ROBUST DISPARITY MAPS WITH UNCERTAINTIES FOR 3D SURFACE RECONSTRUCTION OR GROUND MOTION INFERENCE

A. Jalobeanu[1] **and D. D. Fitzenz**[2]

[1] LSIIT (CNRS - Univ. Strasbourg 1), Illkirch, France
(e-mail: jalobeanu@lsiit.u-strasbg.fr, web: lsiit-miv.u-strasbg.fr/paseo/jalobeanu.php)
[2] EOST (CNRS - Univ. Strasbourg 1), Strasbourg, France
(e-mail: fitzenz@eost.u-strasbg.fr, web: eost.u-strasbg.fr/dfitzenz)

Commission III/2

KEY WORDS: Optical sensors, Remote sensing, Stereo, Uncertainties, Error maps, Change modeling, Digital terrain models

ABSTRACT:

Disparity maps estimated using computer vision-derived algorithms usually lack quantitative error estimates. This can be a major issue when the result is used to measure reliable physical parameters, such as topography for instance. Thus, we developed a new method to infer the dense disparity map from two images. We use a probabilistic approach in order to compute uncertainties as well. Within this framework, parameters are described in terms of random variables. We start by defining a generative model for both raw observed images given all model variables, including disparities. The forward model mainly consists of warping the scene using B-Splines and adding a radiometric change map. Then we use Bayesian inference to invert and recover the a posteriori probability density function (pdf) of the disparity map. The main contributions are: The design of an efficient fractal model to take into account radiometric changes between images; A multigrid processing so as to speed up the optimization process; The use of raw data instead of orthorectified imagery; Efficient approximation schemes to integrate out unwanted parameters and compute uncertainties on the result. Three applications could benefit from this disparity inference method: DEM generation from a stereo pair (along or across track), automatic calibration of pushbroom cameras, and ground deformation estimation from two images at different dates.

1 INTRODUCTION

Computational stereo vision aims at matching pixels from a stereo image pair; a great number of methods have already been developed – see (Brown et al., 2003) for a review. A few calculate dense disparity maps. The majority work along scanlines, assuming images have been rectified beforehand using epipolar geometry to restrict the search space to a single dimension (e.g. horizontal displacements). However, we claim that resampled images are not suitable for a proper probabilistic inference. We advocate the processing of raw data instead, because resampling destroys the independence properties of the noise by creating correlations, whereas only original pixels can be assumed independent variables. Therefore we have to estimate disparities in 2D while most advanced techniques only work in 1D; this is the only way of preserving the integrity of the data and deriving non-biased estimators. Experiments have shown that resampling via classical interpolation (whether bilinear or bicubic) induces systematic errors that can not be neglected when aiming at a sub-pixel accuracy. Working with raw data shall then result in increased accuracy.

The new approach provides a quantitative measure of uncertainty while most methods only compute ad-hoc matching or correlation quality measures. Disparity maps with uncertainties have been computed (Blake et al., 2003), however this was made possible by working in 1D. When it comes to robustness to illumination changes, probabilistic models have been proposed (Zhang et al., 2006) that rely on an illumination ratio map; this was also possible by using 1D disparities. Nonparametric approaches based on mutual information (Hirschmuller, 2005) were developed to handle such changes without having to explicitly model them; however it is unclear how they can be extended to account for spatially adaptive changes due to the terrain reflectance without resorting to a tremendous number of parameters to store the required histograms. In our case, robustness to illumination is achieved via probabilistic change modeling in 2D and does not require extra parameters. Techniques based on correlations, robust or not, usually rely on the choice of a window size which is often arbitrary, and those involving smoothness priors are often fine tuned manually; our technique is designed to be entirely automated as neither window size nor smoothness parameters need to be adjusted.

2 THE FORWARD MODEL

2.1 Deterministic part: rendering

We define the sets of bidimensional indices related to displacement parameters and pixels respectively as $\Omega_\theta = \{1...n_x\} \times \{1...n_y\}$ and $\Omega_p = \{1...N_x\} \times \{1...N_y\}$. There are $N = N_x N_y$ pixels and $n = n_x n_y$ disparity parameters.

The disparity map d is a set of 2D vectors that define a mapping from image 1 to image 2; we use a B-Spline representation (Thévenaz et al., 2000) at scale R parametrized by the coefficients Δ. The disparity can be set at a coarser scale than the image, so as to allow for a multigrid optimization scheme as shown in section 4.2. A coverage factor $\alpha \in [0, 1]$ is defined for each pixel, describing the total contribution of the motion model to the local displacement d, and enabling us to weight each data pixel according to the ability of the model to predict that pixel.

$$d_p = \sum_{j \in \Omega_\theta} \frac{1}{\alpha_p} w_{pj} \Delta_j \quad \text{for} \quad \alpha_p > 0 \tag{1}$$

$$\text{where} \quad w_{pj} = \varphi\left(\frac{1}{R}p - j\right) \quad \text{and} \quad \alpha_p = \sum_{p \in \Omega_p} w_{pj} \tag{2}$$

The 2D kernel φ is separable, i.e. $\varphi(v) = s(v^x)\, s(v^y)$ where s is the 1D B-Spline 3 kernel. We define the sampled disparities $D_j = d_{Rj}$, so that we have $D = S\Delta$ where S is the convolution operator achieving the interpolation for discrete spatial positions (discrete filter $\varphi(p)$, $p \in \Omega_p$).

The image X^1 is the reference for displacement, i.e. $X^1 \equiv X$, whereas the image X^2 is obtained by warping X through the

disparity map using B-Spline interpolation, which is denoted by $W^{\Delta}(X)$. This approximation remains valid as long as the mapping from image 1 to image 2 preserves the area. This is generally the case for disparity maps since the displacements can be seen as local shifts, and the global scaling is rather negligible, especially in along-track stereo. If we denote the Spline coefficients of X by L^0 such that $L^0 = S^{-1}X$, we have:

$$X_p^2 = W^{\Delta}(X)_p = \sum_{k \in \Omega_p} L_k^0 \, \varphi(p - k - d_p) \tag{3}$$

2.2 Self-similar change modeling

The observed images are denoted by Y^1 and Y^2. Image X^1 is assumed to be the reference, whereas X^2 undergoes a radiometric transformation which models reflectance effects, shadows, and relatively large-scale changes that can occur between multi-date observations. We assume all these structured radiometric changes can be embedded in an additive term C. In this work, we also assume that C contains the observation noise. To further simplify the modeling, the noise from both images is embedded in C while the reference image remains noise-free:

$$Y^1 = X^1 \quad \text{and} \quad Y^2 = X^2 + C \tag{4}$$

One could argue that due to surface physics, multiplicative transforms are more likely; in fact, to be really accurate, one should take into account at least 3 phenomena – a multiplicative field related to non-Lambertian reflectance properties, clouds and atmospheric transparency, an additive field related to ambient lighting and atmospheric scattering, and also a nonlinear transfer function that is not necessarily shift invariant. Multidate changes are even more complex to describe. Therefore a simple additive term C shall be a reasonable choice in practice, and we will also assume X^2 and C are independent in the method proposed here.

Change maps can be modeled in various ways. Among all possible models, few are simple enough to allow for an efficient implementation. We want to take into account two essential properties: power spectrum decay (high frequency changes are less likely than the low frequency ones) and spatial adaptivity (there is no reason all areas behave similarly, as there might be occlusions, shadows or more dramatic changes). We propose to use a Gaussian fractal process, motivated by the self-similarity of natural images (Jalobeanu et al., 2003). In the frequency space (where spatial frequencies are denoted by u and v), all coefficients are independent Gaussian variables whose variance obey a self-similar law defined by an energy ε^2 and an exponent q:

$$P(\mathcal{F}[C]_{uv}) = \mathcal{N}\left(0, \varepsilon^2 \left(u^2 + v^2\right)^{-q}\right) \text{ for } u, v \neq 0, 0 \tag{5}$$

We will assume a fractal exponent equal to 1 which is a commonly encountered value in natural scenes. In order to remain in the image space we use Markov Random Fields; this is required by the spatial adaptivity, best parametrized in the image space and not in the frequency space. The expression above corresponds to a diagonal inverse covariance equal to $\varepsilon^{-2}\left(u^2 + v^2\right)$ in the Fourier space. Its counterpart in the image space is proportional to the linear operator H:

$$H = G_x^t G_x + G_y^t G_y \tag{6}$$

where G_x and G_y denote the image intensity gradients (finite differences) in the x and y directions. This is all made possible by setting $q = 1$ (non-integer values require fractional derivatives, whose computational complexity is considerably higher). We set the gradients to zero on the boundaries. We get:

$$P(C \,|\, \lambda) = \frac{1}{Z_\lambda} e^{-\lambda \, (AC)^t \, H \, AC} \tag{7}$$

where λ is a smoothness parameter (related to the factor ε) and Z_λ the corresponding normalization coefficient. Using bounded gradient values enables us to define a proper distribution.

In (7), A is a diagonal matrix that helps achieve spatial adaptivity. The variance of C is divided by A. The simplest choice consists of setting $A_p^2 = \alpha_p$ so as to take into account the coverage factor α_p for each pixel p. Indeed, this factor allows us to put a weight on each data pixel according to the ability of the model to predict this pixel value. Any pixel not related to the model (i.e. that the model can not explain) shall have a weight equal to zero so as to cancel its contribution to the model, which is achieved by an infinite variance, thus allowing the change map value C_p to grow arbitrarily large. Conversely, pixels fully covered by the model (α=1) are not affected.

Other pixel-dependent weights could be included here, such as a mask allowing to exclude some of the pixels (classified as unreliable beforehand) from the inference. The weights could also be updated recursively by analyzing the change map after each inference step to make the method more robust.

2.3 A prior disparity map model

To complete the forward model we have yet to define a prior model for the unknown disparity map. A simple choice is a smoothness prior having the same fractal properties as the change model, particularly suitable for disparities related to 3D surfaces or natural phenomena. To keep the approach as general as possible we do not consider the epipolar constraint as in (Deriche et al., 1994) and we set two separate smoothness parameters ω_x, ω_y for disparities along each direction.

$$P(\Delta \,|\, \omega) = \frac{1}{Z_{\omega_x} Z_{\omega_y}} e^{-\omega_x \, (\Delta^x)^t H^*(\Delta^x) - \omega_y \, (\Delta^y)^t H^*(\Delta^y)} \tag{8}$$

Here Z_{ω_x}, Z_{ω_y} denote normalization coefficients. The operator H^* can be different from H. In principle, the model should apply to the disparities D, not to its Spline coefficients Δ. Therefore we decided to choose $H^* = S^t H S$ in order to use a self-similar prior similar to the change map prior.

We will discuss later how to estimate the smoothness parameters. There is no immediate need for a prior pdf since these parameters are substantially overdetermined. They are related to the underlying structure of the motion field, a special case being along-track stereo where the displacement map along x is particularly smooth, in which case very large values of ω_x are expected.

2.4 Forward model and the related Bayesian network

We just defined the elements of a full generative model describing the formation of an image pair from a single scene X, given a deformation field Δ, a change map C and a noise variance map σ^2 as well as their respective parameters. A Bayesian network displays all causality relations (see Fig. 1) and enables us to write the joint pdf of all model variables as a product of prior pdfs and conditional pdfs. The conditional pdf of the variables given the data is proportional to this joint pdf, since the data is fixed.

3 THE INVERSE PROBLEM

Now we need to solve the inverse problem (with respect to the direct model just defined). In the Bayesian approach (Gelman et al., 1995), we aim at the determination of the a posteriori pdf, i.e. the pdf of the variables of interest Δ given the data Y^1, Y^2, written as $P(\Delta \,|\, Y^1, Y^2)$. Three major steps are involved:

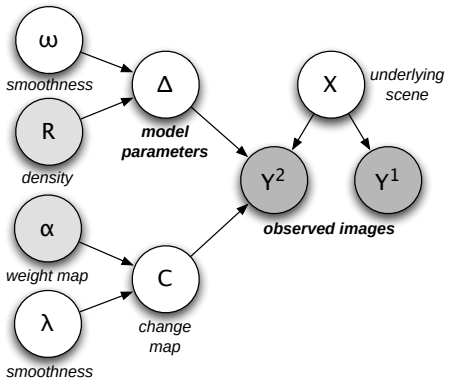

Figure 1: Bayesian network related to the proposed forward model, describing the formation of the observed images given all model variables. Nodes with converging arrows represent conditional pdfs, while all other nodes represent priors. Shaded nodes are fixed (or observed) variables.

- Marginalization, to formally integrate the joint pdf with respect to all variables that are not of interest (also known as nuisance variables), denoted by $\Theta = \{X, C, \lambda, \omega_x, \omega_y\}$ such that: $P(\Delta \,|\, Y^1, Y^2) \propto \int P(Y^1, Y^2, \Delta, \Theta)\, d\Theta$
- Optimization, which aims at finding the values of Δ that maximize the posterior pdf (this step might involve the optimization of other quantities as required by marginalization). In practice an energy function U, defined as the -log of the marginal posterior pdf, is being minimized.
- Gaussian approximation around the optimum to determine parametric uncertainties on the result.

3.1 Image resampling and change map marginalization

The observed image Y^2 is now fixed, so we set $z^2 \equiv Y^2$ and $z^1 = W^{\Delta}(Y^1)$ since we assumed $Y^1 \equiv X$. If we denote by L the B-Spline coefficients of Y^1, we have:

$$z^1_p = W^{\Delta}(Y^1)_p = \sum_{k \in \Omega_p} L_k \, \varphi(p - k - d_p) \tag{9}$$

The difference $z^{12} = z^2 - z^1$ defined this way is equal to the radiometric changes C, according to Eqn. (4). See Fig. 3 for an illustration. If we had a white Gaussian observation noise instead of the structured change model (by replacing H with identity), the obvious procedure would consist of minimizing the classical sum of squared differences (SSD) between Y^2 and $W^{\Delta}(Y^1)$. In general, this SSD-based procedure fails as the changes are spatially structured (hence the use of normalized SSD by some authors, but without explicit radiometric changes). Since we have two deterministic relations (4), integrating out X and C gives

$$P(Y^1, Y^2 \,|\, \Delta, \lambda) \propto \frac{1}{Z_\lambda} e^{-\lambda\, (A z^{12})^t \, H \, A z^{12}} \tag{10}$$

3.2 Estimating the change smoothness parameter

A noninformative, parameter-free prior is assumed for λ since we have little knowledge about the changes; discussing how to choose the prior pdf is beyond the scope of this paper. The integration with respect to λ can be done via the Laplace approximation as explained in (MacKay, 2003), which amounts to evaluating a Gaussian integral. The approximation consists of considering that the -log of the integrand is a quadratic form, defined by an minimum (location and value) and a curvature at the minimum. As there is a single parameter for all pixels, the curvature is very high, therefore the integrand behaves like a Dirac function so this approximation is valid. To evaluate the integral this way, we need to calculate the value at the optimum $\hat{\lambda}$ (the location itself does not matter) as well as the second derivative $-\partial^2 \log P / \partial \lambda^2$; the latter does not depend on Δ so it simply acts as a constant factor and is ignored. The optimum is given by

$$\hat{\lambda} = \frac{N}{2\, \Phi(A\, z^{12})} \quad \text{where} \quad \Phi(X) = X^t H X \tag{11}$$

We finally get the integrated likelihood for Δ:

$$P(Y^1, Y^2 \,|\, \Delta) \propto e^{-\frac{N}{2} \log \Phi(A\, z^{12})} \text{ with } \Phi(A\, z^{12}) \neq 0 \tag{12}$$

We never have $\Phi = 0$ if $R > 1$, because z^1 and z^2 can not be equal for all pixels since the images are noisy and $R > 1$ prevents any over-fitting (there is more than one pixel per parameter).

3.3 Estimating the disparity smoothness

Instead of using the smoothness prior (8) with fixed values of the parameters ω, which would require a rather complex estimation procedure to find optimal values (Jalobeanu et al., 2002), we can use an integrated prior by integrating out ω_x and ω_y from the very beginning. This way we lose the benefits of having a quadratic regularization energy but we make the prior parameter-free. It also makes the full inference method parameter-free since λ was already integrated out.

The pdf (8) can be approximated by a Gaussian around its mode as we did in the previous paragraph, so we apply exactly the same reasoning as we did with λ (noninformative prior, and Laplace approximation), which yields:

$$P(\Delta) \propto e^{-\frac{n}{2} \log \Phi^*(\Delta^x) - \frac{n}{2} \log \Phi^*(\Delta^y)} \tag{13}$$

This enables us to finally express the posterior $P(\Delta \,|\, Y^1, Y^2)$, proportional to the product of expressions (12) and (13):

$$P(\Delta \,|\, Y^1, Y^2) \propto e^{-\frac{N}{2} \log \Phi(A z^{12}) - \frac{n}{2} \log \Phi^*(\Delta^x) - \frac{n}{2} \log \Phi^*(\Delta^y)} \tag{14}$$

A degenerate solution could be found for $\Phi^* = 0$, since nothing prevents us from having a perfectly smooth disparity map (for instance Δ^x is very smooth in along-track stereo, almost linear in our tests as shown in the result section on Fig. 4). The interpretation of the posterior pdf (14) needs special care: we should not seek the global optimum, but rather the most 'significant' one. The peak related to degenerate solutions is the highest (in fact, this posterior is improper, with a singularity at $\Phi^* = 0$, but it can be seen as the limit of a sequence of proper pdfs). However it can be shown by computing the second derivatives that this peak is very narrow compared to the one related to the acceptable solution, so that in terms of probability integrated over the corresponding volume, a meaningful, non-degenerate solution is more probable – even if does not maximize the pdf.

4 PROPOSED ALGORITHM

4.1 A fully unsupervised method

In order to maximize the posterior (14), the energy U needs to be minimized with respect to the disparity map parameters Δ:

$$U = \frac{1}{2} \big(N \log \Phi(A z^{12}) + n \log \Phi^*(\Delta^x) + n \log \Phi^*(\Delta^y) \big) \tag{15}$$

As mentioned above, special care has to be taken to avoid the potential well around the singularity. This can be effectively done by replacing $\log(\Phi^*)$ by $\log(\delta^2 + \Phi^*)$ where δ is strictly positive to avoid the singularity. When the solution is found, δ can be

set to zero, hoping we escaped the sphere of influence of the singularity and will now converge to the acceptable solution. One can also adopt a more careful scheme inspired from graduated non convexity where δ starts large and is progressively decreased until it reaches zero, thus ensuring convergence.

The marginalization that helps make the problem parameter-free does not add much difficulty to the optimization which is already nonlinear. Indeed, z^{12} has a nonlinear dependence on Δ because of the resampling process. The energy (15) is not quadratic (and also not convex). Thus, solving this problem needs special handling to avoid local minima and find the desired solution. Simple Newton-like methods can not be used, and we must resort to a nonlinear gradient descent. This can be sensitive to the initialization, hence the multigrid technique discussed in section 4.2.

We choose to use a nonlinear conjugate gradient method (Press et al., 1993), which only requires the first derivatives of U. The derivatives of the data term are denoted by ζ^x:

$$\zeta_j^x = \frac{\partial}{\partial \Delta_j^x} \frac{N}{2} \log \Phi(Az^{12}) = \frac{N}{\Phi(Az^{12})} (HAz^{12})^t g_j^x \quad (16)$$

where g_j^x is related to the derivative of z^{12} and is defined as:

$$(g_j^x)_p = -\alpha_p \frac{\partial z_p^1}{\partial \Delta_j^x} = w_{pj} W_{\partial x}^{\Delta} (Y^1)_p \quad (17)$$

Here $W_{\partial x}^{\Delta}$ denotes an interpolation based on the derivative of φ, involving the kernel $\varphi'^x(v) = s'(v^x)\, s(v^y)$ instead of φ:

$$W_{\partial x}^{\Delta}(Y^1)_p = \sum_k L_k\, \varphi'^x(p - k - d_p) \quad (18)$$

which is computed at the same time as the usual interpolation W^{Δ}. We get similar equations for Δ^y. The derivatives of the full energy (data term and prior) are finally given by:

$$\frac{\partial U}{\partial \Delta_j^x} = \zeta_j^x + \frac{n}{\Phi^*(\Delta^x)} H^* \Delta_j^x \quad (19)$$

4.2 Multigrid optimization

The most classical approach to multigrid is to start with a coarse disparity model (large R), perform the optimization, and refine recursively by initializing finer scale models with the previous, coarser scale. This has been applied to image registration in (Thévenaz et al., 1998). A dyadic scheme is generally used, starting with $R = R_0$ and dividing by 2 at each step. Coarse estimates are discarded, since they are only used to initialize the next scale.

However, when processing large size images, one had better estimate coarse deformation models from subsampled versions of the input images rather than from the full size images, especially at very coarse scales. A Spline pyramid (Unser et al., 1993) is computed for both images Y^1 and Y^2. At each scale, we apply the inference procedure described above, with a fixed value of R; the range $2 \leq R \leq 8$ achieves a good trade-off between model density and number of data points per parameter. Then, the model is refined through Spline subdivision since the parameters Δ are actually the Spline coefficients of the dense disparity field d. The refined model is used to initialize the next scale.

4.3 Computing uncertainties

The inverse covariance matrix related to a Gaussian approximation of the posterior pdf around the optimum Δ is defined by 4 blocks, related to second derivatives with respect to Δ^x and Δ^y:

$$\Sigma_{\Delta}^{-1} = \left(\begin{array}{c|c} \partial^2 U/\partial \Delta^x \partial \Delta^x & \partial^2 U/\partial \Delta^x \partial \Delta^y \\ \hline \partial^2 U/\partial \Delta^y \partial \Delta^x & \partial^2 U/\partial \Delta^y \partial \Delta^y \end{array} \right)_{\Delta = \hat{\Delta}} \quad (20)$$

Using the first derivatives previously calculated (16)-(18) and assuming a locally linear resampling process W^{Δ}, we get:

$$\frac{\partial^2 U}{\partial \Delta_k^x \partial \Delta_l^x} \simeq \frac{N}{\Phi(Az^{12})} (g_k^x)^t H (g_l^x) - \frac{1}{N} \zeta_k^x \zeta_l^x + \frac{n}{\Phi^*(\Delta^x)} H^* \quad (21)$$

where all expressions are evaluated at $\Delta = \hat{\Delta}$. We obtain a similar expression for Δ_k^y and Δ_l^y. The cross-terms are:

$$\frac{\partial^2 U}{\partial \Delta_k^x \partial \Delta_l^y} \simeq \frac{N}{\Phi(Az^{12})} (g_k^x)^t H (g_l^y) - \frac{1}{N} \zeta_k^x \zeta_l^y \quad (22)$$

The quadratic form is $u^t H v = (G_x u)^t (G_x v) + (G_y u)^t (G_y v)$.

For most applications, we need uncertainties on the disparities $D = S\Delta$ (in practice the optimization is simpler to perform with respect to Δ, this is why we do not use D in the first place). If we need Σ_D^{-1} rather than Σ_{Δ}^{-1}, the following expression can be used for the conversion:

$$\Sigma_D^{-1} = (S^{-1})^t\, \Sigma_{\Delta}^{-1}\, (S^{-1}) \quad (23)$$

where S^{-1} is the operator that transforms a vector into a series of Spline coefficients, which can be implemented very efficiently (Thévenaz et al., 2000). Notice that this is only needed once the optimization procedure has been completed, therefore does not affect the computational cost of the disparity estimation itself.

In principle, uncertainties are expressed through variances and covariances, which have a physical meaning: the former directly relates to confidence intervals on the estimated parameters, and the latter give the correlation between model variables. However, the matrix (20) needs to be inverted, which is difficult in practice because of its size ($2n \times 2n$). Therefore we approximate covariances between neighboring disparity parameters by neglecting long-range interactions.

For each covariance to be computed between variables indexed by i and j, we select a small block of the matrix Σ^{-1} by picking only the entries for variables directly connected to i and j. Obviously, in the inversion, variables that do not interact shall not be involved in the computation. In practice, for a given spatial position j (diagonal elements of each block of Σ, related to Δ_j^x and Δ_j^y) if we restrict to the 8 nearest neighbors $j + (\pm 1, \pm 1)$ for both Δ^x and Δ^y, we only need to invert a 18×18 matrix. With 4 nearest neighbors the size of the matrix reduces to 10×10. This has to be repeated for each spatial location k. Notice that efficient iterative optimization techniques can advantageously replace matrix inversion (Jalobeanu and Gutiérrez, 2007).

If uncertainties do not need to be interpreted, but rather stored and propagated through to other processing algorithms, the inversion can be avoided: only the inverse covariance matrix needs to be propagated. If there are too many terms it can be simplified so as to limit the redundancy of the end result; for instance one can provide inverse covariances related to the 4 nearest neighbor Δ variables, which only requires to store and propagate 7 extra terms for each location j (in addition to the estimates $\hat{\Delta}_j^x$ and $\hat{\Delta}_j^y$); refer to table 1 for details.

Self	$\Sigma_{j,j}^{-1\,xx}$, $\Sigma_{j,j}^{-1\,yy}$
Cross (xy)	$\Sigma_{j,j}^{-1\,xy}$
Horizontal	$\Sigma_{j+(1,0),j}^{-1\,xx}$, $\Sigma_{j+(1,0),j}^{-1\,yy}$
Vertical	$\Sigma_{j,j+(0,1)}^{-1\,xx}$, $\Sigma_{j,j+(0,1)}^{-1\,yy}$

Table 1: Uncertainty terms (inverse covariances, limited to 4 nearest neighbor interactions) produced by the proposed disparity inference algorithm, after inverse covariance simplification.

5 PRELIMINARY RESULTS AND DISCUSSION

5.1 Tests on real Mars Express images

We chose a test area extracted from a raw panchromatic, along-track stereo pair taken by the Mars Express HRSC instrument (ESA). The image was downsampled by a factor 8 using a Spline pyramid (preserving noise statistics) to ensure small displacements ($<$10 pixels) then a study area was selected ($N = 64 \times 64$, see Fig. 2). The initial disparity maps were $D_x =$ linear function of x and $D_y = 0$. A radius $R = 4$ was chosen for the model resolution ($n = 16 \times 16$). Convergence was reached in less than 50 iterations for this particular initialization; a multigrid strategy would significantly reduce this number, however the purpose of the test is to check the validity of the proposed energy functional (15) rather than study how to minimize it most efficiently.

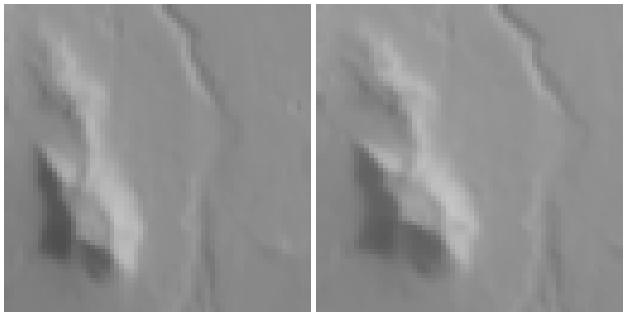

Figure 2: Left: Y^1; right: Y^2. ESA Mars Express HRSC Orbit 0905 (stereo s12/s22), subsampling factor 8, extracted region 64×64 pixels.

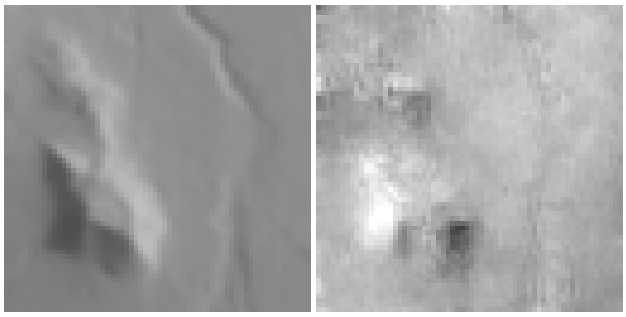

Figure 3: Left: $W^{\Delta}(Y^1)$ (Y^1 warped using the estimated disparity map); right: corresponding change map $C = Y^2 - W^{\Delta}(Y^1)$, contrast enhanced (factor 10) to display the effects of non-Lambertian reflectance.

Fig. 4 displays the estimated disparities D_x and D_y; obviously D_x is unrelated to the topography and appears very smooth. Uncertainties also shown (bottom Fig. 4) are preliminary results; we only show the inverse of the diagonal of Σ_D^{-1} without the cross-terms, however these preliminary error maps already carry valuable information that can be interpreted as error bars on D_x and D_y (standard deviation). It clearly illustrates the spatial variability of errors; the higher the contrast of radiometric features (e.g. texture, edges), the lower the uncertainty. It falls below 0.1 pixel as long as there are enough details, and well below 0.05 pixel near edges, whereas it can reach 0.3 pixel in the smoothest areas. Notice that these estimates do not depend on the observation noise parameters, which are unknown, nonetheless they exploit the available statistics throughout the inference procedure. Notice also that there is no available ground truth for disparities.

5.2 Application to 3D reconstruction

If the imaging geometry is known and reliable (e.g. attitude and position of the satellite computed from the metadata), then 2D displacements can be directly converted into 1D heights, which amounts to projecting the 2D pdfs of disparities to get 1D pdfs of elevations. This will provide a DEM as a height field (longitude, latitude and elevation) where the sampling corresponds to the pixels of image Y^2, most probably irregular on the ground. Therefore it needs to be resampled on a regular grid (again we recommend B-Spline interpolation), and the uncertainties need to be converted accordingly using an equation similar to (23). The same process applies to pixel values from both input images in order to achieve orthorectification, which amounts to providing a textured terrain model using reflected radiance maps. As opposed to traditional orthorectification techniques, this should be done using probability theory, which produces uncertain and correlated pixel values as the estimated elevations are also uncertain. Moreover, the resampling may also produce a blur due to the geometric uncertainty related to probabilistic elevations.

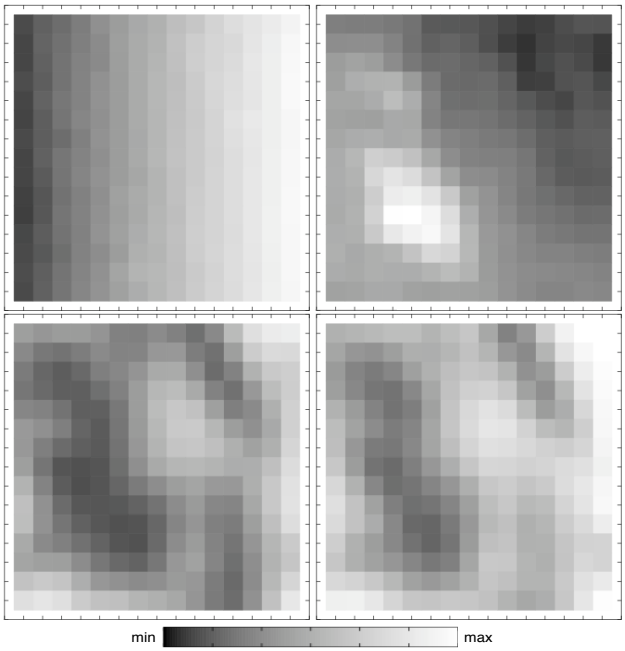

Figure 4: Top: estimated disparities (left: D_x, range=[0,9], right: D_y, range=[3.5,-1.5]); bottom: related standard deviations, range=[0,0.3].

5.3 Application to automatic camera calibration

Remote sensing images are acquired with pushbroom systems whose parameters are not always well-known. Even if a reasonably accurate calibration can be achieved through star-based optical navigation or using onboard sensors, applications aiming at a sub-pixel precision can not rely on it. Indeed, the orbital motion is affected by high-frequency vibrations whose parameters are unknown in general. Imperfect trajectory and attitude models derived from the metadata lead to systematic errors in 3D reconstruction that are often inconsistent with the recorded data.

Why not use the data directly without making complex calculations involving a geometry that is not well constrained? If a dense disparity map D is provided as well as the related errors and correlations summarized by Σ_D^{-1}, it should be possible to recover relative camera motion without any other source of information. This has been achieved for approximate, linear camera models (Gupta and Hartley, 1997). Once the relative motion model has been inferred (parameter values as well as their uncertainties can be computed using a probabilistic approach), a relative DEM can be reconstructed in the camera space. It then needs to be converted into the world space using ground control points; however, even without such absolute knowledge, the relative DEM is still a valuable product that makes use of all the radiometric information contained in the stereo pair and its quality is not dependent on possibly misestimated calibration parameters.

While relative camera calibration for rigid detector arrays has been thoroughly investigated in the field of stereo vision (Deriche et al., 1994), the application to pushbroom sensors is still a open research area, especially from the probabilistic point of view.

5.4 Application to ground deformation

There are two main types of ground deformation that would require precise monitoring: gravitational motions (e.g., landslides) and tectonic-driven motions (e.g., co- and/or post-seismic deformation fields). As reviewed in (Delacourt et al., 2004), currently, most of the techniques for monitoring landslide displacement are derived from measurements of reference stations (triangulation, tacheometry and GPS measurements). The database of movement provided by these techniques is available only for major landslides for a time span not exceeding 20 years for laser measurements and less than 15 years for GPS. Moreover, due to spatial and temporal heterogeneities of the displacements, such ground based measurements are not sufficient to describe fully the velocity field of a landslide. Remote sensing imagery is a powerful tool for landslide monitoring because it offers a synoptic view that can be repeated at different time intervals. This is also the case for very localized slow tectonic motion. Differential SAR interferometry (DINSAR) has shown its capability for deriving high accuracy maps (at centimeter level) of landslide displacement (Fruneau et al., 1996). However, this technique is affected by severe geometrical and environmental limitations (e.g., loss of coherence due to vegetation). Moreover, the SAR image database is limited to 1991, and later.

In recent years, new techniques based on the correlation of satellite optical images for the processing of deformation maps have been developed. Those techniques have been successfully applied to the measurement of coseismic deformation and complement InSar techniques in particular close to the fault surface trace where no interferogram can be computed. The same techniques were applied to the long-term monitoring of landslides by combining aerial images acquired at different times and very high resolution satellite images (QuickBird). Using a pair of SPOT panchromatic images, (Van Puymbroeck et al., 2000) showed that sub-pixel correlation could provide fault slip measurements with an accuracy of 0.1 pixel (1 meter) (Michel and Avouac, 2002). The approach of (Van Puymbroeck et al., 2000) first consists in resampling the SPOT images into a map projection so that the remaining image pixel offsets are only due to the earthquake ground deformation. In a second step, the coseismic offset map and its error estimate are computed from the phase shift of the Fourier transform of a sliding window. In (Binet and Bollinger, 2005), the method was adapted to SPOT 5 images and the correlation window size used was 256 pixels and the window step was 64 pixels in both directions. Attempts in reducing the window size lead to a noisy offset map near the fault because of the temporal decorrelation and of the low contrast of the ground.

What we want with our approach is 1) to break free from the initial assumption of a rigid motion within each window, 2) to take into account the changes, 3) to compute covariance maps. Indeed, this is required to combine the analysis of real surface deformation information and probabilistic geophysical modeling.

6 CONCLUSIONS AND FUTURE WORK

Our long-term goal can be described as fully automated probabilistic Digital Terrain Model (DTM) generation from uncalibrated stereo pairs (possibly from more than two images). This is part of the SpaceFusion project, funded by the French Research Agency (ANR); its goal is to combine multiple data to infer the topography, then to fuse the radiometry information into a well-sampled, single reflectance map. A valuable by-product of this project is the determination of 3D deformation fields with error maps if the 3D reconstruction is performed from stereo pairs taken at different dates. This way one can measure the ground motion or the evolution of the topography for analysis or monitoring purposes. To efficiently handle occlusions and abrupt changes, the forward model will need to be extended to allow for spatially adaptive noise statistics, for a better robustness.

REFERENCES

Binet and Bollinger, 2005. Horizontal coseismic deformation of the 2003 Bam (Iran) earthquake measured from SPOT-5 THR satellite imagery. Geophys. Res. Lett. 32(L02307), pp. doi:10.1029/2004GL021897.

Blake, A., Torr, P., Cox, I. and Criminisi, A., 2003. Estimating uncertainty in dense stereo disparity maps. Microsoft Research Report MSR-TR-2003-93.

Brown, M., Burschka, D. and Hager, G., 2003. Advances in computational stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(8), pp. 993–1008.

Delacourt, C., Allemand, P., Casson, B. and Vadon, H., 2004. Velocity field of the 'La Clapière' landslide measured by the correlation of aerial and QuickBird satellite images. Geophys. Res. Lett. 31(L15619), pp. doi:10.1029/2004GL020193.

Deriche, R., Zhang, Z., Luong, Q.-T. and Faugeras, O., 1994. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. Lecture Notes in Computer Science 800, pp. 567–576.

Fruneau, B., Achache, J. and Delacourt, C., 1996. Observation and modeling of the saint-etienne-de-tinee landslide using SAR interferometry. Tectonophysics 265(3-4), pp. 181–190.

Gelman, A., Carlin, J., Stern, H. and Rubin, D., 1995. Bayesian Data Analysis. Chapman & Hall.

Gupta, R. and Hartley, R. I., 1997. Linear pushbroom cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(9), pp. 963–975.

Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: Proc. of CVPR, pp. II: 807–814.

Jalobeanu, A. and Gutiérrez, J., 2007. Inverse covariance simplification for efficient uncertainty management. In: 27th MaxEnt workshop, AIP Conference Proceedings, Saratoga Springs, NY.

Jalobeanu, A., Blanc-Féraud, L. and Zerubia, J., 2002. Hyperparameter estimation for satellite image restoration using a MCMC Maximum Likelihood method. Pattern Recognition 35(2), pp. 341–352.

Jalobeanu, A., Blanc-Féraud, L. and Zerubia, J., 2003. Natural image modeling using complex wavelets. In: Wavelets X, SPIE Symposium, San Diego, CA.

MacKay, D., 2003. Information Theory, Inference, and Learning Algorithms. Cambridge University Press.

Michel, R. and Avouac, J., 2002. Deformation due to the 17 August Izmit, Turkey, earthquake measured from SPOT images. J. Geophys. Res.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B., 1993. Numerical Recipes in C: The Art of Scientific Computing. 2nd edn, Cambridge University Press.

Thévenaz, P., Blu, T. and Unser, M., 2000. Interpolation revisited. IEEE Trans. on Med. Imaging 19(7), pp. 739–758.

Thévenaz, P., Ruttimann, U. and Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. IEEE Transactions on Image Processing 7(1), pp. 27–41.

Unser, M., Aldroubi, A. and Eden, M., 1993. The L_2-polynomial spline pyramid. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(4), pp. 364–379.

Van Puymbroeck, N., Michel, R., Binet, R., Avouac, J.-P. and Taboury, J., 2000. Mesuring earthquakes from optical satellite images. Applied Optics 39(20), pp. 3486–3494.

Zhang, J., McMillan, L. and Yu, J., 2006. Robust tracking and stereo matching under variable illumination. In: Proc. IEEE CVPR, pp. 871–878.