Dense stereo based on the uniqueness constraint

L. Di Stefano*, M. Marchionni*, S. Mattoccia**, G. Neri*
* DEIS-ARCES, University of Bologna
Viale Risorgimento 2, 40136 Bologna, Italy
** DEIS-ARCES, University of Bologna - Sede di Cesena
Viale Rasi e Spinelli 176, 47023 Cesena, Italy
ldistefano@deis.unibo.it, mmarchionni@litio.it, smattoccia@deis.unibo.it, gneri@deis.unibo.it

Abstract

The paper presents the matching core of a stereo algorithm suitable to real-time applications. Unlike most area-based algorithms, the proposed approach relies on a single matching phase (i.e. do not include the check for left-right consistency). Unreliable disparity measuremnts are primarly detected on the basis of the violation of the uniqueness constraint. In order to further improve the reliability of the matches we enforce additional contraints based on the behaviour of the error function that can be veryfied at a very small computational cost. Experimental results show that the proposed approach provides reliable disparity measurements and that it is significantly fast.

1 Introduction

Numerous examples of dense stereo applications which require real-time performance can be found in teleconferencing, virtual reality, robot navigation and control. According to a recent taxonomy [11], stereo algorithms that generate dense depth measurements can be roughly divided into two classes, namely *global* and *local* algorithms. *Global* algorithms, e.g. [12, 8] yield accurate and dense disparity measurements but exhibit a very high computational cost that renders them unsuited to real-time applications. *Local* algorithms, e.g. [4, 7, 9, 2], 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 can run fast enough to be deployed in many real-time applications.

As far as local matching algorithms are concerned, and considering the more common case of a binocular stereo imaging system, a widely adopted method [4, 9, 2] aimed at detecting unreliable matches, such for example those due to occlusions or photometric distortions, is the so called *left-right consistency constraint* [5]. This method relies on two successive matching phases, left-to-right (direct matching phase) and right-to-left (reverse matching phase). Then, only those matches that turn out to be coherent are retained as valid. The *left-right check* has proven to be particularly effective in detecting and discarding the erroneous matches necessarily yield by area-based algorithms [3, 9, 2]. However, this approach is characterised by a significant computational cost. This paper present a matching approach that detects unreliable matches during the direct matching phase and therefore does not require a reverse matching phase.

2 The basic matching core

We assume 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 (see for example [6]), must be applied to obtain a pair of images in standard form from the original ones.

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 minimize (maximize) an error (similarity) function, ε, representing the degree of dissimilarity (similarity) between two small regions of the stereo pair. Our matching 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 $[0 \dots d_{max}]$ and that the left image is scanned from top to bottom and from left to right during the matching process. The algorithm, starting from one point of left image, say $L(x - d_{max}, y)$, searches for the best candidate by evaluating function ε, within the interval $[R(x - d_{max}, y) \dots R(x, y)]$. Then, for the

1051-4651/02 $17.00 (c) 2002 IEEE

successive point of the reference image $L(x+1-d_{max}, y)$ the procedure is repeated searching for the best match within $[R(x+1-d_{max}, y) \dots R(x+1, y)]$. The process is then iterated for the successive points along the scanline.

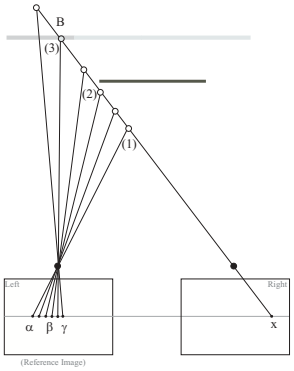

Figure 1. Geometric interpretation.

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 (similarity) 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 signal noise. However, wrong 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 $\alpha \leq \beta$, has 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 wrong 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 wrong and retain the match having the better score. Thus, if the point currently analysed $L(x+\beta-d_{max}, y)$ has a better score than $L(x+\alpha-d_{max}, y)$ (i.e. $\varepsilon(x+\beta-d_{max}, x, y) \leq \varepsilon(x+\alpha-d_{max}, x, y)$) our 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. Considering again the two matches $(1) : L(x+\alpha-d_{max}, y) \rightharpoonup R(x, y)$ and $(2) : L(x+\beta-d_{max}, y) \rightharpoonup R(x, y)$, 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: $(3) : L(x+\gamma-d_{max}, y) \rightharpoonup R(x, y)$, 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. Figure 1 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 locate 3D points lying on the same line of sight originating from a point of the right image. These sort of matches will be referred to as "colliding matches" for a given point of the right image $R(x, y)$.

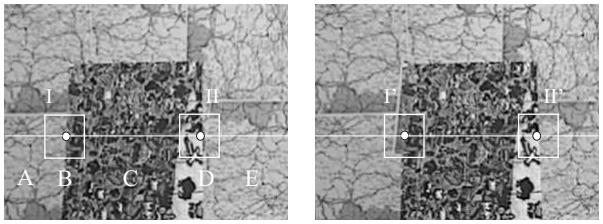

Figure 2. ***Map*** **stereo pair.**

3 Improving match reliability

The reliability of the disparity measurements provided by the basic matching core can be improved by incorporating additional constraints. Since our algorithm is aimed at real-time applications, we introduce new constraints by analysing the behaviour of the error (similarity) function. This allows us to assess the reliability of the matches at a very small computational cost by exploiting the data already computed by the matching core.

Let's consider Figure 3, that plots the error scores (ε *axis*) for each disparity value (d *axis*) along the epipolar line (x *axis*) marked in white in Figure 2. In 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, strong texture does not guarantees 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. 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 Figure 2. Local support for this point comprises background and foreground points. Due to the geometry of stereopsis, when searching

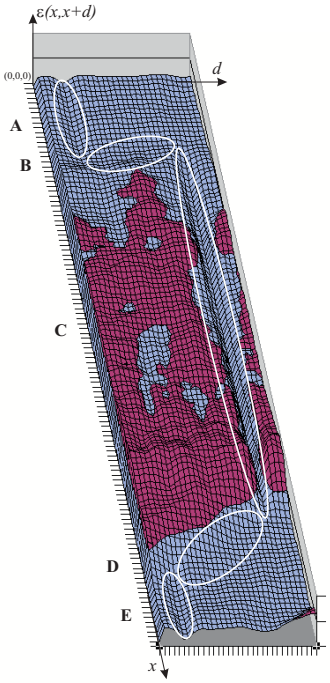

Figure 3. Plot of the error scores along the line marked in white in Figure 2.

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). Figure 2 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 (an high score). Consequently, the overall score for the minimum should result less distinctive than those typically found in regions without depth discontinuities (i.e. A, C and E). The described effect is worsened in region D due to the presence of occluded points that render the correspondence problem even more ambiguous.

In our algorithm the global minimum is located very quickly using a parallel technique [2] that with a few SIMD instructions (i.e. MMX instructions) yields the score (ε_{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, called *distinctiveness test* and *sharpness test*, 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 the match is potentially ambiguous (for example, this happens in presence of repetitive patterns), 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, the match can be considered reliable even though the score of the global minimum turns out not much smaller than those of the *pseudo-minima*.

The following relationship, referred to as *sharpness test*, 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 *sharpness test* we perform an additional test, referred to as *distinctiveness 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 $\frac{\delta \varepsilon}{\varepsilon_{min}}$ to evaluate the distinctiveness of the global minimum, with high ratios indicating distinctive global minima.

4 Experimental results

In this section we show the experimental results obtained with the three stereo pairs *Map*, *Sawtooth* and *Venus*, available at the web site [10]. Additional experimental results can be found at [1]. The three stereo pairs contain planar

slanted objects at different depths presenting several occlusions. The *Map* stereo pair (Figure 2) is highly textured while some regions of *Sawtooth* (Figure 5) and *Venus* (Figure 6) present a lower degree of texture. It is worth noticing that the ground truths available at [10] and our disparity maps are referred to opposite reference images.

Figure 4 shows the ground truth for the *Map* stereo pair (left) and the disparity maps obtained running our algorithm with the sharpness test and distinctiveness test disabled (center) and enabled (right). In these maps the grey levels represent disparity measurements (closer points are brighter) while white points represent unmatched points (i.e. points where the match has been rejected by the matching core or by the two additional test if these are enabled). The map obtained with the two tests disabled shows that the proposed matching approach based on the uniqueness constraint allows for recovering correctly most of the $3D$ structure of the scene and holds the potential to deal with large occlusions (see the wide white area at the right side of the foreground object). The map obtained with the two tests enabled indicates that the additional tests are effective in further rejecting unreliable measurements. These tests allows for discarding several wrong measurements established within the occluded region as well as measurements belonging to regions where the correlation window covers regions at different depths (white points around the object's contour). The disparity maps of Figure 5 and 6, obtained with the tests enabled, show that the proposed algorithm is capable of dealing with complex occlusions (e.g. those found in the *Sawtooth* stereo pair) as well as to discard measurements occurring in low textured areas (e.g. in *Venus* the dark background region and the uniform regions belonging to the foreground object appearing in the left of the image).

Table 1 reports the speed measurements obtained running our stereo algorithm on a Pentium III processor at 800 MHz with different images sizes (i.e. (320×240), (640×480), (800×600) and (1024×768)) and disparity ranges (i.e.16, 32, 48, 64 and 80).

$Image\ Size$	$d = 16$	$d = 32$	$d = 48$	$d = 64$	$d = 80$
(320×240)	39.59	31.25	27.44	25.94	25.96
(640×480)	8.94	6.92	5.77	5.17	4.78
(800×600)	5.56	4.28	3.60	3.18	2.89
(1024×768)	3.32	2.56	2.09	1.86	1.67

Table 1. Speed measurements in terms of frame per second (fps).

5 Conclusion

We have presented a stereo matching approach that relies only on a left-to-right matching phase. It detects unreliable matches via "colliding matches", i.e. matches that violate the uniqueness constraint. Further improvement of the reliability of the matching process is achieved by introducing additional constraints based on the analysis of the behaviour of the error (similarity) function. Our experimental results indicate that the matching algorithm produces dense and reliable disparity maps, detecting and discarding correctly unreliable matches, such as for example those associated with occlusions. Measurements of the execution times show that the algorithm is very fast and hence holds the potential to be employed in real-time applications requiring dense depth maps.

References

[1] *Experimental Results*. `http://labvisione.deis.unibo.it/~smattoccia/stereo.htm`.

[2] L. Di Stefano and S. Mattoccia. Fast stereo matching for the videt system using a general purpose processor with multimedia extensions. In *Fifth IEEE International Workshop on Computer Architecture for Machine Perception*, Padova, Italy, Sept. 2000.

[3] G. Egnal and R. Wildes. Detecting binocular half-occlusions: empirical comparisons of four approaches. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 466 –473, jun 2000.

[4] O. Faugeras et al. Real-time correlation-based stereo: algorithm, implementation and applications. INRIA Technical Report n. 2013, 1993.

[5] P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities. In *12th. International Joint Conference on Artificial Intelligence*, pages 1292–1298, Sydney, Aug. 1991.

[6] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.

[7] T. Kanade, H. Kato, S. Kimura, A. Yoshida, and K. Oda. Development of a video-rate stereo machine. In *Proc. of International Robotics and Systems Conference (IROS '95)*, volume 3, pages 95 – 100, August 1995.

[8] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *Proceedings of International Conference on Computer Vision*, 2001.

[9] K. Konolige. Small vision systems: Hardware and implementation. In *8th Int. Symposium on Robotics Research*, pages 111–116, Hayama, Japan, 1997.

[10] D. Scharstein and R. Szeliski. *Middlebury College Stereo Vision Research Page*. `http://www.middlebury.edu/stereo/`.

[11] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Microsoft Research Technical Report MSR-TR-2001-81, 2001.

[12] C. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE PAMI*, 22(7):675 – 684, July 2000.

1051-4651/02 $17.00 (c) 2002 IEEE

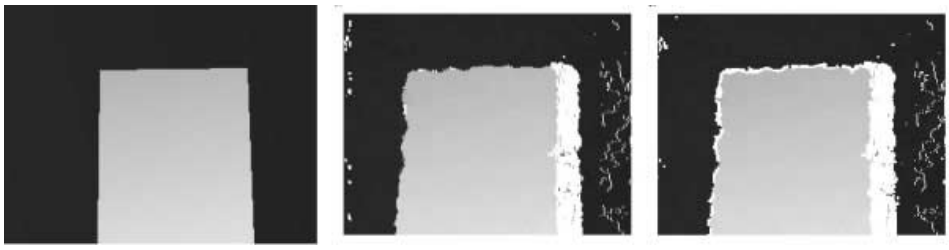

Figure 4. Ground truth for the *Map* stereo pair (left), disparity map obtained with additional tests disabled (center), disparity map obtained with additional tests enabled (right).

Figure 5. Left image of the *Sawtooth* stereo pair (left), ground truth (center), disparity map obtained with additional tests enabled (right).

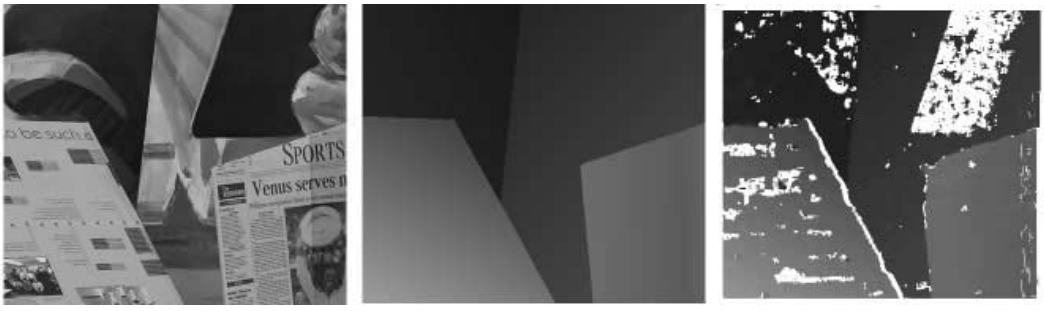

Figure 6. (Left) Left image of the *Venus* stereo pair (left), ground truth (center), disparity map obtained with additional tests enabled (right).

1051-4651/02 $17.00 (c) 2002 IEEE