SUB-PIXEL STEREO MATCHING

N. Sabater, J.M. Morel

CMLA, ENS Cachan, CNRS, UniverSud,
61 Avenue President Wilson, F-94230 Cachan

A. Almansa

LTCI, CNRS & TELECOM ParisTech,
46 rue Barrault, F-75634 PARIS Cedex 13

1. INTRODUCTION

Stereo matching performance has improved significantly in the last decade [1, 2]. The existing techniques can be divided between local and global. Local (block-matching) methods rely on a comparison of a small number of pixels surrounding a pixel of interest and are sensitive to local ambiguities (occlusions, uniform textures, or lack of information). Blocks are usually compared by the normalized cross correlation (NCC) or the sum of squared differences (SSD). These block-matching methods can produce wrong disparities near the intensity discontinuities in the images. This phenomenon is called adhesion or fattening effect. Many papers suggest a solution for this problem using adaptive windows [3, 4, 5], a barycentric correction [6] or feature matching methods [7]. Global methods such as graph cuts [8] and dynamic programming [9, 10] are experimentally less subject to fattening, but they lack any error control. But such global methods can be used anyway to interpolate the disparity after a previous block matching step has given a sparse but reliable set of matches.

To the best of our knowledge, the precision that can be attained in block matching has never been calibrated. The aim of this paper is to discuss how to achieve reliably high precision block matches (up to $1/20$ pixel accuracy and below). This discussion is not valid for completely general stereo matching tools. The observation conditions under which the discussion makes sense are : a wholly and accurately calibrated acquisition system, a small baseline (ranging from 0.2 to 0.05) and a very short time difference between both acquisitions, so that the acquired images are nearly identical.[1] Such a highly accurate system has been recently proposed for information geographic systems in urban areas by [6]. The authors called this new stereo concept “small baseline stereovision” and discussed its feasibility. However, they did not develop a crucial point in the strategy, namely the algorithm by which high precision matches can be obtained from a small baseline stereo pair of images, or the practically attainable precision.

Using small baselines in conjunction with larger ones has been considered in [3], a pragmatic study where different baselines were used to eliminate errors. However, its final sub-pixelian results were computed with the large baselines samples. This strategy was questioned in [6], who give mathematical sensitivity arguments showing that small baseline and high accuracy give a more reliable stereo concept than large baselines.

Let us denote by x an image point in the continuous image domain, and by $u(x)$ and $\tilde{u}(x)$ the images of the stereo pair. The underlying depth map would be fully described by a disparity function $\varepsilon(x)$ which tells how much an observed physical point x shifts along the epipolar direction, from the left image u to the right image $\tilde{u}$. Unfortunately, the physical disparity $\varepsilon(x)$ is not well-sampled. This explains why $\varepsilon(x)$ cannot be recovered at all points, but

[1] The baseline is the ratio B/H where B is the distance between the views and H is the distance between the sensors and the ground.

only essentially at points x around which the depth map is constant or nearly constant. At such points, the following model holds:

$$\tilde{u}(x) = u(x + \varepsilon(x)) + n(x) \; , \tag{1}$$

where n is a Gaussian noise.

The images u and $\tilde{u}$ will be assumed acquired with little aliasing and therefore interpolable. Thus, the errors affecting the computation of ε by correlation are: gross errors ($\sim$ 10 pix.) due to the fact that points are matched at a wrong local maximum of correlation [11], adhesion errors ($\sim$ 1 pix.) due to the fact that some salient image feature is within the comparison window but away from its center [6]. Lack of accuracy due to noise may also be large in very flat zones [6]. Finally we focus on discretization errors ($\sim$ 0.02 pix. on simulated data) due to the accuracy limit in discrete interpolation. This last error is actually the dominant one after the corrections proposed in [11, 6] have been applied.

2. THEORETICAL RESULTS (SUMMARY)

In the main section of the full paper we shall show two main results. The first one establishes the conditions under which the discrete block-matching distance (SSD or NCC) is exactly equal to the continuous one, where the sum is replaced by an integral. The second result establishes how to sample this block-matching distance in such a way that it can be interpolated exactly. This second result, which opens the way to rigorous block matching with sub-pixel accuracy, has been noticed (but not implemented) in [12]. It is also used in the MARC method [13] used by the French space agency (CNES). The simple proof we provide is also new, and so is the first result that was (to the best of our knowledge) not yet formalized.

Finally we show how to take advantage of both results and a judicious choice of the correlation window function, in order to write a block-matching algorithm that is both accurate and computationally efficient. The next section shows that the accuracy of this algorithm meets the fundamental limits given by noise in both images of the stereo pair.

3. EXPERIMENTAL EVALUATION AND CONCLUSION

Three experiments were performed to evaluate the attainable disparity error under realistic noise conditions. The everlasting problem of such evaluations is the reference to a ground truth, that may be questionable. Two ways were found to go around this problem. The first sensible way is to simulate stereo pairs with realistic adhesion and noise features. This was done with a simulated pair of urban aerial images at 25 cm/pixel resolution and low baseline ($B/H = 0.045$), based on a reference image and registered altitude map provided by IGN (French National Geographic). In order to make the noise-dependence of our pointwise disparity estimation more explicit, independently from adhesion, we simulated the secondary image as a translation of 2.5 pixels of a reference Brodatz texture image. The translation was performed using zero-padding, and an independent Gaussian noise has been added independently to both images. Secondly, images from the Middlebury dataset (available on www.vision.middlebury.edu/stereo/) were tested. In that case the noise was estimated, and the manual ground truth was actually improved by cross-validation of the multi-stereo dataset.

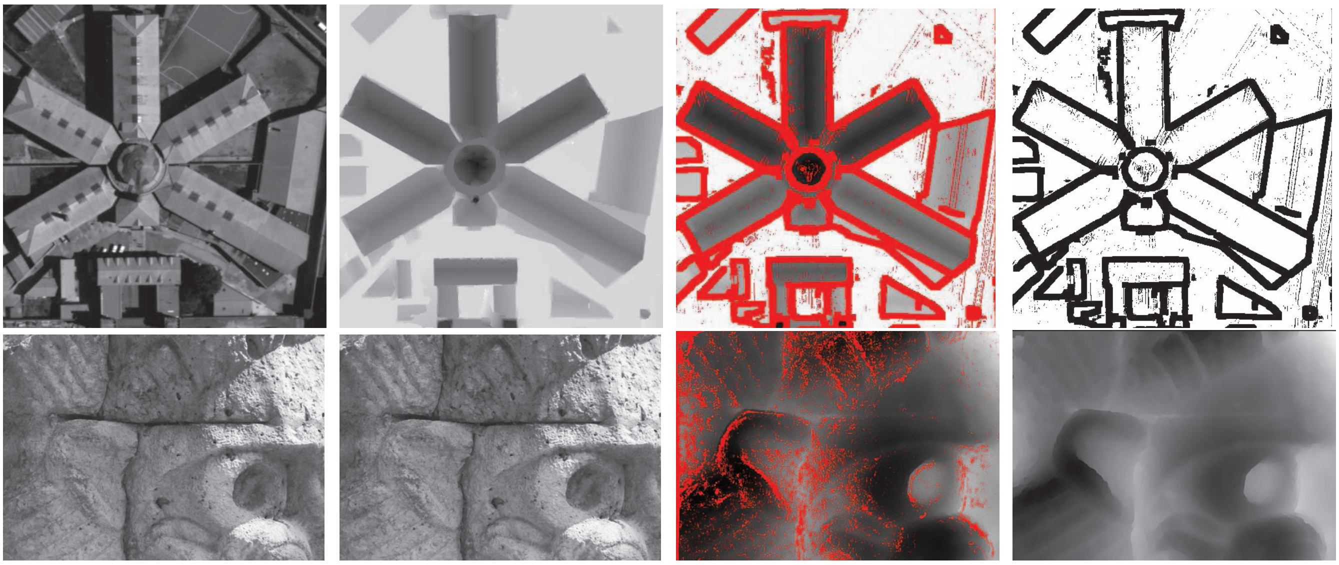

Fig. 1. Top row: simulated stereo pair. Lower row: stereo pair of Lion statue (no ground truth is available). From left to right: Reference image, ground truth disparity in first row (secondary image in second row), sparse disparity map computed with a 9×9 patch (non-matched points in red), and mask of matched points in first row (interpolated disparity map in second row). Statistics in table 1 are computed on the non-red (matched) points.

In all cases, the resulting performance is evaluated by the Root Mean Squared Error (RMSE) measured in pixels on all reliable points, $RMSE = \left(\frac{\Sigma_{i \in I}(d_i - gt_i)^2}{|I|}\right)^{\frac{1}{2}}$, where d_i is the computed disparity and and gt_i is the ground truth value for the pixel i in the set of matched points I. The percentage of bad matches (i.e., $|d_i - gt_i| > 1$) is also given. For the simulated cases the influence of noise in the matching process is studied with several signal to noise ratios $SNR = \frac{\| u \|_2}{\sigma_n}$, where σ_n is the standard deviation of the noise. In each case σ_n is known and the predicted disparity RMSE due to image noise has been computed using a refinement [14] of Delon and Rougé's upper bound [6], which greatly improves the accuracy of this estimate.

A main feature of the experimental setting is the use of a blind *a contrario* rejection method that *does not use the ground truth*. Thus, the percentage of wrong matches is also given, bad matches being those where the computed disparity differs by more than one pixel from the ground truth. As explained in the introduction, the accuracy of matches can only be evaluated on pixels that lie away from disparity edges. These being unknown, security zones where computed by dilating the strong grey level edges by the correlation window. The other pixels were matched only if they passed an *a contrario* test to ensure that the match is meaningful [11]. These two safety filters usually keep more than half the pixels and ensure that the matched pixels are right with very high probability. For all experiments the sub-pixel refinement step goes up to $\frac{1}{64}$ pixel.

Finally we show the results of our algorithm on a real stereo pair of images of a Lion statue. We obtain reliable matches on 86% of the pixels with an estimated disparity accuracy of 0.09 pixels.

The experiments on realistically simulated pairs and real benchmark images show a $1/20$ pixel accuracy to be attained by block-matching, for about half the image points. More interestingly the empirical sub-pixel accuracy in block-matching is close to its predicted limit, which only depends on image noise at regular disparity points. This potentially makes stereo-vision into a highly accurate 3D tool, competitive with laser range scanners, and opens the

Table 1. From left to right: Signal to noise ratio in stereo pair. Predicted and measured disparity error (RMSE) computed in pixels. Percentage of matched points. Percentage of wrong matches.

Dataset	SNR	Predicted RMSE	Measured RMSE	Matches	Bad matches
Simulated- stereo-pair	∞	0.000	0.023	70.6 %	0.00 %
	125.06	0.052	0.058	41.5 %	0.02 %
Translated- Brodatz-texture	48.19	0.010	0.011	99.8 %	0.0 %
	24.09	0.019	0.020	87.1 %	0.0 %
Middlebury Sawtooth	72.39	0.076	0.090	45.2 %	0.1 %
Middlebury Venus	57.74	0.042	0.050	47.2 %	0.1 %
Lion	57.94	0.087	-	86.0 %	-

way for new applications such as highly subpixel change detection and precise measurement of such changes.

4. REFERENCES

[1] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE TPAMI*, vol. 25, no. 8, pp. 993–1008, 2003.

[2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, , no. 47, pp. 7–42, 2002.

[3] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE TPAMI*, vol. 16, no. 9, pp. 920–932, 1994.

[4] J. Lotti and G. Giraudon, "Correlation algorithm with adaptive window for aerial image in stereo vision," 1994.

[5] S. B. Kang, R. Szeliski, and J. Chai, "Handling occlusions in dense multi-view stereo.," *CVPR*, vol. I, pp. 103–110, 2001.

[6] J. Delon and B. Rougé, "Small baseline stereovision," vol. 28, no. 3, pp. 209–223, 2007.

[7] C. Schmid and A. Zisserman, "The geometry and matching of lines and curves over multiple views," *IJCV*, vol. 40, no. 3, pp. 199–234, 2000.

[8] V. Kolmogorov and R. Zabih, "Graph cut algorithms for binocular stereo with occlusions," in *Handbook of Mathematical Models in Computer Vision*. Springer-Verlag, 2005.

[9] Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming," *IEEE TPAMI*, vol. 7, no. 2, pp. 139–154, 1985.

[10] S. Forstmann, Y. Kanou, J. Ohya, S. Thuering, and A. Schmitt, "Real-time stereo by using dynamic programming," in *IEEE CVPR Workshop*, Washington, 2004, vol. 3, pp. 29–36.

[11] N. Sabater, J.-M. Morel, and A. Almansa, "Rejecting wrong matches in stereovision," CMLA Preprint 2008-28. 2008.

[12] R. Szeliski and D. Scharstein, "Sampling the disparity space image," *IEEE TPAMI*, vol. 26, no. 3, pp. 419–425, 2004.

[13] A. Giros, B. Rougé, and H. Vadon, "Appariement fin d'images stéréosc. et instrument dédié avec un faible coeff. stéréoscopique.," French Patent N.0403143, 2004.

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