Available online at www.sciencedirect.com
SCIENCE DIRECT
Image and Vision Computing 22 (2004) 983–1005
image AND vision COMPUTING
www.elsevier.com/locate/imavis

A fast area-based stereo matching algorithm

Luigi Di Stefano[a,b], Massimiliano Marchionni[a,b], Stefano Mattoccia[a,b,*]

[a]*Department of Electronics Computer Science and Systems (DEIS), University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy*
[b]*Advanced Research Center on Electronic Systems for Information and Communication Technologies 'Ercole De Castro' (ARCES), Via Toffano 2/2, 40135 Bologna, Italy*

Received 31 July 2003; received in revised form 26 November 2003; accepted 22 March 2004

Abstract

This paper proposes an area-based stereo algorithm suitable to real time applications. The core of the algorithm relies on the uniqueness constraint and on a matching process that rejects previous matches as soon as more reliable ones are found. The proposed approach is also compared with bidirectional matching (BM), since the latter is the basic method for detecting unreliable matches in most area-based stereo algorithms. We describe the algorithm's matching core, the additional constraints introduced to improve the reliability and the computational optimizations carried out to achieve a very fast implementation. We provide a large set of experimental results, obtained on a standard set of images with ground-truth as well as on stereo sequences, and computation time measurements. These data are used to evaluate the proposed algorithm and compare it with a well-known algorithm based on BM.

Keywords: Stereo; Area-based; Real-time; Uniqueness constraint; Bidirectional matching

1. Introduction

Dense depth measurements are required in many applications, such as teleconferencing, robot navigation and control, exploration and modelling of unstructured environments, virtual reality. According to a recent taxonomy (Scharstein and Szeliski, 2002), stereo algorithms that generate dense depth measurements can be roughly divided into two classes, namely *global* and *local* algorithms. *Global* algorithms [1], rely on iterative schemes that carry out disparity assignments on the basis of the minimisation of a global cost function. These algorithms yield accurate and dense disparity measurements but exhibit a very high computational cost that so far renders them unsuited to real-time applications. *Local* algorithms [2–8], also referred to as area-based algorithms, calculate the disparity at each pixel on the basis of the photometric properties of the neighbouring pixels. Compared to *global* algorithms, *local* algorithms yield significantly less accurate disparity maps but, nowadays, thanks to both research and technology advances, can run fast enough to be deployed in many real-time applications. Numerous examples of real-time dense stereo applications based on the use of local algorithms can be found at the web sites [9,10].

As far as local matching algorithms are concerned, and considering the more common case of a binocular stereo imaging system, a widely adopted method [2,4–8] aimed at detecting unreliable matches, such as for example those due to occlusions or photometric distortions, is the so called *bidirectional matching (BM)* [11], also referred to as *left–right consistency constraint* or *left–right check*. The method can be described as follows. Initially, for each point of the left image find the best match into the right image. Then, reverse the role of the two images and for each point of the right image find the best match into the left image. Finally, keep only those matches that turn out to be coherent when matching left-to-right (direct matching phase) and right-to-left (reverse matching phase). It is worth observing that in both phases the match associated with each pixel is established independently of those found at neighbouring pixels, since the other matching phase will highlight ambiguous matches. BM has proven to be effective in detecting and discarding the erroneous matches necessarily yield by area-based algorithms in presence of occlusions

* Corresponding author.
E-mail addresses: smattoccia@deis.unibo.it (S. Mattoccia), ldistefano@deis.unibo.it (L. Di Stefano), mmarchionni@litio.it (M. Marchionni).

0262-8856/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2004.03.009

[12]. However, this approach is characterised by a significant computational cost. In fact, it requires two matching phases (direct and reverse) and, although some authors have proposed calculation schemes aimed at reducing the impact of the left–right check on the overall stereo execution time [6], in most implementations this implies doubling the computational complexity of the matching process.

We propose a fast *local* algorithm, referred to as *single matching phase (SMP)*, which enables real-time dense stereo applications on a standard Personal Computer. The algorithm is based on a matching core that detects unreliable matches during the direct matching phase and therefore does not require a reverse matching phase.

The paper is organised as follows: Section 2 describes the matching core of the proposed SMP approach and highlights the main difference with respect to BM. Section 3 addresses the issue of improving the reliability of the disparity measurements provided by the basic SMP core. Section 4 outlines the overall stereo matching algorithm based on the SMP approach. Section 5 describes the computational optimisation carried out to minimise the number of calculations required by the stereo algorithm. Section 6 presents and discusses the experimental results obtained with our SMP-based algorithm as well as with a well known BM-based stereo algorithm. Finally, the concluding remarks are drawn in Section 7.

2. The proposed matching approach

We assume here a binocular stereo pair and images in standard form, i.e. with corresponding epipolar lines lying on corresponding image scanlines. Should the latter assumption not be verified, a suitable transformation, known as rectification [13,15] can be applied to obtain a pair of images in standard form from the original ones. In Section 6 we provide experimental results obtained on stereo pairs in standard form as well as on rectified stereo pairs.

Hence, in *local* algorithms, given a point in the reference image the homologous point is selected by searching along the corresponding scanline in the other image, and within a certain disparity range, for the point that minimise (maximize) an error (similarity) function, ε, representing the degree of dissimilarity (similarity) between two small regions centered at the points under examination. Unlike algorithms based on BM, which rely on a direct (i.e. left-to-right) and a reverse (i.e. right-to-left) matching phase, our algorithm uses only a direct matching phase. Our approach relies on the *uniqueness constraint*, which states that a 3D point can be projected at most in one point of each image of the stereo pair, as well as on the ability of modifying disparity measurements dynamically as long as the matching process proceeds.

Let's assume that the left image is the reference, that disparity, d, belongs to the interval $[0,\ldots,d_{\max}]$ and that the left image is scanned from top to bottom and from left to right. Moreover, without loss of generality, we assume that the matching process relies on an error function. The extension to the case of similarity functions is straightforward.

Starting from one point of left image, say $L(x-d_{\max},y)$, the algorithm searches for the best candidate by evaluating function ε within the interval $[R(x-d_{\max},y),\ldots,R(x,y)]$. Then, for the successive point of reference image $L(x+1-d_{\max},y)$ the procedure is repeated searching for the best match within $[R(x+1-d_{\max},y),\ldots,R(x+1,y)]$. The process is then iterated for the successive points along the scanline. This procedure is outlined in Fig. 1, which shows for each point of the left image belonging to the interval $[L(x-d_{\max},y),\ldots,L(x,y)]$ the potential matching points in the right image within the disparity range $[0,\ldots,d_{\max}]$. Note also that in the figure the arcs are marked with the disparity value that brings one point of the left image into the same point $R(x,y)$ of the right image.

Suppose now that the best match found for $L(x+\beta-d_{\max},y)$ is $R(x,y)$, with similarity score $\varepsilon(x+\beta-d_{\max},x,y)$. We adopt the notation $L(x+\beta-d_{\max},y) \rightharpoonup R(x,y)$ to indicate that this match from left to right has been established.

As it is common in area-based algorithms, we use photometric properties, encoded by the error function, as the main cue driving the matching process, even though this cue may be ambiguous due to many causes such as for example photometric distortions, occlusions and noise. However, incorrect matches expose inconsistencies within the set of matches already established that can be deployed to detect and discard them.

Thus, let's suppose that another point of the left image, say $L(x+\alpha-d_{\max},y)$, with $\alpha \le \beta$, has been previously matched with $R(x,y)$ with score $\varepsilon(x+\alpha-d_{\max},x,y)$. This situation, that violates the uniqueness constraint, is used in our approach to detect incorrect matches. In fact, based on the uniqueness constraint we assume that at least one of the two matches, i.e. $L(x+\beta-d_{\max},y) \rightharpoonup R(x,y)$ or $L(x+\alpha-d_{\max},y) \rightharpoonup R(x,y)$, is incorrect and retain the match having the better score. Thus, if the point currently analyzed $L(x+\beta-d_{\max},y)$ has a better score than $L(x+\alpha-d_{\max},y)$ (i.e. $\varepsilon(x+\beta-d_{\max},x,y) \le \varepsilon(x+\alpha-d_{\max},x,y)$) our

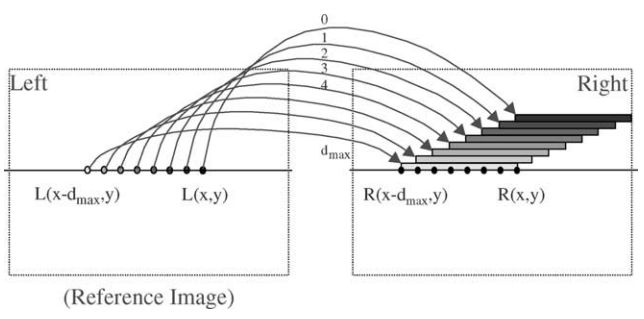

Fig. 1. Matching from left to right.

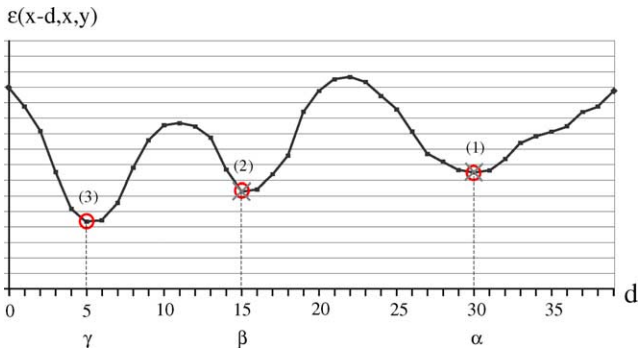

Fig. 2. Scores associated with point $R(x, y)$.

algorithm will reject the previous match and accept the new one. This implies that, although the proposed approach relies on a direct matching phase only, it allows for recovering from possible previous matching errors.

The capability of our algorithm to recover from previous errors as long as better matches are found during the search is also shown in Fig. 2, which plots as a function of $d \in [0, \ldots, d_{\max}]$ all the scores between the point $R(x, y)$ of right image and the points in the reference image $[L(x-d_{\max}, y), \ldots, L(x, y)]$ that are allowed to establish a correspondence with $R(x, y)$ (see Fig. 3).

Recalling the arcs drawn in Fig. 1, we can notice that smaller d values correspond to scores computed more recently while greater d values to scores computed earlier. Considering again the two matches (1) $L(x+\alpha-d_{\max}, y) \rightarrow R(x, y)$ and (2) $L(x+\beta-d_{\max}, y) \rightarrow R(x, y)$, the algorithm will discard the old one, (1) since the new one, (2) has a better score with $R(x, y)$. Moreover, if we find a new 'collision' when analysing the successive points of the left image; (3) $L(x+\gamma-d_{\max}, y) \rightarrow R(x, y)$, the score of this new match will be compared with that associated with the current best match for $R(x, y)$, so as to retain only one single match. That is, as in Fig. 2, since $\varepsilon(x+\gamma-d_{\max}, x, y) \leq \varepsilon(x+\beta-d_{\max}, x, y), L(x+\beta-d_{\max}, y) \rightarrow R(x, y)$ will be discarded and the current match for $R(x, y)$ set to $L(x+\gamma-d_{\max}, y) \rightarrow R(x, y)$.

Fig. 4 shows a geometric interpretation of the proposed matching approach: as long as the left-to-right matching process proceeds, the algorithm disambiguates between all the matches that imply locating 3D points lying on the same

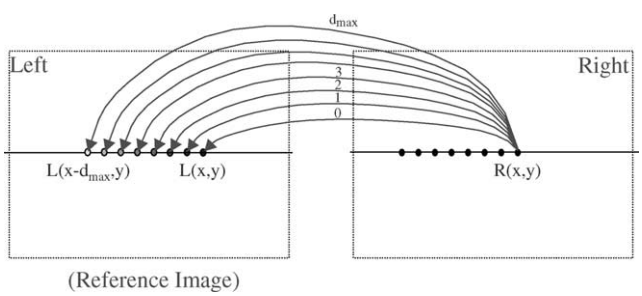

Fig. 3. Potential matches in the left image for $R(x, y)$.

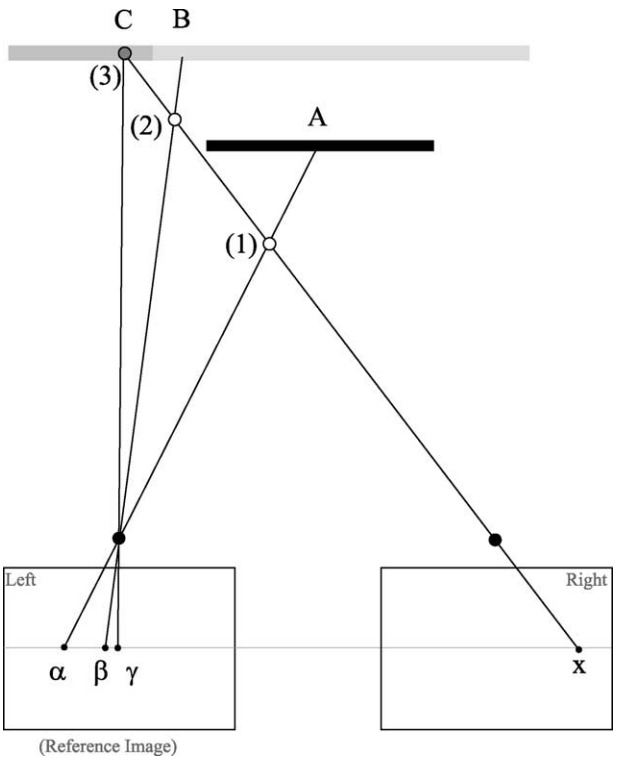

Fig. 4. Geometric interpretation.

line of sight originating from a point of the right image. These sort of matches, referred to as 'colliding matches' for a given point of the right image $R(x, y)$, can be caused by the ambiguity of photometric cues as well as by occlusions. In Fig. 4 the former situation is represented by the incorrect match between α and x, the latter by the incorrect match between β and x.

To conclude this section, we highlight one major difference between the SMP and BM approaches with the aid of Fig. 5 that, for a given scanline, shows on the *x-axis* the x-coordinates of the points of the left image that are

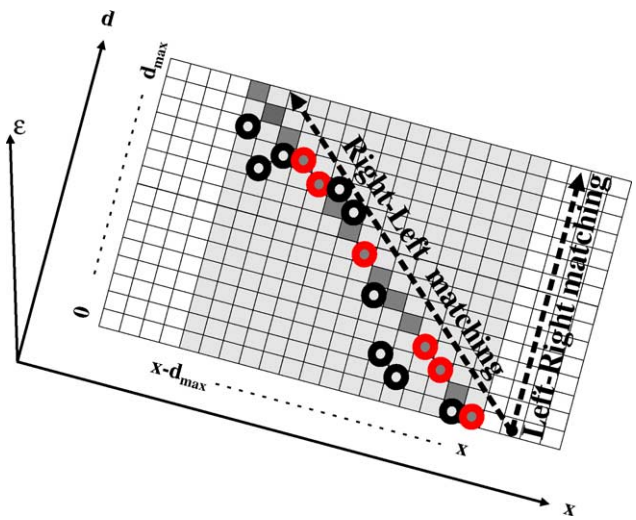

Fig. 5. The search paths of bidirectional matching.

allowed to match with point $R(x, y)$ and on the *d-axis* the disparity range. The score of the error function, not shown in the figure, would be represented on the *ε-axis*. The light-gray area defines all the potential matches for the points in the left image within the interval $[L(x - d_{\max}, y), \ldots, L(x, y)]$ with points of the right image. For each point in the left image, BM chooses in the direct phase the best score along a column in the light-gray area (the matches found when matching left-to-right have been marked with a circle in Fig. 5). Then, in the reverse phase, when matching $R(x, y)$ it chooses the best score along the diagonal path marked with the darker-gray level: a match is accepted only if the match found along this path turns out to be one of those found when matching left-to-right. It is worth noticing that, although during the reverse phase BM checks all of the potential matches along the path in darker-gray, the allowed ones for $R(x, y)$ turn out to be only those that in the direct phase fall in the darker-gray path (i.e. the circles lying in the darker gray path).

Conversely, when matching left-to-right, SMP chooses the best match along a column and at the same time checks if this is also the best match among those already found falling on the darker-gray path (i.e. left-to-right matches that collide on $R(x, y)$). Looking at Fig. 5, the matches cross-checked by SMP are only those represented by the circles lying on the darker-gray path. With SMP a point of the right image, $R(x, y)$, will be certainly matched if there is at least one match in the darker-gray path of Fig. 5 associated with $R(x, y)$. This is not true for BM: even though the direct phase finds a match for $R(x, y)$ lying in the darker-gray path, this may not correspond to the best match found along the darker-gray path in the reverse matching phase and in such a case $R(x, y)$ will not be matched.

Hence, the matching constraint embodied into SMP turns out to be less tight than that of BM and therefore SMP tends to accept more matches. Yet, this implies also that SMP is potentially more prone to mismatches. In the next section we show how to improve the reliability of the matches provided by the basic SMP matching core described so far.

3. Improving match reliability

The goal of improving the reliability of the disparity measurements provided by SMP can be accomplished by introducing additional constraints to the basic matching core. Since we are interested in a fast algorithm, suited to real-time stereo applications, a major guideline of this work has been to introduce new constraints that could also be implemented very efficiently. To this end, rather than carry out additional calculations, we try to exploit the information related to match reliability which is already embodied into the error scores computed by the matching core.

Let's consider Fig. 7, that plots the error scores (*ε-axis*) for each disparity value (*d-axis*) along the epipolar line (*x-axis*) marked in white in Fig. 6. In

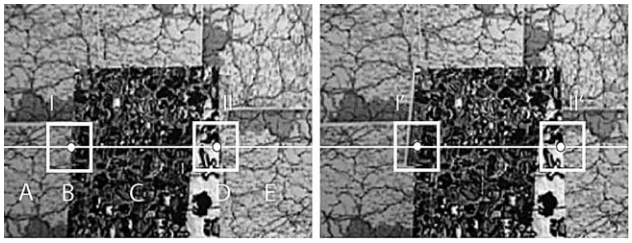

Fig. 6. *Map* stereo pair.

regions A, C and E the minima can be localised without ambiguity. Regions A and E belong to the background (left side and right side of the foreground object, respectively) while region C belongs to the foreground object. The minima in region C are sharper than those in regions A and E. This occurs because the foreground object is more textured than the background and thus the higher signal strength yields higher error scores for non homologous points. However, the sharpness of the minimum does not guarantee match reliability since in presence of repetitive patterns (not the case of the considered example) the global minimum of the error score might be sharp but close to some other local minimum. Hence, another important feature of the global minimum is its distinctiveness compared to the other local minima.

Differently from regions A, C and E, in regions B and D the minimum is ambiguous and therefore the matching process is more prone to mismatches. It can be noticed from Fig. 7 that in regions B and D the global minimum is significantly less sharp and also less distinctive. This occurs because in regions B and D the correlation window covers areas at different depths. Consider for example the point of the left image at the center of correlation window I in Fig. 6. Local support for this point comprises background and foreground points. Due to the geometry of stereopsis, when searching along the epipolar line for a similar area in the right image correlation window I cannot match completely any correlation window of the right image. In fact, since I is located across a depth discontinuity, the points within I do not maintain their spatial relationship when projected into the right image. Thus, only a portion of correlation window I (i.e. its background or foreground portion) can match perfectly a portion of a correlation window in the right image (i.e. the background or the foreground portion). Fig. 6 shows the case in which the foreground portion of the correlation window under examination (i.e. I) matches the foreground portion of correlation window I' in the right image, since they correspond to the same 3D regions, while the background portions of the two windows are very different since they correspond to different regions of the 3D scene. It is worth observing that in such situations the error score computed within the correlation window comprises two components: one is associated with the similar portions of the window (a low score) and the other with the different portions (a high score). Consequently, the overall score for the minimum turns out to be less sharp and also less

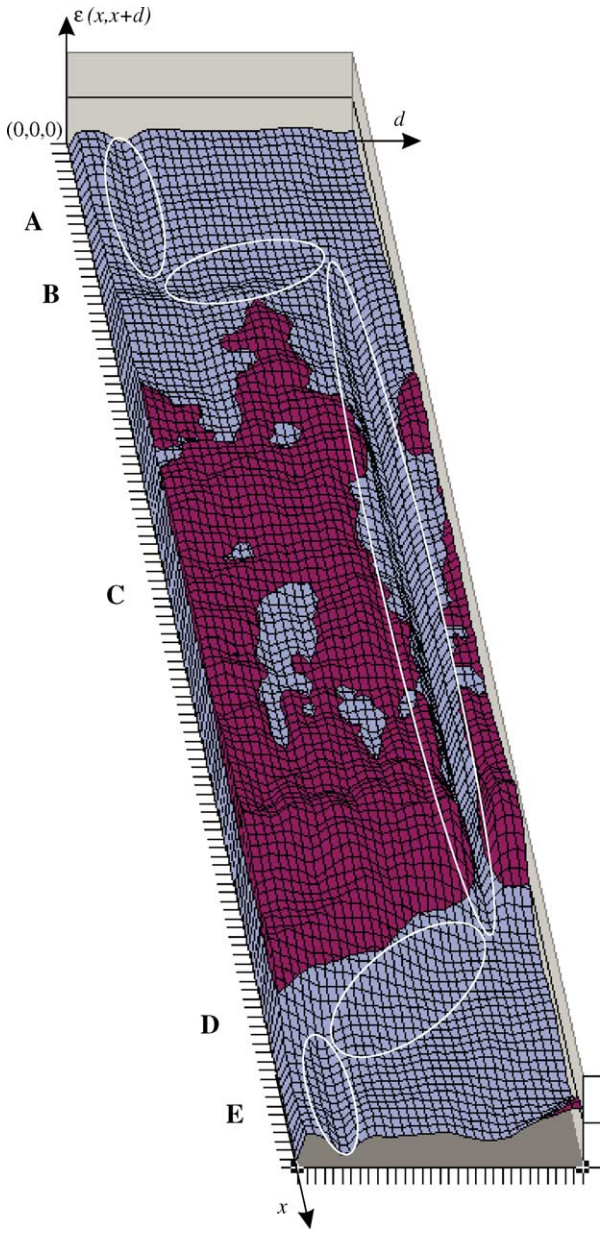

Fig. 7. Plot of the error scores along the line marked in white in Fig. 6.

distinctive than those found in textured regions without depth discontinuities (i.e. A, C and E). Moreover, in region D the presence of occluded points renders the correspondence problem even more ambiguous. The described effect yields matching errors and is responsible for the *border-localisation* problem [7,16]: objects' borders are not localised accurately with respect to their original position due to the ambiguity of the matching process across depth discontinuities.

According to the previous analysis, we have devised a strategy aimed at discarding ambiguous correspondences and based on estimating the sharpness and distinctiveness of the global minimum found by SMP's core.

In our algorithm the global minimum is located very quickly using a parallel technique [6] that with a few SIMD instructions (i.e. MMX instructions [17]) yields the score ($\varepsilon_{\min}$) and the position within the disparity range ($d_{\min}$) of the global minimum as well as the scores ($\varepsilon_1, \varepsilon_2, \varepsilon_3$) and positions ($d_1, d_2, d_3$) of three candidate minima, referred to as *pseudo-minima*. These exhibit small error scores but are not guaranteed to correspond to local minima. To discard ambiguous matches we estimate the behaviour of the error function by means of two tests that are carried out using only the global minimum and the three *pseudo-minima*.

When the three *pseudo-minima* fall far from the position of the global minimum we consider the match as ambiguous, unless the error score of the global minimum is much smaller than those of the *pseudo-minima*. On the other hand, when the *pseudo-minima* are close to the position of the global minimum, we consider the match as reliable.

The following relationship evaluates the degree of aggregation of the *pseudo-minima* in proximity of the global minimum:

$$\delta d = \sum_{i=1}^{3} |d_i - d_{\min}| \tag{1}$$

A low δd value (the lowest value is 4) indicates that the *pseudo-minima* are localized in proximity of the global minimum and thus the match is accepted as reliable. Conversely an high δd value means that the *pseudo-minima* are spread within the disparity range and hence the match is potentially ambiguous. In order to evaluate the reliability of the matches that do not satisfy the previous constraint we perform an additional test aimed at evaluating whether the score of the global minimum is much smaller than those of the *pseudo-minima*. The following relation

$$\delta\varepsilon = \sum_{i=1}^{3} (\varepsilon_i - \varepsilon_{\min}) \tag{2}$$

embodies information about the distinctiveness of the global minimum with respect to the three *pseudo-minima*. Actually, we consider the ratio $\delta\varepsilon/\varepsilon_{\min}$ to evaluate the distinctiveness of the global minimum, with high ratios indicating distinctive global minima.

4. The overall stereo algorithm

The overall stereo algorithm consists of the four main steps shown in Fig. 8.

(A) The input images are normalized by subtraction of the mean values of the intensities computed in a small window centered at each pixel [2]. This allows for compensating potential distortions such as slightly different settings of the cameras and–or variable photometric conditions. Moreover, since matching turns out to be highly unreliable when dealing with poorly textured areas, the variance of the intensities is calculated at each pixel considering a window of

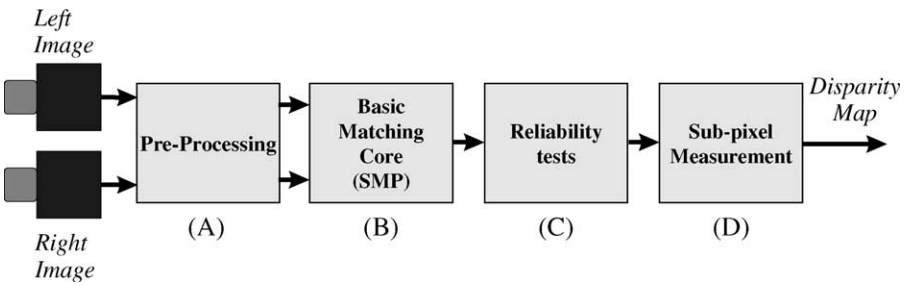

Fig. 8. The four processing steps.

the same size as that used to obtain the means. This information is used in step (C) to detect regions with lack of texture [2].

(B) The normalized images are matched according to the approach described in Section 2, which is independent of the matching function. Currently, we use the sum of absolute differences (SAD) but any other error function, e.g. the sum of squared differences (SSD), or similarity function, e.g. the normalized cross-correlation (NCC) or also non-parametric local transform [19], e.g. the *rank* or *census* transform, could be used.

(C) The reliability of the matches provided by the basic matching core is improved by means of the tests described in Section 3. In addition, this step uses the variance map computed in the pre-processing step to reject the matches found at points belonging to poorly textured areas. These points are detected comparing the values stored in the variance map against a fixed threshold value. The threshold value depends on the noise affecting the imaging process and can be chosen by pointing the stereo system towards a uniform object (e.g. a wall) and then increasing the value until most incorrect disparity measurements disappear.

(D) The final step performs sub-pixel refinement of disparities. Sub-pixel accuracy (up to 1/16 of pixel) is achieved detecting the minimum of a second degree curve interpolating the SAD scores in proximity of the minimum found by the matching core.

5. Computational optimisation

The most expensive task performed by the stereo algorithm is the computation of SAD scores, which are needed to carry out the direct matching phase. In this section we outline the optimisation techniques adopted to avoid redundant calculations. We show first the basic calculation scheme, already described in Ref. [6], and then propose an additional level of incremental calculation aimed at achieving further speed-up.

Suppose that $\mathrm{SAD}(x, y, d)$ is the SAD score between a window of size $(2n+1)(2n+1)$ centered at coordinates (x, y) in the left image and the corresponding window centered at $(x+d, y)$ in the right image:

$$\mathrm{SAD}(x, y, d) = \sum_{i,j=-n}^{n} |L(x+j, y+i) - R(x+d+j, y+i)| \tag{3}$$

Observing Fig. 9, it is easy to notice that $\mathrm{SAD}(x, y+1, d)$ can be attained from $\mathrm{SAD}(x, y, d)$:

$$\mathrm{SAD}(x, y+1, d) = \mathrm{SAD}(x, y, d) + U(x, y+1, d) \tag{4}$$

with $U(x, y+1, d)$ representing the difference between the SADs associated with the lowermost and uppermost rows of

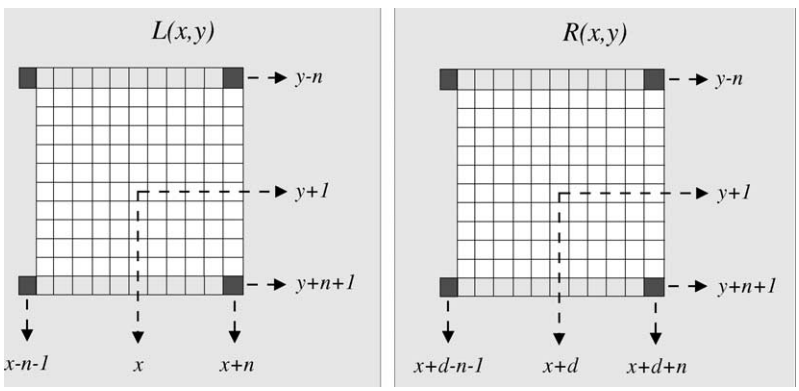

Fig. 9. Incremental calculation scheme.

the matching window (shown in light-gray in Fig. 9):

$$U(x,y+1,d) = \sum_{j=-n}^{n} |L(x+j,y+n+1) - R(x+d+j,y+n+1)| - \sum_{j=-n}^{n} |L(x+j,y-n) - R(x+d+j,y-n)| \quad (5)$$

Furthermore, $U(x,y+1,d)$ can be computed from $U(x-1,y+1,d)$ by simply considering the contributes associated with the four points shown in dark-gray in Fig. 9:

$$\begin{aligned} U(x,y+1,d) = U(x-1,y+1,d) &+ |L(x+n,y+n+1) - R(x+d+n,y+n+1)| \\ &- |L(x+n,y-n) - R(x+d+n,y-n)| \\ &- |L(x-n-1,y+n+1) - R(x+d-n-1,y+n+1)| \\ &+ |L(x-n-1,y-n) - R(x+d-n-1,y-n)| \end{aligned} \quad (6)$$

This allows for keeping complexity small and independent of the size of the matching window, since only four elementary operation are needed to obtain the SAD score at each new point.

The computational scheme of Eqs. (4) and (6) makes use of a *vertical* recursion to obtain the SAD and an *horizontal* recursion to obtain the updating term, U. Hence, it requires storing the SAD scores associated with the previous row (Wd_r values, if W is the width of the image and $d_r = d_{max} + 1$ the disparity range) and the d_r values of U associated with the previous point.

A similar incremental-calculation approach has been adopted in the stereo algorithms developed at INRIA [2] and CMU [3]. However, the scheme described in Refs. [2,3], make use of a *vertical* recursion to obtain the updating term and of an *horizontal* recursion to obtain the similarity (INRIA) or error (CMU) scores. Hence, in order to sustain the recursion the INRIA and CMU scheme requires storing the Wd_r values of the updating term associated with the previous row and the d_r values of the similarity-error scores associated with the previous point.

To implement efficiently our matching algorithm, which is based on disambiguating between the collisions occurring while matching left-to-right along a row, when matching a point of the left image it is necessary to obtain quickly the SAD scores associated with the previous points along the row. Hence, the scheme of Eqs. (4) and (6) is particularly suited to our matching algorithm: since, as the computation proceeds along a row, the scheme requires storing SAD scores to sustain the recursion, when matching a point of the left image the SAD scores of the previous points of the row are already available and can be accessed to disambiguate a collision. Note that this would not be the case of the INRIA and CMU scheme, for which the values stored to sustain the recursion as the computation proceeds along a row are those of the updating term.

As shown in Section 4, the pre-processing step requires computation of the mean and variance of the two images. Considering for example the left image, and posing $N^2 = (2n+1)(2n+1)$, the mean is given by

$$\mu_L(x,y) = \frac{1}{N^2} \sum_{i,j=-n}^{n} L(x+j,y+i) = \frac{1}{N^2} S_1(x,y) \quad (7)$$

while the variance can be expressed [20] as

$$\sigma_L^2(x,y) = \frac{1}{N^2} \sum_{i,j=-n}^{n} L^2(x+j,y+i) - \mu_L^2(x,y) = \frac{1}{N^2} S_2(x,y) - \mu_L^2(x,y) \quad (8)$$

Since Eqs. (7) and (8) rely on the same basic operation, namely scanning the image and summing-up intensities—or squared intensities, it can be easily verified that the computation of mean and variance can be carried out using the following schemes:

$$S_1(x,y+1) = S_1(x,y) + U_{S_1}(x,y+1) \quad (9)$$

$$U_{S_1}(x,y+1) = \sum_{j=-n}^{n} (L(x+j,y+n+1) - L(x+j,y-n)) \quad (10)$$

$$\begin{aligned} U_{S_1}(x,y+1) = U_{S_1}(x-1,y+1) &+ L(x+n,y+n+1) - L(x+n,y-n) \\ &- L(x-n-1,y+n+1) - L(x-n-1,y-n) \end{aligned} \quad (11)$$

$$S_2(x,y+1) = S_2(x,y) + U_{S_2}(x,y+1) \quad (12)$$

$$U_{S_2}(x,y+1) = \sum_{j=-n}^{n} (L^2(x+j,y+n+1) - L^2(x+j,y-n)) \quad (13)$$

$$\begin{aligned} U_{S_2}(x,y+1) = U_{S_2}(x-1,y+1) &+ L^2(x+n,y+n+1) - L^2(x+n,y-n) \\ &- L^2(x-n-1,y+n+1) - L^2(x-n-1,y-n) \end{aligned} \quad (14)$$

In both the matching and pre-processing steps it is possible to introduce a third level of incremental computation aimed at achieving additional speed-up. Both steps use the four pixels at the corners of the correlation window. Formulas (6), (11) and (14) show that these pixels contribute to two terms, say A and B, where A includes the two pixels on the left side of the correlation window and B those on the right side. Observing that term B plays the role of term A when the correlation window is shifted horizontally by $2n+1$ units, we can store, at a very small memory cost, the most recent $2n+1B$ terms so that they can be re-used $2n+1$ units later in place of the A terms. Calling T the array of the B terms, each element can be referenced with the index $\tilde{x} = x \bmod(2n+1)$ and thus all elements are visited each

time the correlation window is shifted horizontally by $2n+1$ units. When shifting the window by one unit, a new B term is calculated while the needed A term is fetched from $T(\tilde{x})$. After both terms have been used, $T(\tilde{x})$ is updated to the newly calculated B term.

To introduce this third level of incremental computation into the pre-processing step, formula (11) for the mean calculation is rewritten as follows:

$$U_{S_1}(x,y+1) = U_{S_1}(x-1,y+1) + L(x+n,y+n+1) - L(x+n,y-n) - T_1(\tilde{x}) \quad (15)$$

where

$$T_1(\tilde{x}) = L(x-n-1,y+n+1) - L(x-n-1,y-n) \quad (16)$$

with $\tilde{x} = x \bmod(2n+1)$

Similarly formula (14) for the variance calculation becomes:

$$U_{S_2}(x,y+1) = U_{S_2}(x-1,y+1) + L^2(x+n,y+n+1) - L^2(x+n,y-n) - T_2(\tilde{x}) \quad (17)$$

where

$$T_2(\tilde{x}) = L^2(x-n-1,y+n+1) - L^2(x-n-1,y-n)$$

with $\tilde{x} = x \bmod(2n+1)$ (18)

In the matching step the third level of incremental computation is applied for each disparity value $d \in [0, d_{\max}]$; thus, the array T grows by one dimension and formula (6) is rewritten as follows:

$$U(x,y+1,d) = U(x-1,y+1,d) + |L(x+n,y+n+1) - R(x+d+n,y+n+1)| - |L(x+n,y-n) - R(x+d+n,y-n)| - T(\tilde{x},d) \quad (19)$$

$$T(\tilde{x},d) = |L(x-n-1,y+n+1) - R(x+d-n-1,y+n+1)| - |L(x-n-1,y-n) - R(x+d-n-1,y-n)|$$

with $\tilde{x} = x \bmod(2n+1), \quad d \in [o, d_{\max}]$ (20)

The extension of the incremental computation scheme described above to the case of other matching functions, such as SSD and NCC, is straightforward.

Finally, the code of the optimised algorithm have been parallelised deploying the SIMD-style instructions [18] provided by the target CPU (i.e. Pentium III processor with MMX [17] and SSE [21] technology).

6. Experimental results

We first discuss and compare the experimental results obtained with SMP and BM on a standard set of stereo images with available ground truth (Section 6.1). Then, in Section 6.2, we discuss the experimental results obtained with SMP on a sequences taken in our laboratory and concerned with a *3D People Tracking* application currently under investigation.

6.1. Experimental results on standard stereo pairs

We discuss here the experimental results obtained on the grayscale stereo pairs available at the Scharstein and Szeliski's web site [22] using the SMP-based algorithm described in Section 4 and a well known BM-based stereo algorithm. We first discuss the results qualitatively, so as to address a number of issues concerning the disparity maps yield by local matching algorithms. Then, we introduce a performance evaluation metric in order to compare the two approaches also in quantitative terms. Finally, we report several measurements aimed at assessing the speed of the two approaches with different image sizes and disparity ranges.

Scharstein and Szeliski's data set [22] consists of the *Tsukuba*, *Map*, *Sawtooth*, *Venus*, *Barn 1*, *Barn 2*, *Bull* and *Poster* stereo pairs and ground-truth disparity maps (Figs. 10–17). For the first four pairs, i.e. *Tsukuba*, *Map*, *Sawtooth* and *Venus*, the occlusion map is also available.

As a representative of BM-based algorithms we consider the well-known SVS algorithm (ver. 2.0) from SRI International [4]. It is worth pointing out that, besides the matching core, further differences between our algorithm and SVS can be found in the pre-processing step as well as in the technique adopted to detect low textured areas. As for the pre-processing step, our algorithm normalises the images by subtraction of the mean value while SVS relies on LoG *filtering* [4]. We detect low textured areas by analysing the variance map while SVS makes use of an *interest operator*. However, since we are mostly interested in evaluating and comparing the SMP and BM matching cores, we carried out the experiments by turning off the detection of low-textured areas for both algorithms. Moreover, for both algorithms we used the same parameters and correlation window (9×9) for all the stereo pairs within the data set.

The *Tsukuba* stereo pair contains complex objects at different depths generating several occlusions, as well as poorly-textured regions in the background, such as for example the wall at the top-right corner. Moreover, this stereo pair contains some specular regions (i.e. the face of the statue and some regions of the lamp) that render quite difficult the stereo matching process. Comparing the output given by SMP (left image of Fig. 18) with the ground truth (right image of Fig. 10) we can observe first that the rough 3D structure has been clearly recovered: the camera an its trestle on the background have been recovered as well as the objects closer to the stereo acquisition system, such as the statue and the lamp's head. Moreover, it is worth observing that several major

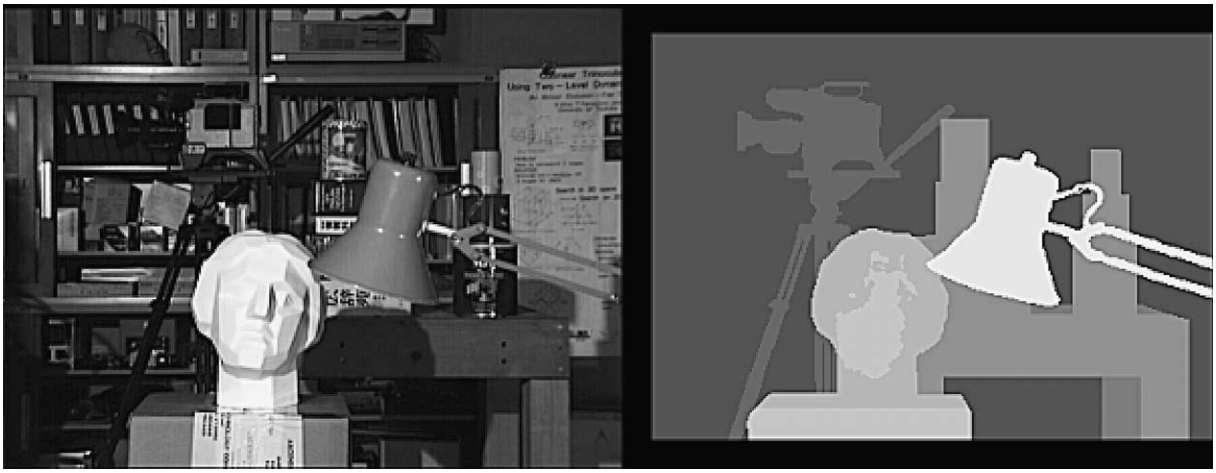

Fig. 10. Left image of the *Tsukuba* stereo pair (left) and ground truth (right).

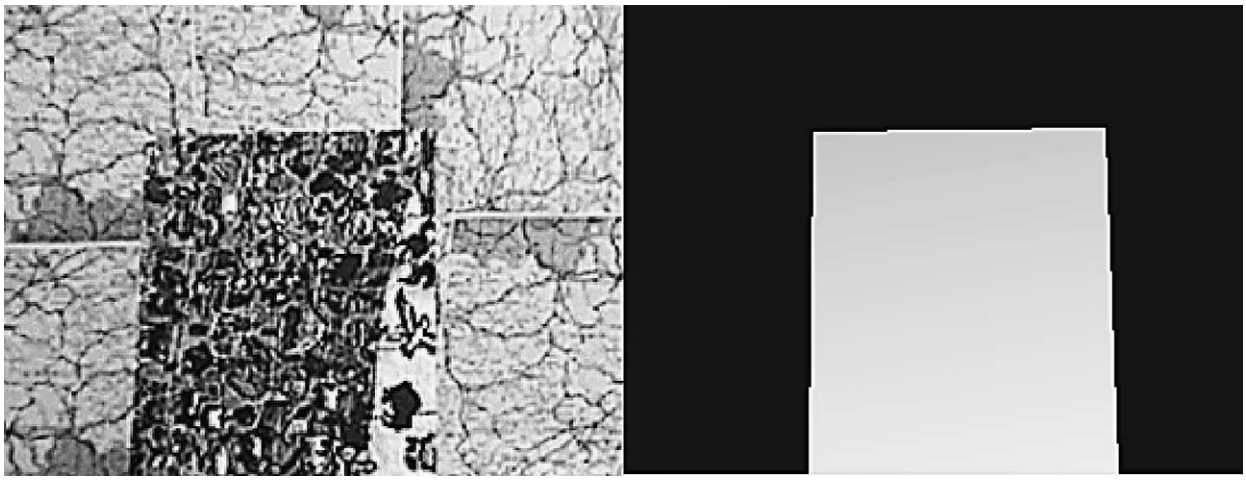

Fig. 11. Left image of the *Map* stereo pair (left) and ground truth (right).

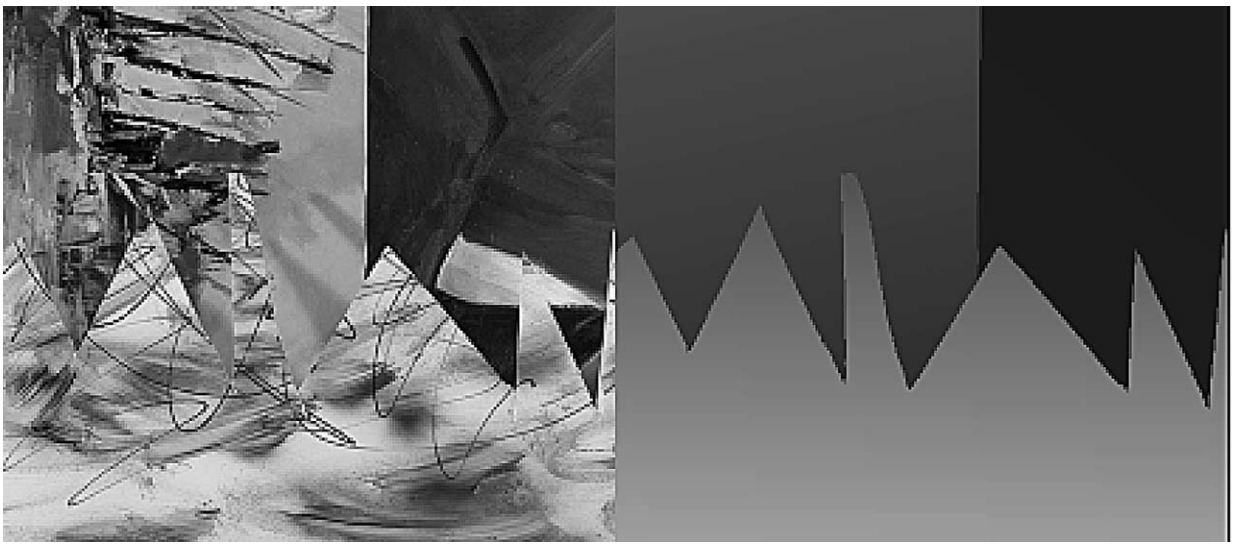

Fig. 12. Left image of the *Sawtooth* stereo pair (left) and ground truth (right).

Fig. 13. Left image of the *Venus* stereo pair (left) and ground truth (right).

occlusions have been discarded (the points left unmatched are represented in white), showing the ability SMP to deal with this problem. However, details such as the lamp's wire, the lamp's switch and the two roads that sustain the lamp, have vanished. Moreover the disparity map is affected by the *border-localisation* problem, as it can be seen comparing the ground truth with the disparity map computed by the algorithm. This causes clearly an unaccurate fitting of the object's silhouette into the disparity map.

This problem is inherent to local algorithms since, as discussed in Section 3, it depends on the method adopted to establish correspondences, which relies on the use of a local support area centered at the point under examination (i.e. the correlation window). Local algorithms behave correctly when the correlation window covers a region at constant depth but are likely to produce artifacts when the correlation window covers regions at different depths. The *vanishing* of details in the disparity map arises when the details are small compared to the size of the correlation window. In this case the signal strength embodied in the texture of the detail can be overcame by the contributions of the other points within the correlation window, resulting in a low-pass filtering effect. Both problems can be mitigated reducing the size of the correlation window; however, this solution has the side-effect of reducing the signal-to-noise ratio leading to more matching errors. Some authors [7,14,16] propose local algorithms aimed at reducing the *border-localisation* problem. However, the results provided by these algorithms are still less accurate that those generated with global, slow algorithms such as the one presented by Kolmogorov and Zabih [1].

The results obtained by the BM-based algorithm on the *Tsukuba* stereo pair (right of Fig. 18) are very similar:

Fig. 14. Left image of the *Barn 1* stereo pair (left) and ground truth (right).

Fig. 15. Left image of the *Barn 2* stereo pair (left) and ground truth (right).

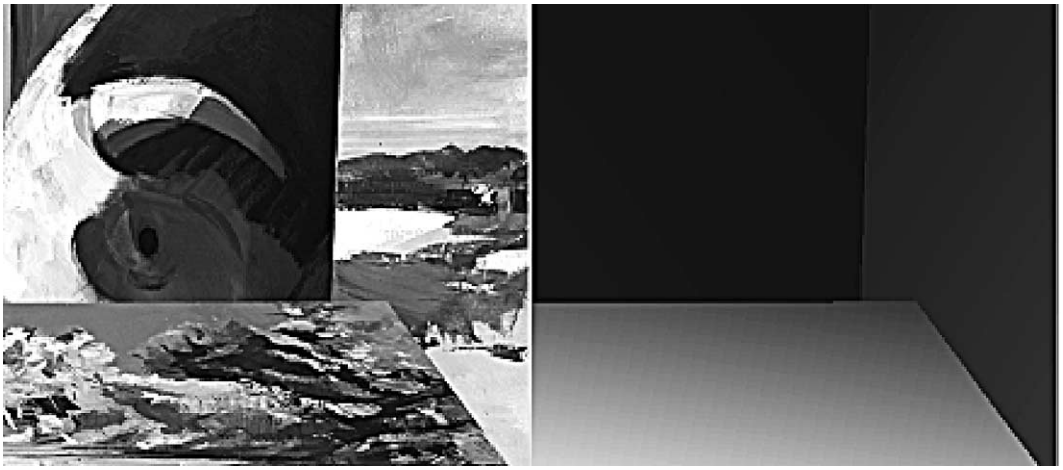

Fig. 16. Left image of the *Bull* stereo pair (left) and ground truth (right).

Fig. 17. Left image of the *Poster* stereo pair (left) and ground truth (right).

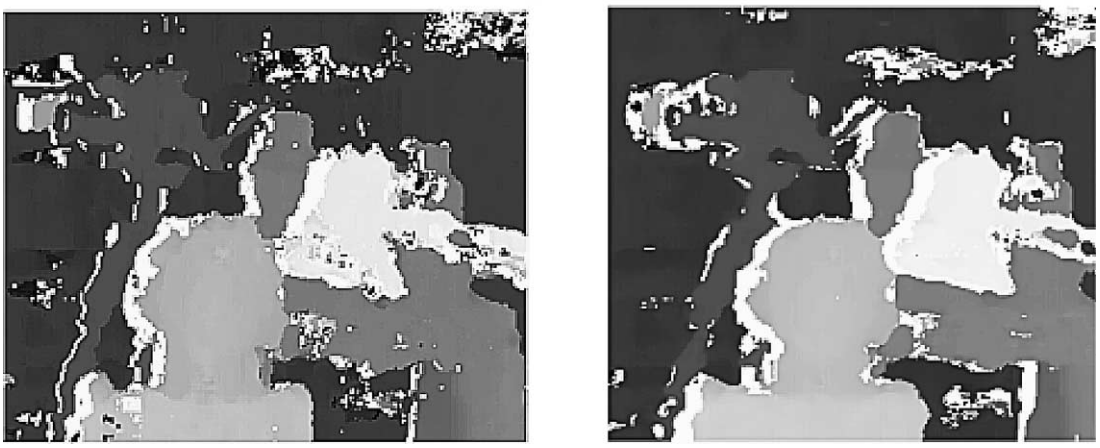

Fig. 18. Disparity map obtained on the *Tsukuba* stereo pair with SMP (left) and BM (right).

the rough 3D structure is recovered, several major occlusions have been discarded, the border are localised with poor accuracy and many small details have been lost.

The other stereo pairs *Map*, *Sawtooth*, *Venus*, *Barn 1*, *Barn 2*, *Bull* and *Poster* present a similar geometrical structure made out of simpler, piecewise planar objects (typically posters or paintings) located at different depths and generating large occlusions. However, compared to *Tsukuba*, these stereo pairs contain very few small details. Moreover, with the exception of *Map*, the other stereo pairs contain several low textured regions.

The disparity maps obtained with SMP and BM are shown in Figs. 19–25. These results show that the algorithms recover very similarly, and quite correctly, the 3D structure of the scene. At depth discontinuities the maps are still affected by the border localisation problem and many of the occluded points are detected and discarded. Moreover, due to the nature of the images, the problem of the vanishing of small details has much less impact on the quality of the results compared to *Tsukuba*.

In order to provide a quantitative assessment of SMP and BM we need to use a performance evaluation metric. Since both approaches mark as void the disparities yield by matches judged unreliable, we cannot use exactly the same metric defined in Ref. [23] since it requires disparities to be defined at each pixel. Though the disparities at unmatched points could have been obtained by interpolation, this would have distorted the results provided by the two matching approaches that we are interested in evaluating. Hence, we define a metric aimed at assessing the performance on the basis of the incomplete disparity maps generated by SMP and BM.

To describe the adopted metric we use a notation where the subscript, called A, identifies the matching approach, i.e. $A \in \{\mathrm{SMP}, \mathrm{BM}\}$. We call, respectively, M_A and U_A the set of points matched and left unmatched by each algorithm within a common region-of-interest (ROI). N_A represents the number of points belonging to set M_A. We define the RMS_A (root mean squared) error between the computed disparity map (d_A) and the ground truth (d_T) as follows:

$$\mathrm{RMS}_A = \sqrt{\frac{1}{N_A} \sum_{(x,y)\in M_A} |d_A(x,y) - d_T(x,y)|^2} \tag{21}$$

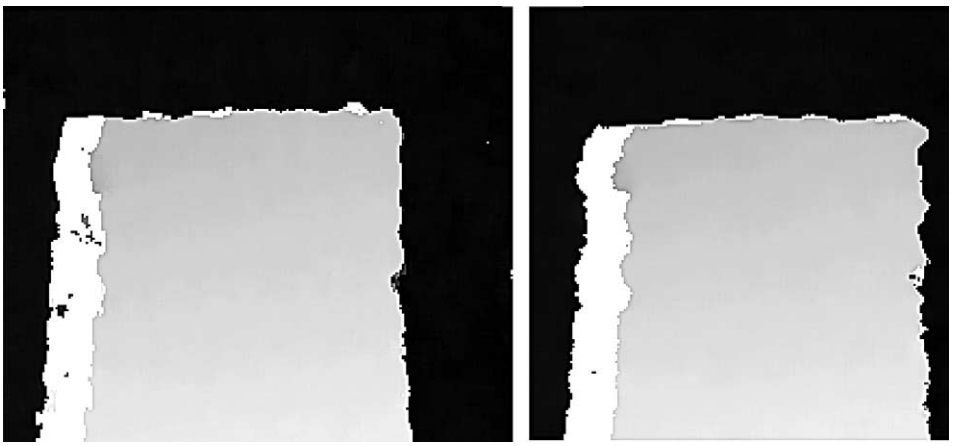

Fig. 19. Disparity map obtained on the *Map* stereo pair with SMP (left) and BM (right).

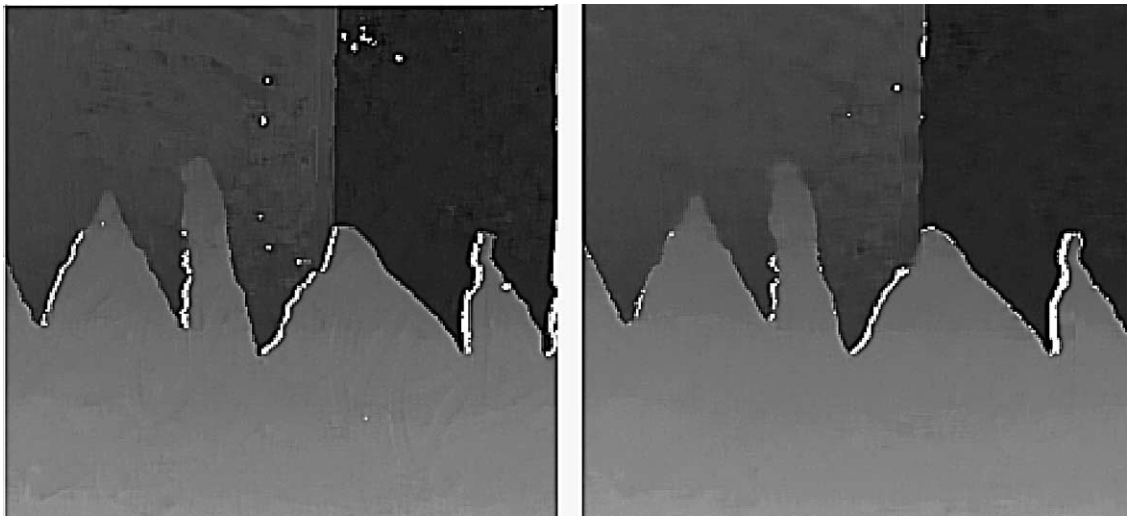

Fig. 20. Disparity map obtained on the *Sawtooth* stereo pair with SMP (left) and BM (right).

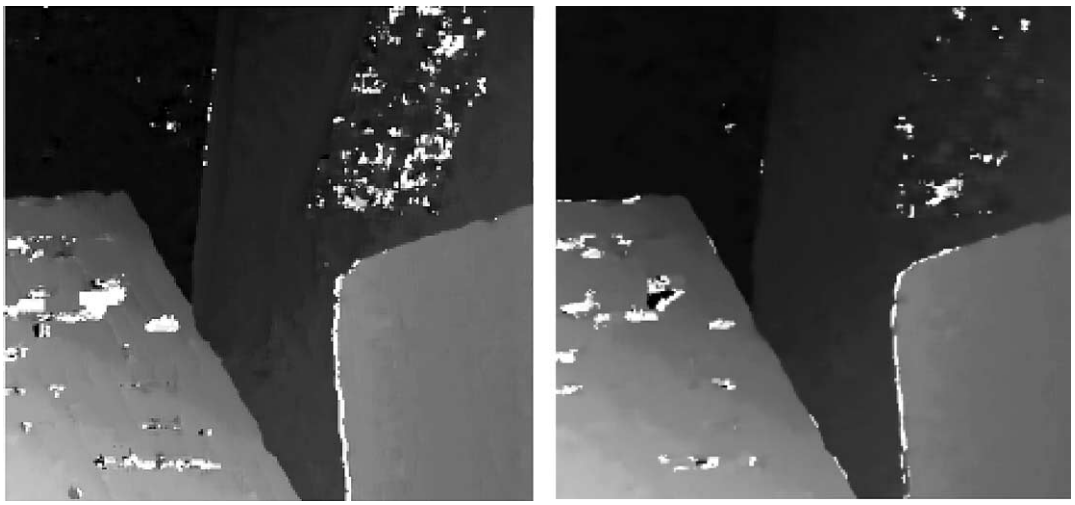

Fig. 21. Disparity map obtained on the *Venus* stereo pair with SMP (left) and BM (right).

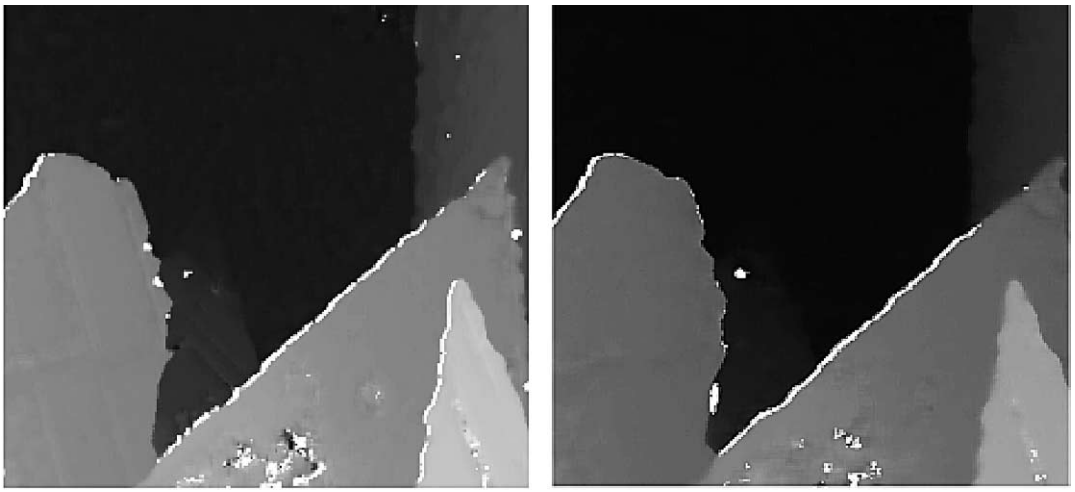

Fig. 22. Disparity map obtained on the *Barn 1* stereo pair with SMP (left) and BM (right).

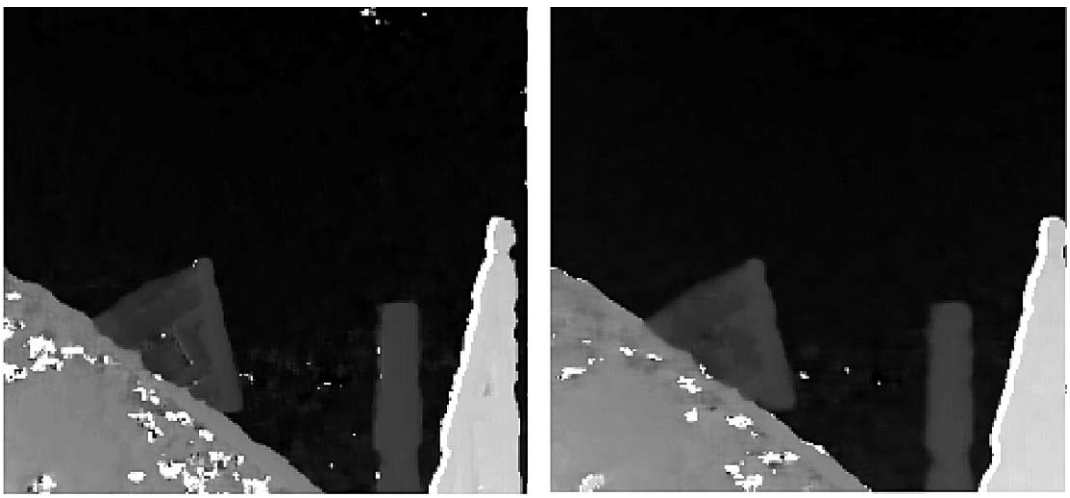

Fig. 23. Disparity map obtained on the *Barn 2* stereo pair with SMP (left) and BM (right).

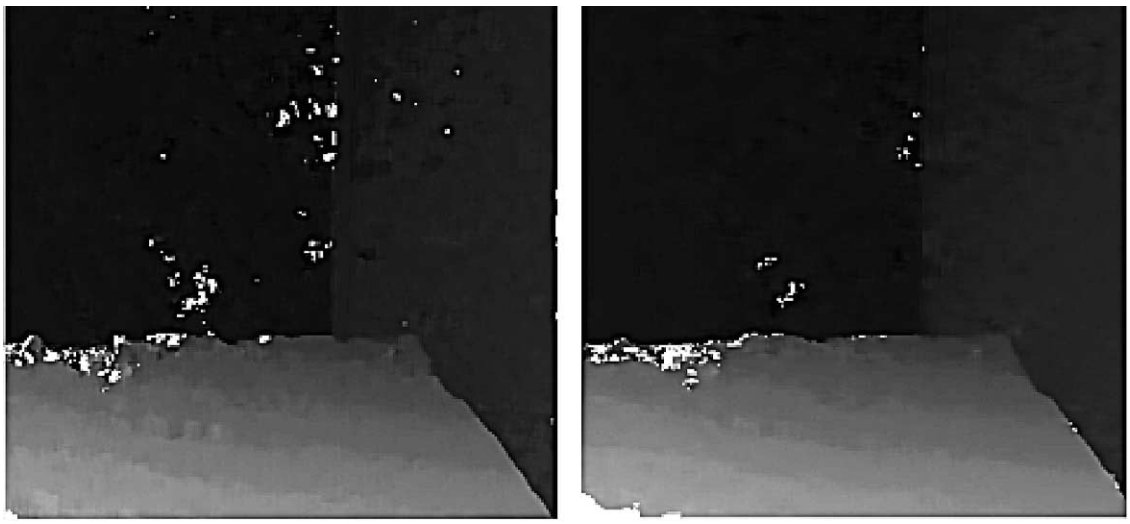

Fig. 24. Disparity map obtained on the *Bull* stereo pair with SMP (left) and BM (right).

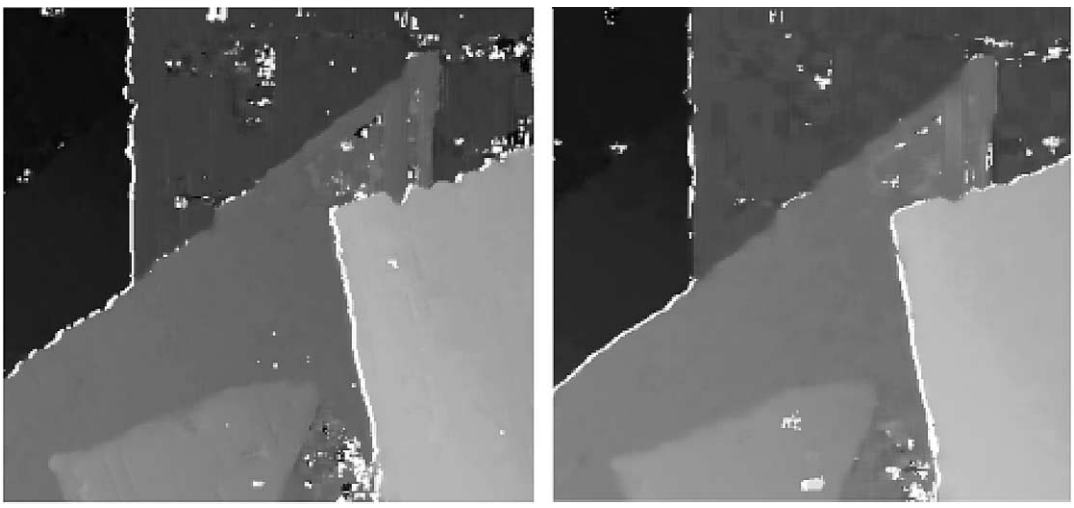

Fig. 25. Disparity map obtained on the *Poster* stereo pair with SMP (Left) and BM (Right).

Table 1
Percentage of matched points (N_A), percentage of incorrect measurements (B_A), percentage of unmatched points (U_A) and root mean squared error (RMS_A) for SMP and BM

Image	N_A (%)	B_A (%)	U_A (%)	RMS_A
Tsukuba (SMP)	90.68	33.77	9.32	5.77
Tsukuba (BM)	89.00	28.23	11.00	5.73
Map (SMP)	92.20	2.78	7.80	3.31
Map (BM)	91.49	2.04	8.51	2.89
Sawtooth (SMP)	99.29	3.67	0.71	0.76
Sawtooth (BM)	99.38	4.43	0.62	0.77
Venus (SMP)	97.98	4.28	2.02	0.97
Venus (BM)	98.87	3.10	1.13	0.65
Barn 1 (SMP)	98.86	2.88	1.14	0.60
Barn 1 (BM)	99.08	2.72	0.92	0.55
Barn 2 (SMP)	98.64	3.79	1.36	0.71
Barn 2 (BM)	98.91	3.31	1.09	0.62
Bull (SMP)	99.42	1.47	0.58	0.59
Bull (BM)	99.61	1.28	0.39	0.44
Poster (SMP)	98.16	3.52	1.84	0.87
Poster (BM)	98.40	2.62	1.60	0.73

We also evaluate the number of incorrect disparity measurements provided by each algorithm. Similarly to Ref. [23], we consider as incorrect measurements those points (x, y) that have been matched by the algorithm and yield $|d_A(x, y) - d_T(x, y)| > 1$. We call B_A the ratio of incorrect measurements to the number of matched points:

$$B_A = \left(\frac{1}{N_A} \sum_{(x,y)\in M_A} \delta_A \right) \tag{22}$$

where $\delta_A = 1$ if $|d_A(x, y) - d_T(x, y)| > 1.0$ and $\delta_A = 0$ otherwise.

As already mentioned, since the two algorithms do not provide necessarily complete disparity measurements within the common ROI, in Eqs. (21) and (22) we consider only those points for which each of the two algorithms provide a disparity measurement (i.e. for SMP we consider the set M_{SMP} and for BM the set M_{BM}).

The capability to deal with occlusions is evaluated using the occlusion maps available Scharstein and Szeliski's web site [22]. For both algorithms we have computed the number of points correctly marked as occluded (*detected occlusions*), O_{R_A}, and the number of points belonging to occluded regions that, incorrectly, are returned as matched (*undetected occlusions*), O_{W_A}.

Table 1 reports the results collected running the two algorithms on the eight stereo pairs. We can notice that the number of points matched by the two algorithms is substantially the same for all the stereo pairs. As for the incorrect measurements, with the exception of *Sawtooth* on which SMP performs better, BM always provides a lower number incorrect disparity measurements. However, the number of incorrect measurements yield by SMP and BM turns out to be very similar throughout the dataset, with the only exception of *Tsukuba* on which BM performs better by approximately 5%. The RMS error, that provides an overall measure of the performance, turns out to be very similar within the whole data set, though BM yields always slightly smaller errors with the exception of the *Sawtooth* stereo pair. It is worth observing that due to the presence of more complex objects and small details, with *Tsukuba* both algorithms yield less satisfactory results in terms of incorrect measurements as well as RMS error.

Figs. 26–33, show visually the points where incorrect measurements are produced by the SMP and BM. To localise the errors we use in these figures the ground truth as background and highlight in white the points where occur incorrect measurements.

Table 2 shows the statistics aimed at assessing the capability of the two algorithms to deal with occluded points. It is worth noticing that with the *Map* stereo pair the two algorithms perform similarly, and pretty well, with a slightly better detection percentage yield by BM. Conversely, with the other three stereo pairs both approaches turns out to be less effective in detecting correctly occluded points, although SMP performs better than BM on the *Venus* and *Sawtooth* pairs. Figs. 34–37 have been obtained by

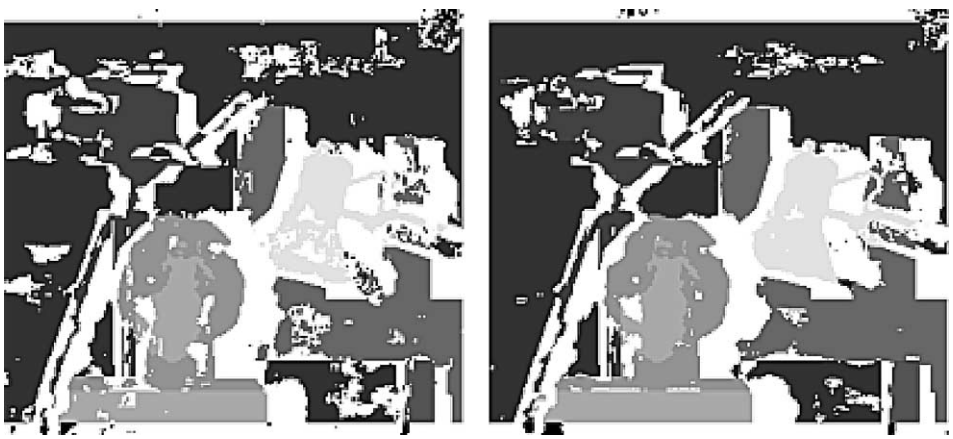

Fig. 26. Incorrect measurements yield by SMP (left) and BM (right) on the *Tsukuba* stereo pair.

Fig. 27. Incorrect measurements yield by SMP (left) and BM (right) on the *Map* stereo pair.

comparing the disparity maps yield by the two algorithms with the occlusion maps available at Scharstein and Szeliski's web site [22]: non-occluded points are shown in black, detected occlusions in white and undetected occlusion in gray.

Finally, we report in Table 3 some measurements aimed at assessing the speed of the two algorithms with different image sizes and disparity ranges. These measurements have been obtained on an Intel Pentium III processor running at 800 MHz. Both algorithms are carefully optimised and deploy the SIMD parallel processing capabilities available on this CPU. From Table 3 we can see that for a small disparity range (i.e. 16) BM is always faster, much faster for small images (i.e. (320 × 240)) and slightly faster for bigger images. Yet, as the disparity range is increased, our algorithm gets faster than BM, significantly faster for big images and large disparity range. For example, with 800×600 stereo pairs and a disparity range of 16 our algorithm runs at 5.56 fps while SVS at 6.96. With this image size and a disparity range of 80 our algorithm is nearly twice faster than BM (i.e. 2.89 fps for SMP and 1.51 for BM).

6.2. Experimental results on stereo sequences

In this subsection we discuss qualitatively the experimental results obtained with SMP on a stereo sequence taken in our laboratory and referred to as *Outdoor*. We are currently using this sequence within a research activity aimed at developing a 3D People Tracking application. The tracking approach is based on first merging the disparity maps extracted by SMP with the information provided by a grayscale change-detection algorithm and then building

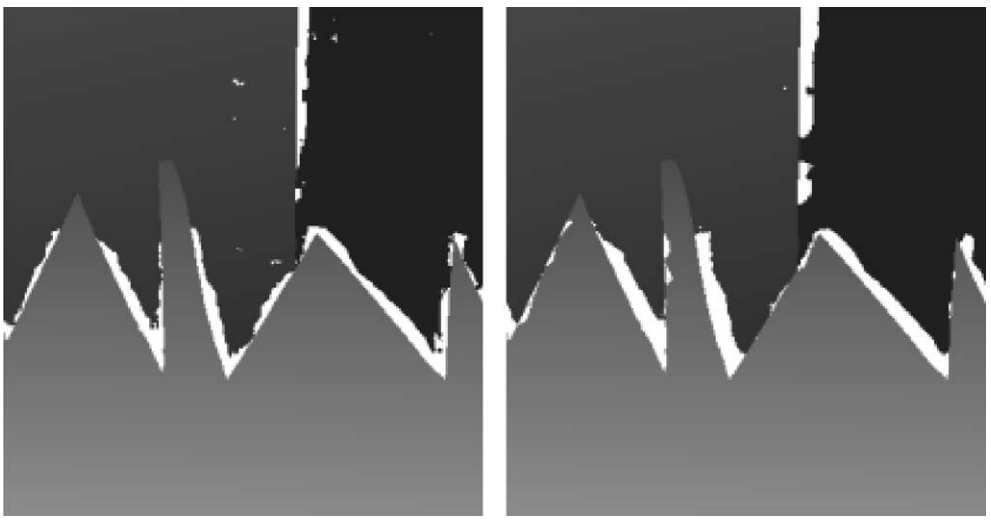

Fig. 28. Incorrect measurements yield by SMP (left) and BM (right) on the *Sawtooth* stereo pair.

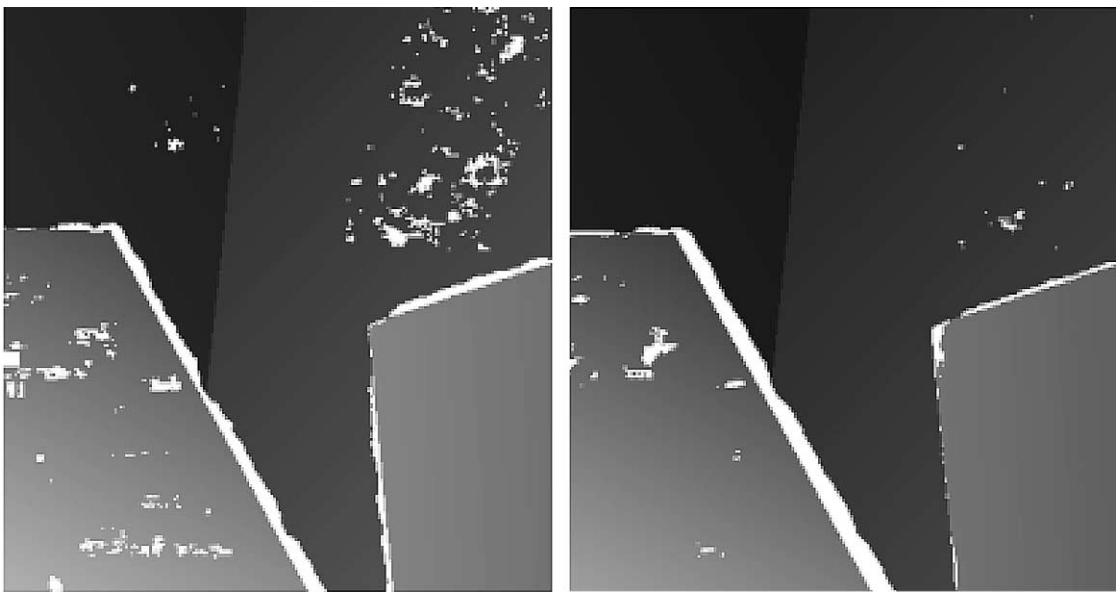

Fig. 29. Incorrect measurements yield by SMP (left) and BM (right) on the *Venus* stereo pair.

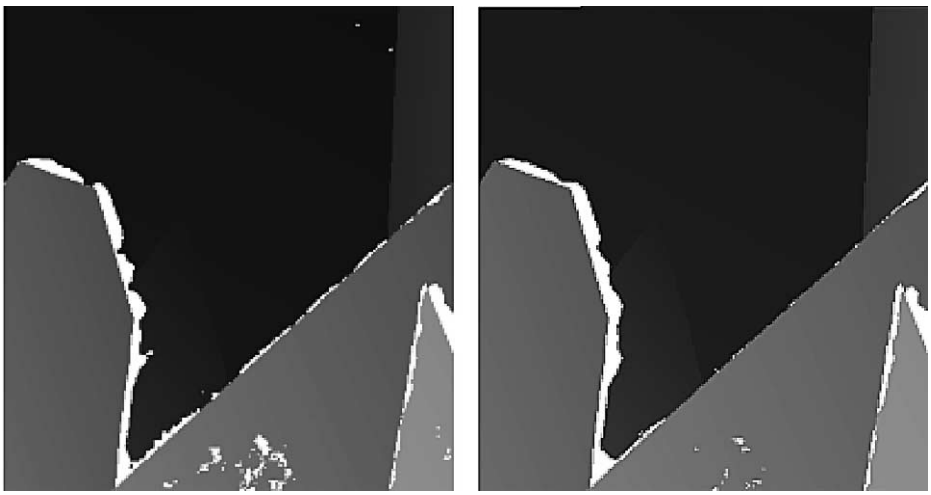

Fig. 30. Incorrect measurements yield by SMP (left) and BM (right) on the *Barn1* stereo pair.

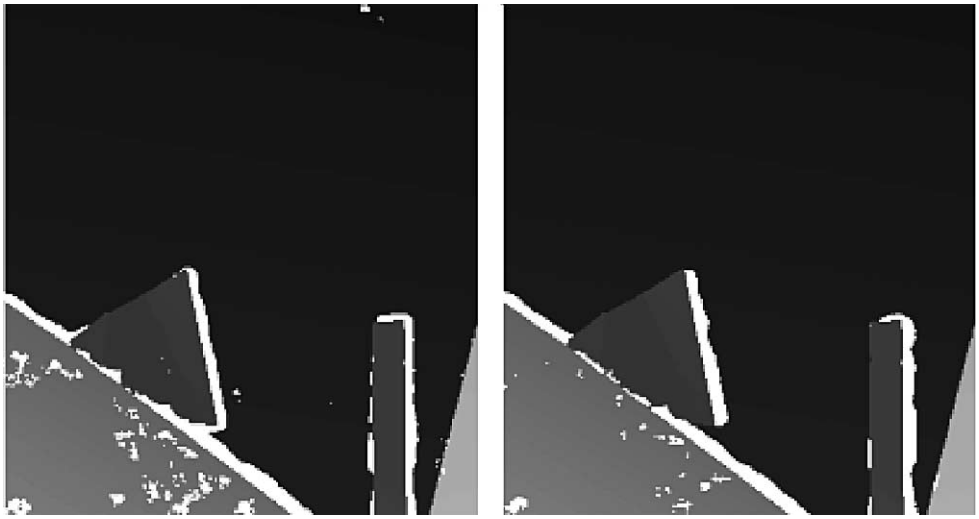

Fig. 31. Incorrect measurements yield by SMP (left) and BM (right) on the *Barn2* stereo pair.

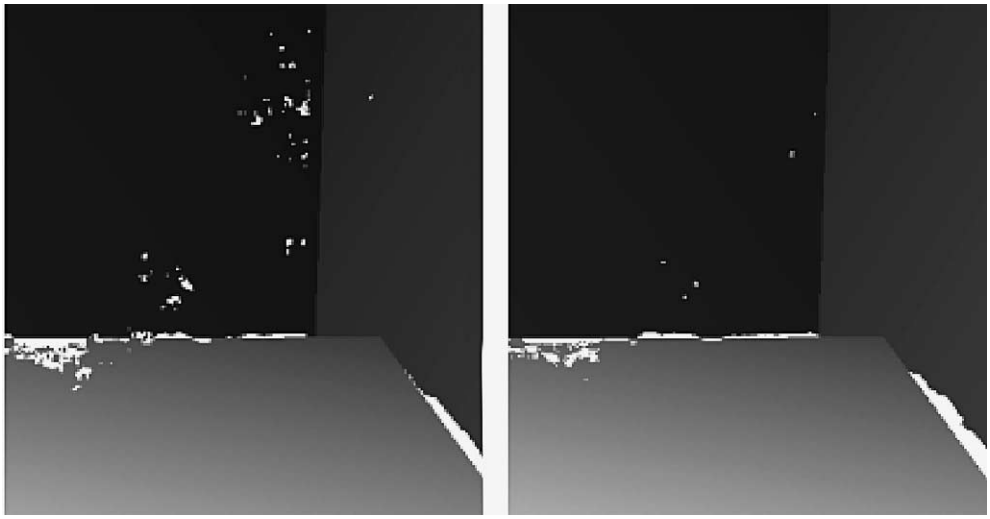

Fig. 32. Incorrect measurements yield by SMP (left) and BM (right) on the *Bull* stereo pair.

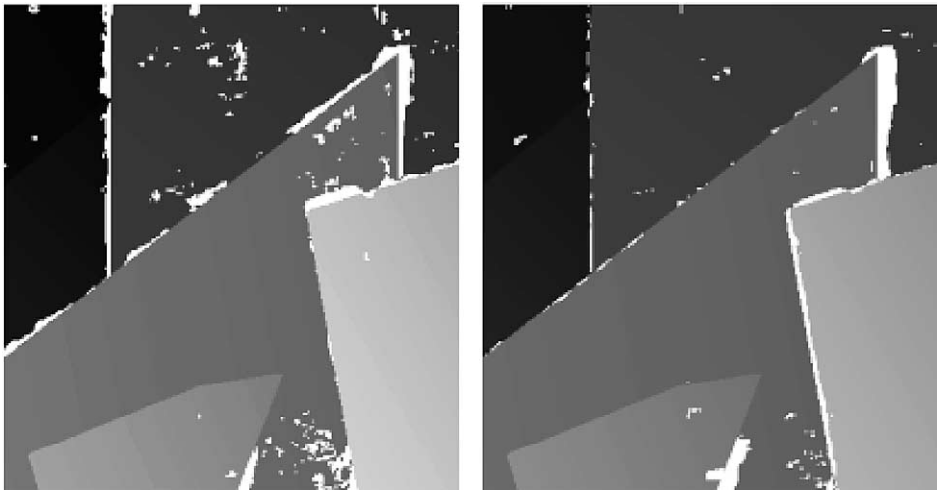

Fig. 33. Incorrect measurements yield by SMP (left) and BM (right) on the *Poster* stereo pair.

a suitable plane-view representation that enables us to track moving objects in the 3D space.

The sequence was acquired with a monocrome MEGA-D digital stereo head [10] equipped with a pair of 4.8 mm lenses. The original stereo pairs were rectified using the method described in Ref. [13] and based on the intrinsic and extrinsic camera parameters estimated with the functions provided by the MATLAB Camera Calibration Toolbox available at [24]. Image size is 640×480 and the rectified sequence was processed using a 15×15 correlation window, a disparity search range of 64 pixels and a subpixel accuracy of 1/8. All the material concerned with this experiment (the original sequence, the rectified sequence, the calibration images, the estimated camera parameters, the disparity-map sequence yield by SMP) is available at the web site [25]. This site contains also the results relative to an *Indoor* sequence taken in our laboratory, not discussed here for the sake of brevity.

In the *Outdoor* sequence the stereo head was mounted at an height of about 2 m from the ground and tilted down so as to look at the entrance of a building, as shown in the uppermost images of Fig. 38 (corresponding to Frame 0030

Table 2
Total number of occluded points (O), percentage of detected occlusions (O_{R_A}) and percentage of undetected occlusions (O_{W_A}) for SMP and BM

Image	O	O_{R_A} (%)	O_{W_A} (%)
Tsukuba (SMP)	2153	9.15	90.85
Tsukuba (BM)	2153	11.75	88.25
Map (SMP)	3086	78.48	21.52
Map (BM)	3086	84.09	15.91
Sawtooth (SMP)	2683	21.77	78.23
Sawtooth (BM)	2683	13.01	86.99
Venus (SMP)	1690	16.51	83.49
Venus (BM)	1690	3.78	96.21

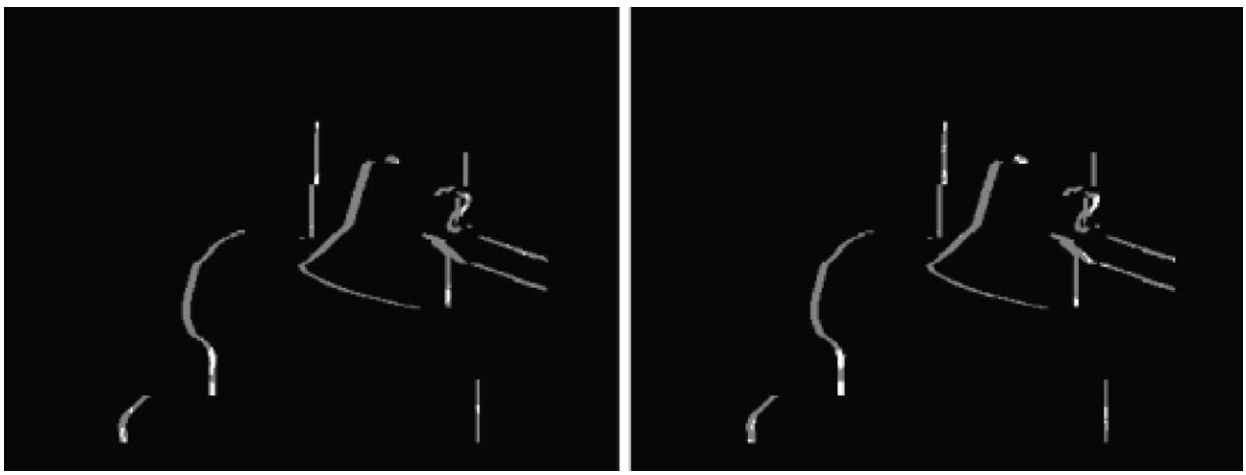

Fig. 34. Occlusions by SMP (left) and BM (right) on the *Tsukuba* stereo pair.

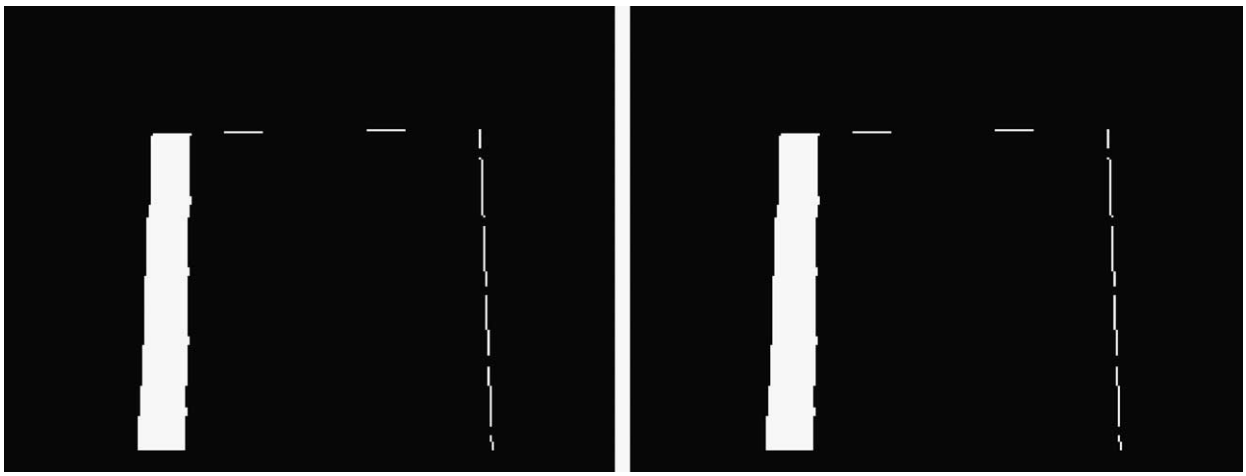

Fig. 35. Occlusions by SMP (left) and BM (right) on the *Map* stereo pair.

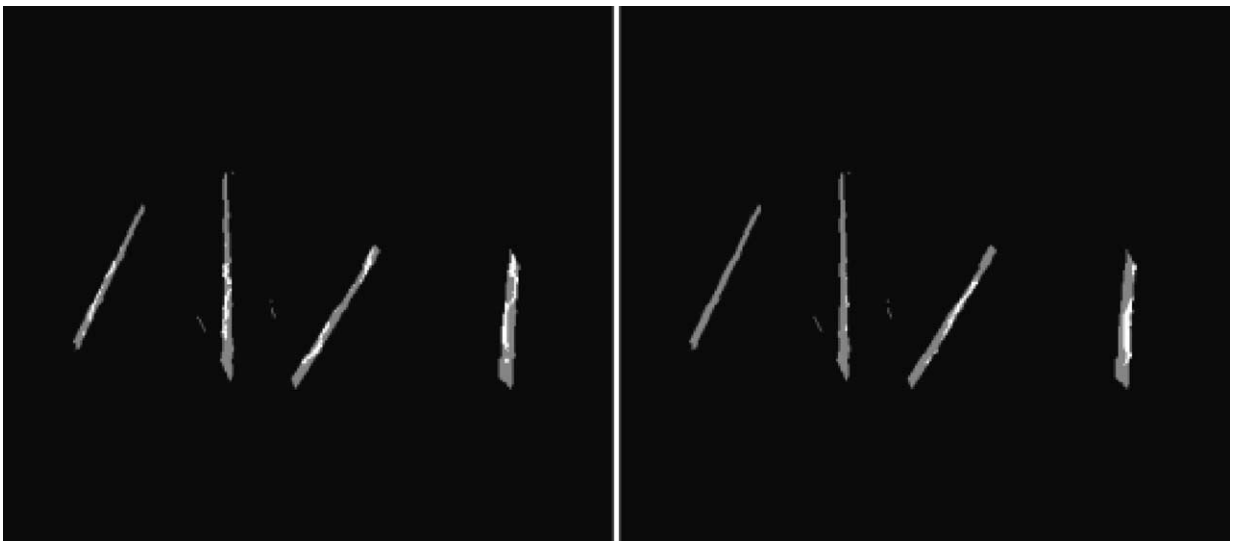

Fig. 36. Occlusions by SMP (left) and BM (right) on the *Sawtooth* stereo pair.

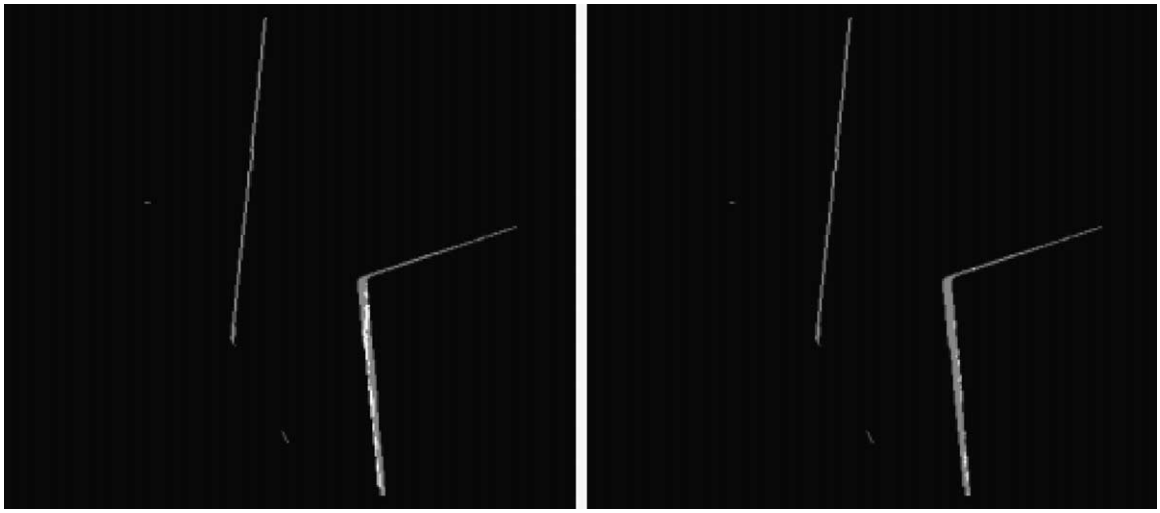

Fig. 37. Occlusions by SMP (left) and BM (right) on the *Venus* stereo pair.

of the stereo sequence). The initial frames of the sequence show the static scene, consisting of a background wall with a long low-textured banister, a staircase with a dozen of steps and a uniform panel sustained by a box; then, two persons enter the scene. The uppermost images of Fig. 39 show frame 0050, with one person in the field of view of the stereo cameras. The lowermost images of Figs. 38 and 39 show the rectified stereo pairs corresponding to the uppermost images. For the two frames of Figs. 38 and 39 we show in Figs. 40 and 41 the disparity maps obtained by the SMP algorithm with four different settings of the threshold that controls the detection of low textured areas.

The top left disparity map of Fig. 40 show that the 3D structure of the scene is correctly recovered in most points. The wall, the banister, the steps at different height and distance, the uniform panel and its support can be clearly noticed. Though the disparity map was obtained disabling the detection of low textured areas (i.e. setting the threshold to 0), the other constraints embodied into SMP allow for classifying as unreliable many disparity measurements occurring in uniform regions, such as the center of the uniform panel, the wall and the banister. Yet, SMP matches incorrectly a few points at the top right border of the uniform panel, as well as some points within the panel. The other three disparity maps of Fig. 40 show that increasing the threshold that controls the detection of low textured areas allows for discarding most of the incorrect matches, but this implies also loosing an increasing amount of correct disparity measurements (e.g. those occurring at the banister).

By observing the disparity maps associated with frame 0050 (Fig. 41) we can notice that most occlusions (i.e. the right sides of the head and of torso of the person) are correctly detected by SMP. Again, even though the detection of low textured areas is disabled, SMP can label as unreliable several points belonging to the uniform regions corresponding to the person's shoulders and pants. Similarly to the previous frame, increasing the threshold allows for discarding most incorrect measurements but significantly reduces the amount of points correctly matched by the algorithm (e.g. the points on the person's shoulders and pants).

The results obtained so far in several outdoor as well as indoor sequences have proven to be adequate for the requirements of 3D People Tracking application currently under development in our laboratory. Though SMP does not deal with the border localisation problem and change detection is sensitive to shadows, the merging of the disparity and grayscale change maps alleviates significantly both problems since the silhouettes of the moving persons are reproduced quite accurately by the change detection algorithm and most shadows can be filtered out based on disparity information. As for the low-textured regions, in a people tracking application most of these regions generally belong to the static background, which is removed by the change-detection step. Hence, with SMP we can generally set a very low threshold for the texture operator so as to recover correctly most of the 3D information associated with moving persons.

Table 3
Speed measurements (in terms of frame per second, fps) for the SMP-based and the BM-based algorithms

Algorithm (size)	$d = 16$ (fps)	$d = 32$ (fps)	$d = 48$ (fps)	$d = 64$ (fps)	$d = 80$ (fps)
SMP (320 × 240)	39.59	31.25	27.44	25.94	25.96
BM (320 × 240)	57.99	33.68	20.49	15.31	12.71
SMP (640 × 480)	8.94	6.92	5.77	5.17	4.78
BM (640 × 480)	11.99	5.93	4.07	3.18	2.54
SMP (800 × 600)	5.56	4.28	3.60	3.18	2.89
BM (800 × 600)	6.96	3.65	2.53	1.94	1.51
SMP (1024 × 768)	3.32	2.56	2.09	1.86	1.67
BM (1024 × 768)	3.79	2.07	1.45	1.06	0.78

Fig. 38. Frame 0030 of the *Outdoor* stereo sequence: (top left) original left image, (top right) original right image, (bottom left) rectified left image, (bottom right) rectified right image.

Fig. 39. Frame 0050 of the *Outdoor* stereo sequence: (top left) original left image (top right) original right image (bottom left) rectified left image (bottom right) rectified right image.

Fig. 40. Results on frame 0030 of the *Outdoor* sequence: (top left) disparity map with threshold set to 0, (top right) Disparity map with threshold set to 1, (bottom left) disparity map with threshold set to 2, (bottom right) disparity map with threshold set to 3. The four disparity maps have been equalized for better visualization.

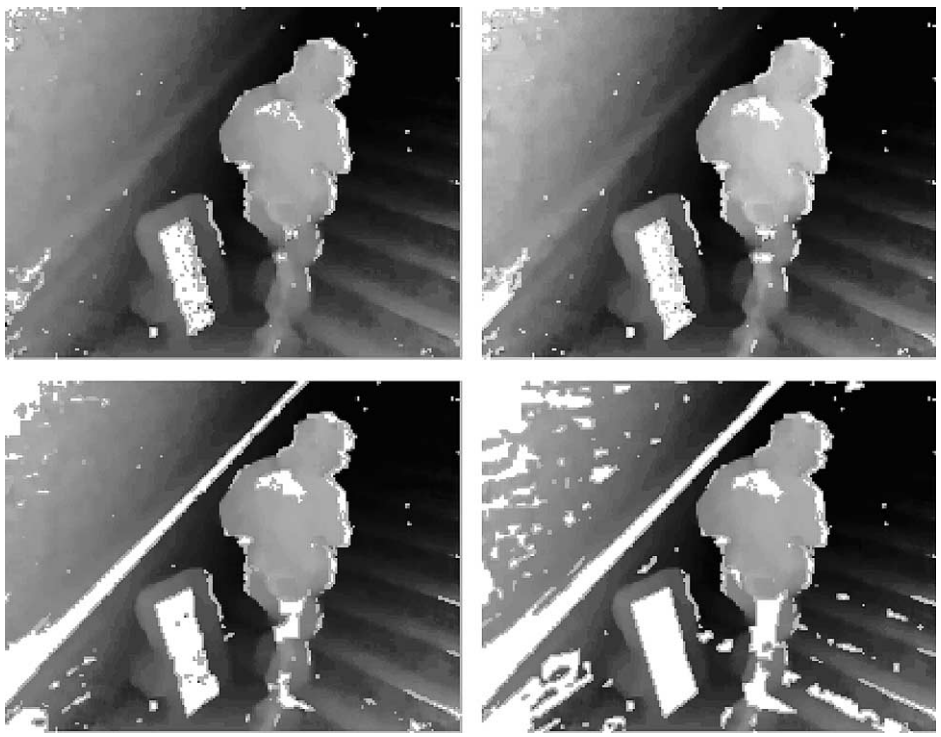

Fig. 41. Results on frame 0050 of the *Outdoor* sequence: (top left) disparity map with threshold set to 0, (top right) disparity map with threshold set to 1, (bottom left) disparity map with threshold set to 2, (bottom right) disparity map with threshold set to 3. The four disparity maps have been equalized for better visualization.

7. Conclusion

We have presented an area-based stereo matching algorithm aimed at real time applications that relies only on a SMP and detects unreliable matches via 'colliding matches', i.e. matches that violate the uniqueness constraint. We have analysed the differences between the proposed SMP approach and BM, i.e. left-to-right matching followed by right-to-left matching, being the latter the method adopted to detect unreliable matches in most area-based stereo algorithms conceived for real time applications. We have shown how to improve the reliability of the matches yield by SMP at a very small computational cost by enforcing simple constraints on the behaviour of the error scores. We have also described an efficient computational scheme that deploys three levels of incremental calculations to avoid unnecessary calculations. We have reported and discussed, both in qualitative and quantitative terms, the experimental results obtained by the proposed algorithm and a well-known BM-based algorithm on a standard stereo data set with ground-truth disparity and occlusion maps. Our results show that in most cases SMP and BM behave quite similarly and that SMP holds the potential for speeding-up significantly dense stereo matching. Eventually, the experimental results obtained on an outdoor rectified stereo sequence indicate that SMP is potentially effective in practical applications such as 3D People Tracking.

Hence, SMP can be considered a viable alternative to BM, in particular with big images and large disparity ranges, as it is the case of many current stereo applications. However, our analysis and results confirm the need for further research aimed at dealing with the typical problems of area-based stereo algorithms. In particular, we plan to embody specific techniques to cope with the border-localisation problem into an SMP based framework.

References

[1] V. Kolmogorov, R. Zabih, Computing Visual Correspondence with Occlusions Using Graph Cuts, Proceedings of International Conference on Computer Vision, 2001.

[2] O. Faugeras, B. Hotz, H. Mathieu, T. Viville, Z. Zhang, P. Fua, E. Thron, L. Moll, G. Berry, Real-time correlation-based stereo: Algorithm. Implementation and Applications, INRIA Technical Report n. 2013, 1993.

[3] T. Kanade, H. Kato, S. Kimura, A. Yoshida, K. Oda, Development of a video-rate stereo machine, Proc. Int. Robotics Syst. Conf. 3 (1995) 95–100.

[4] K. Konolige, Small Vision Systems: Hardware and Implementation, 8th International Symposium on Robostics Research, Hayama, Japan, 1997, pp. 111–116.

[5] G. Van der Val, M. Hansen, M. Piacentino, The ACADIA Vision processor, Proceedings of 5th International Workshop on Computer Architecture for Machine Perception, Padova, Italy, 2001, pp. 31–40.

[6] L. Di Stefano, S. Mattoccia, Fast Stereo Matching for the VIDET System using a General Purpose Processor with Multimedia Extensions, Proceedings of 5th International Workshop on Computer Architecture for Machine Perception, Padova, Italy, 2000, pp. 356–362.

[7] H. Hirschmuller, P. Innocent, J. Garibaldi, Real-time correlation-based stereo vision with reduced border errors, Int. J. Comput. Vision 47 (1–3) (2002) 229–246.

[8] K. Muhlmann, D. Maier, J. Hesser, R. Manner, Calculating dense disparity maps from color stereo images, an efficient implementation, Int. J. Comput. Vision 47 (1–3) (2002) 79–88.

[9] Point Grey Research Home Page, www.ptgrey.com

[10] Videre Design Home Page, www.videredesign.com

[11] P. Fua, Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities, Proceedings of 12th International Joint Conference on Artificial Intelligence, 1991, pp. 1292–1298.

[12] G. Egnal, R. Wildes, Detecting binocular half-occlusions: empirical comparisons of four approaches, Proc. Int. Conf. Comput. Vision Pattern Recognit. (2) (2000) 466–473.

[13] E. Trucco, A. Verri, Introductory techniques for 3D computer vision, Prentice Hall, Englewood cliffs, NJ, 1998.

[14] A. Fusiello, V. Roberto, E. Trucco, Symmetric Stereo with Multiple Windowing, Int. J. Pattern Recognit. Artificial Intell. 14 (8) (2000) 1053–1066.

[15] A. Fusiello, E. Trucco, A. Verri, A compact algorithm for rectification of stereo pairs, Machine Vision Appl. 12 (1) (2000) 16–22.

[16] M. Okutomi, Y. Katayama, S. Oka, A simple stereo algorithm to recover precise object boundaries and smooth surfaces, Int. J. Comput. Vision 47 (1–3) (2002) 261–273.

[17] A. Peleg, U. Weiser, MMX technology extension to the intel architecture, IEEE Micro 16 (4) (1996) 42–50.

[18] R.B. Lee, Multimedia Extensions for General-Purpose Processors, Proceedings of IEEE Workshop on Signal Processing Systems Design and Implementation, Leicester, United Kingdom, 1997, pp. 9–23.

[19] R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence, Proceedings of European Conference on Computer Vision, Stockholm, Sweden, 1994, pp. 151–158.

[20] C. Sun, A fast stereo matching method, Proceedings of Digital Image Computing: Techniques and Applications, Auckland, New Zealand, 1997, pp. 95–97.

[21] T. Shreekant, T. Huff, Implementing streaming SIMD extensions on the Pentium III processor, IEEE Micro 20 (4) (2000) 47–57.

[22] Scharstein D., Szeliski R., Middlebury College Stereo Vision Research Page, www.middlebury.edu/stereo/

[23] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vision 47 (1–3) (2002) 7–42.

[24] Jean-Yves Bouguet Camera Calibraion Toolbox, www.vision.caltech.edu/bouguetj/calib_doc/

[25] Di Stefano L., Marchionni M., Mattoccia S., Experimental results, www.vision.deis.unibo.it/~smattoccia/stereo.htm