Variational approach to interpolate and correct biases in stereo correlation

Gabriele Facciolo[1], Andrés Almansa[1], Alvaro Pardo[2]

[1]InCo, Facultad de Ingenieria, Universidad de la República
Julio Herrera y Reissig 565, C.C. 30, Montevideo, Uruguay

[2]IIE, Facultad de Ingenieria, Universidad de la República
Julio Herrera y Reissig 565, C.C. 30, Montevideo, Uruguay
`facciolo@fing.edu.uy, almansa@fing.edu.uy, apardo@fing.edu.uy`

Résumé − Un défaut bien connu des modèles numériques d'élévation (MNE) obtenus par des techniques de corrélation, est le phénomène d'adhérence, qui apparaît le long des bords très contrastés de l'image comme une dilatation de la partie supérieure (ou inférieure) du terrain estimé. Le phénomène est directement lié à la taille des voisinages utilisés lors de la maximisation de la corrélation normalisée, et la magnitude de ces artefacts ne peut pas être négligée quand une précision très sous-pixellaire est cherchée. Le travail de Delon et Rougé [3] donne une caractérisation de ce phénomène, fournissant un lien entre les disparités mesurées par corrélation et les vraies disparités. Il permet aussi de détecter des régions incorrélables, c'est à dire, des régions ne contenant aucune information utile pour une corrélation suffisamment précise. Comme cette relation (entre disparités vraies et estimées) est exprimée par un système linéaire très mal posé, des nombreuses suppositions simplificatrices ont été adoptées pour sa résolution, conduisant à la *correction barycentrique* du phénomène d'adhérence. Le résultat, bien que beaucoup amélioré par rapport aux disparités brutes, reste légèrement flou et oscillant, ce qui est particulièrement gênant pour le MNE urbain. Dans cet article nous proposons des suppositions simplificatrices moins contraignantes pour l'inversion du système, à savoir, sa régularisation par un terme de surface minimale ou variation totale. Une telle approche permet d'obtenir un terrain avec des bords moins flous, tout en interpolant les régions vides (sans information fiable de corrélation) d'une façon raisonnable.

Abstract − It's well known that DEMs (Digital Elevation Models) obtained by stereo correletion techniques suffer from *adhesion phenomenon*, which is a distortion of the model that appears near strong discontinuties or borders of the image. This phenomenon is directly related to the correlation process, and the magnitudes of the artifacts cannot be neglected when trying to obtain sub-pixel accuracies.

The work by Delon and Rougé [3] characterizes this phenomenon, giving a link between measured and true disparities, and allowing to detect uncorrelatable regions (or regions providing no useful information for correlation). Since this leads to a very ill posed system of equations, many simplifying assumptions have been adopted in order to easily solve it, leading to the so called *barycentric correction* of the adhesion phenomenon. Even though the result is highly improved with respect to the raw correlation disparities, one still observes a slightly blurred disparity map, which is specially annoying in urban areas.

In this work we propose more precise and natural assumptions to solve this system, namely to regularize the solution by a minimal surface or total variation term. Such an approach is naturally expected to allow less blurred edges while still filling in empty areas (without meaningful correlation information) in a reasonable manner.

1 Introduction

The ability of obtaining depth information form an image pair has a wide range of applications. This problem has been studied in depth from multiple approaches during the last decade (see [1] for a complete review). All the most common approaches rely on the fact that the depth of the object is inversely proportional to the disparity of its image projection from two different viewpoints, a phenomenon celled "stereopsis".

The obtention of DEMs (Digital Elevation Models) from aerial or satellite images requires sub-pixel accuracies in the terrain model. A common technique to reach these precisions is the stereo correlation. This method as other block matching methods suffer from the *adhesion phenomenon*, which is directly related to the windowing process and appears near strong discontinuities or borders of the images as a distortion of the elevation map.

The work by Delon and Rougé [3], characterizes this phenomenon, giving a link between measured and true disparities, and allowing to detect uncorrelatable regions (or regions providing no useful information for correlation). Since this leads to a very ill posed system of equations, many simplifying assumptions have been adopted in order to easily solve it, leading to the so called *barycentric correction* of the adhesion phenomenon. Even though the result is highly improved with respect to the raw correlation disparities, one still observes a slightly blurred disparity map, which is specially annoying in urban areas.

In this work we propose more precise and natural assumptions to solve this system, namely to regularize the solution by a minimal surface or total variation term. Such an approach is naturally expected to allow less blurred edges while still filling in empty areas (without meaningful correlation information) in a reasonable manner. We

discuss the technical difficulties of implementing this approach, and compare some results obtained by both methods on synthetic data.

In section 2 we present some fundamental concepts and recent advances by J. Delon and B. Rougé, on the formalization of the adhesion phenomenon and their solution the *barycentric correction*, summarized from [2, ch 4-5]. In section 3 we introduce a more direct and natural way to correct the adhesion, showing in section 4 some details of the implementation. The results will be commented in section 5, followed by the conclusions in section 6.

2 Analytic study of correlation

Assuming that the images have small baseline and are taken from high altitude we can consider a simplified image formation model with parallel projection, where no occlusions (due to vertical structures) can occur. Additionally we assume that the images have been previously rectified to have a horizontal epipolar geometry [4], this reduces the bi-dimensional matching problem to one dimension. Under the hypothesis of subpixel disparities, the relation between the image pair u, $\tilde{u}$ and the disparity ε can be modelled by: $\tilde{u}(x) = u(x + \varepsilon(x))$.

Assuming that the real disparities are small enough to perform first order approximations of the correlation, Delon and Rougé [2] found that the process of maximizing the Normalized Cross Correlation (NCC) [1]

$$\rho_{x_0}(m) = \frac{\int_{\varphi_{x_0}} u(x+m)\tilde{u}(x)dx}{\sqrt{\int_{\varphi_{x_0}} u^2(x+m)dx \int_{\varphi_{x_0}} \tilde{u}^2(x)dx}} \quad (1)$$

produces a disparity map m which is related to the true disparity map ε by the "fundamental equation of correlation" (2):

$$(\varepsilon d_{x_0}) * \varphi = m(d_{x_0} * \varphi) \quad (2)$$

In this equation the function d_{x_0} expresses the "edgeness" of the image u around the point x_0 (center of the window φ):

$$d_{x_0}(x) = \frac{u'^2(x)\int_{\varphi_{x_0}} u^2(y)dy - u(x)u'(x)\int_{\varphi_{x_0}} u'(y)u(y)dy}{\left(\int_{\varphi_{x_0}} u^2(y)dy\right)^2}$$

This is a two variable function of x_0 and x, which is computed over the reference image u, sometimes the notation $d(x_0, x)$ will be used to enforce the fact of being a two variable function.

Equation (2) is very hard to solve for ε, because it only provides information near the edges of u (the regions with high d_{x_0}). In [2] this difficulty was circumvented by a *barycentric correction* which consists of approximating d_{x_0} by a *delta* function at the barycenter of the window: $x_1 = \frac{\int_{\varphi} d_{x_0}(x)xdx}{\int_{\varphi} d_{x_0}(x)dx}$. Then associate the computed disparity $m(x_0)$ to the location of the barycenter x_1, meaning

---

[1] With $\int_{\varphi_{x_0}} u(x)dx$ we denote (in a compact way) the convolution of u with the window function φ_{x_0} centered at point x_0; as the window has a compact support the convolution is well defined. In other cases we use a more extended representation like: $\int u(x)\varphi(x_0 - x)dx$ (where the implicit integration domain is the entire support of the image u).

$\varepsilon(x_1) = m(x_0)$, the resulting irregular sampling of ε is later interpolated on a regular grid.

Delon and Rougé [2] also derived a measure that bounds the residual error in the computation of the disparity map ε, introduced by the fundamental equation of correlation. The measure $N(u, \varphi, x_0)$ (eq. (3)) is a simplified case of the model defined by Delon in [2, ch 5.2]. It's computed over the image u and considers the correlation window φ and the standard deviation of the image noise σ_{noise} (considered gaussian).

$$N(u, \varphi, x_0) = \frac{\sigma_{noise}}{\|u\|_{\varphi_{x_0}} \sqrt{\int_{\varphi_{x_0}} d_{x_0}(x)dx}} < \lambda \quad (3)$$

By imposing a threshold λ this bound can be used to determine the zones where the disparity map ε (resulting from the correlation process) have at least a precision λ (as used in Section 4).

3 Variational Solution

Here we propose an alternative to the barycentric correction while adding regularization. More precisely, given the measured disparity map m (found by maximizing correlation) we shall instead invert equation (2) by minimizing $E(\varepsilon) = \omega D(\varepsilon) + S(\varepsilon)$ with respect to ε, where $D(\varepsilon)$ is the data fitting term and $S(\varepsilon)$ is a surface regularization. To do so we start from the first guess $\varepsilon = \varepsilon_0$ given by barycentric correction and follow the gradient descent path:

$$\frac{\partial \varepsilon}{\partial t} = -\frac{\partial E}{\partial \varepsilon} = -(\omega \frac{\partial D}{\partial \varepsilon} + \frac{\partial S}{\partial \varepsilon}).$$

Data term & its Euler-Lagrange: The data term is taken to minimize an energy based on eq. (2), the fundamental equation of correlation:

$$D(\varepsilon) = \|(\varepsilon \Delta d_{x_0}) * \varphi - m(d_{x_0} * \varphi)\|^2,$$

which is a very ill posed system so its solution will need some regularization to return be solved. To calculate its first derivative we start converting it to matrix notation by introducing the operator K

$$(K\varepsilon)(x_0) = \int \varepsilon(x)d_{x_0}(x)\varphi(x_0 - x)dx$$

$$(K1)(x_0) = \int 1 d_{x_0}(x)\varphi(x_0 - x)dx$$

Then $D(\varepsilon)$ can be written as:

$$
\begin{aligned}
D(\varepsilon) &= |(K\varepsilon) - \overbrace{\text{diag}(m)(K1)}^{=b}|^2 \\
&= <K\varepsilon - b, K\varepsilon - b> \\
&= <K\varepsilon, K\varepsilon> -2 <K\varepsilon, b> + <b, b>
\end{aligned}
$$

And calculating it's first derivative:

$$\frac{\partial D}{\partial \varepsilon} = \frac{\partial}{\partial \varepsilon} <K\varepsilon, K\varepsilon> -\frac{\partial}{\partial \varepsilon}2 <K\varepsilon, b> = 2K^*(K\varepsilon - b)$$
.

The conjugate matrix K^* is defined as:

$$(K^*g)(x) = \int \overline{\varphi(x_0 - x)d_{x_0}(x)}g(x_0)dx_0$$

(observe that it is integrated over the variable x_0, and $d_{x_0}(x)$ is not complex valued). This is demonstrated with:

$$< Kf, g >= \int \left(\int f(x) \varphi(x_0 - x) d(x_0, x) dx \right) \overline{g(x_0)} dx_0 =$$
$$\int f(x) \left(\int \overline{\varphi(x_0 - x) d(x_0, x)} g(x_0) dx_0 \right) dx = < f, K^* g >.$$

The solution for the problem is the following Partial Differential Equation, which can be minimized by a gradient descendent method. [2]

$$\frac{\partial D}{\partial \varepsilon} = 2K^*(K\varepsilon - \text{diag}(m)(K1)) \quad (4)$$

Regularization term & its Euler-Lagrange: The energy for minimizing the surface is similar to the proposed in [5] and [6]:

$$S(\varepsilon) = \int \sqrt{a^2 + |\nabla \varepsilon|^2} dx$$

In this case the value of a^2 is used to control the relative weight of vertical changes; values of $a^2 < 1$ result in sharper edges because the changes of height are less expensive. For our problem the value of a^2 may be selected similar to the proportion $(b/h)^2$, where b is the distance between the cameras and h is the distance between the cameras to the object. This choice balances the horizontal (pixel) and vertical (altitude) scales so that $S(\varepsilon)$ represents the terrain surface measured in pixels. A smaller choice of a makes $S(\varepsilon)$ tend to the total variation of ε.

The calculus of the E-L solution of this type of surface minimization leads to the following equation.

$$\frac{\partial S}{\partial \varepsilon} = \text{div}(\nabla \varepsilon / \sqrt{a^2 + |\nabla \varepsilon|^2})$$

4 Multiple window sizes

In the previous section we presented a variational method to minimize the effects of adhesion artifacts. But the result only takes into account a single window size, so now we extend it by integrating disparities from multiple correlation windows into a single energy. We build a weighted sum of multiple D terms (weighted by ω_i) for all window sizes i, and apply to each term a mask Θ_i to select only one term for each pixel. This mask prevents multiple terms from acting simultaneously over the same pixel and distorting the solution (this happens specially near the image borders where many terms have simultaneously high values). The resulting energy will be:

$$E(\varepsilon) = S(\varepsilon) + \sum_i \omega_i \, \Theta_i \, D(m_i, \varphi_i, \varepsilon) \quad (5)$$

$$\Theta_i(x) = \begin{cases} 1 &, \quad \text{if} \quad i = i_{min}(x) \\ 0 &, \quad \text{otherwise.} \end{cases}$$
$$i_{min}(x) = \arg\min_i \{size(\varphi_i) : N(u, \varphi_i, x) < \lambda\}$$

---

[2]The numerical implementation of this differential equation, can be efficiently computed with six convolutions by the window function φ (plus some pixel-by-pixel additions and multiplications), resulting from developing the operator K and d_{x_0}.

For the determination of the masks Θ_i we use eq. (3) to select the areas of sufficient precision $\{x : N(u, \varphi_i, x) < \lambda\}$, for each correlation window size. As the term $N(u, \varphi_i, x)$ allows to compare directly the correlation curvatures of different window sizes, we select the smallest admissible window size for each pixel of the image.

5 Experimental results

Here we present some early results of an implementation of the minimization of equation (5). We computed the "correlation maps" with seven sizes of correlation windows for our images. The initial condition of the algorithm is the DEM obtained with the barycentric correction (fig. 2), and the computed correlation maps are used as parameters.

In fig. 3 we can observe that regularization (with low a value) sharpens the borders and reduces the oscillatory artifacts of the structures, but as expected from a minimal surface term the corners tend to be rounded.

The adhesion artifacts can be noticed between the ground truth (fig. 1) and the initial condition (fig. 2), noticing that the central structures suffer particulary from horizontal dilation in the initial condition. But the regularized DEM (fig. 3) is qualitatively similar to the ground truth, the borders are straight lines and less dilated, this is due to the data term that emphasizes the border values of the object.

The small objects present at the right side of the images, are more blurry in the regularized DEM because the data term is small in zones poorly contrasted. Increasing the relative weight of the data term with respect of the regularization will conduct to instability of the first one. This is due to the nature of the data term which has very large range of values, from high values in zones with highly contrasted to very small values in zones with low contrast.

Most areas of the image don't have any valid data term, so the regularization is predominant there, but this isn't necessarily true at lower resolutions. The actual implementation of our method only operates at one resolution, while the result of the barycentric corrected DEM [2], is generated with a multi-resolution algorithm. With a multi-resolution implementation the influence areas of each window will be magnified at lower resolution, allowing to correct the elevation model in zones where the actual implementation doesn't act.

6 Conclusion and future work

We presented a method to correct the adhesion phenomenon that also prevents the oscillatory artifacts result of barycentric correction and allows to interpolate in non feasible areas, while sharpening the edges. In some cases the resulting DEM is not better than the original (the barycentric correction), but we believe that most of these problems may be addressed in the future works.

The problems related with the data term's range of values, can be addressed by re-scaling eq. (4) depending
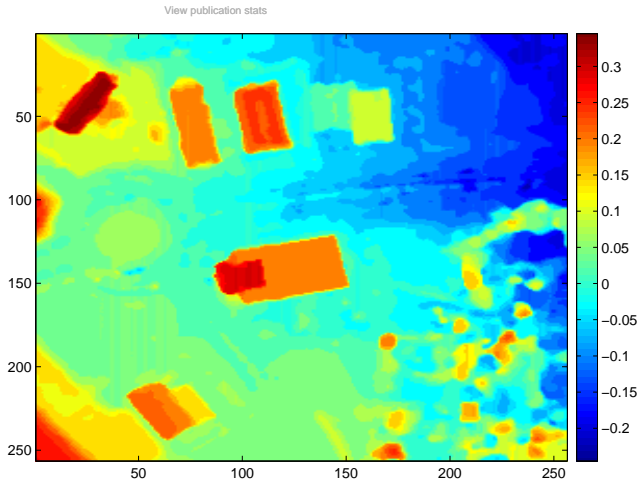
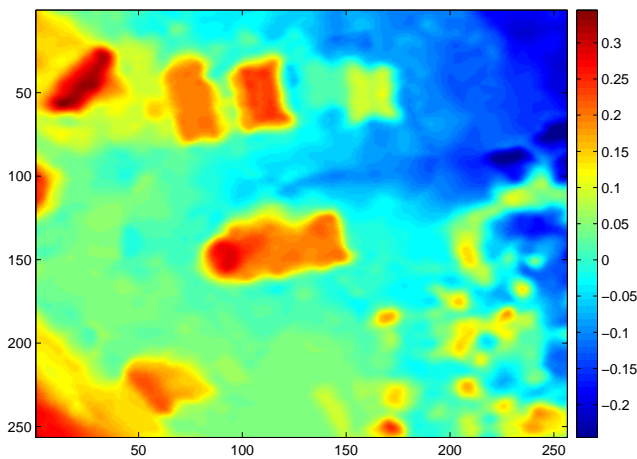FIG. 1: Ground truth elevation model.

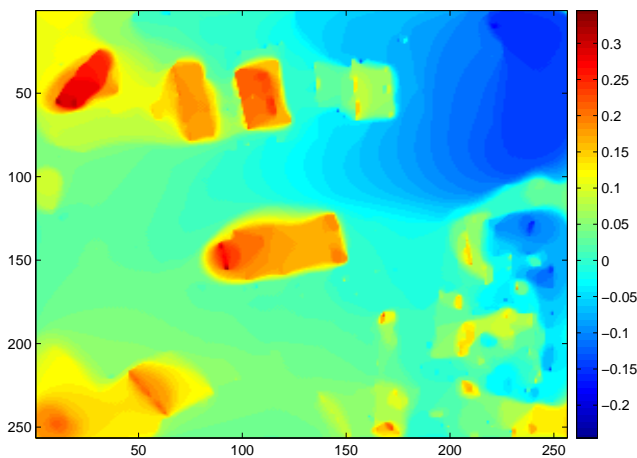FIG. 2: Elevation model barycentric corrected.

FIG. 3: Computed elevation model applying 2000 steps of the algorithm to the result of the barycentric correction.

the contrast of each pixel of the image. Using a more anisotropic regularization we can avoid the rounding of the corners proper of the surface minimization term, but produces elevation maps more similar to the reference image, so this term needs to be controlled carefully.

Lastly as mentioned in section 5 a multi-resolution implementation is more robust to the lack of valuable information at a certain resolution. The non-trivial aspect of this extension concerns how to integrate data-fitting terms at multiple resolutions into a single energy. The possible solutions are applying the resulting ε from every scale to the images and change scale, or adding the ε obtained so far to process the next scale. Of course recovering the exact urban shapes would involve a later detection stage where straight borders of buildings, planar walls and roofs are explicitly detected and fitted to the image data as for instance in [7]. We are also working on similar ideas. Anyhow, a DEM restoration like the one proposed here is still more general (and hence also useful in regions where planar or straight-line shapes cannot be significantly detected), and serves also as a more accurate initial guess that can guide such later detection stages.

Acknowledgements

References

[1] Myron Z. Brown, Darius Burschka, and Gregory D. Hager. Advances in Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.

[2] J. Delon. *Fine comparison of images and other problems.* PhD thesis, Ecole Normale Supérieure de Cachan, 2004.

[3] J. Delon and B. Rougé. Analytic study of the stereoscopic correlation, 2004. Personal Communication.

[4] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images : The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications.* MIT Press, 2001.

[5] O.D. Faugeras and R. Keriven. Variational-principles, surface evolution, pdes, level set methods, and the stereo problem. *IP*, 7(3):336–344, March 1998.

[6] Ron Kimmel. 3d shape reconstruction from autostereograms and stereo. *Journal of Visual Communication and Image Representation*, 13(1-2):324–333, 2002.

[7] X. Descombes M. Ortner and J. Zerubia. Building Extraction from Digital Elevation Model. *In Proc. of ICASSP'03*, April 2003.