Symmetric Stereo with Multiple Windowing

A. Fusiello, V. Roberto and E. Trucco

Abstract

We present a new, efficient stereo algorithm addressing robust disparity estimation in the presence of occlusions. The algorithm is an adaptive, multi-window scheme using left-right consistency to compute disparity and its associated uncertainty. We demonstrate and discuss performances with both synthetic and real stereo pairs, and show how our results improve on those of closely related techniques for both accuracy and efficiency.

Key-words: Computer Vision, Stereo; Depth and Shape Recovery; Area based; Multiple-Window.

1 Introduction

The aim of *computational stereopsis* [4, 2] is to reconstruct the 3-D geometry of a scene from two (or more) views, which we call *left* and *right*, taken by pinhole cameras. A well-known problem is *correspondence*, i.e., finding which points in the left and right images are projections of the same scene point (a *conjugate pair*). This is approached as search: finding the element in the right image which is most similar, according to a similarity metric, to a given element in the left image (a point, region, or generic feature).

Several factors make the correspondence problem difficult: (i) its inherent *ambiguity*, which requires the introduction of physical and geometric constraints [5, 7, 6], the most important being the *epipolar* constraint (see for example [5]); (ii) *occlusions*, i.e., points in one image with no corresponding point in the other; (iii) *photometric distortions* [3] arising when the pixels, projection of the same world point on the two images, have different intensities; and (iv) *figural distortion* [12], i.e., perspective images of the same objects taken from different views are in general different.

Correspondence algorithms can be grouped into two broad classes, area-based and feature-based. *Area-based* algorithms [12, 1, 7, 6, 11, 9] match small image windows centered at a given pixel, assuming that the grey levels are similar. They yield dense depth maps, but fail within occluded areas and/or poorly textured regions. Several correlation-related measures have been proposed, the *Sum of Squared Differences* (SSD) measure being a choice adopted most widely. *Feature-based* [10, 16, 15, 14] algorithms match local cues (e.g., edges, segments, corners) and can provide robust, but sparse, disparity maps requiring interpolation. These algorithms depend on feature extraction to locate reliable features in the two images.

This paper presents a new *symmetric, multi-window algorithm* (henceforth SMW) which addresses problems (i)-(iv) listed above, and outperforms closely related methods. SMW's assumptions are clearly stated in Section 2. SMW is based on the SSD measure (Section 3); it employs an adaptive, multi-window scheme to cure distortions and yield accurate disparities (Section 4), associated to uncertainty estimates. Robustness in the presence of occlusions is achieved thanks to the *left-right consistency constraint* (Section 5). A consistent uncertainty estimation mechanism (Section 6) guarantees that the depth maps produced can be used by data fusion schemes like [17]. To facilitate the reproduction of our results, we give a pseudocode

summary of the SMW algorithm (Section 7) as well as the Internet address of our public-domain implementation. A detailed experimental evaluation, including a comparison with similar methods reported in the literature, is reported in Sections 8 and 9. The paper is closed by a brief discussion of our work (Section 10).

2 Assumptions

With no loss of generality, we assume that conjugate pairs lie along raster lines, that is, the stereo pair has been *rectified* [5, 8], after appropriate calibration, to achieve parallel and horizontal epipolar lines in each image.

We also assume that the image intensities $I(x, y)$ of corresponding points in the two images are the same. If this is not true, the images can be *normalised* by a simple algorithm [3] which computes the parameters α, β of the gray-level *global* transformation

$$I_l(x, y) = \alpha I_r(x, y) + \beta \qquad \forall (x, y)$$

by fitting a straight line to the plot of the left cumulative histogram versus the right cumulative histogram. This normalisation fails if images are taken from too far viewpoints.

3 The SSD Algorithm

The basic structure of SSD can be outlined as follows. For each pixel in the image chosen as *reference* (e.g., the left one, I_l), similarity scores are computed by comparing a fixed, small window centered on the pixel to a window in the other image (here, I_r), shifting along the raster line. Windows are compared through the *normalised SSD measure*, which quantifies the difference between intensity patterns:

$$C(x,y,d) = \frac{\sum_{(\xi,\eta)} [I_l(x+\xi, y+\eta) - I_r(x+\xi+d, y+\eta)]^2}{\sqrt{\sum_{(\xi,\eta)} I_l(x+\xi, y+\eta)^2 \sum_{(\xi,\eta)} I_r(x+\xi+d, y+\eta)^2}} \quad (1)$$

where $\xi \in [-n, n]$, $\eta \in [-m, m]$. The disparity estimate for pixel (x, y) is the one that minimises the SSD error:

$$d_o(x,y) = \arg\min_d C(x,y,d). \quad (2)$$

Subpixel accuracy can be achieved by fitting a parabola to the SSD error function $C(d)$ in the neighbourhood of the minimum d_0 [1]:

$$s(x,y) = \frac{1}{2} \frac{C(x,y,d_o-1) - C(x,y,d_o+1)}{C(x,y,d_o-1) - 2C(x,y,d_o) + C(x,y,d_o+1)} \quad (3)$$

A basic SSD correlation algorithm has an asymptotic complexity of $O(N^2nm)$, with N the image size. However we can observe that squared differences need to be computed only once for each disparity, and the sum over the window needs not be recomputed from scratch when the window moves by one pixel. The optimised implementation that follows from this observation[6] has a computational complexity of $O(4N^2)$, which is independent of the window size.

4 The Need for Multiple Windows

As observed by Kanade and Okutomi [12], when the correlation window covers a region with non-constant disparity, area-based matching is likely to fail, and the error in the depth estimates grows with the window size. Reducing the latter, on the other hand, makes the estimated disparities more sensitive to noise.

To overcome such difficulties, Kanade and Okutomi proposed a statistically sound, adaptive technique which selects at each pixel the window size that minimises the uncertainty in the disparity estimates.

In the present work we take the multiple-window approach, in the simplified version proposed by [11, 9]. For each pixel we perform the correlation with nine different windows (showed in Figure 1), and retain the disparity with the smallest SSD error value. The idea is that a window yielding a smaller SSD error is more likely to cover a constant depth region; in this way, *the disparity profile itself drives the selection of an appropriate window.*

Consider the case of a piecewise-constant surface: points within a window close to surface discontinuities come from two different planes, therefore a single "average" disparity cannot be assigned to the whole window without making a gross error. The multiple windows approach can be regarded as a robust technique able to fit a constant disparity model to data consisting of piecewise-constant surface, that is, capable of "drawing the line" between two different populations (see Figure 3).

5 Occlusions and Left-Right Consistency

Occlusions create points that do not belong to any conjugate pairs. In many cases, occlusions involve depth discontinuities: indeed, occlusions in one image correspond to disparity jumps in the other [9].

A key observation to address the occlusion problem is that *matching is not a symmetric process*: taking different images (right or left) as reference, one obtains, in general, different sets of conjugate pairs, in which some points are involved in more than one conjugate pairs. Such pairs are not invariant to the choice of the reference image. As each point in one image can match at most one point in the other (the *uniqueness constraint*), such pairs can be

discarded (*left-right consistency*) [7, 6].

Consider for instance point B of Figure 2 and take the left image, I_l, as reference. Although B has no corresponding point in the right image, I_l (its conjugate point is occluded), the SSD minimisation returns a match anyhow (C'). If I_r is taken as reference, instead, C' is correctly matched to its conjugate point (C) in the left image. Therefore the conjugate pairs (B,C') and (C,C') violate left-right consistency; in other words, C' does not satisfy the uniqueness constraint. Notice that point B is *recognised* as occluded (strictly speaking, its conjugate point is occluded); our approach takes advantage of left-right consistency to detect occlusions and suppress the resulting unfeasible matches.

For each point (x, y) in the left image, the disparity $d_l(x, y)$ is computed as described in Section 3. The process is repeated with the right image as reference. If $d_l(x, y) = -d_r(x + d_l(x, y), y)$ the point is assigned the computed disparity; otherwise it is marked as occluded and a disparity is assigned heuristically. Following [13], we assume that occluded areas, occurring between two planes at different depth, take the disparity of the deeper plane.

6 Uncertainty Estimates

Area-based algorithms are likely to fail not only in occluded regions, but also in poorly-textured regions, which make disparity estimates more uncertain; it is therefore essential to assign it confidence estimates to disparities. Several uncertainty estimation schemes have been proposed for SSD, mostly based on the shape of the SSD error function [1, 17].

Our approach takes advantage of the multiple windows. Disparity estimation is sensitive to window shape in two cases: first, near a disparity jump (as discussed in Section 4) and, second, where the texture is poor, or the

signal-to-noise ratio (SNR) is low. Consequently, we define uncertainty as the estimated variance of the disparity measures obtained with the various windows (see algorithm summary in next section); occluded points are assigned infinite variance. Experimental results show that such our uncertainty measure is consistent, i.e., it grows as the SNR decreases (Section 8).

7 The SMW Algorithm

To facilitate the reproduction of our work, we summarise the SMW algorithm in pseudocode. An on-line demonstration is available from `http://www.dimi.uniud.it/~fusiello/demo-smw/smw.html`, where our C implementation can be downloaded as well.

Let $C(x, y, d; I_l, I_r, w)$ be the SSD error computed from I_l to I_r according to Eq. (1) at point (x, y), with disparity d and window w. Let s_l be the subpixel correction defined by Eq. (3). The y coordinate is omitted for the sake of simplicity, since we assume horizontal epipolar lines.

for all (x, y) in I_l **do**
  for all $w = 1 \dots K$ **do**
    $d_{l,w}(x) = \arg\min_d C(x, y, d; I_l, I_r, w)$
    $d_{r,w}(x) = \arg\min_d C(x, y, d; I_r, I_l, w)$
  end for
  $\sigma_d^2(x) = \frac{1}{K-1} \sum_{w=1}^{K} (d_{l,w}(x) - \bar{d}_{l,w}(x))^2.$
  $d_l(x) = \arg\min_w C(x, y, d_{l,w}; I_l, I_r, w)$
  $d_r(x) = \arg\min_w C(x, y, d_{r,w}; I_r, I_l, w)$
  $d(x) = d_l(x) + s_l(x)$
end for
for all (x, y) in I_l **do**
  if $(d_l(x) \neq -d_r(x + d_l(x))$ **then**

$\sigma_d^2(x) = +\infty$

end if

end for

8 Experimental Evaluation

This section reports the main results of experimental evaluation of SMW. The evaluation was aimed at assessing

- the accuracy of disparity computation,
- robustness against occlusion,
- the consistency of uncertainty estimation,
- the performance of SMW when compared to similar algorithms.

We used synthetic data sets commonly found in the stereo literature and controlled amounts of noise. We also reproduced patterns used for testing algorithms used in our comparative evaluation. The next section reports further tests with real stereo pair sof size 128×128.

Random-dot stereograms

We first performed experiments on noise-free random-dot stereograms (RDS), shown in Figure 4. In the disparity maps, the gray level encodes the disparity, that is the depth (the brighter the closer). Images have been equalised to improve readability; subpixel-accuracy values have been computed and rounded to integers. Following [12], the estimated Mean Absolute Error (MAE), that is the mean of absolute differences between estimated and ground true disparities, has been computed.

Simple SSD correlation applied to the RDS shows how most of the problems outlined in Sections 4 and 5 affect disparity computation. Figure 5

shows the disparity maps computed by SSD with fixed windows 3×3 and 7×7. Both pictures show the effect of disparity jumps (near the left and horizontal borders of the square patch) and of occlusions (near the right border of the square patch). The SMW algorithm with a 7×7 window was applied to the square RDS of Figure 4 and to a circular RDS (not shown here). Figure 6 show the disparity maps computed by SMW and the estimated uncertainty maps (the darker the lower) in both cases.

The MAE is negligible, and may be ascribed to subpixel estimation only. The occluded points, shown in white in the uncertainty maps, are recovered with 100% accuracy in both cases. The circle RDS shows that the algorithm is not biased toward square disparity patterns, as the shape of the SSD windows might suggest. The reader may want to compare the present results to those reported in [3].

Experiments with noisy RDSs show a graceful degradation when noise increases. Gaussian noise with zero mean and increasing variance was added independently to both images of the square RDS. Figure 7 plots the MAE against the standard deviation of the noise for SMW and SSD correlation. Each point depicts the average result of 20 independent trials.

In order to assess the uncertainty estimator incorporated in SMW, we plotted the average uncertainty computed over a square patch of uniform disparity against the SNR (Figure 8). The results show that the computed uncertainty consistently increases as the SNR decreases.

Gray-level ramps

We performed a systematic, quantitative comparison between SMW, our implementation of the Adaptive Window (AW) algorithm [12] (possibly the closest method to SMW in the literature), and fixed-window SSD with different window sizes. The evaluation was based on the main test pattern used

by [12]: an input stereo pair of an intensity ramp in the horizontal direction, warped according to a given disparity pattern. The left disparity jump creates a "disocclusion" area which is filled with random dots (Figure 9). Gaussian noise with zero mean and unit variance (gray level) was added to both images independently.

Figure 10 illustrates a comparison of the three algorithms using the ramp stereo pair.

Qualitative comparisons are illustrated in Table 1, which summarises the results of our comparison of the MAE for SSD, AW, and SMW, using input pairs with different noise levels and different window sizes.

Results with fixed-window SSD (Figure 10) confirm that too small a window (e.g., 3×3) increases sensitivity to noise, whereas larger windows (e.g., 7×7) act as low-pass filters and are likely to blur depth discontinuities.

More interestingly, Figure 10 shows that AW is the most accurate (since it reduces simultaneously both random and systematic errors along the disparity edges), but performs poorly within occluded areas, leading to large local errors, as it does not exploit the uniqueness constraint. Subpixel corrections are smooth since this algorithm is essentially a complex, iterative subpixel adjustment. SMW yields a depth map that is globally more reliable, as it enforces left-right consistency: occluded points are detected with 100% accuracy.

Further experiments with larger disparities (not reported here) show that the improvement in accuracy achieved by SMW with respect to AW increases with disparity, owing to the increasingly large areas of occlusion[1].

Another advantage of SMW with respect to AW is *efficiency*. Running

[1]Notice that our implementation of AW failed to converge to a solution with RDSs, probably because this algorithm relies on intensity derivatives, which are ill-defined for random dot patterns.

on a SUN SparcStation 4 (110MHz) under SunOS 5.5, our implementation of the SMW takes 8 seconds, on average, to compute the depth maps in Figure 10 (128×128 input images), while AW takes 32 minutes on average.

9 Experiments with Real Data

We report the results of the application of the SMW algorithm on standard image pairs from the JISCT (JPL-INRIA-SRI-CMU-TELEOS) stereo test set, and from the CMU-CIL (Carnegie-Mellon University—Calibrated Imaging Laboratory) in Figure 11. In the disparity maps, the gray level encodes disparity, that is depth (the brighter the closer). Images have been equalised to improve readability. Subpixel-accuracy values have been rounded to integer values for display. We also report the estimated variance maps (the darker the lower). Small values cannot be appreciated in spite of histogram equalisation, due to the large difference between high-uncertainty occlusion points and the rest of the image. Although a quantitative comparison with other methods was not possible with real images, the quality of SMW results seems perfectly comparable to that of the results reported, for example, in [18, 9, 3].

In Figure 12, we report the result of SMW on the "Head" stereo pair (from the Multiview Image Database, University of Tsukuba), for which the disparity ground truth is given. In this case we could compute the error image and the Mean Absolute Error, MAE=0.6194.

Running on a Sun SparcStation 4 (110MHz) under SunOS 5.5, our current implementation takes 50 seconds, on average, to compute depth maps from 256×256 pairs, with a disparity range of 10 pixels.

10 Discussion

We have introduced a new, efficient algorithm for stereo reconstruction, SMW, based on a multi-window approach, and taking advantage of left-right consistency. Our tests have shown the advantages offered by SMW. The adaptive, multi-window scheme yields robust disparity estimates in the presence of occlusions, and clearly outperforms fixed-window schemes. If necessary, the slight amount of noise caused by subpixel interpolation can be kept small by increasing the baseline, which does not worsen performance significantly thanks to the robust treatment of occlusions. This is an advantage over several stereo matching schemes, often limited by the assumption of small baselines.

Left-right consistency proves effective in eliminating false matches and identifying occluded regions (notice that this can be regarded as a segmentation method in itself). In addition, disparity is assigned to occluded points heuristically, thereby achieving reasonable depth maps even in occluded areas. Uncertainty maps are also computed, allowing the use of SMW as a module within more complex data fusion frameworks ([17]). As for any area-based correspondence method, SMW's performance is affected adversely by poorly-textured regions, but areas of low texture are associated consistently with high uncertainty values.

The efficiency of SMW is clearly superior to that of similar adaptive-window methods, and direct comparisons with [12] have been reported. The reason is that SMW performs a one-step, single-scale matching, with no need for interpolation and optimisation. The main disadvantage is that the window size remains a free parameter; notice, however, that adaptive-window schemes are much slower in achieving comparable accuracies.

Acknowledgements

This work was partially supported by a British Council-MURST/CRUI grant. The "Castle" images were provided by the Calibrated Imaging Laboratory at Carnegie Mellon University (CMU-CIL), supported by ARPA, NSF, and NASA. The "Head" images with ground truth, from the Multiview Image Database, are courtesy of Dr. Y. Otha, University of Tsukuba.

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[2] R. C. Bolles, H. H. Baker, and M. J. Hannah. The JISCT stereo evaluation. In *Proceedings of the Image Understanding Workshop*, pages 263–274, Washington, DC, April 1993. ARPA, Morgan Kaufmann.

[3] I. J. Cox, S. Hingorani, B. M. Maggs, and S. B. Rao. A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3):542–567, May 1996.

[4] U. R. Dhond and J. K. Aggarwal. Structure from stereo – a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, November/December 1989.

[5] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, Cambridge, MA, 1993.

[6] O. Faugeras, B. Hotz, H. Mathieu, T. Viéville, Z. Zhang, P. Fua, E. Théron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, and C. Proy.

Real-time correlation-based stereo: algorithm, implementation and applications. Technical Report 2013, Unité de recherche INRIA Sophia-Antipolis, August 1993.

[7] P. Fua. Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1292–1298, Sydney, Australia, August 1991.

[8] A. Fusiello, E. Trucco, and A. Verri. Rectification with unconstrained stereo geometry. In A. F. Clark, editor, *Proceedings of the British Machine Vision Conference*, pages 400–409. BMVA Press, September 1997.

[9] D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. *International Journal of Computer Vision*, 14(3):211–226, April 1995.

[10] W.E.L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(1):17–34, January 1985.

[11] S. S. Intille and A. F. Bobick. Disparity-space images and large occlusion stereo. In Jan-Olof Eklundh, editor, *European Conference on Computer Vision*, pages 179–186, Stockholm, Sweden, May 1994. Springer-Verlag.

[12] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, September 1994.

[13] J. J. Little and W. E. Gillett. Direct evidence for occlusions in stereo and motion. *Image and Vision Computing*, 8(4):328–340, November 1990.

[14] G. Medioni and R. Nevatia. Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31(1):2–18, July 1985.

[15] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154, March 1985.

[16] S. B. Pollard, J.E.W. Mayhew, and J.P. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient constraint. *Perception*, 14:449–470, 1985.

[17] E. Trucco, V. Roberto, S. Tinonin, and M. Corbatto. SSD disparity estimation for dynamic stereo. In R. B. Fisher and E. Trucco, editors, *Proceedings of the British Machine Vision Conference*, pages 342–352. BMVA Press, 1996.

[18] Y. Yang and A. L. Yuille. Multilevel enhancement and detection of stereo disparity surfaces. *Artificial Intelligence*, 78(1-2):121–145, October 1995.

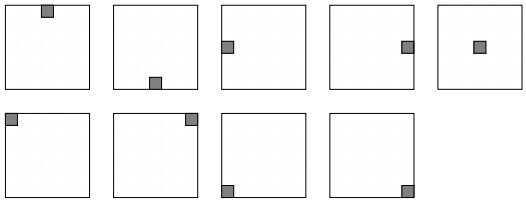

Figure 1. The nine correlation windows. The pixel for which disparity is computed is highlighted.

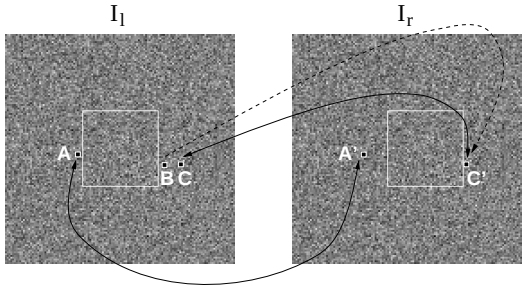

Figure 2. Left-right consistency. Point A is correctly matched to A'. Point B is given C' as a match, but C' matches C≠B.

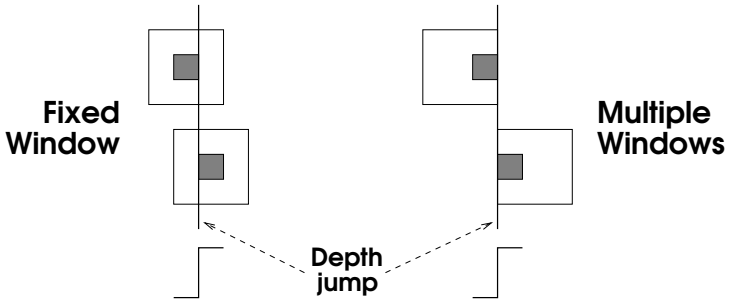

Figure 3. Multiple windows approach. If one use windows of fixed size with different centers, it is likely that one of them will cover a constant depth area.

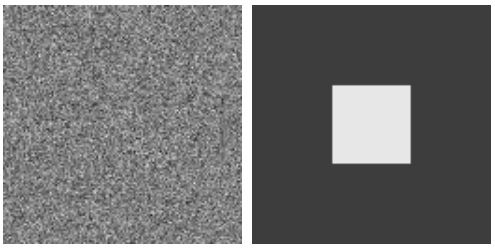

Figure 4. Square RDS. The right image of the stereogram is computed by warping the left one, which is a random texture (left), according to a given disparity pattern (right): the square has disparity 10 pixel, the background 3 pixel.

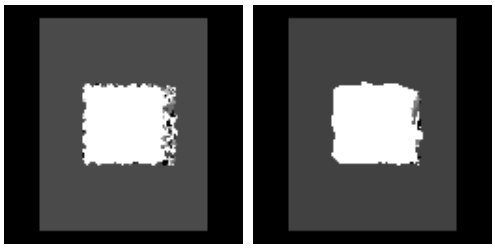

Figure 5. Computed disparity map by SSD correlation for the square RDS with 3×3 window (left) and 7×7 window (right); MAE is 0.240 and 0.144, respectively.

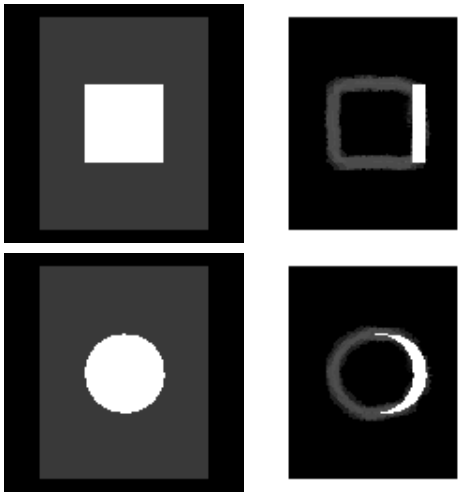

Figure 6. Computed disparity map (left) and uncertainty (right) by SMW for the square RDS (top) and for the circle RDS (bottom). MAE is 0.019 and 0.026 respectively.

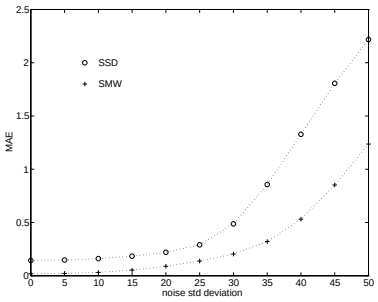

Figure 7. MAE of SMW and SSD vs noise standard deviation for the square RDS. Window is 7×7.

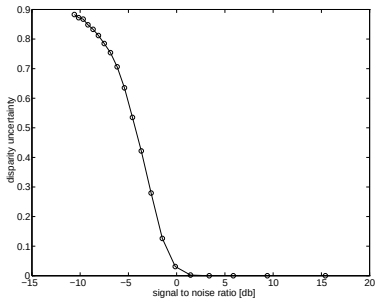

Figure 8. Mean uncertainty vs SNR for a constant disparity region of the square RDS.

Figure 9. Stereo pair. The central square has disparity 5 pixel, the background 2 pixel.

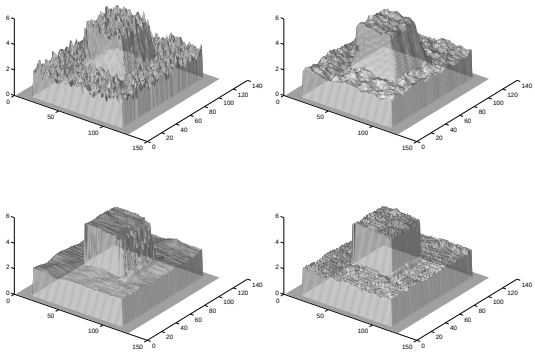

Figure 10. Isometric plots of the disparity maps computed with: SSD correlation 3×3 window (top left) and 7×7 window (top right), AW (bottom left) and SMW 7×7 algorithms (bottom right), with $\sigma^2 = 1.0$. The orientation is chosen to show occluded points.

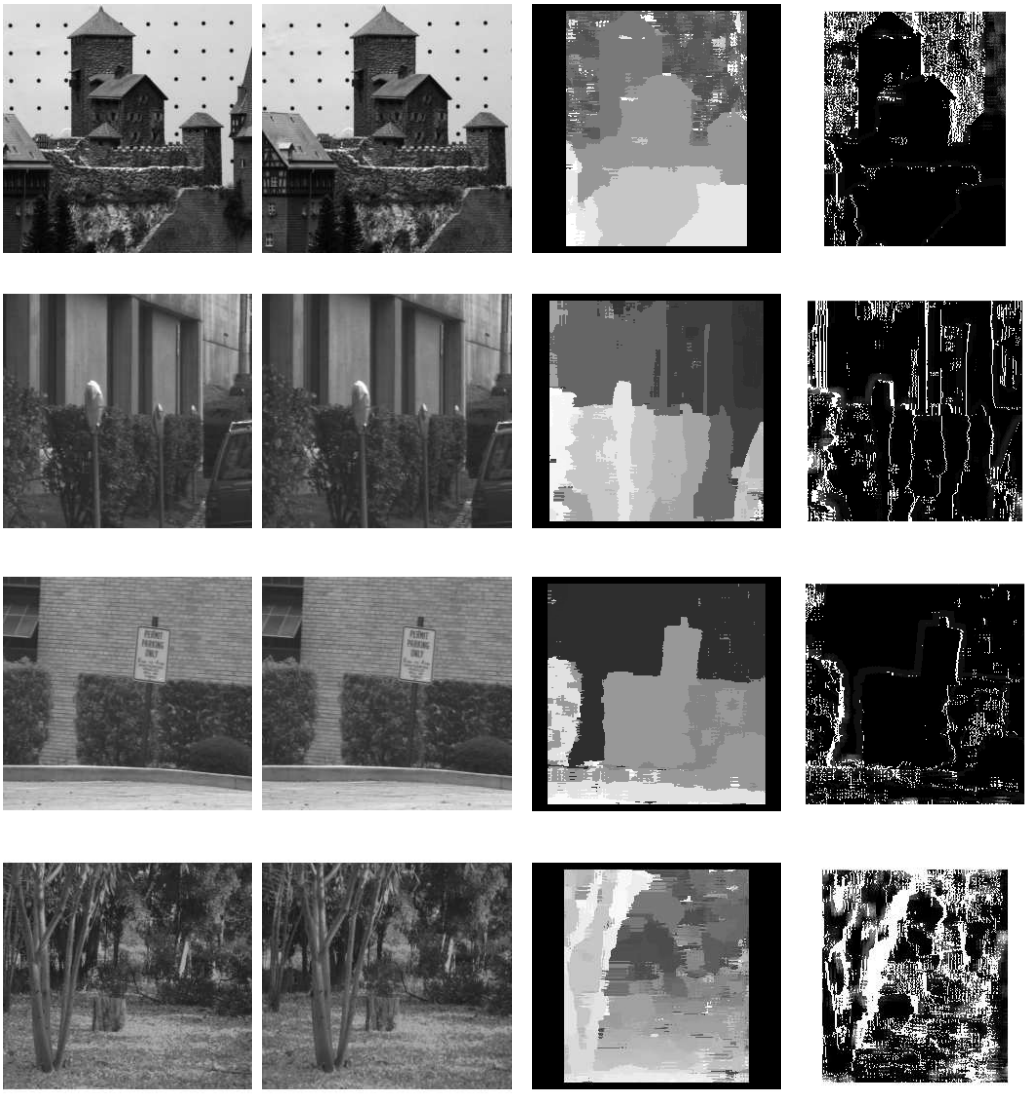

Figure 11. Disparity (left) and uncertainty maps (right) fo the “Castle”, “Parking meter”, “Schrub” and “Trees” stereo pairs

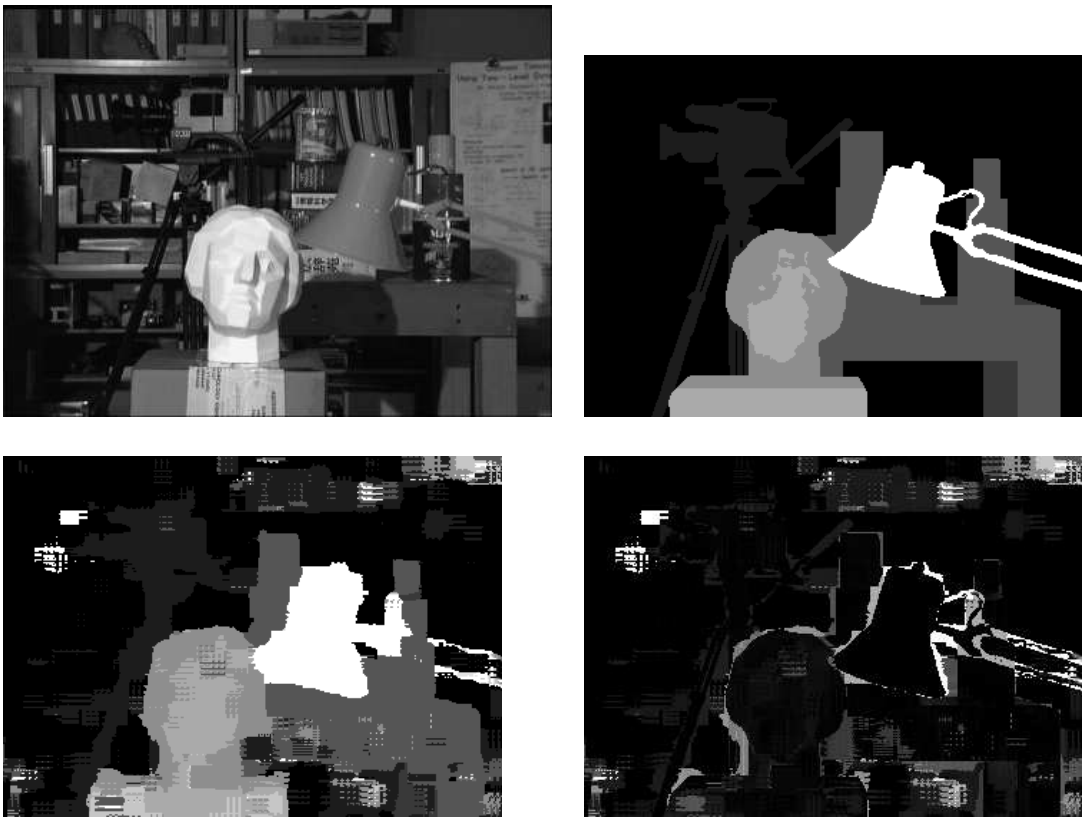

Figure 12. "Head" image and disparity ground truth (top row); disparity computed by SMW and error image(bottom row).

Algorithm	MAE		
	$\sigma^2 = 1.0$	$\sigma^2 = 3.0$	$\sigma^2 = 10.0$
SSD 7x7	0.182	0.468	1.235
SSD 15x15	0.284	0.392	0.988
AW	0.101	0.244	1.045
SMW 7x7	0.082	0.318	0.979
SMW 15x15	0.059	0.235	0.819

Table 1. Comparison of estimated errors: mean absolute (MAE) for different noise variances. Notice that 15×15 is the maximum window size allowed for AW.