Small Baseline Stereovision

Julie Delon LTCI (Télécom Paris, France), julie.delon@enst.fr
Bernard Rougé CNES (Toulouse, France), bernard.rouge@cnes.fr

Abstract

This paper presents a study of small baseline stereovision. It is generally admitted that because of the finite resolution of images, getting a good precision in depth from stereovision demands a large angle between the views. In this paper, we show that under simple and feasible hypotheses, small baseline stereovision can be rehabilitated and even favoured. The main hypothesis is that the images should be band limited, in order to achieve sub-pixel precisions in the matching process. This assumption is not satisfied for common stereo pairs. Yet, this becomes realistic for recent spatial or aerian acquisition devices. In this context, block-matching methods, which had become somewhat obsolete for large baseline stereovision, regain their relevance. A multi-scale algorithm dedicated to small baseline stereovision is described along with experiments on small angle stereo pairs at the end of the paper.

Keywords: Stereo, Discrete correlation, Shannon sampling, Digital Elevation Model (DEM), Numerical Elevation Model (NEM).

1 Introduction

Stereopsis is the process of reconstructing depth from two images of the same scene. This relies on the following fact: if two images of a scene are acquired [1] from different angles, the depth of the scene creates a geometric disparity between them. If the acquisition system is calibrated, the knowledge of this disparity function ε allows one to determine the digital elevation model (DEM) of the observed scene. In this paper, we focus mainly on matching stereo pairs of satellite or aerial images, that have been rectified to epipolar geometry (see [7]). If the altitude of the cameras is high enough for the parallel projection model to be accurate, ε and the depth function z are linked at a first approximation by the relation $z = \frac{\varepsilon}{b/h}$, where b/h is a stereoscopic coefficient [2], only dependent on the acquisition conditions. This coefficient roughly represents the tangent of the angle between the views (see Figure 1). The precision dz of the depth measurement is consequently linked to the precision $d\varepsilon$ of the disparity measurement by

$$dz = \frac{d\varepsilon}{b/h}. \tag{1}$$

[1] For example, in the satellite case, images are acquired by CCD retina matrices.

[2] This coefficient is the ratio between the baseline b (*i.e.* the distance between the camera centers) and the distance h between the scene and the camera system. In reality, b/h changes slowly in space.

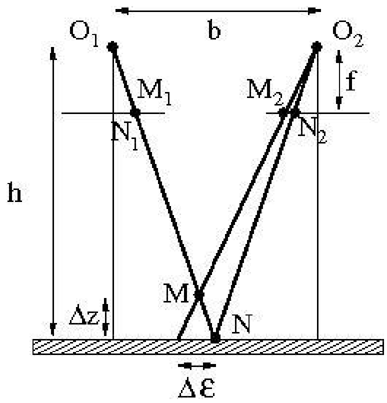


Figure 1: *Stereopsis principle. O_1 and O_2 are the centers of the cameras. The projections of the ground points M and N in the first image are M_1 and N_1, and M_2 and N_2 in the second one. We see that the position of M_1 in the first image is not the same as the position of M_2 in the second image. Let us denote with ΔM the shift between these positions (resp. ΔN for N). The difference of shifts $\Delta M - \Delta N$ is proportional to the disparity $\Delta\varepsilon$ (the proportionality coefficient is actually the image resolution) and $\Delta\varepsilon$ itself is roughly proportional to the depth difference Δz (the proportionality coefficient is b/h).*

It follows that for a given accuracy $d\varepsilon$ of the disparity measurement, the larger the coefficient $\frac{b}{h}$, the smaller the depth error. It is commonly admitted that $d\varepsilon$ does not depend on b/h, but only on the image resolution. For this reason, high stereoscopic coefficients have always been preferred in stereoscopy (typically, $\frac{b}{h} = 1$, which corresponds to an angle of approximately 53^o). However, a large coefficient also means more changes between the images (more different hidden surfaces, differences in radiometry, larger geometrical deformations, moving objects, etc...), hence more difficulties in the matching process. This is especially true in the case of urban images, where buildings create a large amount of occluded areas, which change fast with the observation angle. Hence, a smaller angle between the views should naturally yield a more accurate disparity measurement. The choice of the coefficient b/h should result from a compromise between these effects.

The objective of this paper is a mathematical study of small baseline stereovision. The efficiency of the human visual system clearly supports the use of small angles. Yet, this kind of stereovision makes sense only under specific acquisition conditions and with specific matching methods. First, the acquisition device needs to be perfectly known and calibrated. In addition, and this hypothesis is essential, the images sampling must be controlled. Most stereo correspondence algorithms only compute integer disparities. This may be completely adequate for a variety of applications but is clearly insufficient for small baseline stereovision. Indeed, when both views are separated by a very small angle, the disparities observed on the images can be quite small in comparison with the pixel size. A matching method of pixelian precision is unable to find any interesting depth information in such a case [3]. Hence, small baseline stereovision requires matching methods specific for subpixelian disparities. We have mentioned that the depth precision dz is linked to the disparity precision $d\varepsilon$ by the relation $dz = \frac{d\varepsilon}{b/h}$. Matching two frames with a small b/h coefficient makes sense only if the precision loss due to the angle is compensated by a better accuracy on ε. Now, for subpixelian

[3] For instance, if $b/h = 0.04$ and if the image resolution is $50cm$, the best elevation accuracy of a matching method of pixelian precision is $0.5/0.04 = 12.5m$.

precision to be achieved, the images of the pair have to be interpolated perfectly. For this reason, they must be well sampled according to Shannon [15] theory [4]. These conditions (small baseline and well sampling) are generally not satisfied by benchmark stereo pairs. Yet, these assumptions are becoming valid with recent satellite acquisition systems.

The strategies used over the years to resolve the matching problem between both images can be roughly divided in local and global methods. Local approaches compute the disparity of a given element by observing only its close neighborhood. Among these methods, area-based (also called "block-matching") approaches estimate the disparity at x by comparing a patch around x in the first frame with similar patches in the second frame, for a given metric or "matching cost". The most standard cost, the normalized cross correlation [6], is merely a scalar product between normalized image patches. Block-matching methods can produce dense subpixel maps, but are hardly reliable in non-textured regions and suffer from adhesion artifacts [5]. However, these methods remain very popular, especially in the industrial community. In contrast, global approaches solve optimization problems on the entire disparity map ε, by making global smoothness assumptions. They involve sophisticated energy minimization methods [1], dynamic programming [12, 4], belief propagation [17], or graph-cuts [10]. These methods show very good performance for standard large baseline stereovision and common stereo pairs (see [14] for an instructive and documented comparison of stereo algorithms). However, they remain computationally too expansive to be applied with subpixel accuracy. In addition, graph-cuts based methods produce strong staircasing artifacts whenever the depth is not piecewise constant, like in the case of urban areas with pitched roofs.

Our focus here is to study the feasibility of small angle stereovision. Hence, for the sake of simplicity, we'll concentrate on the most traditional local matching cost, namely the normalized cross correlation. Correlation matching being both locally and analytically formulated, it allows one to estimate at each point the matching error. Once this feasibility is demonstrated, this will open the way to the use of more sophisticated global methods. The central result of next section is a mathematical formulation of the correlation matching error. We show that this error can be divided in two terms. One is due to the noise and is divided by the b/h coefficient, and the other one is inherent to the method and **independent** of b/h. In other words, the first part of the error is smaller with large stereoscopic angles, but the second part is independent of the angle. Since small baseline generates less occlusions and much more similar images, this independence result gives strong support to small baseline stereovision. To the best of our knowledge, this fact, obvious in animal and human vision, was never pointed out. The comparison of these two terms indicates that in non homogeneus, informative image regions, the noise term can be neglected before the other even for very small baselines. Several questions linked to correlation will be addressed under this new perspective, in particular the question of the size of the window used in block-matching methods and the discrete formulation and interpolation of the correlation coefficient. A multi-scale algorithm based on these results and dedicated to small baseline stereovision will be described.

[4] Shannon sampling theory shows that well sampled images can be completely recovered from their samples, hence interpolated with infinite precision.

2 Analytic Study - Continuous Case

2.1 Notations, model and hypotheses

Let us denote with u and $\tilde{u}$ the images of the stereoscopic pair. One assumes without loss of generality that the images are $2\pi \times 2\pi$ periodic and known on $[-\pi, \pi] \times [-\pi, \pi]$. Only discrete versions of u and $\tilde{u}$ are available. Thus, in what follows, the images u and $\tilde{u}$ are supposed to be band limited. According to Shannon sampling theory [15], this implies that the continuous functions u and $\tilde{u}$ can be reconstructed from their samples, provided that the sampling rate is high enough[5]. The images are supposed to be well sampled, on a regular $2N \times 2N$ grid [6]. Under these hypotheses, it becomes easy to show that u (respectively $\tilde{u}$) can be written as a trigonometric polynomial

$$u(x, y) = \sum_{n=-N}^{N-1} \sum_{m=-N}^{N-1} \hat{u}(m, n) e^{i(nx+my)}, \tag{2}$$

where the coefficients $\hat{u}(m, n)$ represent the discrete Fourier transform (DFT) of the discrete version of u ($\hat{u}$ can be obtained by FFT). Under these simple and realistic hypotheses, the discrete images u and $\tilde{u}$ can be interpreted as continuous periodic functions. As trigonometric polynomials, they are smooth, bounded, and so are all their derivatives.

Suppose that u and $\tilde{u}$ satisfy the classical model

$$\tilde{u}(x) = \lambda(x) u(x + \varepsilon(x)), \tag{3}$$

where λ variates slowly in space and where the disparity function ε describes the geometrical deformation between u and $\tilde{u}$. The function ε is assumed to be bounded.

This model is of course false if the angle between the snapshots is too large (see Figure 2), but is quite reasonable if b/h is small. Indeed, the model assumes that the differences between u and $\tilde{u}$ are purely geometrical, up to a multiplicative function λ with slow spatial variations, and that almost no occlusion or radiometric change occurs. Ultimately, the model is more and more accurate when b/h becomes small. Human eyes [13] almost satisfy these hypotheses.

Normalized Cross Correlation

Consider a smooth, positive, normalized and compactly supported window function φ. We shall use the following notations:

- φ_{x_0} the shifted function $\varphi_{x_0} : x \to \varphi(x_0 - x)$,
- $\int_{\varphi_{x_0}} f = \int_{\varphi_{x_0}} f(x)dx = \int \varphi(x_0 - x) f(x) dx$ for every integrable function f,

[5] More precisely, in one dimension, the Shannon-Whittaker theorem tells us that if $\hat{f}$ is supported in $[-\pi/A, \pi/A]$, then

$$f(t) = \sum_{n=-\infty}^{\infty} f(nA) \frac{\sin(\pi(t - nA)/A)}{\pi(t - nA)/A}.$$

[6] This hypothesis is not satisfied in any real acquisition system, but becomes valid for instance in the case of SPOT5 satellites (two linear CCD arrays allow to create a quincunx grid adapted to the modulation transfer function spectrum [3]).

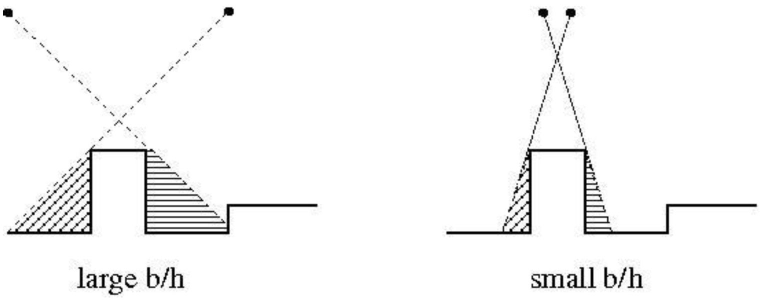


Figure 2: *Differences of occlusion zones in function of b/h. Occlusions in the left image of the stereo pair are signaled by horizontal lines, occlusions in the right image of the stereo pair are signaled by slanted lines. We observe that with a large b/h, the occlusion differences are much more critical than with a small coefficient. This is especially true in urban zones, where the depth can change very fast.*

- $\|f\|_{\varphi_{x_0}}$ the weighted norm $\sqrt{\int \varphi_{x_0}(x) f^2(x)dx}$ for every square integrable function f,
- $< .,. >_{\varphi_{x_0}}$ the corresponding scalar product $< f, g >_{\varphi_{x_0}} = \int \varphi_{x_0}(x)f(x)g(x)dx$ (the effective way to compute correctly discrete scalar products and norms will be discussed in the last section).

We also note $\tau_m u$ the shifted image $x \to u(x+m)$. For each point x_0 of $\tilde{u}$, the normalized cross correlation computes the disparity $m(x_0)$ between u and $\tilde{u}$ at x_0 by maximizing a local similarity coefficient between the images:

$$m(x_0) = \arg\max_m \rho_{x_0}(m), \quad \text{where} \tag{4}$$

$$\rho_{x_0}(m) = \frac{< \tau_m u, \tilde{u} >_{\varphi_{x_0}}}{\|\tau_m u\|_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}}. \tag{5}$$

This function ρ_{x_0} is called the correlation product at x_0 and φ is called the correlation window. The value $\rho_{x_0}(m)$ measures the similarity between the neighborhood of x_0 in the image $\tilde{u}$ and the neighborhood of $x_0 + m$ in the image u. Schwarz inequality [7] ensures that ρ_{x_0} is always between -1 and 1. It is not ensured, though, that the shift $m(x_0)$ at which ρ_{x_0} is maximum is exactly equal to the real disparity $\varepsilon(x_0)$ at x_0. The relation between the functions m and ε is the heart of the next section. In the following, we set $\lambda = 1$ in model (3) since its slow variations hardly alter the correlation coefficient.

Traditionally, most authors consider a centered correlation coefficient, which means that u (resp. $\tilde{u}$) becomes $u - \int_{\varphi_{x_0}} u$ (resp. $\tilde{u} - \int_{\varphi_{x_0}} \tilde{u}$) around x_0. The results of the following sections can be easily generalized to this case, but one will see in paragraph 2.4.1 why this choice is not always judicious.

[7] Schwarz inequality tells us that for any square-integrable functions f anf g, one has

$$\left| \int f(x)g(x)dx \right| \leq \sqrt{\int |f(x)|^2 dx \ . \ \int |g(x)|^2 dx}.$$

2.2 Analytic formulation of the correlation. Case without noise.

In all the following, one makes the classical assumption that the images have been rectified to epipolar geometry: the search for corresponding points can be reduced to one dimension. Hence, all the derivatives used (and written as 1-D derivatives) must be understood "along the direction" of these epipolar lines.

Definition 1 *The following function is called* **correlation density of u at x_0**

$$d^u_{x_0} : x \longrightarrow \frac{\|u\|^2_{\varphi_{x_0}} u'^2(x) - < u, u' >_{\varphi_{x_0}} u(x)u'(x)}{\|u\|^4_{\varphi_{x_0}}}. \tag{6}$$

The function $d^u_{x_0}$ only depends on the image u, the window φ and the correlation point x_0. We will see why this function indicates where the correlation is sensible and can be accurate. The next proposition formulates the relation between the measured disparity $m(x_0)$ at x_0 and the real disparity function ε.

<u>Central equation of correlation.</u>

Proposition 1 *Assume that the disparity function ε and the shift $m(x_0)$ which maximizes ρ_{x_0} satisfy $|\varepsilon(x) - m(x_0)| \ll 1$ on the support of φ_{x_0}. Then $m(x_0)$ is linked to ε by the first order approximation*

$$\boxed{< d^{\tau_{m(x_0)}u}_{x_0}, m(x_0) >_{\varphi_{x_0}} \simeq < d^{\tau_{m(x_0)}u}_{x_0}, \varepsilon >_{\varphi_{x_0}} .} \tag{7}$$

Proof : The first derivative of ρ_{x_0} is

$$\rho'_{x_0}(m) = \frac{< \tau_m u', \tilde{u} >_{\varphi_{x_0}}}{\|\tau_m u\|_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}} - \frac{< \tau_m u, \tilde{u} >_{\varphi_{x_0}} < \tau_m u', \tau_m u >_{\varphi_{x_0}}}{\|\tau_m u\|^3_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}}. \tag{8}$$

Consequently,

$$\rho'_{x_0}(m) = 0 \Leftrightarrow \|\tau_m u\|^2_{\varphi_{x_0}} < \tau_m u', \tilde{u} >_{\varphi_{x_0}} = < \tau_m u, \tilde{u} >_{\varphi_{x_0}} < \tau_m u', \tau_m u >_{\varphi_{x_0}} . \tag{9}$$

Now, let $m(x_0)$ be the shift which maximizes ρ_{x_0}, then $\rho'_{x_0}(m(x_0)) = 0$. Under the assumption that $|\varepsilon - m(x_0)|$ is small enough, a first order approximation gives

$$\tilde{u}(x) = u(x + \varepsilon(x)) \simeq u(x + m(x_0)) + u'(x + m(x_0))(\varepsilon(x) - m(x_0)).$$

Thus, the first order development of the equality $\rho'_{x_0}(m(x_0)) = 0$ gives

$$\|\tau_{m(x_0)} u\|^2_{\varphi_{x_0}} < \tau_{m(x_0)} u'^2, \varepsilon - m(x_0) >_{\varphi_{x_0}} \simeq \tag{10}$$

$$< \tau_{m(x_0)} u, \tau_{m(x_0)})u' >_{\varphi_{x_0}} < \tau_{m(x_0)} u \; \tau_{m(x_0)} u', \varepsilon - m(x_0) >_{\varphi_{x_0}}, \tag{11}$$

which can be rewritten

$$< d^{\tau_{m(x_0)}u}_{x_0}, m(x_0) >_{\varphi_{x_0}} \simeq < d^{\tau_{m(x_0)}u}_{x_0}, \varepsilon >_{\varphi_{x_0}} . \tag{12}$$

□

We will call equation (7) the **central equation of correlation**. For a given image u, this equation clarifies the relation between the disparity function ε and the shift $m(x_0)$ measured by correlation at x_0 between u and $\tilde{u}$ when $|\varepsilon(x) - m(x_0)|$ is small enough in the neighborhood of x_0. This hypothesis means that the variations of ε are small on the window φ_{x_0} and that the shift $m(x_0)$ is a close approximation of the values of ε on this window. Of course, this hypothesis is all the more true since b/h is small.

Interpretation of Proposition 1: *This equation shows that ε is linked with m via a deconvolution relation. If ε is constant on the support of φ_{x_0}, i.e. on $x_0 + supp(\varphi)$, the equation becomes $m(x_0) = \varepsilon(x_0)$, which means that the shift computed by correlation is equal to the local shift between the views. Now, if ε is not constant on the support of φ_{x_0}, (7) shows that the values $\varepsilon(x)$ that matter the most in the measurement $m(x_0)$ are taken at points x at which $d_{x_0}^{\tau_{m(x_0)}u}(x)$ is large.*

This property can be interpreted as an **adhesion** phenomena, as we will see in the consequences section. In zones where u and $\tilde{u}$ are flat (constant), the correlation density is null, which means that no reliable relation between m and ε can be recovered from (7). This confirms the intuition that correlation needs texture information in order to succeed.

Second derivative.

Equation (7) characterizes the point $m(x_0)$ at which ρ_{x_0} is maximum. Now, it is interesting to look more closely at ρ''_{x_0} in the neighbourhood of $m(x_0)$ in order to get an idea of the behaviour of ρ_{x_0} around its maximum.

Proposition 2 *(See Appendix for the proof) Under the hypotheses of Proposition 1, the first order development of ρ''_{x_0} at $m(x_0)$ is*

$$\boxed{\rho''_{x_0}(m(x_0)) \simeq - < d_{x_0}^{\tau_{m(x_0)}u}, 1 >_{\varphi_{x_0}} .} \tag{13}$$

As expected, this approximation satisfies $\rho''_{x_0}(m(x_0)) \leq 0$ (see footnote 7 on Schwarz inequality), which is coherent with the fact that $\rho_{x_0}(m(x_0))$ is maximum. It is interesting to note that this equation can also be approximated by

$$\boxed{\rho''_{x_0}(m(x_0)) \simeq - < d_{x_0}^{\tilde{u}}, 1 >_{\varphi_{x_0}} .} \tag{14}$$

This approximation, that we will call **correlation curvature**, just relies on the knowledge of $\tilde{u}$, independently of ε. This expression gives an *a priori* information about the locations where the maximum can be accurate. The larger the absolute second derivative is, the sharper the maximum is, and the more precisely localized it can be. We will see in the next section the importance of this quantity when noise is added to the images.

Weighted L^2-distance.

Instead of maximizing the correlation coefficient, we can try to minimize the L^2-distance

$$m \to e_{x_0}(m) = \|u(x+m) - \tilde{u}(x)\|_{\varphi_{x_0}}. \tag{15}$$

This minimization can work as soon as the images radiometries are close enough, which is all the more true since b/h is small. In this case, the analytic link between $m(x_0)$ and ε becomes

$$< \tau_{m(x_0)} u'^2, m(x_0) >_{\varphi_{x_0}} \simeq < \tau_{m(x_0)} u'^2, \varepsilon >_{\varphi_{x_0}} . \tag{16}$$

This equation is similar to the correlation one, except that the function $d^u_{x_0}$ is replaced by $u'^2 / \|u\|^2_{\varphi_{x_0}}$.

2.3 Case with noise

We suppose here that white Gaussian noises [8] are added to the images of the pair. The noisy images are denoted with u and $\tilde{u}$. In order to regularize the problem, a convolution with a small and smooth normalized function g (a prolate or a Gaussian) is applied to both images. For the sake of simplicity, we will still denote the regularized images with u and $\tilde{u}$. The model becomes

$$\tilde{u}(x) = u(x + \varepsilon(x)) + g_b(x), \tag{17}$$

where we denote with g_b the convolution $g * b$ between a Gaussian noise b of standard deviation σ_b and the function g.

2.3.1 Central and morphological equations

Before studying the influence of noise on the correlation process, let us start with a more simple case.

Weighted L^2-distance.
Assume that we try to minimize the weighted L^2-distance $m \to e_{x_0}(m) = \|u(x+m) - \tilde{u}(x)\|_{\varphi_{x_0}}$. As we have seen, it makes sense as soon as u and $\tilde{u}$ are radiometrically similar enough, *i.e.* as soon as b/h is small enough.

Proposition 3 *Assume that u and $\tilde{u}$ satisfy relation (17), that ε and the location $m(x_0)$ at which e_{x_0} is minimal satisfy $|\varepsilon - m(x_0)| \ll 1$ on the support of φ_{x_0} and that the noise satisfies the relation $\frac{\|g_b\|_{\varphi_{x_0}}}{\|\tau_{m(x_0)} u'\|_{\varphi_{x_0}}} \ll 1$. Then, equation (16) holds.*

Proof : If $m(x_0)$ is the location at which e_{x_0} is minimal, then $e'_{x_0}(m(x_0)) = 0$, *i.e.*

$$< \tau_{m(x_0)} u', \tilde{u} >_{\varphi_{x_0}} - < \tau_{m(x_0)} u', \tau_{m(x_0)} u >_{\varphi_{x_0}} = 0. \tag{18}$$

If $|m(x_0) - \varepsilon|$ is small enough on the support of φ_{x_0}, a first order expansion of $\tilde{u}$ gives

$$\tilde{u}(x) - \tau_{m(x_0)} u(x) \simeq (\varepsilon(x) - m(x_0)) \tau_{m(x_0)} u' + g_b(x). \tag{19}$$

It follows that

$$m(x_0) \simeq \frac{< \tau_{m(x_0)} u'^2, \varepsilon >_{\varphi_{x_0}}}{\|\tau_{m(x_0)} u'\|^2_{\varphi_{x_0}}} + \frac{< \tau_{m(x_0)} u', g_b >_{\varphi_{x_0}}}{\|\tau_{m(x_0)} u'\|^2_{\varphi_{x_0}}}. \tag{20}$$

[8] For a sake of simplicity, the formulations are continuous. In the discrete case the noise is supposed to be a Shannon white noise (see [16]).

Schwarz inequality (footnote 7) tells us that the second term is smaller than $\frac{\|g_b\|_{\varphi_{x_0}}}{\|\tau_{m(x_0)}u'\|_{\varphi_{x_0}}}$. If this quantity, due to the noise, is smaller than the desired precision on the measure $m(x_0)$, equation (16) holds.

□

The previous proof tells us that $m(x_0)$ and ε are linked in first approximation by relation (20). The computation of the value $m(x_0)$ is distorted by a noise term. Now, $\|g_b\|_{\varphi_{x_0}}$ can be estimated by its expectation

$$E(\|g_b\|^2_{\varphi_{x_0}}) = E\left(\int_{\varphi_{x_0}} \left(\int g(x-t)b(t)dt\right)^2 dx\right) = \int \varphi_{x_0}(x)\|g\|^2_{L^2}\sigma_b^2 dx = \|g\|^2_{L^2}\sigma_b^2. \tag{21}$$

Thus, $\|g_b\|_{\varphi_{x_0}}/\|\tau_{m(x_0)}u'\|_{\varphi_{x_0}}$ can be approximated by $\sigma_b\|g\|_{L^2}/\|\tilde{u}'\|_{\varphi_{x_0}}$, which just depends on σ_b, $\|g_b\|$ and $\tilde{u}$. This term is an approximation of the error made in the estimation of $m(x_0)$. As a consequence, equation (16) is seen as valid if this additive term can be neglected in comparison with the desired precision on the measurement $m(x_0)$.

Order of magnitude: if we take $\sigma_b \simeq 1$, $\|\tilde{u}'\|_{\varphi_{x_0}} \simeq 10$ and $\|g\|_{L^2} \simeq 0.5$ (which is the case if g is a 2-D Gaussian of standard deviation $\sigma = 0.56$), then $\sigma_b\|g\|_{L^2}/\|\tilde{u}'\|_{\varphi_{x_0}} \simeq 0.05$. In this case, equation (20) tells us that we cannot hope a better precision on $m(x_0)$ than 0.05 pixels. We can also remark that the lower the slope of $\tilde{u}$ is, the more g has to be spread in order to neglect this additive term. This confirms the property that the more constant the image, the more influent the noise.

Correlation.

The generalization of the previous proposition to the correlation case is obvious if we remark that the role played by the function $|\tilde{u}'|^2$ is now played by the density function $\|\tilde{u}\|^2_{\varphi_{x_0}} d^{\tilde{u}}_{x_0}$. Let us make things a little more precise.

Proposition 4 *Assume that u and $\tilde{u}$ satisfy relation (17), that ε and the location $m(x_0)$ at which ρ_{x_0} is maximal satisfy $|\varepsilon - m(x_0)| \ll 1$ on the support of φ_{x_0} and that*

$$\frac{\|g_b\|_{\varphi_{x_0}}}{\|\tau_m u\|_{\varphi_{x_0}}\left(< d^{\tau_m u}_{x_0}, 1 >_{\varphi_{x_0}}\right)^{1/2}} \ll 1. \tag{22}$$

Then, equation (7) holds.

Proof : See proof in appendix. □

Again, the computation of $m(x_0)$ is distorted by a noise term. In practice, the error due to the noise in the computation of $m(x_0)$ can be approximated by

$$N(\tilde{u}, g, \sigma_b, \varphi, x_0) := \frac{\sigma_b\|g\|_{L^2}}{\|\tilde{u}\|_{\varphi_{x_0}}\sqrt{< d^{\tilde{u}}_{x_0}, 1 >_{\varphi_{x_0}}}}. \tag{23}$$

This approximation of the additive "bias" indicates where the correlation makes sense, where it can be accurate, and allows one to decide which window size should be used at these locations. One can recognize the correlation curvature (defined in (14)) in the denominator of this term. This curvature plays the same role as $\|\tilde{u}'\|_{\varphi_{x_0}}$ in the L^2 case. For a given amount of noise, the larger the correlation curvature in (23), the smaller the error induced by the noise bias at x_0.

2.4 Consequences of the central equation.

2.4.1 Matching costs and reliability

The previous results point out the link between the form of the matching cost and the reliability of the disparity measured by block-matching methods. If the matching cost is reduced to a local weighted L^2-distance, relations (16) and (20) underline the importance of the image derivatives in the matching reliability. It confirms the idea that block-matching needs contrast in order to make sense.

In the case of the normalized cross correlation, the image derivatives are replaced in the equations by the correlation density $d_{x_0}^{\tilde{u}}$ (defined in (6)). The values of $d_{x_0}^{\tilde{u}}$ not only depend on $\tilde{u}$, but also on the local geometry of the pair $(\tilde{u}, \tilde{u}')$ in the neighborhood of the point x_0: the more $\tilde{u}$ and $\tilde{u}'$ are orthogonal for the scalar product $<,>_{\varphi_{x_0}}$, the larger $d_{x_0}^{\tilde{u}}$. This is not easy to interpret. For that reason, the results obtained by correlation can be considered as somewhat less reliable than those of L^2-minimization. The weaker the constraint of similarity between the images is, the less reliable the results will be when the images are only geometrically shifted. This conclusion also applies to the question of the centering of the correlation coefficient.

2.4.2 Optimal matching window.

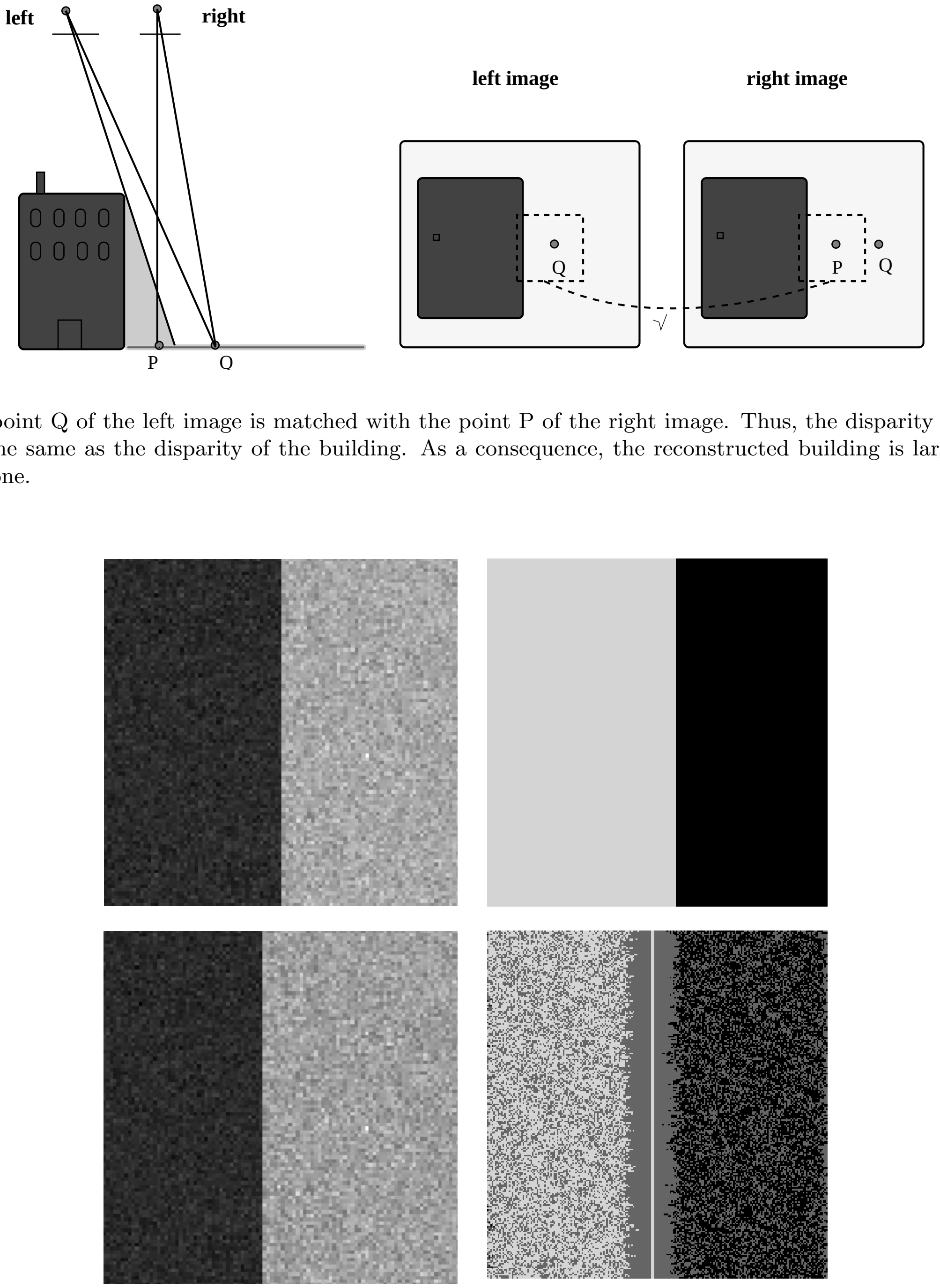
Assume that the noise standard deviation of the image is known (it can be deduced from the knowledge of the acquisition system). The images being given, we want to restrict ourselves to points x_0 at which (23) is small. In this prospect, the size of the correlation window φ can be chosen at each point in order to minimize the term (23). At the same time, this size must be as small as possible if we want the measurement $m(x_0)$ to be a good approximation of $\varepsilon(x_0)$. If all the windows used are of the form $s\varphi(sx)$ where φ is a given function (a Gaussian or a prolate spheroidal function, for instance), s can be chosen at x_0, when it is possible, as the smallest size s such that

$$N(\tilde{u}, g, \sigma_b, \varphi_s, x_0) < \alpha, \tag{24}$$

where α is the desired precision on the measurement $m(x_0)$ and N the function defined in (23). Points where this inequality can be achieved for a given size s are called **valid points**. These points are those at which the results of the correlation can be considered as reliable. We can expect the chosen size to be small at points of information (near edges) and larger in flat zones.

2.4.3 Adhesion effect reduction

Adhesion is a well-known artefact of block-matching methods. This artefact appears in the neighbourhood of depth discontinuity, especially when this discontinuity is strengthened by a grey level discontinuity. It results in a dilatation of the upper-grounds in the disparity map. It can be illustrated by the following example (see Figure 3 (a)): a textured building lies on a textured ground, in such a way that a part of the ground is occluded by the building in the left frame. One assumes that the grey level difference between the ground and the building is larger than the intensity variations in the textured areas. Let Q be a point whose distance to the building is less than half of the matching window. If we look in the right image for the best correspondent for Q, a block-matching method will probably choose P, which means that the disparity accorded to Q will be the same as the one of the building. As a consequence, the reconstructed building will be dilated by the size of a half window.



(a) The point Q of the left image is matched with the point P of the right image. Thus, the disparity assigned to Q is the same as the disparity of the building. As a consequence, the reconstructed building is larger than the real one.

(b) On the left: synthetic stereo pair, the dark part of the images corresponds to the "upper-ground". This part is shifted to the left in the second image. As a consequence, a small central strip of the ground appears in the second image and is occluded in the first one. Top right: disparity measured by correlation. The adhesion around the edge is clear, the reconstructed "upper-ground" is dilated by a half-window. Bottom right: same disparity after a barycentric correction. The gray points are those where no information remains.

Figure 3: ***Adhesion phenomenon.***

Equation (7) gives a very simple analytic explanation to the adhesion artifact. Indeed, assume that the density function $d_{x_0}^{\tau_{m(x_0)}u}$ at x_0 is in reality concentrated around a point x_1, such that $d_{x_0}^{\tau_{m(x_0)}u}$ can be well approximated by the delta function δ_{x_1}, then equation (7) yields

$$\varepsilon(x_1) \simeq m(x_0). \tag{25}$$

This means that the shift measured by correlation at x_0 is in reality the disparity of the point x_1. In two dimensions, if the neighbourhood of x_0 is composed of flat zones on both sides of an edge, the shift measured at x_0 is an average of the real disparities on the edge. This fact has no effect if the elevation has no variations over the window φ_{x_0}, but it obviously produces adhesion if the grey level edge coincides with an elevation discontinuity. This confirms the previous intuitive explanation of adhesion, and explains the dilatation of the upper-ground which can be often observed in numerical elevation models (NEM). This drawback is inherent to any block-matching process, but is particularly strong in the correlation case [5]: the L^2 similarity measure favours naturally points at which information is located, *i.e.* near the edges or in textured areas.

The explanation of this phenomena allows one to propose a practical correction: instead of assigning the measurement $m(x_0)$ to x_0, it can be assigned to the point $G(x_0)$ which is the most likely to have the disparity $m(x_0)$. This point $G(x_0)$ is computed as the barycenter of all the points of the correlation window, weighted by the values of the density function,

$$G(x_0) = \frac{< d_{x_0}^{\tilde{u}}, M >_{\varphi_{x_0}}}{< d_{x_0}^{\tilde{u}}, 1 >_{\varphi_{x_0}}}, \tag{26}$$

where $M(x)$ covers all the physical points of the support of φ_{x_0} and where $d_{x_0}^{\tilde{u}}$ is used as an approximation of the density $d_{x_0}^{\tau_{m(x_0)}u}$. In the case considered previously (when the density is concentrated at x_1), it gives $G(x_0) = M(x_1)$ and the shift measured at x_0 is correctly attributed to x_1. This procedure, called **barycentric correction**, is illustrated in a very simple case in Figures 3 (b). This correction shifts the disparities to informative points. As a consequence, some points loose their disparity, but the disparities so assigned are much more reliable.

2.4.4 On the link between baseline and precision

Let us denote by z_{real} the real depth function, and with z_{meas} the depth recovered by the correlation process. We have seen that z_{real} and the disparity function ε are linked by the relation $z_{real} = \frac{\varepsilon}{b/h}$. According to this, the equation (7) can be rewritten as

$$z_{meas}(x_0) = \frac{m(x_0)}{b/h} \simeq \frac{< z_{real}, d_{x_0}^{\tau_{m(x_0)}u} >_{\varphi_{x_0}}}{< 1, d_{x_0}^{\tau_{m(x_0)}u} >_{\varphi_{x_0}}} \simeq \frac{< z_{real}, d_{x_0}^{\tilde{u}} >_{\varphi_{x_0}}}{< 1, d_{x_0}^{\tilde{u}} >_{\varphi_{x_0}}}. \tag{27}$$

It follows that in the absence of noise, the accuracy of the measured depth does not depend on the angle between the views (hence on the b/h value). The only error encountered in this measurement is due to the bad estimation of the disparity by the correlation process and can be written

$$E_1(x_0) := \left| z_{real}(x_0) - \frac{< z_{real}, d_{x_0}^{\tilde{u}} >_{\varphi_{x_0}}}{< 1, d_{x_0}^{\tilde{u}} >_{\varphi_{x_0}}} \right|. \tag{28}$$

This ideal case clearly advocates for weak b/h, which reduce all the matching difficulties encountered with high stereoscopic coefficients.

In the real world (where images are altered with additive noise), if the angle b/h decreases too much, the results lose precision. Indeed, in the previous relation, a term due to the noise is added, and divided by b/h. Proposition 4 shows that this term can be approximated by

$$E_2(x_0, b/h) := \frac{\sigma_b \|g\|_{L^2}}{b/h \; \|\tilde{u}\|_{\varphi_{x_0}} \sqrt{< 1, d^{\tilde{u}}_{x_0} >_{\varphi_{x_0}}}}. \tag{29}$$

Two errors appear in the estimation of z_{real}: the error E_1, inherent to the block-matching process and the error due to the noise, bounded by E_2. Only the second one depends on the value b/h.

Proposition 5 *Let x_0 be a point of u, and b_0/h_0 an angle which satisfies the relation $E_1(x_0) >> E_2(x_0, b_0/h_0)$. Then, as long as $b/h \geq b_0/h_0$, the precision of the depth measured by correlation at x_0 is independent of the value of b/h.*

Following this proposition, it is absurd to increase the angle b/h while $E_1 >> E_2$. The value of E_1 clearly depends of the variations of the function z_{real}. If these variations are large on the support of φ, E_1 will predominate.

Order of magnitude: we do not have access to z_{real}, so the comparison between E_1 and E_2 is not possible in general. However, for a fixed expected accuracy on z_{meas}, the evaluation of E_2 tells us where b/h should stand. If x_0 is such that $\|\tilde{u}'\|_{\varphi_{x_0}} \simeq 10$ and if $\sigma_b \|g\|_{L^2} \simeq 0.5$, we see that $E_2 \simeq \frac{0.05}{b/h}$. This means that for a given image resolution λ (meters by pixel), the error due to the noise at x_0 will be less than $\frac{0.05\lambda}{b/h}$ meters. If the resolution of u is fifty centimeters by pixel, this error in depth will be approximately fifty centimeters for $b/h = 0.05$. This b/h value is already very small. We will see in experiments that the "acceptable" values of b/h for a given precision are much smaller than the values generally used in aerial stereocopy (where $b/h \simeq 0.8$).

In a way, this idea can be linked with some aspects of human vision. Indeed, the human eyes are very close (let say approximately $7cm$). If we look at a scene located 70cm from our eyes, the stereoscopic coefficient is already 0.1. If the distance increases to $7m$, b/h becomes 0.01. Even if stereopsis is not the only process used by the brain for reconstructing depth, the efficiency of the visual system is also supporting the use of small angles ([13]).

3 Discrete formulation and experiments.

The previous analytic study tends to rehabilitate small baseline stereovision, at least theoretically. In order to support these results, a multi-scale algorithm dedicated to small baseline stereo pairs was developed.

This section presents the outline of this algorithm and its most significant points. The discrete aspects of the procedure (that is to say sampling and interpolation) are described in depth because of their decisive influence on the matching process. Experiments on simulated and real stereo pairs follow.

3.1 Multi-Scale Algorithm

The central hypotheses of this study are that the deformation between the images of the stereo pair is purely geometric, of the form $\tilde{u}(x) = u(x + \varepsilon(x))$ with an eventual additive noise, and that the disparity function ε has small variations on the correlation window support.

This hypothesis on ε is not true in full generality. Even for small baseline stereo pairs, the variations of ε on the correlation window can still be relatively important.

In order to make this assumption valid, the correlation procedure is embedded into a discrete scale-space framework. The scale-space theory has already been used for stereovision, for example by Jones and Malik [9] or by Alvarez *et al.* [1]. The main idea here is to replace the regularization function g by a family $(g_s)_s$, where $g_s(x) = \frac{1}{s}g(x/s)$, and to refine through the scales the computation of the disparity function. At each scale s the images are sampled on an adapted grid Π_{Γ_s} such that the remaining disparity function at this scale is everywhere smaller than the pixel size.

The complete algorithm can be splitted in two phases: a learning phase, devoted to the computation of window sizes at every scale and every point, and a muti-scale matching phase which uses a sequence of given scales $(s_k)_{k=1\ldots n}$ and corresponding grids $(\Gamma_k)_{k=1\ldots n}$, Γ_0 being the roughest grid and Γ_n the finest one.

Learning phase

1. *Compute the bound (23) for each scale s, each size of window φ and each point of the grid at scale s;*

2. *For each point x_0, use this bound to determine the minimum size of the correlation window at x_0. Compute also the validity of x_0 at each scale (clearly, the larger the scale is, the larger the number of valid points is);*

3. *Compute the barycentric correction (26) at each point and each scale. This correction just depends on the images and on the optimal window computed previously.*

Multi-scale algorithm

1. *Start with the roughest scale s_0 and let $\varepsilon_0 = 0$ and $k = 0$;*

2. *Compute the image $u_k(x) = u(x + \varepsilon_k(x))$;*

3. *Use a correlation algorithm to compute the disparity map $\tilde{\varepsilon}_{k+1}$ between $(g_{s_k} * u_k)$ and $(g_{s_k} * \tilde{u})$ at each valid point of Γ_k. This step requires to use the images sampled on the grid $\Pi_{\Gamma_{k+1}}$ (see next section);*

4. *Correct $\tilde{\varepsilon}_{k+1}$ with the barycentric correction;*

5. *Let $\varepsilon_{k+1} = \tilde{\varepsilon}_{k+1} + \varepsilon_k \circ (Id + \tilde{\varepsilon}_{k+1})$. The values of the function ε_{k+1} are not known everywhere. Interpolate it (for instance by isotropic diffusion). At this point, $u \circ (Id + \varepsilon_{k+1}) = u_k \circ (Id + \tilde{\varepsilon}_{k+1})$ should be closer to $\tilde{u}$ than $u \circ (Id + \varepsilon_k)$ was.*

6. *Replace k by $k + 1$ and repeat steps 2 to 6 until the finer scale is reached.*

The actual algorithm works with dyadic scales. At each scale corresponds a sampling grid. If the sampling grid of the finest scale is Γ, the previous scale is sampled in 2Γ, etc... The largest scale, which corresponds to $2^n\Gamma$, is chosen such that $2^{n-1} \leq \|\varepsilon\|_\infty < 2^n$. This way, the real shift at the first scale is everywhere smaller than one pixel. We assume that the correction made at each scale is such that the shift map is always everywhere smaller than one pixel. Note that the finer the scale is, the larger the noise is, thus the less points will be considered as valid (in proportion). Now, all the informative points (corners, edges...) should remain valid through the scales if the density information at these points is large enough to override the noise.

3.2 Sampling and subpixelian disparities computation.

Sampling and interpolation are two critical points in stereovision. These aspects are often disregarded in spite of their decisive influence on the matching process. Subpixelian disparities can be obtained by computing the correlation map on a grid finer than the sampling grid Γ of the images. To this purpose, many algorithms estimate the correlation map at points of Γ and compute a local continuous (parabolic for instance) fit in order to refine the disparity.

Now, as shown in [18], this direct interpolation is not adequate. The correlation coefficient is not well sampled on this grid and interpolating it directly may result in the apparition of false maxima.

In real acquisition systems, the continuous image before sampling is of the form $h * O$ where O is the landscape and h the impulse response of the camera. Let S be the compact support of $\hat{h}$. Then, $h * O$ is also spectrally supported on S. Let Γ be the sampling grid and let Π_Γ be the Dirac Comb $\sum_{\gamma \in \Gamma} \delta_\gamma$. The sampled image is $u = (h * O).\Pi_\Gamma$. If we suppose that S is contained in a cell R of the dual grid, the weak form of the Shannon-Whittaker theorem [15] tells us that $h * O$ can be recovered from u via the interpolation formula:

$$h * O = u * \frac{1}{|R|} \overline{\mathcal{F}}(\mathbf{1}_R), \tag{30}$$

where $\overline{\mathcal{F}}(\mathbf{1}_R)$ denotes the inverse Fourier transform of the caracteristic function of the cell R.

Numerical consequence: Let N and D be respectively the numerator and denominator of the continuous correlation coefficient ρ. $N(m) = (\varphi_{x_0}\tilde{u}) * u(m)$, thus $\hat{N} = \widehat{\varphi_{x_0}\tilde{u}}\hat{u}$. Now, if we assume that the window φ_{x_0} has a spectral support included in the reciprocal cell R, the support of $\widehat{\varphi_{x_0}\tilde{u}}$ is in $R + R = \{x + y,\ (x, y) \in R^2\}$. It follows that if the numerator N is computed in the spectral domain, its accurate computation must be done on the grid $\Gamma/2$. This means that both images must be oversampled at least by a factor 2 before computing N in the Fourier domain. In the same way, the spectral support of D^2 is included in $S + S$. Thus, to properly reconstruct the continuous version of D^2, its discrete version must be computed in $\Gamma/2$. Finally, we can recover the continuous versions of N and D^2 thanks to their values on $\Gamma/2$, and the continuous correlation at x is just the division of $N(x)$ by $\sqrt{D^2(x)}$.

3.3 Results

A multi-scale algorithm, called MARC (Multiresolution algorithm for refined correlation) has been tested on both simulated and real stereo pairs.

The first experiments are realized from a one meter sampled orthophoto of Marseille and a precise numerical terrain model of the same area (see Figure 4) provided by the society ISTAR. In this experiment, the Shannon principle is satisfied. Indeed, the modulation transfer function of the orthophoto is spectrally supported on the reciprocal cell of the sampling grid. From this single image, several stereoscopic pairs are simulated with different $\frac{b}{h}$ values. A Gaussian white noise of standard deviation $\sigma = 1$ is added to the pairs (the images are 8 bits coded). The resulting disparity computed by the multiscale algorithm for $\frac{b}{h} = 0.025$ is shown on Figure 4. The interest of this academic example is the possibility to compare the method accuracy in function of the $\frac{b}{h}$ value.

Figure 5 shows the altitude accuracy in function of $\frac{b}{h}$ for three different versions of the correlation algorithm: the standard correlation with a rectangular window (top line), the correlation with a prolate spheroidal window (middle line) and the multiscale algorithm presented above (bottom line). Since the multiscale algorithm computes a validity value at each point of the grid image, the elevation mean square error is only computed on points which remain valid at the finest scale. As expected, if b/h becomes too small ($b/h < 0.01$), the noise error becomes dominant and the elevation error increases, then explodes when $b/h \to 0$. If b/h becomes too large ($b/h > 0.2$), the differences between the images become too large and the elevation precision also decreases. In this experiment, the accuracy of the measured depth on valid points is minimal around the value $b/h = 0.1$.

The next experiment uses a pair of 25cm aerial images of Toulouse with a b/h factor of 0.045. This pair presents a disadvantage: the large interval of time between the shots (more than 20 minutes) results in several changes due to motion or shadow shiftings. Besides, the ground truth of the area is known but incomplete, several depth informations are missing, in particular the wall surrounding the prison. For these reasons, another secondary image is simulated using the first one and the ground truth with the same b/h ratio. The new pair is shown on the first line of Figure 6. With such a small baseline, the images are very similar, they present disparities with values between -2 and 2 pixels. In a way, this similarity makes the matching process easier. On the other hand, the matching needs to be applied with subpixel accuracy, since a traditional matching algorithm computing integer disparities would yield a depth map with only 5 levels of depth. The second line of Figure 6 shows the ground truth and the result of the multiscale algorithm MARC on this pair (the third line shows the corresponding 3D projections). We can observe that the depth map computed is smoother than the ground truth. This property is the main drawback of the previous modelization and may be the price to pay to get a good accuracy almost everywhere. Figure 7 shows the result of the graph-cuts algorithm proposed by Kolmogorov et al. in [10] with a smoothness parameter $\lambda = 5$. In many global stereo algorithms, the data term naturally favours piecewise constant depth maps. This property has a noticeable advantage since it permits to get precise and sharp discontinuities. Yet, in the case of small baseline stereovision, global optimization is faced with two shortcomings. First, in order to get relevant elevation levels, the algorithm must be applied with sub-pixel precision, which is computationally expensive. This can eventually be done by oversampling images before matching (in the example of Figure 7, both images have been oversampled by a factor 2, which yields a resulting map with 7 depth levels). However, this also increases greatly the computing time. Secondly, in the case of urban areas, which present slanted surfaces, this kind of algorithm produces severe staircase effects. Now, global optimization yields all the same a very good first estimation of the elevation map and its application to small baseline stereovision should be further studied.

Figure 8 shows the results of MARC on two excerpts of a real aerial pair of images of Marseille. The images have been taken with less than a minute of difference, with a $50cm$ resolution and a b/h ratio of 0.04. The ground truth of the area is unknown. This case is particularly difficult. Indeed, a $10m$ elevation difference (which is large, even for urban areas) corresponds in these images to a $40cm$ ground disparity, which is smaller than the pixel size. The results obtained on these images are visually good, apart from a few zones of motion: several cars or buses have moved in the interval of time and result in isolated peaks in the elevation map.

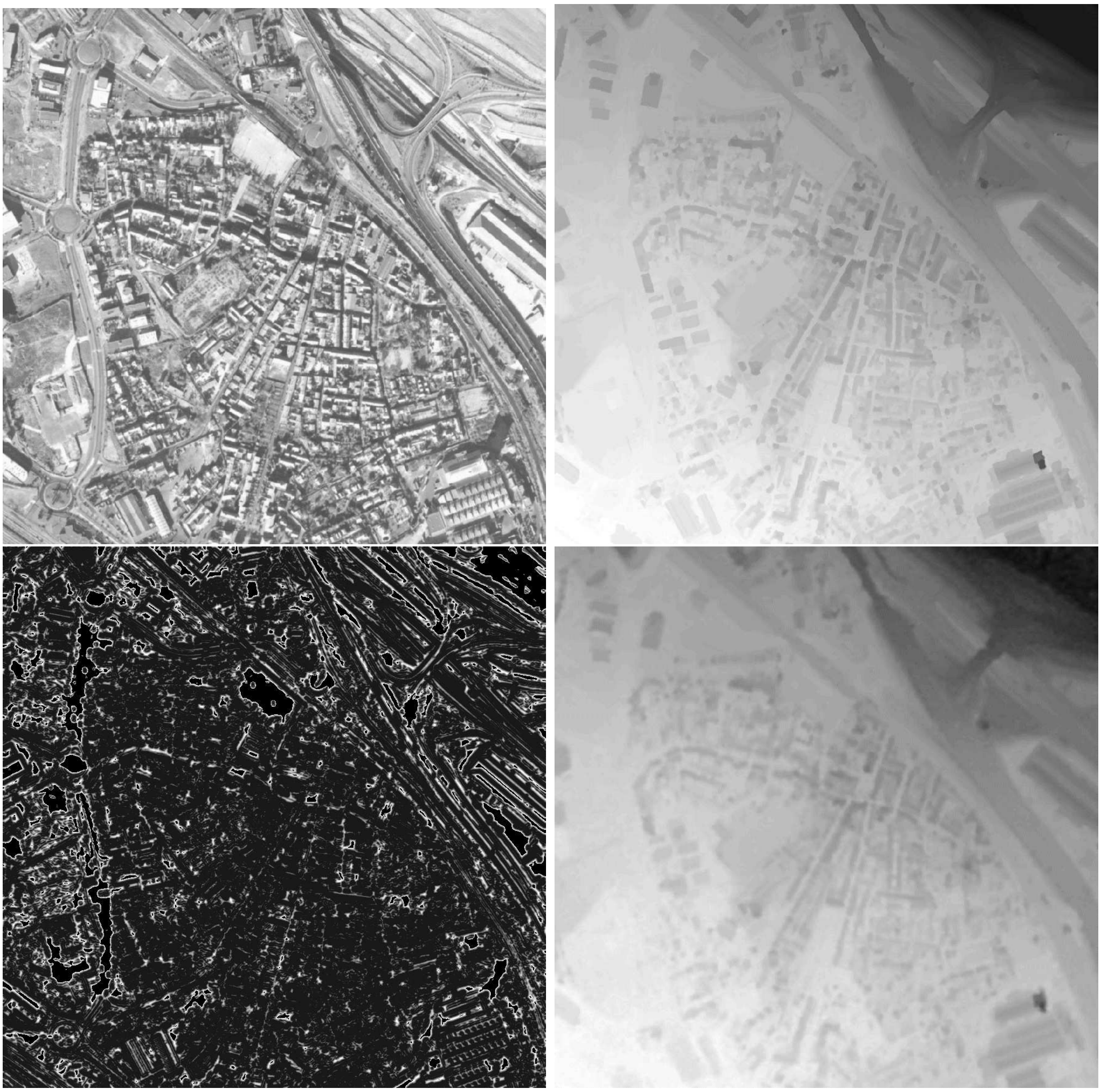


Figure 4: *First line: one meter sampling photo of Marseille and numerical terrain model of the same area, provided by ISTAR. Several stereoscopic pairs are simulated from these two images for different b/h ratios. Second line: Results of the multiscale algorithm on the pair simulated with $b/h = 0.025$. The left image shows the size of the window used at each point. The lighter the point is, the larger the window used by the algorithm was. The black points correspond to the zones where the correlation process is considered as not reliable. At these points, the disparity map is completed by isotropic diffusion. The right image shows the resulting disparity map computed by MARC.*

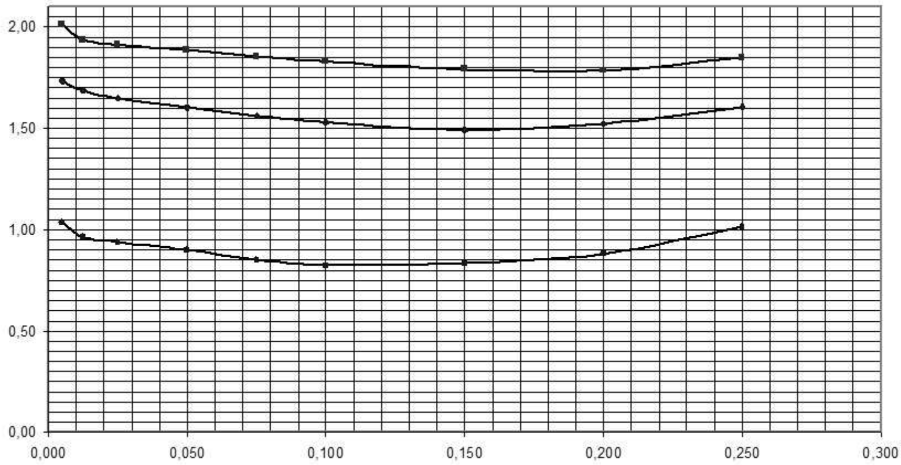


Figure 5: *Mean square error in elevation in function of the $\frac{b}{h}$ ratio for the experiment of Figure 4. The top line correspond to the results of the standard correlation algorithm, the middle line to the results obtained with a prolate spheroidal window and the bottom line to the multiscale algorithm. We observe that the simple use of the prolate spheroidal window in the standard correlation algorithm improves the results. The multiscale algorithm MARC improves the precision by a factor 2.*

4 Conclusion

Stereoscopic vision relies on the fact that when a scene is observed from two different viewpoints, the depth of a point is approximately proportional to the difference of position between its projections in both views. The proportionality coefficient is actually the tangent of the angle between the views, also called b/h. Usually, the b/h ratios used in stereo are equal to 1 or have this order of magnitude. Indeed, it is always assumed that the angle between the views has to be large to yield a good depth reconstruction. However, the difficulties of the matching process increase rapidly with the b/h factor.

In this paper, the difference between the images of the pair was assumed to be purely geometrical, up to a proportionnality coefficient variating slowly in space. This hypothesis is sound in small baseline stereovision. An analytic study of the correlation process shows that it is possible to predict where the matching results can be reliable and which range of b/h values can yield optimal results. In this range, the precision obtained by correlation matching is independent of b/h. This conclusion supports the idea that among these acceptable angles, the smallest ones, which generate fewer occlusions and much more similar images, both from the geometrical and radiometric viewpoints, are preferable. These results have given rise to a multi-scale correlation algorithm, tested on simulated and real aerial pairs. Such pairs will be available in the next satellite generation.

Acknowledgement The multi-scale algorithm, called MARC (Multiresolution algorithm for refined correlation), has been coded by Nathalie Camlong ([2]) and Vincent Muron ([11]). It is a part of the CNES patent [8]. Many thanks to Jean-Michel Morel and Andrès Almansa for their constant help, to Vincent Muron and Nathalie Camlong for the realisation of the multi-scale algorithm MARC. This work has been partially financed by the Centre National d'Etudes Spatiales (CNES),

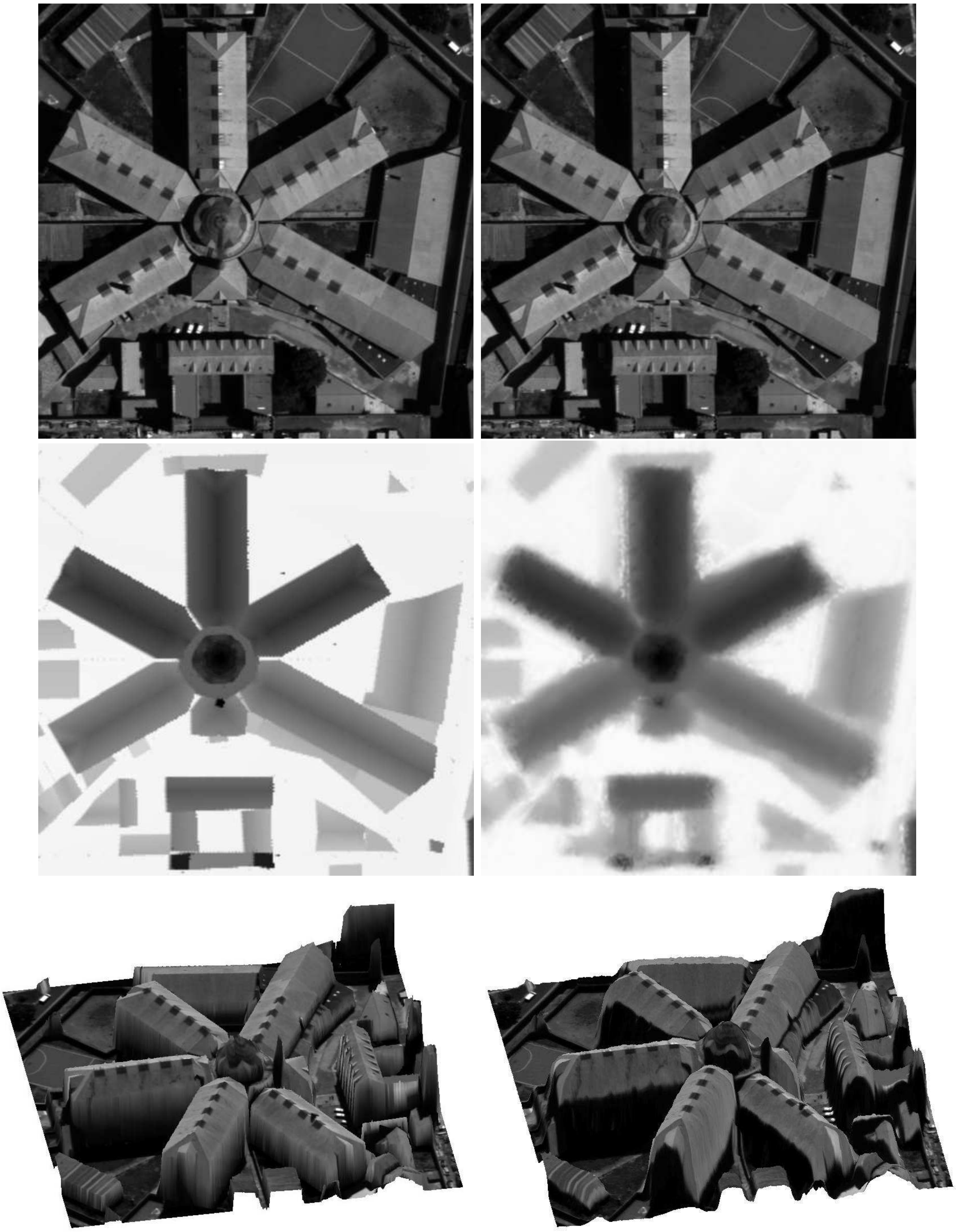


Figure 6: *First line: pair of* 512×512 *aerial images of Toulouse with a resolution of 25cm. The second image is simulated using the first one and the ground truth with a b/h ratio of* 0.045*. Second line: ground truth of the pair and result of the MARC algorithm. Third line : corresponding 3D projections.*

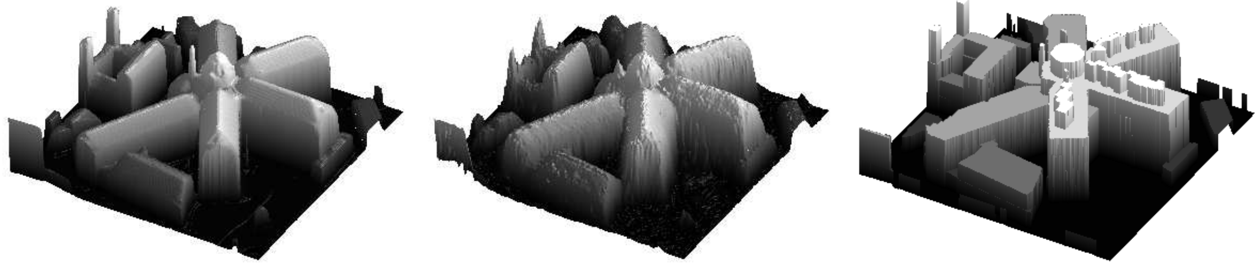


Figure 7: *Left: 3D projection of the Toulouse pair ground truth ; Middle: 3D projection of the MARC result ; Right: result of the Graph-Cuts algorithm presented by Kolmogorow et al. in [10] with a smoothness factor $\lambda = 5$ (the software used here is kindly provided by V.Kolmogorov on its web page www.adastral.ucl.ac.uk/∼vladkolm/software.html).*

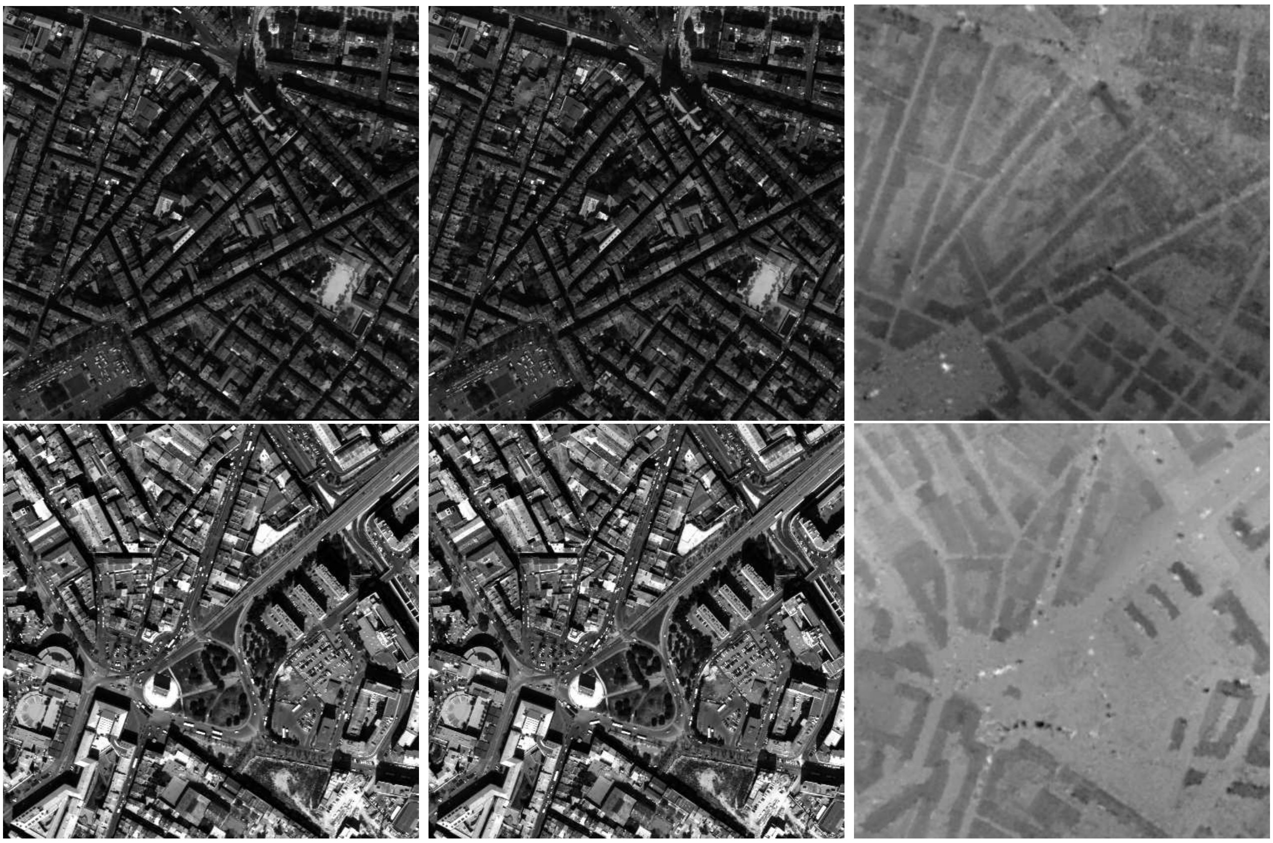


Figure 8: *First and second columns: two* 1000×1000 *excerpts of a real pair of aerial images of Marseille with a resolution of 50cm and a b/h ratio of 0.04. Third column: the corresponding MARC results.*

the Office of Naval research under grant N00014-97-1-0839, the Ministère de la Recherche et de la Technologie (projet RNRT ISII) and the Direction Générale des Armements (DGA).

References

[1] L. Alvarez, R. Deriche, J. Sanchez and J. Weickert, "Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-Space Based Approach", INRIA, 2000.

[2] N. Camlong, "Report CSSI/111-1/COR-ET-MARC-2, Description de l'algorithme MARC (Multiresolution Algorithm to Refine Correlation)", CNES, 2001.

[3] E. Carfantan and B. Rougé, "Estimation non biaisée de décalages subpixellaires sur les images SPOT", GRESTSI'01 on Signal and Image Processing, Toulouse, 2001.

[4] Ingemar J. Cox and Sunita L. Hingorani and Satish B. Rao and Bruce M. Maggs, "A maximum likelihood stereo algorithm", journal = Comput. Vis. Image Underst., vol 63, no 3, pp. 542–567, 1996.

[5] J. Delon and B. Rougé, "Le phénomène d'adhérence en stéréoscopie dépendant du critère de corrélation", GRESTSI'01 on Signal and Image Processing, Toulouse, 2001.

[6] O. Faugeras and B. Hotz and H. Metthieu and T. Vieville and Z. Zhang and P. Fua and E. Theron and L. Moll and G. Berry and J. Vuillemin and P. Bertin and C. Proy, "Real-Time Correlation Based stereo: algorithm, implementations and applications", Research report INRIA-2013, 1993.

[7] O. Faugeras and Q. - T. Luong, "The geometry of multiple images", The MIT Press, 2001.

[8] A. Giros, B. Rougé and H. Vadon, "Appariement fin d'images stéréoscopiques et instrument dédié avec un faible coefficient stéréoscopique", French Patent N° 0403143, march 2004.

[9] D.G. Jones and J. Malik, "A computational framework for determining stereo correspondence from a set of linear spatial filters", Lecture Notes in Computer Science, Vol. 588, pp. 395-410, 1992.

[10] Vladimir Kolmogorov and Ramin Zabih, "Computing Visual Correspondence with Occlusions using Graph Cuts" In IEEE International Conference on Computer Vision (ICCV), July 2001.

[11] V. Muron, "Report CSSI/111-1/COR-ET-MARC-5, Manuel utilisateur de la chaîne de calcul de décalages entre images par l'algorithme MARC", CNES, 2003.

[12] Y. Ohta and T. Kanade, "Stereo by intra-scaline search using dynamic programming", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 7, pp.139–154, 1985.

[13] S. E. Palmer, Vision Science: Photon to Phenomenology, chapter Stereoscopic Information, Bradford Book, pp. 206-221, 1999.

[14] Daniel Scharstein and Richard Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", Int. J. Comput. Vision, vol 47, no 1-3, pp 7–42, 2002.

[15] C. E. Shannon, "A mathematical theory of communication", The Bell System technical journal, Vol.27, pp.379-423, 1948.

[16] C. E. Shannon, "Communication in the Presence of Noise", Proceedings of the Institution of Radio Engineers, Vol.37, pp. 10-21, 1949.

[17] J. Sun, N. Zheng, and H. Shum. Stereo matching using belief propagation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(07):787–800, 2003.

[18] R. Szeliski and D. Scharstein, "Sampling the Disparity Space Image", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004.

5 Appendix

Proof of Proposition 2

We have

$$\rho'_{x_0}(m) = \frac{< \tau_m u', \tilde{u} >_{\varphi_{x_0}}}{\|\tau_m u\|_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}} - \frac{< \tau_m u, \tilde{u} >_{\varphi_{x_0}} < \tau_m u', \tau_m u >_{\varphi_{x_0}}}{\|\tau_m u\|^3_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}}. \tag{31}$$

Thus,

$$\begin{aligned}
\rho''_{x_0}(m) &= \frac{< \tau_m u'', \tilde{u} >_{\varphi_{x_0}}}{\|\tau_m u\|_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}} - 2\frac{< \tau_m u', \tilde{u} >_{\varphi_{x_0}} < \tau_m u, \tau_m u' >_{\varphi_{x_0}}}{\|\tau_m u\|^3_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}} \\
&- \frac{< \tau_m u, \tilde{u} >_{\varphi_{x_0}} < \tau_m u, \tau_m u'' >_{\varphi_{x_0}}}{\|\tau_m u\|^3_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}} - \frac{< \tau_m u, \tilde{u} >_{\varphi_{x_0}} < \tau_m u', \tau_m u' >_{\varphi_{x_0}}}{\|\tau_m u\|^3_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}} \\
&+ 3\frac{< \tau_m u, \tilde{u} >_{\varphi_{x_0}} < \tau_m u, \tau_m u' >^2_{\varphi_{x_0}}}{\|\tau_m u\|^5_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}}.
\end{aligned}$$

But at $m = m(x_0)$, $\rho'_{x_0}(m) = 0$, consequently

$$\begin{aligned}
\rho''_{x_0}(m) &= \frac{< \tau_m u'', \tilde{u} >_{\varphi_{x_0}} \|\tau_m u\|^2_{\varphi_{x_0}} + < \tau_m u', \tilde{u} >_{\varphi_{x_0}} < \tau_m u, \tau_m u' >_{\varphi_{x_0}}}{\|\tau_m u\|^3_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}} \\
&- \frac{< \tau_m u, \tilde{u} >_{\varphi_{x_0}} < \tau_m u, \tau_m u'' >_{\varphi_{x_0}} + < \tau_m u, \tilde{u} >_{\varphi_{x_0}} \|\tau_m u'\|^2_{\varphi_{x_0}}}{\|\tau_m u\|^3_{\varphi_{x_0}} \|\tilde{u}\|_{\varphi_{x_0}}}.
\end{aligned}$$

Replacing $\tilde{u}$ by its first order approximation $\tau_{m(x_0)}u$, it finally gives

$$\rho''_{x_0}(m(x_0)) \simeq \frac{< \tau_{m(x_0)}u, \tau_{m(x_0)}u' >^2_{\varphi_{x_0}} - \|\tau_{m(x_0)}u\|^2_{\varphi_{x_0}} \|\tau_{m(x_0)}u'\|^2_{\varphi_{x_0}}}{\|\tau_{m(x_0)}u\|^4_{\varphi_{x_0}}} = -< d_{x_0}^{\tau_{m(x_0)}u}, 1 >_{\varphi_{x_0}} . \tag{32}$$

Proof of Proposition 4

The proof is similar to the one of Proposition 3. Indeed,

$$\rho'_{x_0}(m) = 0 \iff \|\tau_m u\|^2_{\varphi_{x_0}} < \tau_m u', \tilde{u} >_{\varphi_{x_0}} = < \tau_m u, \tilde{u} >_{\varphi_{x_0}} < \tau_m u', \tau_m u >_{\varphi_{x_0}} .$$

Let us define the function

$$w^m(x) = \frac{\|\tau_m u\|^2_{\varphi_{x_0}} \tau_m u'(x) - \tau_m u(x) < \tau_m u, \tau_m u' >}{\|\tau_m u\|^2_{\varphi_{x_0}} < \tau_m u', \tau_m u' >_{\varphi_{x_0}} - < \tau_m u, \tau_m u' >^2_{\varphi_{x_0}}}. \tag{33}$$

The function w^m clearly satisfies

$$< w^m, \tau_m u >_{\varphi_{x_0}} = 0, \quad < w^m, \tau_m u' >_{\varphi_{x_0}} = 1, \quad \text{and} \tag{34}$$

$$\|w^m\|^2_{\varphi_{x_0}} = \frac{1}{\|\tau_m u\|^2_{\varphi_{x_0}} < d^{\tau_m u}_{x_0}, 1 >_{\varphi_{x_0}}}. \tag{35}$$

Now, the equation $\rho'_{x_0}(m) = 0$ can be rewritten

$$< w^m, u(x + \varepsilon(x)) >_{\varphi_{x_0}} + < w^m, g_b >_{\varphi_{x_0}} = 0. \tag{36}$$

The first order expansion of this equality gives

$$< w^m, \tau_m u + (\varepsilon(x) - m)\tau_m u' >_{\varphi_{x_0}} + < w^m, g_b >_{\varphi_{x_0}} \simeq 0. \tag{37}$$

Thus

$$m \simeq \frac{< d^{\tau_m u}_{x_0}, \varepsilon(x) >_{\varphi_{x_0}}}{< d^{\tau_m u}_{x_0}, 1 >_{\varphi_{x_0}}} + < w^m, g_b >_{\varphi_{x_0}}, \tag{38}$$

The second term can be bounded from above thanks to Schwarz inequality (footnote (7)),

$$< w^m, g_b >_{\varphi_{x_0}} \leq \frac{\|g_b\|_{\varphi_{x_0}}}{\|\tau_m u\|_{\varphi_{x_0}} \left(< d^{\tau_m u}_{x_0}, 1 >_{\varphi_{x_0}}\right)^{1/2}}. \tag{39}$$

If this quantity is small enough in comparison with the desired precision on m, equation (7) holds.