Stereoscopic Images Matching Using a Localized Correlation Function

Jérôme PERRIN , Bruno TORRÉSANI , Philippe FUCHS

Abstract— **We propose an algorithm of stereo matching based upon a multiscale decomposition of the input images. Contrary to "coarse-to-fine"-type algorithms, the proposed method processes information from different frequency bands simultaneously. This task is achieved by a localized correlation function defined for every pixel of the image, obtained by integration with respect to a scale variable. This is equivalent to defining a correlation kernel, which is extremely well localized with respect to position and disparity variables.**

Keywords— **stereo vision, matching, correlation, frequency, scale, wavelet transform.**

I. Introduction

The matching of stereoscopic images is one of the classical problems of computer vision. Several approaches have been developed (matching of characteristics or primitives [9], [11], global matching [12], object-oriented matching,...). Correlation-based approaches are among the most popular [5], [11], but generally suffer from the lack of locality of correlation functions.

The goal of this note is to present a localized version of correlation functions, based upon multiresolution decompositions of the images, i.e. two-dimensional wavelet transform. The localized correlation function (LCF) amounts to compare (fixed-pixel) wavelet transforms of two images at different points. The LCF takes advantage of the locality of wavelet transforms to provide estimates for disparity between images. Given a pixel in the first image, the corresponding disparity is defined as the argument which maximizes the LCF. From the estimated disparities it is possible to recover three-dimensional information in the way one perceives depth from information of both eyes [6]. Indeed, assuming parallel optical axes (which we do from now on), the depth z is inversely proportional to the disparity d: $z = ef/d$, where e is the distance between optical axes and f is the focal length of the cameras.

This note is organized as follows. We describe this method in Section II, and give numerical results in Section III. Finally, Section IV is devoted to conclusions.

II. Localized Correlation Functions

Throughout this note, we focus on the case of the spaces $L^2(\mathbb{R})$ and $L^2(\mathbb{R}^2)$ of square-integrable functions of one and two variables respectively. Even though we are mainly interested in the case of images, we mostly discuss the one-dimensional situation for the sake of simplicity. As we shall see, this is sufficient for our purpose as long as the optical axes of the cameras are horizontally shifted, which will be the case in the simulations presented here. The generalization to the two-dimensional case is straightforward, and some possible extensions will be discussed at the end of the paper. We limit the discussion to continuously-defined signals for the sake of simplicity. Discretization issues are briefly discussed at the end of the section.

A. Preliminaries, Notations

One of the standard approaches involves correlation functions. The correlation function introduces itself naturally when one considers signals $f(x)$ which are merely perturbations of a shifted copy $g(x+z)$ of some reference signal $g(x)$. As is well known, the minimization of $\int |f(x) - g(x+\delta)|^2 dx$ wrt δ is equivalent to the maximization of the correlation

$$h(\delta) = (f \star g)(\delta) = \int f(x)\bar{g}(x+\delta)\, dx \ . \tag{1}$$

Correlation function-based approaches to stereo-matching have been developed by several authors (see [4], [9] and references therein.)

The more recently introduced *wavelet analysis* [7] may be used to provide a local version of Fourier analysis. The main idea is the following (there are in fact several alternatives, see e.g. [1], [8]. We stick here to the simplest version.) Given a *mother wavelet*, i.e. a function $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ such that the following admissibility condition holds $0 < C_\psi = \int_0^\infty |\hat{\psi}(\omega)|^2 \frac{d\omega}{\omega} < \infty$ (where the Fourier transform is defined by $\hat{f}(\omega) = \int f(x) e^{-i\omega x} dx$), one associates with it the following family of *wavelets*, labeled by $b \in \mathbb{R}, a \in \mathbb{R}^+$

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\, \psi\left(\frac{x-b}{a}\right) \ , \quad x \in \mathbb{R} \ . \tag{2}$$

We stick to real-valued wavelets for simplicity, but to keep in mind extensions to the complex case we keep complex-conjugates signs in the equations. To any $f \in L^2(\mathbb{R})$ is associated its *continuous wavelet transform* (*CWT* for short) $\mathcal{W}[f]$,

$$\mathcal{W}[f](a,b) = (f, \psi_{a,b}) = \int_{-\infty}^{+\infty} f(x)\bar{\psi}_{a,b}(x)\, dx \ . \tag{3}$$

The CWT is invertible on its range, the original signal may be recovered from $\mathcal{W}[f](a,b)$: for all $f \in L^2(\mathbb{R})$, $f(x) = \frac{1}{C_\psi} \int_0^\infty \int_{-\infty}^{+\infty} \mathcal{W}[f](a,b)\psi_{a,b}(x)\, \frac{da\, db}{a^2}$. In fact, there exist infinitely many inversion formulae of that type. In particular, if $k_\psi = \int \overline{\hat{\psi}}(\omega) d\omega/\omega$ is finite and nonzero, one has

$$f(x) = \frac{1}{k_\psi} \int_0^\infty \mathcal{W}[f](a,x)\, \frac{da}{a^{3/2}} \ . \tag{4}$$

From now on, we assume that this is true, and we normalize $\psi(x)$ so that $k_\psi = 1$.

The CWT may be discretized in several different ways. We choose to stay as close as possible to translation invariance, and retain a wavelet transform discretized on a grid of the form $(b_k, a_j) = (kb_0, a_0^j), j,k \in \mathbb{Z}$ (i.e. an *uniform grid*, as opposed to the usual *dyadic grid* $(kb_0 a_0^j, a_0^j), j,k \in \mathbb{Z}$). Therefore, the discrete wavelet transform of a shifted signal is a shifted copy of the wavelet transform of the original signal. In addition, we periodize the images we consider to simplify the treatment of the boundaries.

J. Perrin is with the Centre de Robotique of ENSMP, 60 Bvd St Michel, Paris, France, and the Centre d'étude et de Recherche en Médecine Aérospatiale; email address: jerome@caor.ensmp.fr

B. Torrésani is with CPT, CNRS Luminy, Case 907, 13288 Marseille Cédex 09, and LATP, Marseille; email address: torresan@cpt.univ-mrs.fr

P. Fuchs is with the Centre de Robotique of ENSMP, 60 Bvd St Michel, Paris, France; email address: fuchs@caor.ensmp.fr

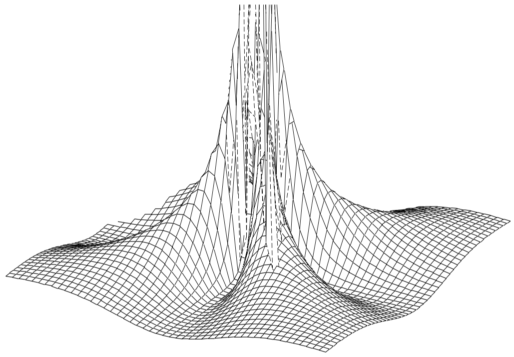

Fig. 1. Kernel $\mathcal{K}_\psi(x, y)$ of the LCF in the case of LOG wavelet (see text)

B. Localized Correlation

The purpose of this section is to introduce and study a localized alternative to the correlation function given in (1). Our starting point is the CWT described above. Given two signals $f, g \in L^2(\mathbb{R})$, we define their *localized correlation function* (*LCF* for short) as follows

$$s(b, b') = (f \overset{loc}{\star} g)(b, b') = \int \mathcal{W}[f](a, b)\, \overline{\mathcal{W}[g]}(a, b')\, \frac{da}{a} \ , \quad (5)$$

that is, as the scalar product of the scale representations of the signals in b and b'. It is simpler to think of the integration in (5) as being performed between 0 and ∞, as this makes explicit computations possible. In fact, the integration is done practically between two reference scales A_1 and A_2, which depend on the size of the image. The reference scales may be adapted locally to the image. We will not consider this more general approach here.

It is readily shown that the LCF may be written with the help of a family of kernels $K_{b,b'}(x, y)$:

$$s(b, b') = \langle K_{b,b'} \cdot f, g\rangle = \int f(x)\bar{g}(y)\, K_{b,b'}(x, y)\, dx\, dy \quad (6)$$

$$K_{b,b'}(x, y) = \mathcal{K}_\psi(x - b, y - b') = \int \bar{\psi}\left(\frac{x-b}{a}\right) \psi\left(\frac{y-b'}{a}\right) \frac{da}{a^2} \ . \quad (7)$$

It is worth comparing the kernel so-obtained with the equivalent object in the case of usual correlation methods.In the latter case, the correlation function may be written as in Eq. (6), with a (distribution) kernel of the form $\mathcal{K}_0(x, y) = \delta(x - y)$, therefore uniquely enforcing localization near $x = y$. The kernel $\mathcal{K}_\psi(x, y)$ has additional localization features: in addition to being localized near $x = y$ as before, it enforces localization near $x = y = 0$. Therefore, the kernel $K_{b,b'}(x, y)$ appearing in the LCF focuses the analysis near the points $x = b$ and $y = b'$. This is essentially how localization is added to the correlation. An example of kernel is given in Fig. 1, in the case of the LOG wavelet (see the discussion below).

The LCF may be given an alternative interpretation. Expressing the scalar product in (6) in the Fourier domain, we see that the LCF may be written

$$s(b, b') = \langle \mathcal{C}T_{-b}f, T_{-b'}g\rangle \ , \quad (8)$$

where T_b denotes translation operator ($T_b f(x) = f(x - b)$), and $\mathcal{C}$ is an operator, which may be written in the Fourier domain as

$$\widehat{\mathcal{C}f}(\omega) = \int \hat{f}(k) H\left(\frac{\omega}{k}\right) \frac{dk}{k} \ , \quad (9)$$

with

$$H(\lambda) = \int \overline{\hat{\psi}}(u)\hat{\psi}(\lambda u) du \ . \quad (10)$$

Due to the localization properties of wavelets in the Fourier domain, the function $H(\lambda)$ is peaked near the scale $\lambda = 1$. Hence, the operator $\mathcal{C}$ in Eq. (9) may be thought of as a *multiplicative smoothing*, in other words a smoothing (i.e. a convolution) with respect to the variable $\log(\omega)$. Therefore, computing the LCF amounts to consider shifted copies of f and g, and take a scalar product of the Fourier transform of f with a (multiplicatively) smoothed copy of the Fourier transform of g. Would the smoothing be additive (i.e. the usual smoothing) instead of multiplicative, we would obtain in that way a scalar product of $T_{-b}f$ with the inverse Fourier transform of a smoothed $\widehat{T_{-b'}g}$, i.e. a product of $T_{-b'}g$ with a function localized near the origin. Since the smoothing is multiplicative, the situation is more complex, but the analysis above shows that the qualitative behavior of the LCF is similar. A quantitative analysis of the performances of the LCF, given in Section III below, shows that the LCF is a suitable tool for matching.

We have tested the LCF algorithm using several choices for the mother wavelet. Particularly interesting are the wavelets which are derivatives $\phi^{(n)}(x)$ of a certain order n of a *smoothing function* $\phi(x)$, for the number of vanishing moments of the wavelet (which controls the frequency localization) may be easily controlled. In this note, we limit ourselves to the family of *DOG wavelets*

$$\phi(x) = C \exp\{-x^2/2\} \ ; \quad (11)$$

other possible choices involve the (complex-valued)*Cauchy wavelets* ($\phi(x) = C/(x + i)$). In both cases the kernel $\mathcal{K}(x, y)$ may be computed explicitely. We display in Fig. 1 the kernel $\mathcal{K}_\psi(x, y)$ corresponding to the LOG wavelet, i.e. the derivative of order two of a Gaussian function. Notice the localization near $x = y$ and the sharp peak at $x = y = 0$ mentioned above (the kernel has been cut at $x, y \to 0$ since it diverges.) The restriction of the kernel to a line $y = \alpha x$ deals with the comparison of signals $f(x)$ and $g(\alpha x)$; not only does the LCF measure the similarity between $f(x)$ and $g(x)$, but also possible distortions, as is the case of the rotation induced by a change of point of view. One can see that the kernel presents *attractive* regions (positive values of the kernel) and repulsive regions (negative values, far from the diagonals). This suggests to design kernels respectful for the spatial organization of the scene, i.e. vanishing identically for opposite directions of similarity measure (in the quadrants where $xy < 0$). We shall not go into such details, which are discussed in [13].

C. Preprocessing, Contrast Coefficients, Normalization

The LCF turns out to be extremely sensitive to the (local) mean value of the signals. This may result in artifacts for the matching problem. Indeed, given two signals, it may happen that a given pattern at a point say x_1 in the first signal is closer to a pattern at a point x_2 in the second signal, but the LCF gives a maximum at a point x_0 in the second signal simply because of a greater intensity of the latter. To overcome such a shortcoming, it makes sense to perform a "local normalization" of the wavelet transforms before computing the LCF (such a treatment is fairly close to the *phase correlation* techniques discussed in [5]). The *multiresolution contrast*, which was proposed in a computer vision context in [2] and studied later in [3] provides an elegant way of doing such a local normalization, in addition to exhibiting similarities with human visual system. In this way, Duval-Destin shows that the multiresolution contrast

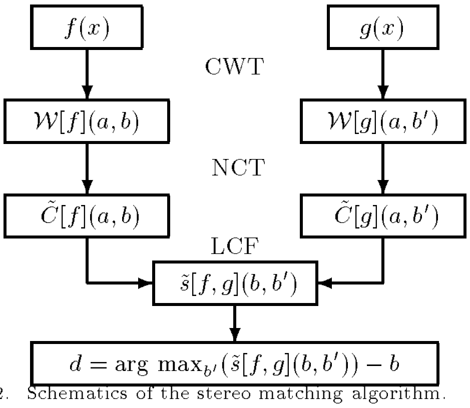

Fig. 2. Schematics of the stereo matching algorithm.

enables to predict human contrast thresholds for complex visual stimulations [2]. Contrast analysis amounts to locally normalize wavelet coefficients $\mathcal{W}[f](b,a)$ by an average intensity at same location b and scale a. More precisely, given $f \in L^2(\mathbb{R})$, we consider its *smoothing at scale a*

$$P[f](a,b) = \int_a^{+\infty} \mathcal{W}[f](\alpha,b)\,\frac{d\alpha}{\alpha^{3/2}} \tag{12}$$

Clearly, $\lim_{a\to 0} P[f](a,b) = f(b)$. The *contrast coefficient* of $f(x)$ at (b,a) is then defined by

$$C[f](a,b) = \frac{\mathcal{W}[f](a,b)}{P[f](a,b)} = -a^{3/2}\,\frac{\partial \log P[f](a,b)}{\partial a} \tag{13}$$

Contrast coefficients may be shown to yield a *multiplicative multiresolution decomposition*: the signal is expressed as a (continuous or discrete) product of contributions from different scales, which are closely related to exponentials of contrast functions (see [3] for mathematical details.)

It turns out that even better results are obtained if, prior to computing the LCF, the contrast coefficients are normalized to have length unity. This is achieved by defining

$$\left\{\begin{array}{lcll} \tilde{C}[f](a,b) & = & C[f](a,b)/\sqrt{N[f](b)}\ , & \text{where} \\ N[f](b) & = & \int_0^\infty |C[f](a,b)|^2\,\frac{da}{a}\ . & \end{array}\right. \tag{14}$$

The coefficients $\tilde{C}[f](a,b)$ form the *normalized contrast transform* (NCT for short) of $f(x)$.

Our description here is based on the continuous contrast decomposition methods. The corresponding discrete theory is a simple issue, and is described in [3].

D. General Organization of the Algorithm

The algorithm may be described as follows: given a pair of signals f, g, compute the corresponding NCT coefficients $\tilde{C}[f](a,b)$ and $\tilde{C}[g](a,b)$, and

$$\tilde{s}(b,b') = \int_0^\infty \tilde{C}[f](a,b)\overline{\tilde{C}[g]}(a,b')\frac{da}{a}\ . \tag{15}$$

The LCF in (15) takes values in $[-1,1]$. Given a point b in the first signal, it may be matched to another point in the second signal if there exists b' such that $\tilde{s}(b,b')$ is larger than a certain threshold σ_c. The threshold σ_c is determined from the histogram of the values of $\tilde{s}(b,b')$. When matching is possible, the disparity at point b in $f(x)$ is defined by

$$d = \arg\max_{b'}(\tilde{s}[f,g](b,b')) - b \tag{16}$$

The algorithm is summarized in Figure 2.

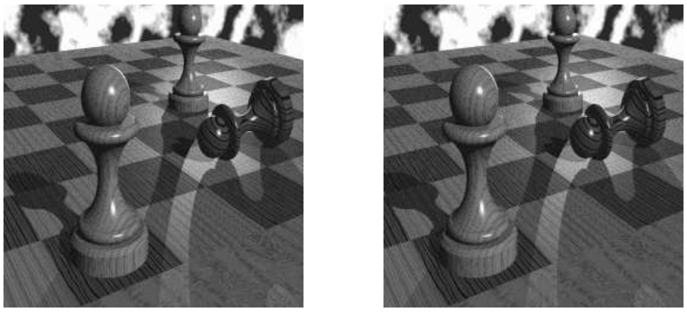

Fig. 3. Initial images: left and right artificial images, created assuming parallel optical axes (see text).

Fig. 4. Preprocessed (i.e. obtained from normalized contrast coefficients) image.

III. Results

We present the results obtained on a test image of size 256×256 (see Figure 3.) The (artificial) images were generated using the software PoV-Ray 3.0. The rightmost image was obtained by an horizontal shift of the camera. Therefore, the disparities between points of the two images are purely horizontal, and the problem reduces to a one-dimensional one. The left and right images of Figure 3 were used as input of the algorithm.

The wavelet was the DOG wavelet of order 2. Scales were discretized as indicated at the end of Section II-A, with $a_0 = \sqrt{2}$. The effect of the contrast and the quadratic normalization of the contrast coefficients may be interpreted as a non-linear transform of the image, very close to a logarithm [3] (see Fig. 4.) The structures of the images (in particular the edges) are the same. However, the contrast of the image is emphasized in less energetic regions. For example, the edges due to shades are now as strong as those coming from true edges of the scene.

The histogram of the values of the LCF for the considered image is given in Figure 5. The threshold has been set to $\sigma_c = 0.9$, which seems to strike a reasonable breakpoint between the

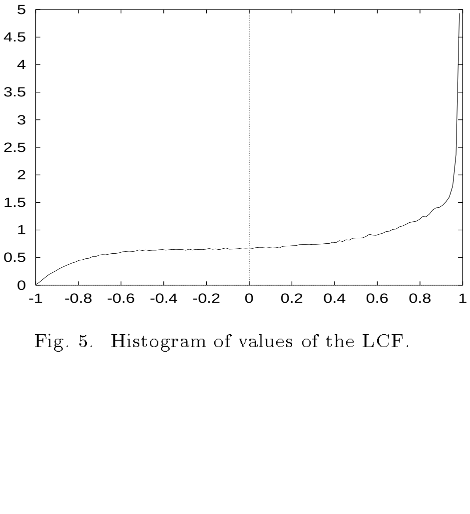

Fig. 5. Histogram of values of the LCF.

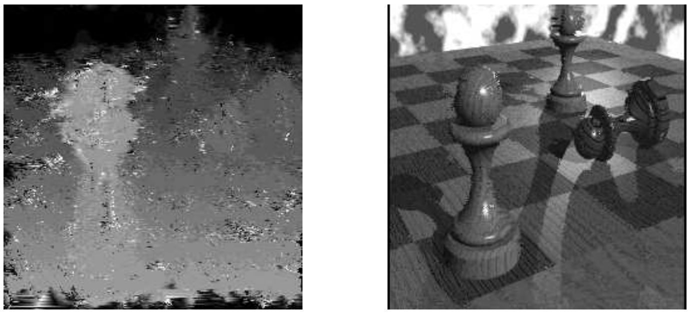

Fig. 6. Result images: estimated disparities (left) and reconstructed right image from the left image and the disparity (right.)

flat behavior and the peak of the histogram, and corresponds to a 85% critical value.

The quality of the results may be estimated on a visual basis first. The right image, reconstructed from the left image and the disparities, is hardly distinguishable from the original (see Figure 6, right image, to be compared with Figure 3, right image). The main defects come from immediate neighborhoods of masked regions, which obviously cannot be reconstructed. A more quantitative measure of quality is provided by a closer look at the estimated disparities (since we were working with artificial images, the true disparity is known). In our simulation, the difference between estimated and true disparity was less than 1 pixel in more than 85% of pixels, therefore yielding a correct matching. The average value of the latter difference was 0.86 pixels, which is in the range of the image precision. The disparities are shown in Fig. 6 (left image). According to the relation between disparity and depth, the brighter the pixel, the nearer the point in the scene.

The computations presented here have been performed on a "standard" UNIX workstation (a Silicon Graphics 02), using a computer code developed in the C++ programming language. The whole procedure takes approximately 3 minutes (half of which is taken by contrast decomposition). Even though optimization of the LCF algorithms has not been a main goal so far (see the remarks in the conclusion below), these results compare well with other methods (for example, global matching methods take several hours on similar computing environments, and standard coarse-to-fine correlation methods are not faster than LCF, and yield poorer results.)

IV. Conclusions

We have described a new method for matching stereoscopic images, based upon multiresolution decomposition of the images. The LCF, constructed from multiresolution contrast transform of the images has been shown to provide an efficient tool for estimating disparity and therefore obtaining 3-dimensional information from stereo images. In addition to the test case presented here, other tests have been performed on several synthetic images, leading to similar conclusions.

Our analysis was limited to horizontal disparity, requiring only one-dimensional analysis tools. Similar methods may be developed in the two-dimensional situation. It is necessary to use an adapted version of wavelet analysis. Two choices are possible. The simplest one amounts to generalize directly the approach presented here, with one-dimensional scale variable and two-dimensional space variable, with a *radial wavelet*. An alternative, using *directional wavelets*, would require the introduction of an extra rotation degree of freedom.

It is worth pointing out that all the operations performed in the present approach may be done in a computationally efficient way. Indeed, wavelet, smoothing and contrast coefficients may be computed using pyramidal algorithms of complexity $O(N \log N)$ (the algorithms are exact for certain wavelets, and only approximate with others, remaining of high precision as long as the wavelets have fast decay in the Fourier domain.) The results presented here were obtained using a FFT-based implementation, of complexity $O(N \log^2 N)$.

References

[1] R. Carmona, W.L. Hwang and B. Torrésani (1997): *Practical Time-Frequency Analysis: Wavelet and Gabor Transforms.* Academic Press, to appear.

[2] M. Duval-Destin (1991): *Analyse spatiale et spatio-temporelle de la perception visuelle à l'aide de la transformée en ondelettes.* PhD Thesis, Marseille (in French).

[3] M. Duval-Destin, M.A. Muschietti and B. Torrésani (1993): Continuous wavelet decompositions, multiresolution and contrast analysis. *SIAM J. Math. An.* **24** 2195-2207.

[4] M.A. Fichler and O. Firschein Eds (1987): *Readings in computer vision.* Kauffmann, Los Altos.

[5] D.J. Fleet (1994): Disparity from local weighted phase-correlation, *IEEE International Conference on Systems, Man and Cybernetics,* **1**, 48-54.

[6] B. Julesz (1971): *Foundations of the cyclopean perception.* University of Chicago Press.

[7] S. Mallat (1989): A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. PAMI,* **11**, 674-693.

[8] S. Mallat (1997): *Wavelet Signal Processing.* Academic Press, to appear.

[9] D. Marr (1982): *Vision: a computational investigation into the human representation and processing of visual information.* Freeman, New-York.

[10] D. Marr and T. Poggio (1979): Computational theory of human stereoscopic vision. *Proc. Royal Soc. London* **B204** 301-328.

[11] H.K. Nishihara (1987): Practical real-time imaging stereo matcher, in *Readings in Computer Vision,* Fichler, M.A. and Firschein, O. Eds, Kauffman, Los Altos.

[12] M.H. Ouali, H. Lange and C. Laurgeau (1996): An energy minimization approach to dense stereovision. *Proceedings of the IEEE International Conference on Image Processing* **2** (Shape and 3D), 841-846.

[13] J. Perrin (1997): *Conception de noyaux de corrélation localisée pour la mise en correspondance d'images stéréoscopiques.* Technical report ENSMP, Paris.