DISCARDING MOVING OBJECTS IN QUASI-SIMULTANEOUS STEREOVISION

N. Sabater, J.-M. Morel*, A. Almansa**, G. Blanchet†*

*ENS Cachan, CNRS-CMLA
61 Avenue President Wilson
94230 Cachan, France

**Telecom ParisTech, CNRS-LTCI
46 rue Barrault
75634 Paris, France

†CNES
18 av. Edouard Belin
31401 Toulouse, France

ABSTRACT

This paper proposes a statistical rejection rule, designed for small baseline stereo satellites. The method learns an *a contrario* model for image blocks and discards the casual matches between the images of the stereo pair. A formula estimating the expected number of false alarms under the background model is proved. Comparative experiments on *quasi-simultaneous stereo* in aerial imagery demonstrate the elimination of all incoherent motions.

Index Terms— Image matching, number of false alarms, *a contrario* models, Stereo vision, Satellite applications

1. INTRODUCTION

Rejecting false matches is a key problem in stereovision, especially when computing digital elevation models from aerial or satellite imaging. Indeed, using the same instrument for both views yields non-simultaneous snapshots. In the case of a large baseline these changes are dramatic, because the instrument has to travel a long distance between shots. Fortunately, thanks to recent advances in subpixel stereo [1, 2, 3], lower baselines have become tractable, thus enabling quasi-simultaneous single-instrument stereo. For instance, the upcoming Pleiades satellite shall allow for stereo pairs at about three seconds interval by decreasing the baseline/height ratio to 0.15. This mission, which we are currently preparing, requires an automatic and faultless stereo reconstruction. Even lower baselines and lower time intervals are planned for future Earth observation satellites. At such small time intervals illumination changes become negligible, but moving vehicles or even pedestrians remain a serious problem in urban areas. Their reliable detection is crucial for the high resolution low baseline systems proposed at CNES by [4, 2].

The shortcomings of block-matching have led to the overall dominance of global stereo reconstruction methods such as graph-cuts [5]. Yet, a false match rejection is required for all methods. Thus, even for global methods, a local rejection rule involving block-matching seems necessary. The elimination of mismatches in the block-matching framework was considered in [6], where the two causes of mismatch are considered, namely the mismatches on weakly textured objects, and on periodic structures. A *confidently stable matching* was defined in order to establish the largest possible unambiguous matching at a given confidence level. The method has two parameters that control the compromise between the percentage of bad matches and the density of the map, but the match density falls dramatically when the percentage of mismatches decreases. Similarly, [7] tries to eliminate errors on repeated patterns using a self-similarity function, the auto-SSD function (see Section 3), but with this method the matches seem to concentrate mainly on image edges. This is problematic, because edges are prone to fattening errors [2]. It is therefore important to confirm meaningful matches, even in flat textured regions. Since such regions may look like noise, a fine statistical decision is required to accept or reject matches in texture. The auto-SSD strategy does not eliminate errors due to moving objects, as will be shown in Section 3.

The method proposed here will be based on the *a contrario* approach [8], an adaptation to image analysis of hypothesis testing. The basic assumption of *a contrario methods* is the *Helmholtz principle*, according to which all perceptions can be characterized as having a low expectation of occurring just by chance. This expectation is estimated by a number of false alarms (NFA). The *a contrario* method has been successfully applied in shape matching [9], Scale Invariant Feature Transform (SIFT) [10], change detection [11, 12], and region similarity [13, 14, 15], but not yet to block-matching. The main ingredient of the *a contrario* method for blocks will be the learning of an accurate probability distribution for image blocks which generates a statistical *background model* for blocks. Precursors on statistical background modeling are [16, 17].

Section 2 describes this learning and proves the main result, Theorem 1, which guarantees a false alarm control. Section 3 contains comparative experiments and a conclusion.

2. THE *A CONTRARIO* MODEL

In this section the ACBM (*a contrario* block-matching) model is presented. Consider a stereo image pair I, I' in epipolar geometry. With a low baseline, the deformations between the images of the stereo pair are minor. Thus, one can reject matches $\mathbf{q} \in I \to \mathbf{q}' \in I'$ by comparing a block $B_{\mathbf{q}}$ around $\mathbf{q}$ with a block $B_{\mathbf{q}'}$ around $\mathbf{q}'$. Realistically the blocks are $s \times s$ squares with s ranging from 5 to 11. The *a contrario* model for blocks will be defined after a dimension reduction by standard Principal Componet Analysis (PCA) of the s^2-dimensional set of all image blocks. By keeping the first $N < s^2$ components with larger eigenvalues, the dimension is reduced but the most significant information retained. A global ordering of PCA eigenvectors is used to select the main components. A local ordering will instead be used for the statistical matching rule. From now on the N PCA coordinates of each block $B_{\mathbf{q}}$ will be ordered in decreasing absolute value. In that way, comparisons of the PCA components of $B_{\mathbf{q}}$ to those of $B_{\mathbf{q}'}$ will be made from the most relevant to the least relevant one for this particular block. Let $(c_{\sigma_{\mathbf{q}}(1)}(\mathbf{q}), \dots, c_{\sigma_{\mathbf{q}}(N)}(\mathbf{q}))$ be the PCA coefficients of $B_{\mathbf{q}}$, ordered in that way. By a slight abuse of notation, we will write $c_i(\mathbf{q})$ instead of $c_{\sigma_{\mathbf{q}}(i)}(\mathbf{q})$.

Let $\mathbf{q}$ be a point in the reference image I. We look for a pixel $\mathbf{q}'$ in the secondary image I' such that $B_{\mathbf{q}}$ and $B_{\mathbf{q}'}$ are similar and wish to establish a rejection criterion. The idea is to estimate accurately the probability that "just by chance" a block $B_{\mathbf{q}'}$ in I' looks like a block $B_{\mathbf{q}}$ in I. This will be done by learning from I' a realistic

random model for the PCA components $c_i(\mathbf{q}')$ of blocks of I'.

Definition 1 (empirical probability) *Let $B_{\mathbf{q}}$ be a block in I. We call empirical probability that an observed block $B_{\mathbf{q}'}$ in I' be similar to $B_{\mathbf{q}}$ for the feature i,*

$$\widehat{p^i}_{\mathbf{q}\,\mathbf{q}'} = \begin{cases} H_i(\mathbf{q}') & \text{if } H_i(\mathbf{q}) < |H_i(\mathbf{q}) - H_i(\mathbf{q}')| \\ 1 - H_i(\mathbf{q}') & \text{if } 1 - H_i(\mathbf{q}) < |H_i(\mathbf{q}) - H_i(\mathbf{q}')| \\ 2 \cdot |H_i(\mathbf{q}) - H_i(\mathbf{q}')| & \text{otherwise} \end{cases}$$

where $H_i(\mathbf{q}) := H_i(c_i(\mathbf{q}))$ is the normalized cumulative histogram of $c_i(\mathbf{q})$ for I' (see Fig. 1).

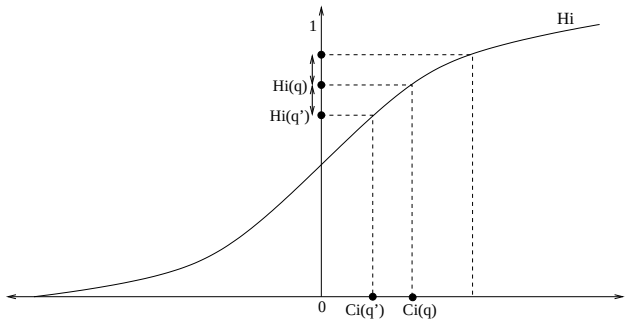

Fig. 1. Computation of the empirical probability.

The first principal components contain the more relevant information in the block. Thus, if two blocks are not similar for one of the first components, they should not be matched, even if their next components are similar. Due to this fact, the components will be compared with a nondecreasing exigency level. Furthermore, a quantized definition of the empirical probabilities will be needed to limit the number of tests.

Definition 2 (quantized probability) *Let $B_{\mathbf{q}}$ be a block in I. Let $\Pi := \{\pi_j = 1/2^{j-1}\}_{j=1,\dots,Q}$ be a set of probability thresholds and let*
$\Upsilon := \{\, p = (p_1, \dots, p_N) \mid p_i \in \Pi, \quad p_i \leqslant p_j \text{ if } i < j\}$
be the family of non-decreasing N-tuples in Π^N. The quantized empirical probability that $B_{\mathbf{q}'}$ be similar to $B_{\mathbf{q}}$ for the feature i, is defined by $p^i_{\mathbf{q}\,\mathbf{q}'} = \inf_{t \in \Pi} \{t \geqslant \sup_{j \leqslant i} (\widehat{p^j}_{\mathbf{q}\,\mathbf{q}'})\}$.

In short, $(p^1_{\mathbf{q}\,\mathbf{q}'}, \dots, p^N_{\mathbf{q}\,\mathbf{q}'})$ is the smallest upper bound in Υ of the empirical probabilities $(\widehat{p^1}_{\mathbf{q}\,\mathbf{q}'}, \dots, \widehat{p^N}_{\mathbf{q}\,\mathbf{q}'})$.

Definition 3 (*a contrario* model) *We call* a contrario model *associated with a reference image a vectorial random field defined on the image domain, with values in $\mathbb{R}^N$, $\mathbb{C}(\mathbf{q}) = (\mathbb{C}_1(\mathbf{q}), \dots, \mathbb{C}_N(\mathbf{q}))$ such that*
• for each $\mathbf{q} \in I$, the components $\mathbb{C}_i(\mathbf{q})$, $i = 1, \dots, N$ are independent random variables;
• for each i, the law of $\mathbb{C}_i(\mathbf{q})$ is the empirical histogram of $c_i(\cdot)$ for the reference image.

Thanks to the independence of the $\mathbb{C}_i(\mathbf{q})$, the above definition of the *a contrario* model will allow one to compute a block resemblance probability as the product of the marginal resemblance probabilities of the $\mathbb{C}_i(\mathbf{q})$. There is a strong adequacy of this independence assumption to the empirical model. Indeed, the PCA transform ensures that the $c_i(\mathbf{q})$ are empirically uncorrelated.

Definition 4 (Number of false alarms) *Let $B_{\mathbf{q}} \in I$ and $B_{\mathbf{q}'} \in I'$ be two observed blocks. Define the Number of False Alarms of the event* "a random block $\mathbb{B}_{\mathbf{q}'}$ is as similar to $B_{\mathbf{q}}$ as $B_{\mathbf{q}'}$ is" *by*

$$NFA_{\mathbf{q}\,\mathbf{q}'} = N_{test} \cdot Pr_{\mathbf{q}\,\mathbf{q}'},$$

where N_{test} is the number of tested matches, and $Pr_{\mathbf{q}\,\mathbf{q}'}$ the probability that $\mathbb{B}_{\mathbf{q}'}$ be as similar to $B_{\mathbf{q}}$ as $B_{\mathbf{q}'}$ under the a contrario model for $\mathbb{B}_{\mathbf{q}'}$.

Since by Def. 3, the principal components are independent under the *a contrario* model, the probability that $\mathbb{B}_{\mathbf{q}'}$ is that similar to $B_{\mathbf{q}}$ is equal to $Pr_{\mathbf{q}\,\mathbf{q}'} = \prod_{i=1}^{N} p^i_{\mathbf{q}\,\mathbf{q}'}$. Therefore, $NFA_{\mathbf{q}\,\mathbf{q}'} = N_{test} \cdot \prod_{i=1}^{N} p^i_{\mathbf{q}\,\mathbf{q}'}$.

Definition 5 (ϵ-meaningful match) *A pair of pixels $\mathbf{q}$ and $\mathbf{q}'$ in a stereo pair of images is an ϵ-meaningful match if*

$$NFA_{\mathbf{q}\,\mathbf{q}'} \leqslant \epsilon\,.$$

This generic definition from [8] gives here a tool to decide whether a match is meaningful or not. The NFA of a match actually also gives a security level: the smaller the NFA, the more meaningful the match.

We now address the computation of N_{test}, the number of performed tests for comparing all the blocks. It is the product of three terms. The first one is the image size $\#I$. The second one is the size of the search region which we denote by $S' \subset I'$. We mentioned before that the search is done on the epipolar line. In practice, a segment of this line is enough. If $\mathbf{q} = (q_1, q_2)$ is the point of reference we look for $\mathbf{q}' = (q_1', q_2) \in I'$ such that $q_1' \in [q_1 - R, q_1 + R]$ where R is a fixed integer larger than the maximal possible disparity. The third and most important factor is the number of tested nondecreasing probability distributions $FC_{N,Q} = \#\Upsilon$. This number is a function of the number N of principal components and of the number Q of probability quanta. Thus
$N_{test} = \#I \cdot \#S' \cdot \#\Upsilon = n\,(2R+1)\,FC_{N,Q}$, and it is an easy check that $FC_{N,Q} = \sum_{t=0}^{Q} (t+1) \cdot \binom{N+Q-t-3}{Q-t-1}$.

Theorem 1 *Let $\Gamma = \Sigma_{\mathbf{q},\mathbf{q}'} \chi_{B_{\mathbf{q}},\mathbb{B}_{\mathbf{q}'}}$ be the random number of occurrences of an ϵ-meaningful match between a deterministic block $B_{\mathbf{q}}$ in I and a random block satisfying the* a contrario *model in I'. Then the expectation of Γ is smaller than ϵ.*

Proof 1 *Using the linearity of the expectation to add:*
$\chi_{B_{\mathbf{q}},\mathbb{B}_{\mathbf{q}'}} = \left\{ \begin{array}{ll} 1, & \text{if } NFA(B_{\mathbf{q}}, \mathbb{B}_{\mathbf{q}'}) \leqslant \epsilon \\ 0, & \text{if } NFA(B_{\mathbf{q}}, \mathbb{B}_{\mathbf{q}'}) > \epsilon \end{array} \right.$ *yields*
$\mathbb{E}[\Gamma] = \sum_{\mathbf{q},\mathbf{q}'} \mathbb{E}[\chi_{\mathbf{q},\mathbf{q}'}] = \sum_{\mathbf{q},\mathbf{q}'} \mathbb{P}\left[NFA(B_{\mathbf{q}}, \mathbb{B}_{\mathbf{q}'}) \leqslant \epsilon\right]$. *The probability inside the expectation can be computed using definitions 4 and 1 as $\mathbb{E}[\chi_{\mathbf{q},\mathbf{q}'}] := \mathbb{P}\left[NFA(B_{\mathbf{q}}, \mathbb{B}_{\mathbf{q}'}) \leqslant \epsilon\right]$*
$= \mathbb{P}\left[\prod_i^N p^i(B_{\mathbf{q}}, \mathbb{B}_{\mathbf{q}'}) \leqslant \frac{\epsilon}{N_{test}}\right]$. *The probability of the non-disjoint union of events can be upper-bounded by their probability sum, and the intersection below involves only independent events*

according to our background model. Thus:

$$\begin{aligned}
&\mathbb{E}[\chi_{\mathbf{q},\mathbf{q}'}] = \\
&= \mathbb{P}\left[\bigcup_{\substack{p\in\Upsilon \\ \prod_i p_i \leqslant \epsilon/N_{test}}} \bigcap_i \left(2\cdot|H_i(c_i(\mathbf{q})) - H_i(\mathbb{C}_i(\mathbf{q}'))| \leqslant p_i\right)\right] \\
&\leqslant \sum_{\substack{p\in\Upsilon \\ \prod_i p_i \leqslant \epsilon/N_{test}}} \prod_i \mathbb{P}\left[2\cdot|H_i(c_i(\mathbf{q})) - H_i(\mathbb{C}_i(\mathbf{q}'))| \leqslant p_i\right] \\
&= \sum_{\substack{p\in\Upsilon \\ \prod_i p_i \leqslant \epsilon/N_{test}}} \prod_i p_i \leqslant \frac{\epsilon}{\#I\,\#S'}.
\end{aligned}$$

In the last line we used the fact that $H_i(\mathbb{C}_i(\mathbf{q}'))$ follows a $\mathrm{Uni}[0,1]$ *distribution, since the random variable $\mathbb{C}_i(\mathbf{q}')$ is drawn from the cumulative distribution H_i. Finally, recalling that $N_{tests} = \#I\,\#S'\,\#\Upsilon$, this last sum can be upper bounded by $\frac{\epsilon}{\#I\,\#S'}$. So we have shown that*

$$\mathbb{E}[\Gamma] = \sum_{\mathbf{q},\mathbf{q}'} \mathbb{E}\left[\chi_{B_{\mathbf{q}},\mathbb{B}_{\mathbf{q}'}}\right] \leqslant \sum_{\mathbf{q},\mathbf{q}'} \frac{\epsilon}{\#I\,\#S'} = \epsilon.$$

Remark 1 *The NFA indicates the level of similarity between two points: the smaller the NFA, the more meaningful the match. In fact, given ϵ, Def. 5 gives a tool to decide whether a match is meaningful or not. Thanks to the Theorem 1 the ϵ parameter can be fixed once and for all. If for instance the desired number of admissible mismatches is 1, then fix $\epsilon = 1$. This will mean that on average not more than one mismatch will occur (provided the a contrario model $\mathbb{B}$ for the blocks in I' is faithful). Other fixed parameters are: the size of the patch (9×9), the number of components ($N = 9$) and the number of probability thresholds ($Q = 5$). Since the dependency on these parameters is very low, they are fixed for all the images. Then, the presented accept/reject decision rule (ACBM) can be seen as a parameterless method.*

3. EXPERIMENTAL RESULTS, CONCLUSION

The *a contrario* test permits to eliminate systematically all unreliable matches, in particular all wrong matches caused by moving objects or poorly textured regions. Fig. 2 provided by CNES is a typical urban aerial scene where vehicles and pedestrians have moved. The rejected matches are left in black. The corresponding mask of accepted matches shows that the moving objects have not caused a single mismatch. The rejection criterion of MARC (Multiresolution Algorithm for Refined Correlation, CNES patented algorithm [4]) fails in the images having moving objects. Its final disparities on streets have dark (too low) or white (too high) blobs corresponding to vehicles and pedestrians with wrong disparities caused by their motion.

Rejecting false matches in stereovision has already been addressed by computing the similarity of matches in one of the images. This similarity is usually measured with SSD (Sum of Squared Differences). This is the case of [7] which used the so-called auto-SSD function. In [18] a similar criterion is proposed in the context of matching of local image descriptors (SIFT).

More precisely, the test consists in matching $\mathbf{q} \rightarrow \mathbf{q}'$ if $SSD(B_{\mathbf{q}}, B_{\mathbf{q}'}) < min\{SSD(B_{\mathbf{q}}, B_{\mathbf{r}})|\ r \in I \cap S(\mathbf{q})\}$, where $S(\mathbf{q})$ is a neighborhood of $\mathbf{q}$.

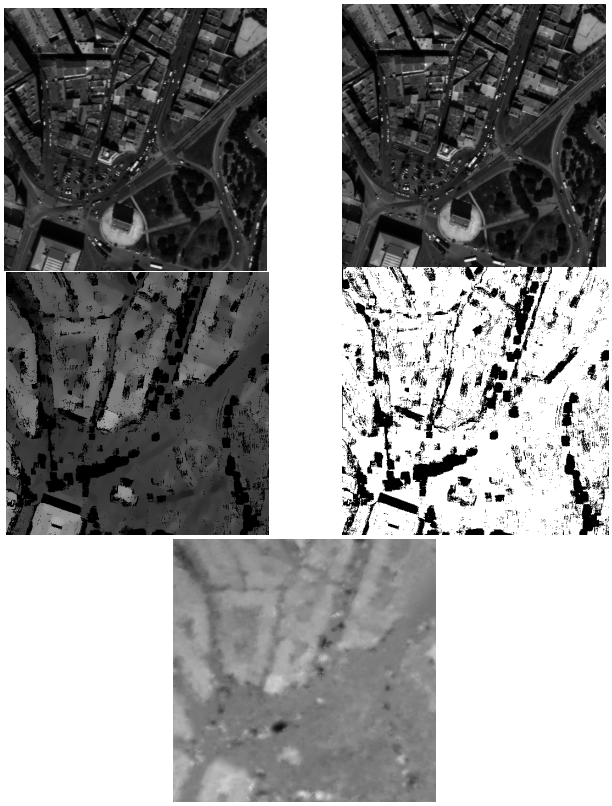

Fig. 2. Top: Left and right aerial images. Several vehicles have moved. Middle-left, disparity obtained with ACBM (black points are rejected. The brighter the disparity the higher the point). Right, mask of accepted (white) points. Bottom: Disparity map obtained by MARC. Mismatches occur with moving vehicles.

The straightforward auto-SSD threshold is able to reject ambiguous matches due to periodic structures in the epipolar direction (Class 2 errors defined in [6]). But this test is not enough for rejecting matches in moving objects. Fig. 3 shows two examples where there are pedestrians and vehicles moving. ACBM rejects the wrong matches due to motion, and the auto-SSD test fails to do so. MARC obtains several mismatches in its disparity map.

The ACBM method has also been compared to other *non-dense* methods rejecting false matches in simultaneous stereo. Table 1 compares the ACBM results with the results in [6], [19] and [20], published on the first Middlebury benchmark dataset [21].

The ACBM threshold is able to detect moving objects and poor or periodic textured regions by performing a rigorous selection of meaningful, reliable matches. This is particularly important in quasi-simultaneous stereo. The ACBM method with fixed threshold $\varepsilon = 1$ can be seen as a completely automatic validation procedure. In future research, it will be used either as a non dense stereo algorithm which must be completed by global interpolation, or as an *a posteriori* check of matches obtained by other stereo matching methods.[1]

[1]Work was supported by the French Space Agency (CNES). Input images in this paper were provided and copyrighted by CNES.

	Tsukuba		Sawtooth		Venus		Map	
	Error	Density	Error	Density	Error	Density	Error	Density
ACBM	**0.31**	45.6	**0.09**	65.7	**0.02**	54.1	**0.0**	84.8
Sara [6]	1.4	45	1.6	52	0.8	40	0.3	74
Veksler [19]	0.36	75	0.54	87	0.16	73	0.01	87
Mordohai [20]	1.18	74.5	0.27	78.4	0.20	74.1	0.08	94.2

Table 1. Quantitative results on the first Middlebury benchmark dataset. The error statistics (percentage) are computed on the mask of non occluded pixels and a mismatch is an error bigger than 1 pixel. ACBM obtains many less mismatches in the four images with a comparable proportion of good matches.

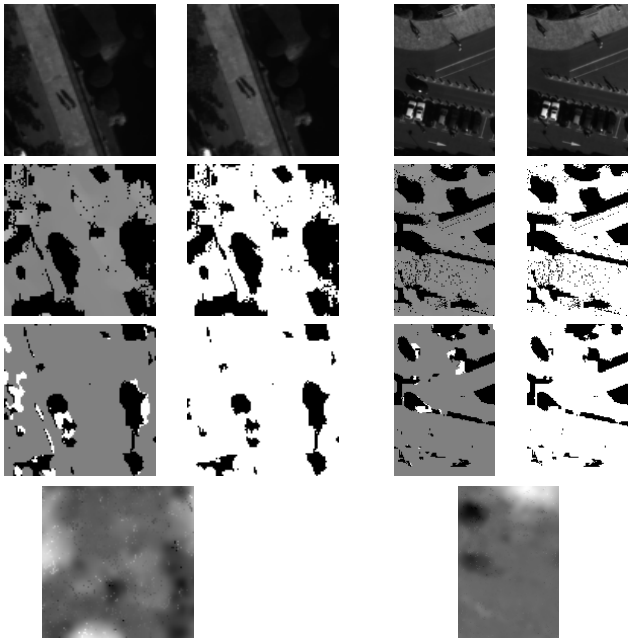

Fig. 3. Top: Left and right quasi-simultaneous aerial images. In the first example, a pedestrian has moved several meters. In the second one, a car disappears and several pedestrians have moved. 2nd line: Disparity map obtained with ACBM and its associated mask of accepted points. 3d line: Disparity map and mask obtained with the self-similarity threshold. Several mismatches (white points in the disparity map) occur. Bottom: MARC disparity map. All bright or dark disparities are wrong.

4. REFERENCES

[1] R. Szeliski and D. Scharstein, "Symmetric sub-pixel stereo matching," *ECCV*, vol. 2, pp. 525–540, 2002.

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

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

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

[5] V. Kolmogorov and R. Zabih, *Graph Cut Algorithms for Binocular Stereo with Occlusions*, Mathematical Models in Computer Vision: The Handbook, Springer-Verlag, 2005.

[6] R. Sara, "Finding the largest unambiguous component of stereo matching," in *ECCV*, 2002, pp. 900–914.

[7] R. Manduchi and C. Tomasi, "Distinctiveness maps for image matching," *ICIP*, pp. 26–31, 1999.

[8] A. Desolneux, L. Moisan, and J.M. Morel, *From Gestalt Theory to Image Analysis. A probabilistic Approach*, Springer, 2007.

[9] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel, "An a contrario decision method for shape element recognition," *IJCV*, vol. 69, no. 3, pp. 295–315, 2006.

[10] J. Rabin, J. Delon, and Y.Gousseau, "A contrario matching of sift-like descriptors," 2008.

[11] A. Robin, L. Moisan, and S. Le Hégarat-Mascle, "An a-contrario approach for sub-pixel change detection in satellite imagery," *Tech. Report MAP5 2009-15*, 2009.

[12] F. Dibos, S. Pelletier, and G. Koepfler, "Real-time segmentation of moving objects in a video sequence by a contrario detection," in *ICIP*, 2005.

[13] G. Née, S.e Jehan-Besson, L. Brun, and M. Revenu, "Significance tests and statistical inequalities for region matching," *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 350–360, 2008.

[14] L. Igual, J. Preciozzi, L. Garrido, A. Almansa, V. Caselles, and B. Rougé, "Automatic low baseline stereo in urban areas," *Inverse Problems and Imaging*, vol. 1, no. 2, pp. 319–348, 2007.

[15] N. Burrus, T. M. Bernard, and J.-M. Jolion, "Image segmentation by a contrario simulation," *Pattern Recognition*, vol. 42, no. 7, pp. 1520–1532, July 2009.

[16] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," *CVPR*, vol. 2, pp. 302–309, 2004.

[17] K. A. Patwardhan, G. Sapiro, and V. Morellas, "Robust foreground detection in video using pixel layers," *IEEE PAMI*, vol. 30, no. 4, pp. 746–751, 2008.

[18] D. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[19] O. Veksler, "Extracting dense features for visual correspondence with graph cuts," in *CVPR*, 2003, vol. 1, pp. 689–694.

[20] P. Mordohai and G. Medioni, "Stereo using monocular cues within the tensor voting framework," *IEEE T-PAMI*, vol. 28, pp. 968–982, 2006.

[21] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, no. 47(1/2/3), pp. 7–42, 2002.