Fattening free correlation algorithms

Gwendoline Blanchet, Antoni Buades, Bartomeu Coll, Jean-Michel Morel, Bernard Rougé

Fattening free correlation algorithms

G. Blanchet * A. Buades † B. Coll ‡ J.M Morel §
B. Rouge ¶

March 23, 2010

Abstract

Block matching along epipolar lines is the core of most stereovision algorithms in geographic information systems. The usual distances between blocks are the sum of squared distances in the block (SSD) or the correlation. These distances suffer the adhesion (or fattening) effect, a defect by which the center of the block inherits the disparity of the more contrasted pixels in the block. This report shows that there is a simple and universal solution to this problem. It is enough to use an adaptive weight in the SSD. This weight is nothing but the square of the gradient of the first image in the epipolar direction. This magic adaptive weight yields a computed disparity which is the result of a convolution of the real disparity with a fixed kernel. The choice of the kernel is left to the user. Experiments on simulated and real pairs prove that the formula applies really, and eliminates surface bumps clearly due to the adhesion phenomenon. [1]

1 Introduction

Stereovision consists in finding the depth of a scene from several views of it. This is one of the central problems in computer vision, and it has been the object of research for the last thirty years. Stereovision is based on the fact that differences of depth in a 3D scene create geometrical disparities between different views of the same scene if they are taken from different points of view.

The corresponding problem or disparity computation between two stereo pair images u and v, reduces to the search of a disparity functions ϵ such that $u(x) = v(x + \epsilon(x))$. As in motion estimation, the above equation presents the aperture problem, namely the ambiguity of the solution, even when some regularity is demanded for the disparity. For this reason, many stereovision algorithms do not look for the function ϵ matching the grey level intensity of each pixel but match the grey level of an entire block around each pixel. The

*CNES, 18 avenue Edouard Belin, Toulouse 31401, France
†MAP5, CNRS - Universite Paris Descartes, 45 rue Saints Peres, 75270 Paris Cedex 06, France
‡Dpt Matematiques Informatica, Universitat Illes Balears, Ctra Valldemossa km 7.5, Palma de Mallorca, Spain
§CMLA, ENS Cachan, 61 av. President Wilson, Cachan 94235, France
¶CESBIO, 18 Av E. Belin, Toulouse 31400, France
[1] A Spanish patent has been applied for by Universitat de Illes Balears [1]

resulting algorithm is known as block matching, SSD (sum of squared distances) or correlation.

The most important drawback of correlation is the well known "fattening effect". According to Kanade et Okutomi [6], "A central problem in stereo matching by computing correlation or sum of squared differences (SSD) lies in selecting an appropriate window size. The window size must be large enough to include enough intensity variation for reliable matching, but small enough to avoid the effects of projective distortion. If the window is too small and does not cover enough intensity variation, it gives a poor disparity estimate, because the signal (intensity variation) to noise ratio is low. If, on the other hand, the window is too large and covers a region in which the depth of scene points (i.e. disparity) varies, then the position of maximum correlation or minimum SSD may not represent correct matching due to different projective distortion in the left and right images. The fattening effect occurs when the selected window contains pixels at different depth. In that case we cannot find exactly the same window and the obtained disparity depends on the different disparities of the window and not only the central pixel itself."

The usual way to cope with the fattening effect is to use adaptive windows that avoid image discontinuities as was first proposed by Kanade et al [6]. Similar works pre-computing edge points and recursively growing a comparison window avoiding them were proposed by Lotti et al. [7] and recently by Wang et al. [16]. Patricio et al. [10] and Yoon et al. [18] select an adaptive window containing only pixels with a grey level similar to the reference one, like in neighborhood and bilateral filters [14, 17].

Other approaches do not try to avoid the discontinuities of the image. They select an adaptive window with a minimum distance criterium. The subjacent idea is that windows which do not contain discontinuities will be matched with a small window distance. Fusiello et al. [4] choose among all the windows containing the reference pixel the one which has a minimal distance with its corresponding one in the second image. Veksler [15] applied the same strategy but including square windows of different sizes. A more elaborated version by Hirschmuller et al [5] adapts the shape of the window by dividing the correlation window into small sub-windows and taking those which attain the minimum distance. The Delon et al. [3] paper proposes a different strategy, the barycentric correction attributing the disparity of a window to the window barycenter pondered by the image gradients.

Point feature matching methods overcome the fattening problem at the cost of a drastic reduction of the match density. Matched features can also be curvilinear, which also circumvents the fattening problem to some extent. For instance, Schmid [13] describes a set of algorithms for automatically matching individual line segments and curves. Robert [11] presents an edge-based stereovision algorithm, where the primitives to be matched are cubic B-splines approximations of the 2-D edges. Muse et al. [9] and Cao et al. [2] discuss how to automatically match pieces of level lines and extract coherent groups of such matches. Matas et al. [8] solves the problem by matching stable and homogeneous image regions, but their match set is again sparse. Even if features may seem more local, they depend anyway on a broad neighborhood. It is true that the fine scale Laplacian extrema used (e.g.) in the SIFT method are very local, but their descriptor around involves anyway a 8×8 window. Thus, if this window contains some edge, the fattening problem can occur anyway.

The fattening effect is not the sole obstacle to a correct disparity computation. Occlusions and moving objects make it a very difficult and sometimes ill-posed problem. Taking simultaneous snapshots with a low baseline avoids partially these drawbacks. However, when using a low baseline a larger precision in the disparity computation is needed to get the same depth precision. The use of a low B/H (where B is the baseline and H is the altitude) was proposed in satellite imaging by Delon and Rouge [3].

2 Mathematical analysis of correlation algorithm

Let us denote by $\mathbf{x} = (x, y)$ an image point in the continuous image domain, and by $u_1(\mathbf{x}) = u_1(x, y)$ and $u_2(\mathbf{x})$ the images of an ortho-rectified stereo pair. Assume that the epipolar direction is the x axis. The underlying depth map can be deduced from the disparity function $\varepsilon(\mathbf{x})$ giving the shift of an observed physical point $\mathbf{x}$ from the left image u_1 in the right image u_2. The physical disparity $\varepsilon(\mathbf{x})$ is not well-sampled. Therefore, it cannot be recovered at all points, but only essentially at points $\mathbf{x}$ around which the depth map is continuous. Following the formulation by Delon and Rouge [3] and Sabater [12], around such points, a deformation model holds:

$$\begin{aligned} u_1(\mathbf{x}) &= u(x + \varepsilon(\mathbf{x}), y) + n_1(\mathbf{x}) \\ u_2(\mathbf{x}) &= u(\mathbf{x}) + n_2(\mathbf{x}). \end{aligned} \tag{1}$$

Block matching amounts to finding the disparity at $\mathbf{x}_0$ minimizing

$$e_{\mathbf{x}_0}(\mu) = \int_{[0,N]^2} \varphi(\mathbf{x} - \mathbf{x}_0)\big(u_1(\mathbf{x}) - u_2(\mathbf{x} + (\mu, 0))\big)^2 d\mathbf{x}. \tag{2}$$

where $\varphi(\mathbf{x}-\mathbf{x}_0)$ is a soft window function centered at $\mathbf{x}_0$. For a sake of compactness in notation, $\varphi_{\mathbf{x}_0}(\mathbf{x})$ stands for $\varphi(\mathbf{x}-\mathbf{x}_0)$, $\int_{\varphi_{\mathbf{x}_0}} u(\mathbf{x})d\mathbf{x}$ will be an abbreviation for $\int \varphi(\mathbf{x} - \mathbf{x}_0)u(\mathbf{x})d\mathbf{x}$; we will write $u(\mathbf{x} + \mu)$ for $u(\mathbf{x} + (\mu, 0))$ and ε for $\varepsilon(\mathbf{x})$. The minimization problem (2) rewrites

$$\min_{\mu} \int_{\varphi_{\mathbf{x}_0}} \big(u(\mathbf{x} + \varepsilon(\mathbf{x})) + n_1(\mathbf{x}) - u(\mathbf{x} + \mu) - n_2(\mathbf{x} + \mu)\big)^2 d\mathbf{x}.$$

Differentiating this energy with respect to μ implies that any local minimum $\mu = \mu(\mathbf{x}_0)$ satisfies

$$\int_{\varphi_{\mathbf{x}_0}} \Big(u(\mathbf{x}+\varepsilon(\mathbf{x}))+n_1(\mathbf{x})-u(\mathbf{x}+\mu)-n_2(\mathbf{x}+\mu)\Big)\times\Big(u_x(\mathbf{x}+\mu)+(n_2)_x(\mathbf{x}+\mu)\Big)\, d\mathbf{x} = 0. \tag{3}$$

One has by Taylor-Lagrange formula $u_x(\mathbf{x} + \mu) = (u_x(\mathbf{x} + \varepsilon)) + O_1(\mu - \varepsilon)$, with

$$O_1(\mu - \varepsilon) \leq |\mu - \varepsilon| \max |u_{xx}(\mathbf{x} + \varepsilon)| \tag{4}$$

and $u(\mathbf{x} + \varepsilon(\mathbf{x})) - u(\mathbf{x} + \mu) = u_x(\mathbf{x} + \varepsilon)(\varepsilon - \mu) + O_2((\varepsilon - \mu)^2)$, where

$$|O_2((\varepsilon - \mu)^2)| \leq \frac{1}{2} \max |(u_{xx}(\mathbf{x} + \varepsilon))|(\varepsilon - \mu)^2 .$$

Thus equation (3) yields

$$\int_{\varphi_{\mathbf{x}_0}} \Big(u_x(\mathbf{x}+\varepsilon)(\varepsilon-\mu)+O_2((\varepsilon-\mu)^2)+n_1(\mathbf{x})-n_2(\mathbf{x}+\mu)\Big)\times \Big(u_x(\mathbf{x}+\varepsilon)+O_1(\mu-\varepsilon)+(n_2)_x(\mathbf{x}+\mu)\Big)\,d\mathbf{x}=0. \quad (5)$$

and therefore

$$\mu\int_{\varphi_{\mathbf{x}_0}}(u_x(\mathbf{x}+\varepsilon))^2 d\mathbf{x}=\int_{\varphi_{\mathbf{x}_0}}(u_x(\mathbf{x}+\varepsilon))^2\,\varepsilon(\mathbf{x})\,d\mathbf{x}+\tilde{\mathcal{A}}+\tilde{\mathcal{B}}+\mathcal{O}_1+\mathcal{O}_2, \quad (6)$$

where

$$\begin{aligned}
\tilde{\mathcal{A}} &= \int_{\varphi_{\mathbf{x}_0}} u_x(\mathbf{x}+\varepsilon)\big(n_1(\mathbf{x})-n_2(\mathbf{x}+\mu)\big)d\mathbf{x}; && (7)\\
\tilde{\mathcal{B}} &= \int_{\varphi_{\mathbf{x}_0}} \big(n_1(\mathbf{x})-n_2(\mathbf{x}+\mu)\big)(n_2)_x(\mathbf{x}+\mu)d\mathbf{x}; && (8)\\
\mathcal{O}_1 &= \int_{\varphi_{\mathbf{x}_0}} u_x(\mathbf{x}+\varepsilon)(\varepsilon-\mu)(n_2)_x(\mathbf{x}+\mu)d\mathbf{x}\\
&\quad+\int_{\varphi_{\mathbf{x}_0}} O_1(\mu-\varepsilon)\big(n_1(\mathbf{x})-n_2(\mathbf{x}+\mu)\big)d\mathbf{x}; && (9)\\
\mathcal{O}_2 &= \int_{\varphi_{\mathbf{x}_0}} O_2(\varepsilon-\mu)^2(u_x(\mathbf{x}+\varepsilon))d\mathbf{x}\\
&\quad+\int_{\varphi_{\mathbf{x}_0}} O_2(\varepsilon-\mu)^2[O_1(\mu-\varepsilon)+(n_2)_x(\mathbf{x}+\mu)]d\mathbf{x}\\
&\quad+\int_{\varphi_{\mathbf{x}_0}} O_1(\mu-\varepsilon)(u_x(\mathbf{x}+\varepsilon))(\varepsilon-\mu)d\mathbf{x}. && (10)
\end{aligned}$$

Denote by $\overline{\varepsilon}$ the average of ε on the support of $\varphi(\mathbf{x}-\mathbf{x}_0)$, denoted by $B_{\mathbf{x}_0}$. By the Taylor-Lagrange theorem we have

$$\tilde{\mathcal{A}}=\mathcal{A}+\mathcal{O}_{\mathcal{A}}$$

where

$$\mathcal{A}=\int_{\varphi_{\mathbf{x}_0}} u_x(\mathbf{x}+\varepsilon)\big(n_1(\mathbf{x})-n_2(\mathbf{x}+\overline{\varepsilon})\big)d\mathbf{x} \quad (11)$$

and

$$\mathcal{O}_{\mathcal{A}}=(\overline{\varepsilon}-\mu)\int_{\varphi_{\mathbf{x}_0}}(u_x(\mathbf{x}+\varepsilon))(n_2)_x(\mathbf{x}+\tilde{\varepsilon}(\mathbf{x}))d\mathbf{x}, \quad (12)$$

where $\tilde{\varepsilon}(\mathbf{x})$ satisfies $\tilde{\varepsilon}(\mathbf{x})\in[\min(\mu,\overline{\varepsilon}),\max(\mu,\overline{\varepsilon})]$. In the same way,

$$\tilde{\mathcal{B}}=\int_{\varphi_{\mathbf{x}_0}} \big(n_1(\mathbf{x})-n_2(\mathbf{x}+\mu)\big)(n_2)_x(\mathbf{x}+\mu)d\mathbf{x}$$

so that $\tilde{\mathcal{B}}=\mathcal{B}+\mathcal{O}_{\mathcal{B}}$, where

$$\mathcal{B} = \int_{\varphi_{\mathbf{x}_0}} \big(n_1(\mathbf{x}) - n_2(\mathbf{x}+\overline{\varepsilon})\big)(n_2)_x(\mathbf{x}+\overline{\varepsilon})d\mathbf{x} \tag{13}$$

and

$$\mathcal{O}_{\mathcal{B}} = (\mu - \overline{\varepsilon}) \int_{\varphi_{\mathbf{x}_0}} n_1(\mathbf{x})(n_2)_{xx}(\mathbf{x}+\tilde{\varepsilon}(\mathbf{x})) - (n_2(n_2)_x)_x(\mathbf{x}+\tilde{\varepsilon}(\mathbf{x}))d\mathbf{x}. \tag{14}$$

The terms $\mathcal{A}$ and $\mathcal{B}$ are stochastic and we must estimate their expectation and variance. The terms $\mathcal{O}_1$, $\mathcal{O}_2$, $\mathcal{O}_{\mathcal{A}}$, $\mathcal{O}_{\mathcal{B}}$ are higher order terms with respect to $\varepsilon - \mu$ and are negligible if $\varepsilon - \mu$ is small, and the noise samples bounded.

Lemma 1 *Consider the main error terms*

$$\mathcal{A} = \int_{\varphi_{\mathbf{x}_0}} u_x(\mathbf{x}+\varepsilon(\mathbf{x}))\big(n_1(\mathbf{x}) - n_2(\mathbf{x}+\overline{\varepsilon})\big)d\mathbf{x}$$

and

$$\mathcal{B} = \int_{\varphi_{\mathbf{x}_0}} \big(n_1(\mathbf{x}) - n_2(\mathbf{x}+\overline{\varepsilon})\big)(n_2)_x(\mathbf{x}+\overline{\varepsilon})d\mathbf{x}$$

as defined above. One has $\mathbf{E}\mathcal{A} = \mathbf{E}\mathcal{B} = 0$ *and*

$$\begin{aligned}
\mathrm{Var}(\mathcal{A}) &= 2\sigma^2 \int [\varphi(\mathbf{x}-\mathbf{x}_0)u_x(\mathbf{x}+\varepsilon)]_N^2 \, d\mathbf{x} \\
&\leq 2\sigma^2 \int \varphi(\mathbf{x}-\mathbf{x}_0)^2 u_x(\mathbf{x}+\varepsilon)^2; \\
\mathrm{Var}(\mathcal{B}) &\leq \frac{2\pi^2\sigma^4}{3} \int \varphi(\mathbf{x}-\mathbf{x}_0)^2 d\mathbf{x} + \sigma^4 \int \varphi_x(\mathbf{x}-\mathbf{x}_0)^2 d\mathbf{x}.
\end{aligned}$$

Proof: Notice that $n_1(\mathbf{x})$ and $n_2(\mathbf{x}+\overline{\varepsilon})$ are independent Gaussian noises with variance σ^2. Thus their difference is again a Gaussian noise with variance $2\sigma^2$. It therefore follows that

$$\mathrm{Var}(\mathcal{A}) = 2\sigma^2 \int [\varphi(\mathbf{x}-\mathbf{x}_0)u_x(\mathbf{x}+\varepsilon))]_N^2 \, d\mathbf{x} \leq 2\sigma^2 \int \varphi(\mathbf{x}-\mathbf{x}_0)^2 (u_x(\mathbf{x}+\varepsilon))^2 d\mathbf{x}.$$

$$\begin{aligned}
\mathrm{Var}(\mathcal{B}) &\leq 2\left[\mathrm{Var}(\int_{\varphi_{\mathbf{x}_0}} n_1(\mathbf{x})(n_2)_x(\mathbf{x}+\overline{\varepsilon}) + \mathrm{Var}(\int_{\varphi_{\mathbf{x}_0}} n_2(\mathbf{x}+\overline{\varepsilon})(n_2)_x(\mathbf{x}+\overline{\varepsilon}))\right] \\
&\leq 2\left[\sigma^2 \times \frac{\pi^2\sigma^2}{3} \int \varphi^2(\mathbf{x}-\mathbf{x}_0) + \frac{\sigma^4}{2} \int \varphi_x(\mathbf{x}-\mathbf{x}_0)^2\right] \\
&= \frac{2\pi^2\sigma^4}{3} \int \varphi(\mathbf{x}-\mathbf{x}_0)^2 + \sigma^4 \int \varphi_x(\mathbf{x}-\mathbf{x}_0)^2.
\end{aligned}$$

Theorem 1 (Main disparity formula and exact noise error estimate) *Consider an optimal disparity* $\mu(\mathbf{x}_0)$ *obtained as any absolute minimizer of* $e_{\mathbf{x}_0}(\mu)$ *(defined by* (2)*). Then*

$$\mu(\mathbf{x}_0) = \frac{\int_{\varphi_{\mathbf{x}_0}} [u_x(\mathbf{x}+\varepsilon(\mathbf{x}))]^2 \varepsilon(\mathbf{x}) d\mathbf{x}}{\int_{\varphi_{\mathbf{x}_0}} [u_x(\mathbf{x}+\varepsilon(\mathbf{x}))]^2 d\mathbf{x}} + \mathcal{E}_{\mathbf{x}_0} + \mathcal{F}_{\mathbf{x}_0} + \mathcal{O}_{\mathbf{x}_0} \tag{15}$$

where

$$\mathcal{E}_{\mathbf{x}_0} = \frac{\int_{\varphi_{\mathbf{x}_0}} (u_x(\mathbf{x}+\varepsilon(\mathbf{x}))\big(n_1(\mathbf{x}) - n_2(\mathbf{x}+\overline{\varepsilon})\big)d\mathbf{x}}{\int_{\varphi_{\mathbf{x}_0}} [u_x(\mathbf{x}+\varepsilon(\mathbf{x}))]^2 d\mathbf{x}}$$

is the dominant noise term,

$$\mathcal{F}_{\mathbf{x}_0} = \frac{\int_{\varphi_{\mathbf{x}_0}} \big(n_1(\mathbf{x}) - n_2(\mathbf{x}+\overline{\varepsilon})\big)(n_2)_x(\mathbf{x}+\overline{\varepsilon})d\mathbf{x}}{\int_{\varphi_{\mathbf{x}_0}} [u_x(\mathbf{x}+\varepsilon(\mathbf{x}))]^2 d\mathbf{x}}$$

and $\mathcal{O}_{\mathbf{x}_0}$ is made of smaller terms. In addition the variances of the main error terms due to noise satisfy

$$\mathrm{Var}(\mathcal{E}_{\mathbf{x}_0}) = 2\sigma^2 \frac{\int [\varphi(\mathbf{x}-\mathbf{x}_0)u_x(\mathbf{x}+\varepsilon)]_N^2 \, d\mathbf{x}}{\Big(\int \varphi(\mathbf{x}-\mathbf{x}_0)u_x(\mathbf{x}+\varepsilon)^2 d\mathbf{x}\Big)^2}; \tag{16}$$

$$\mathrm{Var}(\mathcal{F}_{\mathbf{x}_0}) \leq \frac{\frac{2\pi^2}{3}\sigma^4 \int \varphi(\mathbf{x}-\mathbf{x}_0)^2 d\mathbf{x} + \sigma^4 \int \varphi_x(\mathbf{x}-\mathbf{x}_0)^2 d\mathbf{x}}{\Big(\int \varphi(\mathbf{x}-\mathbf{x}_0)u_x(\mathbf{x}+\varepsilon)^2 d\mathbf{x}\Big)^2}. \tag{17}$$

Finally,

$$\mathcal{O}_{\mathbf{x}_0} = \frac{\mathcal{O}_1 + \mathcal{O}_2 + \mathcal{O}_{\mathcal{A}} + \mathcal{O}_{\mathcal{B}}}{\int_{\varphi_{\mathbf{x}_0}} [u_x(\mathbf{x}+\varepsilon(\mathbf{x}))]^2 d\mathbf{x}},$$

and

$$\mathbf{E}\mathcal{O}_{\mathbf{x}_0} = O(\max_{\mathbf{x}\in B_{x_0}} |\varepsilon(\mathbf{x}) - \mu|),$$

$$\mathrm{Var}(\mathcal{O}_{\mathbf{x}_0}) = O(\max_{\mathbf{x}\in B_{x_0}} |\varepsilon(\mathbf{x}) - \mu|^2).$$

Proof: This result is an immediate consequence of (6) completed with the variance estimates in Lemma 1. The estimates for the higher order terms $\mathcal{O}$ are a straightforward application of Cauchy-Schwartz inequality.

Remark Theorem 1 makes sense only when the optimal disparity $\mu(\mathbf{x}_0)$ is consistent, namely satisfies for $\mathbf{x}$ in the support $B_{\mathbf{x}_0}$ of $\varphi(\mathbf{x}-\mathbf{x}_0)$,

$$|\varepsilon(\mathbf{x}) - \mu(\mathbf{x}_0)| << 1. \tag{18}$$

Thus, one of the main steps of block matching must be to eliminate inconsistent matches.

Remark In all treated examples, it will be observed that $\mathrm{Var}(\mathcal{B}) \ll \mathrm{Var}(\mathcal{A})$, which by Lemma 1 directly follows from

$$\sigma^2 \left[\frac{2\pi^2}{3}\int \varphi(\mathbf{x}-\mathbf{x}_0)^2 + \int \varphi_x(\mathbf{x}-\mathbf{x}_0)^2\right] \ll 2\int [\varphi(\mathbf{x}-\mathbf{x}_0)u_x(\mathbf{x}+\overline{\varepsilon})]_N^2. \tag{19}$$

3 Proposed strategy: avoiding the fattening problem

The previous mathematical formulation tells us that the obtained minimizer for the correlation problem satisfies

$$\mu(\mathbf{x}_0) = \frac{\int_{\varphi_{\mathbf{x}_0}} [u_x(\mathbf{x} + \varepsilon(\mathbf{x}))]^2 \varepsilon(\mathbf{x}) d\mathbf{x}}{\int_{\varphi_{\mathbf{x}_0}} [u_x(\mathbf{x} + \varepsilon(\mathbf{x}))]^2 d\mathbf{x}} \tag{20}$$

up to the noise terms. That is, the obtained minimizer will be an barycenter of the disparities at each point in the correlation window weighted by its square gradient.

Therefore, whenever a pixel or a cluster of pixels have a large gradient with respect to their neighboring ones, the estimated disparity for these neighboring pixels will be obtained by combining mainly the disparities of these few very contrasted pixels. It can even happen that a single pixel dominates the estimated disparity for all of its neighboring ones. This effect is mainly noticeable near image edges, where a line of pixels dominates the correlation of all their neighboring ones. Yet, the fattening effect happens everywhere in some degree, because the gradient barycenter is never exactly the center of the correlation window. Even if this is not very noticeable when looking at the disparity image, this effect becomes conspicuous when looking at the 3D reconstruction of the estimated depth, (Fig. 4).

There are not many ways to avoid this: to remove the disparity imbalance in the window, we shall compensate the effect of the squared gradients in the above integral by directly modifying the values of the window function φ, making it adaptive. By taking $\varphi_{x_0}(\mathbf{x}) = \frac{\rho_{\mathbf{x}_0}(\mathbf{x})}{u_x(\mathbf{x}+\varepsilon(\mathbf{x}))^2}$ in equation (20) we obtain

$$\mu(\mathbf{x}_0) = \frac{\int_{\rho_{\mathbf{x}_0}} \varepsilon(\mathbf{x}) d\mathbf{x}}{\int_{\rho_{\mathbf{x}_0}} d\mathbf{x}}, \tag{21}$$

which is equivalent to

$$\mu(\mathbf{x}_0) = \int \rho(\mathbf{x} - \mathbf{x}_0) \varepsilon(\mathbf{x}) d\mathbf{x},$$

since the function ρ is normalized to have the integral equal to one. Thus, we obtain a weighted average of all disparities in the correlation neighborhood, but not weighted by the image gradient. Therefore, the computed disparity is the convolution of the ground truth disparity ε with an isotropic kernel which can be fixed at will. By using this weighted correlation window, the dominant noise term in Theorem 1 rewrites

$$\mathrm{Var}(\mathcal{E}_{\mathbf{x}_0}) = 2\sigma^2 \int \frac{\rho(\mathbf{x} - \mathbf{x}_0)^2}{u_x(\mathbf{x} + \varepsilon(x))^2} d\mathbf{x}. \tag{22}$$

The discrete implementation of such an algorithm faces the problem of computing the true derivatives $u_x(\mathbf{x} + \varepsilon(\mathbf{x}))$ from the two available images u_1 and u_2. We actually compute the derivative on the first image obtaining

$$u_1'(\mathbf{x})^2 = (u'(\mathbf{x} + \varepsilon(\mathbf{x}))(1 + \varepsilon'(\mathbf{x})) + n_1'(\mathbf{x}))^2.$$

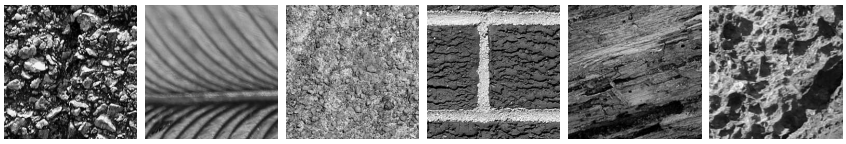

Figure 1: Reference image warped by a known disparity to obtain an image pair.

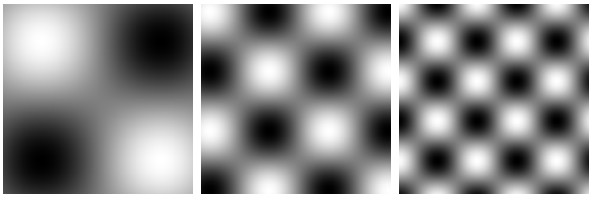

Figure 2: Ground truth disparities applied to images in Fig. 1

As this is a stochastic term, the right choice must be indicated by its mean

$$Eu_1'(\mathbf{x})^2 = u'(\mathbf{x}+\varepsilon(\mathbf{x}))^2(1+\varepsilon'(\mathbf{x}))^2 + 2\sigma^2.$$

This identity shows that, because of the noise term, we will be only able to compute the actual derivatives if and when $\varepsilon'(\mathbf{x})$ is small. Making this assumption, which means that either the relief is smooth, and in order to avoid too small gradients due mainly to noise, we shall use the following weighting function

$$\varphi_{\mathbf{x}_0}(\mathbf{x}) = \frac{\rho_{\mathbf{x}_0}(\mathbf{x})}{\max(u_x(\mathbf{x}+\varepsilon(\mathbf{x}))^2, 6\sigma^2)},$$

being σ the noise standard deviation.

4 Experiments

In order to illustrate and compare the performance of the classical correlation strategy and the proposed adaptive algorithm, several tests were performed on synthetic and real stereo pairs.

The first experiments simulated pairs with a smooth disparity function. The disparity ε in Fig. 2 was applied to the reference texture images u of Fig. 1. Each image was warped by ε to obtain the image pair. Gaussian white noise was added to both images of the pair. Texture images were used to make sure that around each pixel there was enough information to permit its correct matching. The first ground truth disparity varies slowly and smoothly while the other two are more oscillatory.

Fig. 3 presents the disparity maps obtained by both strategies for the first image of the data base. In this case, a noise with standard deviation 1 has been added, yielding a signal to noise ratio of about one hundred. The results with correlation and with the proposed strategy are shown with prolate functions supported by 7×7 and 11×11 pixels. Observe that the disparity obtained with the proposed strategy is more similar to the ground truth than the classical correlation algorithm. This improvement is conspicuous when the 11×11 prolate

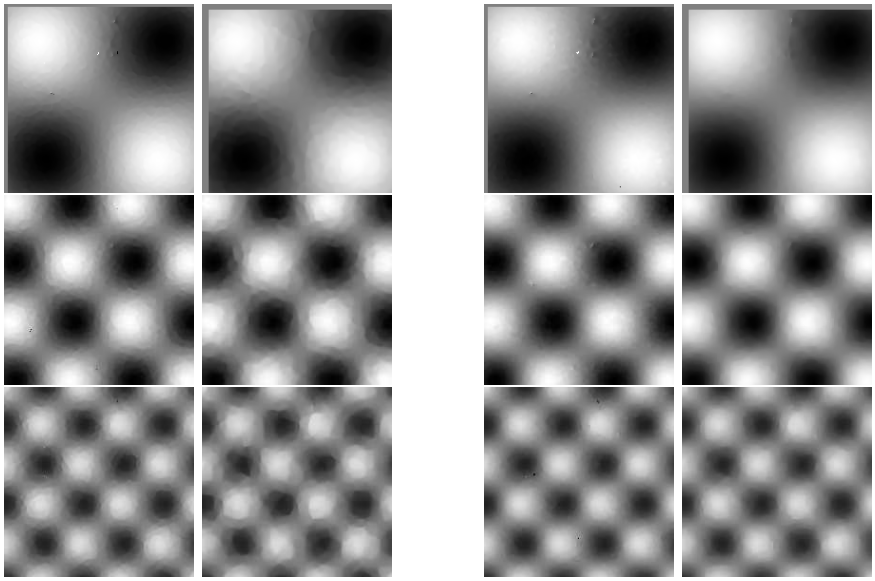

Figure 3: Obtained disparities for the first image in Fig. 1 and the three ground truth disparities in Fig. 2. The left column shows the disparities obtained with a classical correlation algorithm with an isotropic weighting window of size 7×7 and 11×11. In the right column, same experiments but with the proposed algorithm.

is used or when the disparity map is more oscillatory. This experimental fact is in agreement with the mathematical arguments and formulas developed in the previous section. The obtained disparity for the classical correlation strategy depends on the true disparity on the 7×7 or 11×11 neighborhood and is weighted by the square of the gradient. Thus, with a larger window the probability of having large gradients on the window is increased and the favored disparity by these large gradient points can be more different than the one of the reference pixel.

In Fig. 4 are displayed the three-dimensional representations of the central raw in Fig. 3 with a 7×7 prolate function. One better appreciates, with this representation, the differences between the classical and the adaptive correlation. The surface obtained by the adaptive correlation is smooth and very similar to the ground truth. However, the surface by the classical correlation strategy presents many irregularities due to its dependence on the image gradients.

Table 1 shows the average Euclidean distance between the obtained disparity and the ground truth for the six images in Fig. 1. The error values are very similar when the prolate is small or when the disparity varies slowly, while it increases for the classical correlation algorithm when a larger prolate or an oscillating ground truth is applied. Table 2 shows the error committed by comparing the true normals to the surface of the ground truth with the normals to the surfaces of the obtained disparities. Are shown the ratio of points of the surface for which the normal has an error of more than 10 degrees with respect to the original normal. These results are absolutely impressive, since the accuracy gain has a factor as large as 50! Notice that the distance of normals is the

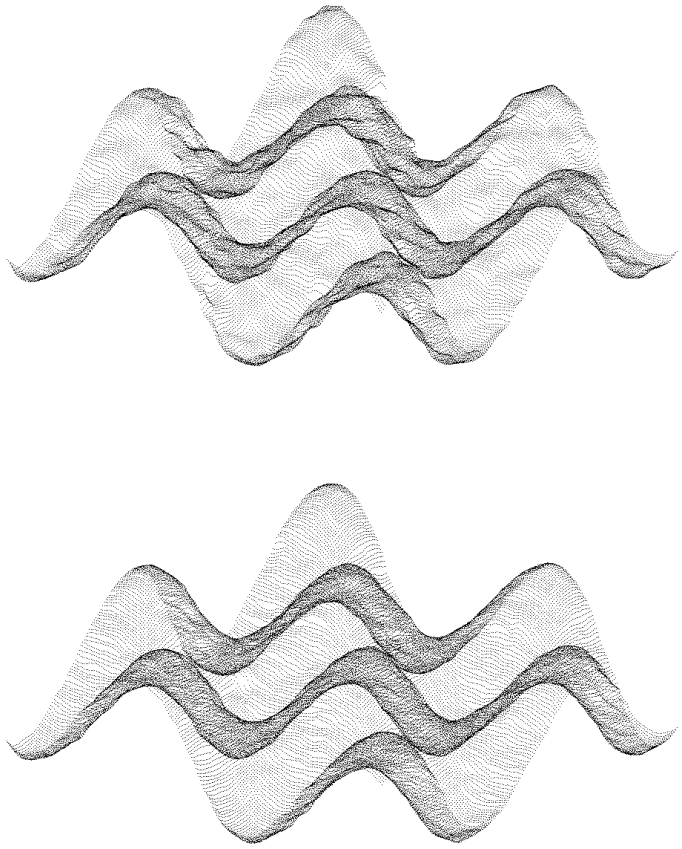

Figure 4: Three dimensional representation of the estimated disparity from the middle row of Fig. 3. Top: estimated disparity by correlation with a 7×7 correlation window. Bottom: proposed adaptive correlation with the same 7×7 correlation window. The fattening effect creates evident irregularities in the reconstructed surface.

7×7	$\sigma = 0.0$	$\sigma = 1.0$ (SNR=100)	$\sigma = 2.0$ (SNR=50)
Correlation	0.118	0.121	0.138
Proposed	0.108	0.113	0.139

11×11	$\sigma = 0.0$	$\sigma = 1.0$ (SNR=100)	$\sigma = 2.0$ (SNR=50)
Correlation	0.135	0.136	0.139
Proposed	0.107	0.109	0.116

Table 1: Average error on the disparity computation on the six images of Fig. 1 and the middle ground truth of Fig. 2. For the proposed method the distance is computed to the convolved ground truth as predicted by the formulas. The first table is obtained by using a correlation window of 7×7 pixels while the second table is obtained by using a correlation prolate of size 11×11. We observe that the correlation error increases when using a larger window. By using a larger window the ground truth disparity varies more and the possibility of having a large gradient increases, therefore making correlation more sensitive to adhesion. The obtained errors are quite similar for both algorithms, showing that the use of an adaptive correlation does not diminish the precision of correlation.

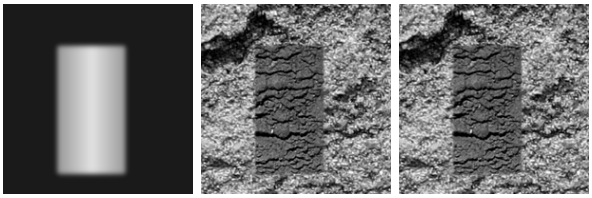

Figure 5: Synthetic image pair. Left: the disparity ground truth, the background has uniform disparity while the building simulates the slope of a roof. Center and right: image pair.

right measure to estimate how two renderings of the same object differ visually. Indeed, most 3D visualizations are done by a Lambertian model. The grey level of the rendered image is the scalar product of the surface normal with the solar direction. Thus the above error measure is the right one to estimate the visual gain.

The next experiment is with a synthetic disparity map applied to a building image. The background has uniform disparity while the building simulates the slope of a roof. Since the background has uniform disparity, we can only observe the fattening effect in and near the building. We show the ground truth disparity and the simulated image pair in Fig. 5. Fig. 6 shows the estimated disparities with the classical correlation algorithm and with the proposed adaptive correlation, using again a prolate window of 7×7 and 11×11 pixels. The same figure shows the error image difference between the estimated disparities and the ground truth. With the proposed strategy the obtained image difference stands between the estimated disparities and the convolved ground truth by the same prolate. This is consistent with the formulation in previous section, where we showed that the adaptive correlation estimates a convolved disparity, independent of the gradient of the image. For the correlation algorithm, we observe a prominent error near the boundaries of the building, while for the

7×7	$\sigma = 0.0$	$\sigma = 1.0$ (SNR=100)	$\sigma = 2.0$ (SNR=50)
Correlation	0.35	0.54	1.27
Proposed	0.04	0.20	1.26

11×11	$\sigma = 0.0$	$\sigma = 1.0$ (SNR=100)	$\sigma = 2.0$ (SNR=50)
Correlation	0.48	0.50	0.64
Proposed	0.01	0.01	0.11

Table 2: Average on the six images of Fig. 1 and the middle ground truth of Fig. 2 of the percentage of points with an angular difference of the surface normal to the ground truth normal larger than 10 degrees. For the proposed method the distance is computed to the convolved ground truth as predicted by the formulas. The first table is obtained by using a correlation window of 7×7 pixels while the second table is obtained by using a correlation prolate of size 11×11. Observe that with a larger correlation window a surface more similar to the original one is obtained. This result is particularly impressive: the obtained percentage of points with a very different normal to the surface is much higher for the classical correlation than the proposed algorithm.

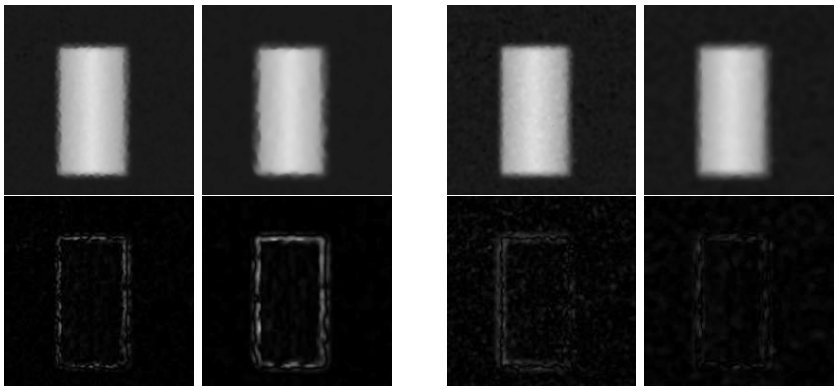

Figure 6: Obtained disparities for the synthetic image pair in Fig. 5. The top left columns displays the disparities obtained with a classical correlation algorithm with an isotropic weighting window of size 7×7 and 11×11. The top right columns show the same experiments but with the proposed algorithm. Bottom: image difference between the estimated disparities and the ground truth. For the proposed strategy the displayed image difference stands between the estimated disparities and the convolved ground truth by the same prolate.

Figure 7: Synthetic image pair. Left: the disparity ground truth. The background has uniform disparity while the building simulates the slope of a roof. Center and right: image pair.

proposed strategy this error passed unnoticed.

The last experience displays a more complicated case with occlusion and shadows containing nearly no information. Fig. 7 shows the image pair and its ground truth. In Fig. 8 are displayed the estimated disparities and the error image difference between the estimated disparities and the ground truth. For the proposed strategy the image difference stands again between the estimated disparities and the convolved ground truth by the same prolate. We observe that the committed error is mainly concentrated near the edges of the building where the fattening effect is more severe. Although in the synthetic case of Fig 6 we were able to nearly eliminate the error near the edges with the proposed strategy, this is not the case for this pair. The error committed by the correlation algorithm is reduced but not eliminated. This is due to the occlusions which make ε discontinuous, and to the fact that near most of the building boundaries the shadow has removed all possible information that could by used to correct the match. It is curious to observe, however, that the error is much smaller at non shadowed edges, even if occlusions and discontinuities of the disparity are still present.

5 Conclusion

We presented a novel approach to adaptive correlation for image matching and depth reconstruction. The presented approach naturally avoids one of the main drawbacks in correlation namely the fattening effect. The proposed experimentation illustrates how mathematical arguments are soundly reflected in practice.

The weak point of the proposed approach is the requirement of a smooth ground truth disparity function. Thus, the method brings no miracle on occlusions. Dealing with occlusions in the same spirit is the current object of research and will be the natural continuation of the current work.

References

[1] A. Buades, B. Coll, JM Morel, and B. Rouge. Procedimiento de establecimiento de correspondencia entre una primera imagen digital y una segunda imagen digital de una misma escena para la obtencion de disparidades. *Spanish Patent, Reference P25155ES00, UIB*, 2009.

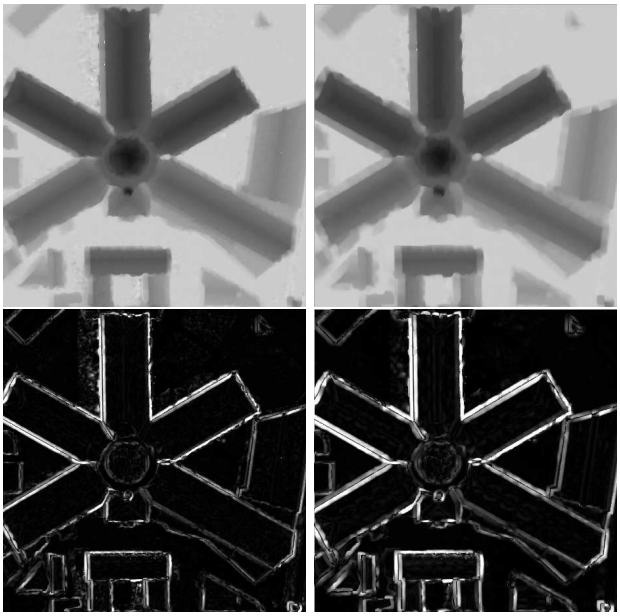

Figure 8: Obtained disparities with the classical correlation algorithm for the synthetic image pair in Fig. 7. Top: disparities obtained with an isotropic weighting window of size 7×7 and 11×11. Bottom: image difference between the estimated disparities and the ground truth.

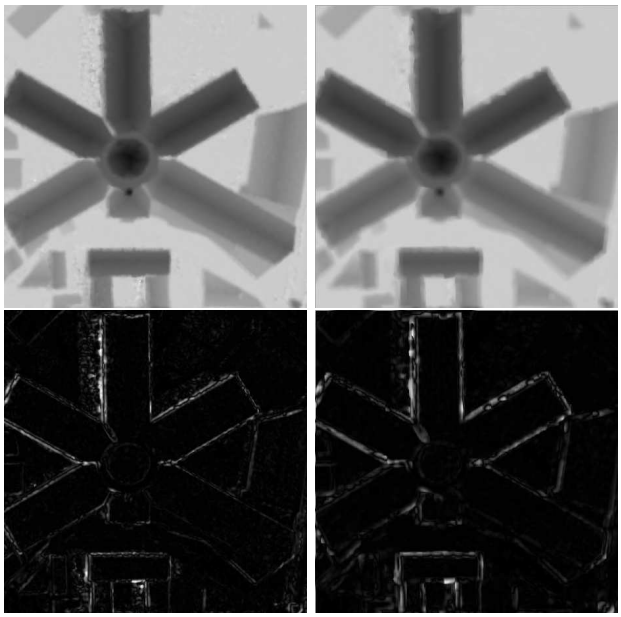

Figure 9: Obtained disparities with the proposed strategy for the synthetic image pair in Fig. 5. Top: disparities obtained with an isotropic weighting window of size 7×7 and 11×11. Bottom: image difference between the estimated disparities and the convolved ground truth by the same prolate.

[2] F. Cao, J. Delon, A. Desolneux, P. Muse, and F. Sur. A unified framework for detecting groups and application to shape recognition. *Journal of Mathematical Imaging and Vision*, 27(2):91–119, 2007.

[3] J. Delon and B. Rougé. Small baseline stereovision. *Journal of Mathematical Imaging and Vision*, 28(3):209–223, 2007.

[4] A. Fusiello, V. Roberto, and E. Trucco. Symmetric stereo with multiple windowing. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(8):1053–1066, 2000.

[5] H. Hirschmuller, P.R. Innocent, and J. Garibaldi. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, 47(1-3):229–246, 2002.

[6] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, 1994.

[7] J. Lotti and G. Giraudon. Correlation algorithm with adaptive window for aerial image in stereo vision. *In Image and Signal Processing for Remote Sensing*, 1:2315–10, 1994.

[8] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[9] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel. An a contrario decision method for shape element recognition. *International Journal of Computer Vision*, 69(3):295–315, 2006.

[10] M.P. Patricio, F. Cabestaing, O. Colot, and P. Bonnet. A similarity-based adaptive neighborhood method for correlation-based stereo matching. In *International Conference on Image Processing*, volume 2, pages 1341–1344, 2004.

[11] L. Robert and O.D. Faugeras. Curve-based stereo: figural continuity and curvature. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991.

[12] N. Sabater. *Reliability and accuracy in stereovision. Application to aerial and satellite high resolution images.* Ph.D. thesis, ENS Cachan, December 2009.

[13] C. Schmid and A. Zisserman. The geometry and matching of lines and curves over multiple views. *International Journal of Computer Vision*, 40(3):199–234, 2000.

[14] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision*, volume 846. Citeseer, 1998.

[15] O. Veksler. Fast variable window for stereo correspondence using integral images. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:556–561, 2003.

[16] Liang Wang, Miao Liao, Minglun Gong, Ruigang Yang, and David Nister. High-quality real-time stereo using adaptive cost aggregation and dynamic programming. In *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission*, pages 798–805, 2006.

[17] L. Yaroslavsky and M. Eden. Fundamentals of digital optics, 2003.

[18] S. Yoon, K.-J.and Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.