prior distribution, we adopt a gaussian random function framework [15, 1]. This model, through the choice of the covariance, is well adapted to the case of mid-resolution stereo systems, where the disparity varies relatively smoothly. For high-resolution applications, such as urban or indoor scenes, models which directly take into account the form of objects may be chosen: for example, in [1], Belliveau proposes to model discontinuities at the edges through Poisson processes, or in [14] stochastic geometric models are used to extract road networks.

3 Sampling algorithm

The form of the likelihood, which exhibits a non-linear relation between $(I_l, I_r)$ and $d$, does not permit for sampling with the Gibbs sampler [8]. Moreover, when the posterior distribution cannot be easily approximated to compute a convenient proposal transition kernel, the Metropolis-Hastings algorithm [10, 3] can prove very inefficient [9]. A solution is to make directly use of the form of the conditional distribution to generate the transitions.

Let us consider the sampling problem in the case where the prior distribution $\eta$ is gaussian. Let $Z$ and $W$ be two independent gaussian processes with zero mean and covariance $C$. For each $\theta \in [0, 2\pi]$, the linear combination $Z(\theta) = Z \cos \theta + W \sin \theta$ is also a gaussian process with zero mean and covariance $C$.

The idea of the following algorithm is to make use of this relation, together with an appropriate choice of $\theta$ to generate the transitions of a Markov chain. We have the following property:

**Property 3.1** Let $Z = (Z_k, k \geq 0)$ be a Markov
chain whose transitions are built according to:

\[
W \sim g \quad (7)
\]

\[
\pi(\theta|z, w) = \frac{\mathcal{L}(y|z \cos \theta + w \sin \theta)1_{[0,2\pi]}(\theta)}{\int_0^{2\pi} \mathcal{L}(y|z \cos \theta + w \sin \theta) d\alpha} \quad (8)
\]

\[
\theta|z, w \sim \pi(\theta|z, w) \quad (9)
\]

\[
\tilde{z} = z \cos \theta + w \sin \theta \quad (10)
\]

Then \( Z \) has \( \pi \) as stationary distribution.

A proof of this result can be found in [13]. Note that in practice the computation of \( \pi(\theta|z, w) \) can be done only for a finite number of values \( \theta_i \). However, if a regular discretization of \([0, 2\pi]\) is chosen, the previous property still holds.

The interesting feature of this algorithm is the construction at each iteration of a continuous path of possible transitions, which increases the transition rate and mixing within the chain. Moreover, the computation of this continuous path, based on linear combinations, can be performed at a low cost, regardless of the dimension of the system.

However, Markov chain computations can be quite expensive when applied to large systems. The solution proposed consists in splitting the sampling in two consecutive phases. This idea is motivated by the following decomposition of the posterior distribution, where \( S \) is a sub-grid of the total grid \( G \), and \(-S\) is its complementary in \( G \):

\[
\pi(d|y) = \frac{\mathcal{L}(y|d)}{\mathcal{L}(y|d_S)} g(d_{-S}|d_S) \mathcal{L}(y|d_S) g(d_S) \quad (11)
\]

The term \( \mathcal{L}(y|d_S) g(d_S) \) is the posterior distribution on \( S \), whereas \( g(d_{-S}|d_S) \) can be interpreted as the density of the gaussian vector \( d_{-S} \) given \( d_S \). Sampling from it can be done directly and very efficiently using an algorithm such as conditioning by kriging [5]. The first term \( w = \frac{\mathcal{L}(y|d)}{\mathcal{L}(y|d_S)} \) is a correction factor which shall be used to weight the samples in the Monte Carlo computations.

Therefore, we end up with the following algorithm. First, we run the Markov chain algorithm on the sub-grid \( S \), then, for each simulated state, we generate the complementary grid \(-S\) by direct simulation. Finally, according to importance sampling theory [9], the Monte Carlo computation in equation 2 becomes:

\[
E_{\pi} f \approx \sum_{i \leq n} w_i f(d_i) \quad (12)
\]

4 Experimental results

We have applied the previous model and sampling algorithms to the study of a stereoscopic pair of SPOT images of the Marseille area, of size 512 x 512 pixels.
Figure 2: Probability over a domain of positive errors larger than 3 pixels - Scale graduating is 10%.

In a previous step, a disparity map \( \hat{d} \) has been computed with a correlation based algorithm, and we would like to estimate the probability of large errors. A total of \( 10^5 \) iterations have been performed. Using the importance sampling algorithm, the size of the sampling grid has been reduced to \( 64 \times 64 \).

To illustrate the efficiency of our sampling algorithm, we compared the computation times required to generate 100 samples with at least 10 state transitions between two samples, assuming that these conditions ensure similar convergence properties. Three algorithms are compared: the independence Metropolis-Hastings sampler (denoted MH) [9], our algorithm run on the total grid (denoted TCP for “transitions along continuous paths”) and the same algorithm, run on the sub-grid (\( 64 \times 64 \)) and followed by the importance sampling step to generate the whole grid (denoted TCPIS). The results, obtained with a SUN Ultra 60 computer, are shown on Table 1. The independence sampler uses as transition kernel the prior distribution \( g \). Hence, the transition rate is very low, and the chain must be run much longer to ensure a minimum number of transitions between two states. In comparison, TCP, which additionally uses the expression of the likelihood to generate the transitions, allows to divide the computation time by a factor 10. It becomes really efficient when it is coupled with the size reduction step and the importance sampling (TCPIS): the reduction factor in computation time is larger than 500.

<table>
<thead>
<tr>
<th>CPU time (sec.)</th>
<th>MH</th>
<th>TCP</th>
<th>TCPIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>642571</td>
<td>55454</td>
<td>1326</td>
<td></td>
</tr>
</tbody>
</table>

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

We address now the following practical problem: consider an object in motion in a 3D environment, for which one must guarantee that no collision occurs along its trajectory. Hence, we need to determine the critical areas, where the probability that a disparity error greater than a given threshold \( s \) occurs is significant. Note that in this type of problems marginal probabilities, computed for each pixel, are not appropriate, since errors are spatially correlated and the probability of an error anywhere in a domain is not equal to the product of the marginal probabilities. The spatial aspect is hence crucial. This justifies the use of computing demanding Monte Carlo methods, which alone permit to compute the entire (multidimensional) distribution. As an illustration, we divided the disparity map in \( 64 \) sub-domains and computed the probability that errors larger than \( 3 \) pixels occur anywhere in each sub-domain. The probability map displayed on figure 2 can then be used to determine the safe areas, where the probability of errors are below a given risk value.

In order to validate the model and the method, we constructed a test based on the computation of confidence intervals (this approach is followed in a different context in [12]). In this situation, the risk value \( \alpha \) is fixed, and one needs to determine the values \( d_{inf} \) and \( d_{sup} \) for which \( P(d \in [d_{inf}, d_{sup}]) \geq 1 - \alpha \). One can easily show that, given \( n \) independent samples \( d_i \), \( d_{inf} = \min_i d_i \) and \( d_{sup} = \max_i d_i \) provide a (non-centered) confidence interval with \( \alpha = \frac{2}{n+1} \).

The computed interval has the form of two embedding surfaces. Hence, to test the reliability of the model, we used a reference disparity map. If the stochastic model is well adapted, the percentage of reference disparity values falling outside the confidence interval should be approximately equal to \( \alpha \). We used 2000 samples to construct the interval based on the maximum and the minimum, and found that 0.11% of the reference disparity values fell outside this interval, which is in very good agreement with the theoretical value \( 2 \times 10^{-3} \approx 0.1\% \).

5 Conclusion

In this paper, we have considered the problem of assessing the uncertainty of a disparity map as a sampling problem from a posterior distribution under a bayesian framework. This requires the setting of a stochastic model which describes the spatial structure of the disparity (prior model) and the properties of the stereoscopic pair (likelihood model). The computation problem can be solved using Monte Carlo sampling, but requires efficient sampling algorithms which can reduce the computational burden. Our sampling algorithm, which relies on Markov chain theory and assumes gaussian priors, is based on the
construction of a continuous path of possible transitions. Moreover, we have shown that using importance sampling results, it is possible to split the sampling problem in two steps and to run the Markov chain algorithm on a reduced size system, which results in a largely reduced computation time.

The bayesian model used in this first approach assumes gaussian additive white noise for the images, which may not be quite realistic. The next step will be to take into account a possible spatial structure. We also think that the choice of the spatial structure of the prior model is an important issue. These remarks draw the line for future work.

References


