Efficient Stereo with Multiple Windowing

Andrea Fusiello, Vito Roberto
Machine Vision Laboratory, Dept. Informatics
University of Udine, Italy
{fusiello,roberto}@dimi.uniud.it

Emanuele Trucco
Dept. of Computing and Electrical Eng.
Heriot-Watt University, UK
mtc@cee.hw.ac.uk

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 robustness and efficiency.

1. Introduction

The aim of computational stereopsis 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 (for a comprehensive review on computational stereo, see [3]). 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).

Area-based (or *correlation-based*) algorithms [6, 5, 10] 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. *Feature-based* algorithms [13, 8, 16] match local cues (e.g., edges, lines, corners) and can provide robust, but sparse, disparity maps requiring interpolation. These algorithms rely on feature extraction.

Several factors make the correspondence problem difficult: (i) its inherent *ambiguity* requires the introduction of physical and geometrical constraints; (ii) *occlusions*, i.e., points in one image with no corresponding point in the other; (iii) *photometric distortions* [2] arising when the pixels, projection of a world point on the two images, have different intensities; and (iv) *projective distortion* [11] that makes the projected shapes different in the two images.

This paper presents a robust area-based algorithm, addressing all problems (i)-(iv) listed above. The basic area-based method, SSD correlation, is sketched first (Section 2), followed by our adaptive, multi-window scheme (Section 3), which contrasts distortions and yields accurate disparities. Robust disparity estimates in the presence of occlusions are achieved thanks to the *left-right consistency constraint* (Section 4); the associate uncertainty is estimated too (Section 5). Our algorithm is presented in Section 6. Experimental results, improved performance with respect to Kanade and Okutomi's adaptive-window scheme [11], and a brief comparison with other methods are presented in Sections 7 and 8.

2. The SSD Correlation Algorithm

We assume that conjugate pairs lie along raster lines. This implies no loss of generality, as general images can be rectified after appropriate calibration to achieve epipolar lines parallel and horizontal in each image [4] We also assume that the image intensity of a 3D point is the same on the two images. If this is not true, the images must be *normalised* [2] to cure such distortion.

Similarity scores are computed, for each pixel in the left image, by comparing a fixed small window centered on the pixel to a window in the right image, shifting along the raster line. As a similarity measure we adopt the well-known (normalised) SSD (*Sum of Squared Differences*) error:

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

where $\xi \in [-n,n]$, $\eta \in [-m,m]$. The computed disparity is the one that minimises the SSD error:

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

1063-6919/97 $10.00

Subpixel precision is achieved by fitting a curve to the errors in the neighbourhood of the minimum [1]:

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

A basic SSD correlation algorithm has an asymptotic complexity of $O(N^2nm)$, with N the image size. Following [5], we implemented an optimised version making computational complexity independent of the window size.

3. The Need for Multiple Windows

As observed by Kanade and Okutomi [11], 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 computed disparity more noise-sensitive. 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 this work we take the multiple window approach in the simplified version proposed by [7, 10]. For each pixel we perform the correlation with nine different windows (showed in Fig. 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*.

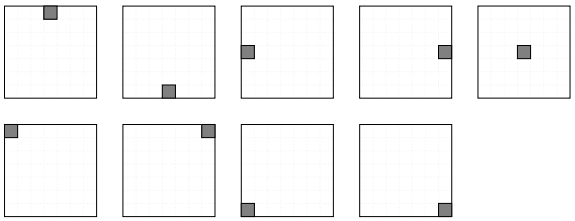

Figure 1. The nine asymmetric correlation windows. The pixel for which disparity is computed is highlighted.

4. Occlusions and Left-Right Consistency

Occlusions create points that do not belong to any conjugate pairs. In many cases, occlusions occur at depth discontinuities: indeed, one may observe [7] that occlusions on one image correspond to disparity jumps on the other. Although evidences have been reported [15] that occlusions help the human visual system in detecting object boundaries, in computational stereo they are a major source of errors.

A key observation to address the occlusion problem is that *matching is not a symmetric process*: when searching for conjugate pairs, only the visible points in one image are matched. If the role of left and right images is reversed, new conjugate pairs are found. The so-called *left-right consistency constraint* [6, 5] states that feasible conjugate pairs are those found with both direct and reverse matchings. It is worthwhile noting that the latter is equivalent to the *uniqueness constraint*, which states that each point on one image can match at most one point on the other image. Consider for instance an occluded point, e.g., B, in the left image of Fig. 2: although it has no corresponding point in the right image, the SSD minimisation matches it to some point (C') anyhow. One can see that the latter point, in turn, corresponds to a different point in the left image, but this information is available only by searching from right to left.

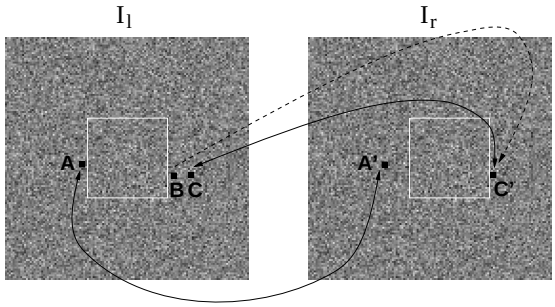

Figure 2. Left-right consistency. Point B which is occluded, is given C' as a match, but C' matches C $\neq$ B. The pair (B,C') can be suppressed.

In our approach, occlusions are detected by checking the left-right consistency, and suppressing unfeasible matches accordingly. For each point on the left image the disparity $d_l(x)$ is computed as described in Section 2. The process is repeated after reversing the two images. If $d_l(x) = -d_r(x + d_l(x))$ the point keeps its computed left disparity, otherwise it is marked as occluded and a disparity is assigned heuristically: following [12], we assume that occluded areas, occurring between two planes at different depth, take the disparity of the deeper plane.

5. 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 compute confidence measures for disparities,

which enables solutions to fill in gaps of the depth maps (e.g., by fusing multiple views [14, 17]). Several techniques are available to estimate uncertainty, most of them based on the shape of the SSD error function [1, 14, 17].

In our approach we take advantage of the multiple windows used for computing disparities. We define uncertainty as the estimated variance of the disparity measures obtained with the various windows (see algorithm below); occluded points are assigned infinite variance.

6. The Algorithm

We now present our symmetric, multi-window algorithm, henceforth SMW. The symbol w refers to the nine windows used; $C(x, y, d; I_l, I_r, w)$ is the SSD error computed from I_l to I_r according to Eq. 1 in the point (x, y), with disparity d, window w; $subpixel$ refers to the subpixel correction computed according to Eq. 3. The y coordinate will be omitted for sake of simplicity, since it does not vary.

for all (x, y) in I_l **do**
  for all w **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
  $\sigma_d^2(x) = \frac{1}{N-1} \sum_{w=1}^{N} (d_{l,w}(x) - \bar{d}_{l,w}(x))^2$.
  $d_l(x) = \arg min_w C(x, y, d_{l,w}; I_l, I_r, w)$
  $d_r(x) = \arg min_w C(x, y, d_{r,w}; I_r, I_l, w)$
  $d(x) = d_l(x) + subpixel$
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

To facilitate reimplementations and experimentations with our algorithm, the C code of the algorithm is available via anonymous ftp at `taras.dimi.uniud.it/pub/sources/smw.tar.gz`

7. Experiments with Synthetic Data

We first performed experiments on uncorrupted random-dot stereograms, generated with the disparity pattern of Fig. 3, in order to verify the correct behaviour in the absence of noise. Fig. 4 shows the output of SMW; the estimated mean absolute error (MAE) is 0.019 pixel and the maximum absolute error is 0.107 pixel. We conclude that the computed depth map is accurate, and the errors may be ascribed to the subpixel accuracy only. The occluded points, shown in white in the variance map, are recovered with 100% accuracy, in this case. Further experiments with noisy random-dot stereograms (not included for reason of space) also showed very good performances.

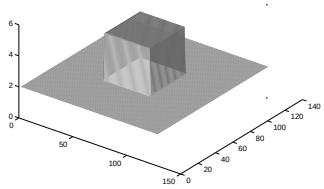

Figure 3. Disparity pattern used to generate the right images: the central square has disparity 5 pixel, the background 2 pixel.

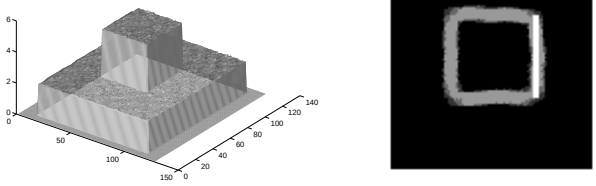

Figure 4. Computed disparity map (left) for the random-dot stereogram and its variance (right).

In order to perform a quantitative comparison between SMW and our implementation of the Adaptive Window (AW) algorithm [11], possibly the closest method to SMW in the literature, we created the same pattern used by [11], an input stereo pair of a ramp in the horizontal direction, deformed according to the disparity pattern in Fig. 3. The left disparity jump creates a "disocclusion" area which is filled with a random texture. Gaussian noise with zero mean and unit variance (gray level) was added independently to both images.

Fig. 5 illustrates a comparison of four different algorithms using this stereo pair. It shows the isometric plots of the disparity maps computed by simple SSD correlation and fixed windows of 3×3 and 7×7 pixels; by our implementation of the AW (after three iterations), which refines the initial estimates obtained with a 3×3 window; and by our SMW algorithm with 7×7 windows.

The results of the fixed-window SSD confirm that a window too small (3×3) is noise-sensitive, whereas a large one (7×7) acts as a low-pass filter, and is likely to miss depth discontinuities. The AW algorithm is more accurate, since it simultaneously reduces both the random errors and the

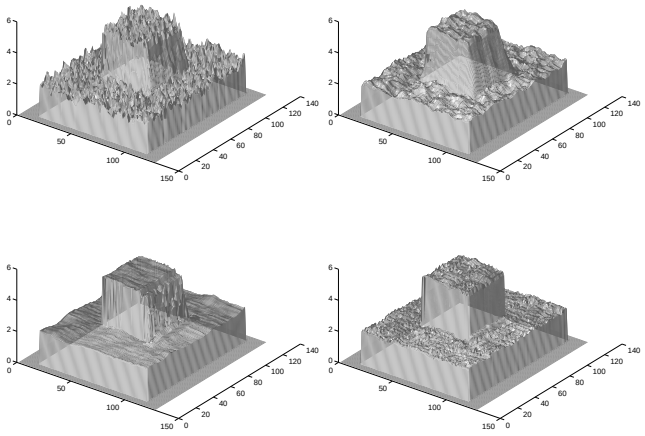

Figure 5. Isometric plots of the disparity maps computed with: SSD correlation 3×3 window (top left) and 7×7 window (top right), AW (bottom left) and SMW 7×7 algorithms (bottom right), with $\sigma^2 = 1.0$.

systematic ones, along the disparity edges. However, it performs poorly within occluded areas. Subpixel corrections are smooth, since this algorithm is essentially a complex, iterative subpixel adjustment. The SMW algorithm addresses the occlusion problem explicitly, and therefore yields a depth map that is even more accurate. The slight amount of noise across the disparity surface is a product of the simple subpixel-accuracy method adopted (see the conclusions). Fig. 6 compares qualitatively the isometric plots of the absolute errors (difference between true and reconstructed surface) for the AW and SMW algorithms.

Further comparisons are illustrated in Table 1, which summarises the results of our comparison of the MAE for the three algorithms (SSD, AW, SMW), using input pairs with different noise levels and different window sizes. The table shows that SMW algorithm performs better than the others. In fact, the AW algorithm is not effective in detecting occlusions, since neither symmetry nor uniqueness constraints are exploited. SMW is effective thanks to the left-right consistency check applied: the occluded points are detected with 100% accuracy. The main source of error for AW are the occluded points, leading to large local errors; for SMW, the errors are mainly due to the inaccuracy of the subpixel method used, which results only in small errors throughout the disparity map.

Further experiments with larger disparities show that the improvement in accuracy achieved by SMW with respect to AW increases with disparity, owing to the larger occluded regions.

Another advantage of SMW with respect to AW is the efficiency. Running on a SUN SparcStation 4 with SunOS 5.5, 110MHz, our implementation of the SMW takes 8 seconds, on the average, to compute the depth maps in Fig. 5 (128×128 input images), while AW takes 32 minutes on the average.

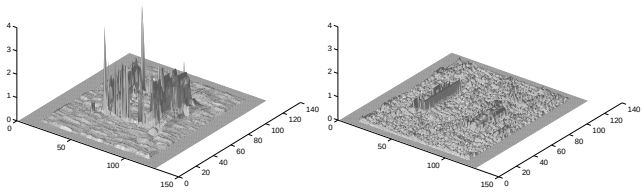

Figure 6. Isometric plots of estimated errors, as differences between computed and true disparities for the AW (left) and SMW algorithm (right). Note the different vertical scale.

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.

8. 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 Figures 7, 8, 9, 10. 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 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 was not possible with real images, the quality of

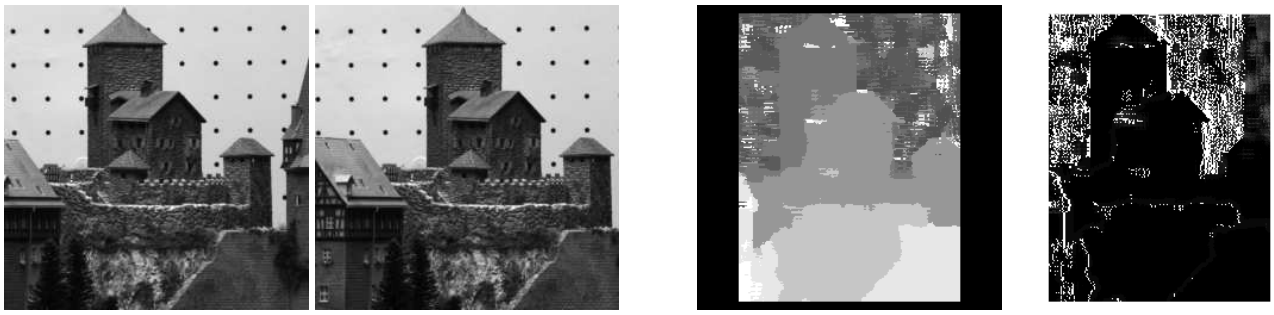

Figure 7. The “Castle” stereo pair; the disparity (left) and variance maps (right).

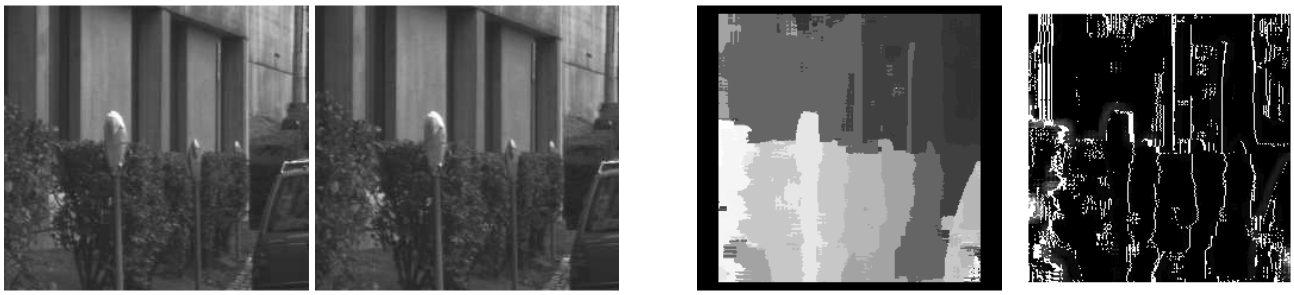

Figure 8. The “Parking meter” stereo pair; the disparity (left) and variance maps (right).

Figure 9. The “Shrub” stereo pair; the disparity (left) and variance maps (right).

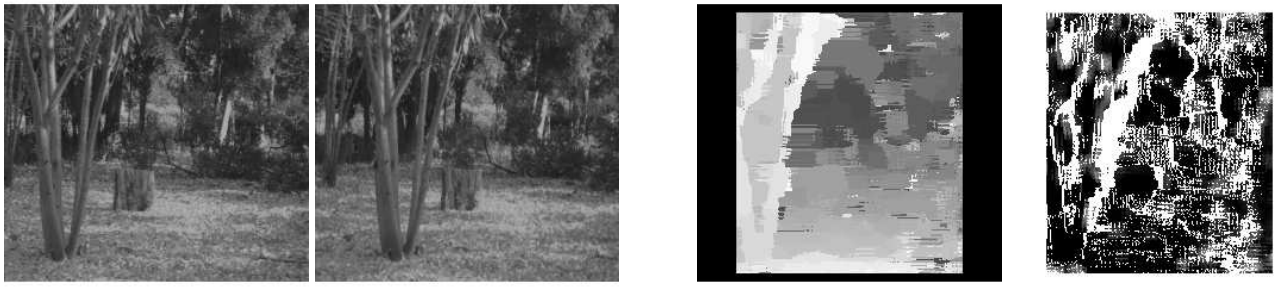

Figure 10. The “Trees” stereo pair; the disparity (left) and variance maps (right).

SMW results seems perfectly comparable to that of results reported, for example, in [7, 2, 9].

Running on the same hw/sw platform, our current implementation takes 50 seconds, on the average, to compute depth maps from 256×256 pairs, and a disparity range of 10 pixels.

9. Conclusions

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. Notice that the slight amount of noise resulting from the simple subpixel interpolation used may be made to correspond to small depth errors by increasing the baseline, thanks to the robust treatment of occlusions. This is an advantage over several stereo matching schemes, often limited by the assumption of small baselines.

The left-right consistency check proves very 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 (e.g., [14, 17]).

The efficiency of SMW is clearly superior to that of similar adaptive-window methods, and direct comparisons with [11] have been reported. This is due to the fact 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-windows schemes are much slower, and the quality of our results is good anyway. There are also problems with poorly-textured regions, but these are typical of any area-based approach, and cannot be regarded as a disadvantage of SMW. Moreover, our uncertainty treatment marks consistently areas of low texture with high uncertainty [17].

Work is in progress to embed the SMW module in a dynamic stereo fusion system.

Acknowledgements

This work was partially supported by a British Council-MURST/CRUI grant. The data for this research were partially provided by the Calibrated Imaging Laboratory at Carnegie Mellon University (CMU-CIL), supported by ARPA, NSF, and NASA.

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] 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.

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

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

[5] 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, Août 1993.

[6] 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*, Sydney, Australia, August 1991.

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

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

[9] R. D. Henkel. Hierarchical calculation of 3d-structure. Technical Report 5/94, Zentrum fur Kognitionswissenschaften, Universitat Bremen, 1994.

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

[11] 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.

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

[13] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976.

[14] L. Matthies, T. Kanade, and R. Szelisky. Kalman filter based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3:209–236, 1989.

[15] K. Nakayama and S. Shimojo. Da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research*, 30:1811–1825, 1990.

[16] 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, 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.