DENSE AND RELIABLE DSM GENERATION
FROM VHR STEREO PAIRS IN URBAN ENVIRONMENTS

T. Bekaert(1), S. Gautama(1), W. Philips(1), R. Goossens(2)

(1) TELIN, Ghent University, St.Pietersnieuwstraat 41, B-9000 Gent, Belgium
(2) Dept. Geography, Ghent University, Krijgslaan 281 (S8), B-9000 Gent, Belgium

KEY WORDS: stereo vision, urban environment, object borders, correlation

ABSTRACT:

Photogrammetric terrain reconstruction from aerial and space stereopairs of images occupies a prominent place in cartography and remote sensing. Stereo vision systems determine depth from two or more images using automated techniques. The most important and time consuming task for a stereo vision system is the registration of both images, i.e. the matching of corresponding pixels. Area based stereo attempts to determine the correspondence for every pixel, which results in a dense depth map. Correlation is the basic method used to find corresponding pixels. However, correlation assumes that the depth is equal for all pixels of a correlation window, which is violated at depth discontinuities. The result is that object borders are blurred and small details or objects are removed, depending on the size of the correlation window. In this paper, we focus on the generation of reliable surface models using dense techniques. An overview is given of correlation-based techniques using adaptive windows. These adaptive techniques separate fore- from background information in a correlation window using structure specific information (e.g. gradient, segmentation) and are able to compensate for the blurring effect that occurs at object boundaries. The result is a dense surface model with an emphasis on reliability. A case study that compares the different techniques is presented.

1. INTRODUCTION

Stereo vision systems aim at reconstructing 3D scenes by matching two or more images taken from slightly different viewpoints. The main problem that has to be solved is the identification of corresponding pixels, i.e. pixels that represent the same point in the scene. In the area of computer vision, this correspondence problem has been studied extensively. A great diversity of algorithms has been developed, often in a specific context like the automatic extraction of digital elevation models or vision systems for autonomous robots. Almost all algorithms contain a large number of parameters and all suffer problems caused by a lack of knowledge about correctly extracted information.

The variety of different algorithms can be divided into local and model based techniques. Model based stereo algorithms treat pixels in some manner according to a specific predefined model that has to be fitted. Information is added that cannot be extracted from pixel space in the strict sense. Local techniques only use information from the pixel space to determine correspondences between images. Systematic comparisons of algorithms and quantisations of the quality of the matching results are less well represented in literature (Scharstein & Szeliski [1], Kostkova [2]). Existing test benches are based only on a small number of stereo couples for which the highly parametrised algorithms are finely tuned.

The focus of this paper is on the available raw data in pixel space. We are concerned with determining the feasibility of extracting as much information as possible out of pixel space and the quality of this information. No complex and highly parametrised models are used to catch a lack of information. It is not our intention to develop a specific algorithm for object extraction. Instead, a generic local matching algorithm with subpixel resolution, hierarchic matching and symmetric filtering has been implemented.

Local algorithms try to establish correspondences by selecting some pixels in the neighborhood of the pixel of interest and by correlating those pixels to possible corresponding ones in a second image. The fundamental problem that arises with stereo matching algorithms is caused by the geometrically different positions of the objects in each image due to different viewpoints of the sensors. Local algorithms implicitly assume that the depth of the scene is constant in the areas or windows around the pixels of interest, which is not always the case. Knowledge about the weaknesses of correlation based local algorithms allows to formulate improvements that give better results near object borders where a constant scene depth is generally not guaranteed. A number of techniques try to solve this problem by not taking into account the pixels that do not correspond to the same depth as the pixel of interest when determining the correspondence measure.

In this paper, correspondence problems that arise near object borders are described. In the context of remote sensing, these object borders typically correspond to the edges of buildings. The possible matching situations are described qualitatively. For this qualitative evaluation a classic local matching technique based on rectangular windows with fixed size is systematically compared with the solution proposed by Cord *et al.* [3]. This method is based on the assumption that object borders correspond to variations in pixel values in such a way that it is possible to extract the object borders with an edge detector. After evaluation of the obtained correspondences certain specific cases are selected. Based on those case studies strengths and weaknesses of both methods are discussed and an improvement is proposed. This improvement is qualitatively discussed and compared with the original technique. This research is a first step in the systematic evaluation of different techniques that want to extract a maximum of useful information in pixel space in order to improve the quality of the matching. In a next step, the quality of the matching should be described quantitatively.

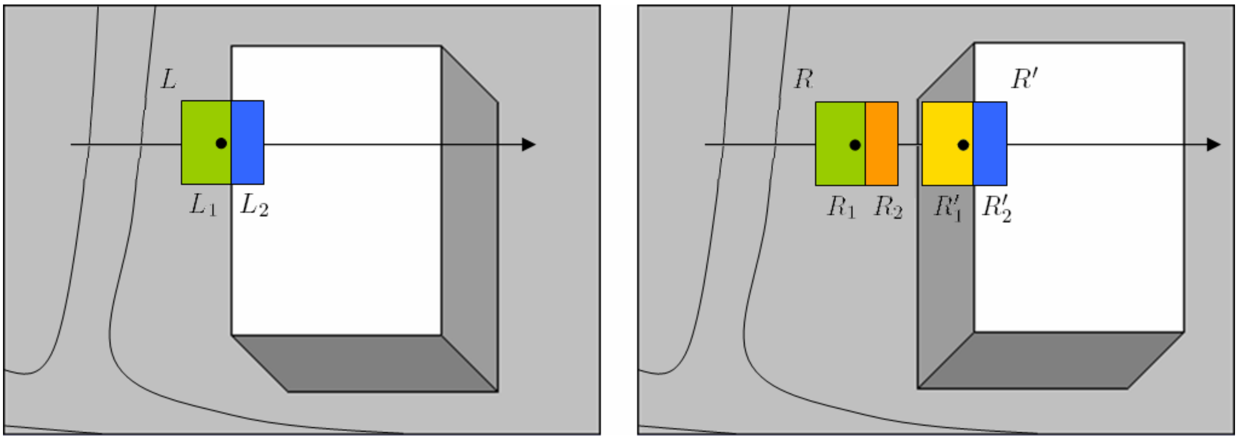


Fig.1 Illustration of the problem of depth discontinuities that occur when using correlation windows for matching.

2. FUNDAMENTAL WEAKNESS OF CORRELATION BASED LOCAL ALGORITHMS

Figure 1 shows the difficulty in determining the correspondences near object borders. Window L around the pixel of interest in the left image is correlated with a second window R, which is moved over all possible positions in the right image. The position where the correlation measure has the highest value determines the pixel in the right image that corresponds to the pixel of interest. Note that the fixed size window L overlaps a depth discontinuity. In the right image two possible second windows R and R' are shown. The corresponding right pixel is defined by its window, let C(L,R) be the cost of correspondence between the two pixels, i.e. the correlation value of the windows L and R. This cost expresses the similarity between the windows and is usually determined by using normalized cross correlation. A high cost corresponds to a high similarity.

The part L2 of the window L covers an area with a clearly different height as the one of the point represented by the pixel of interest. If the pixel values of window L2 differ from those of R2, this part of the window will cause a lower correlation value. The value C(L2,R'2) on the other hand is high since both windows cover the same area.

When the value C(L1,R'1) is high enough, C(L,R') might be higher than C(L,R) in which case a wrong correspondence is established: the pixel of interest in the left image is matched with the center pixel of window R'. This way reconstructed buildings tend to be wider than they are: pixels of which the correlation window partially overlaps the buildings are given the same height as the building itself.

Different propositions have been made to change the shape of the rectangular correlation window in order to improve the quality of the matching results. A first method is by changing the size of the window: smaller windows reduce the problems near borders but increase the effect of noise.

Fusiello et al. [5] use different positions for the pixel of interest, eg. in every corner, in the middle of every side and in the center. Correlation is done for all window positions but only the results of the best window is used. This way one tries to select the window with the greatest number of pixels that represent the same height as the pixel of interest. Kanade & Okutomi [4] proposed an iterative algorithm in which the size and shape of the rectangular windows is changed based on variations in pixel intensities and previously found disparities. Tao et al. [6] leave the idea of classic windows and use (rather large) windows based on image segmentation. These techniques rely heavily on the previous segmentation: errors in segmentation cause errors in large areas.

Cord et al. [3] proposed using adjusted windows based on edge detection. In a first step the edges in the left image are extracted. Next, starting from a classic rectangular window V centered around pixel of interest p(xc,yc) the adjusted window V' is constructed by removing a number of pixels, depending on the extracted contours.

All pixels (x,y) within the square window for which there is no 4-connected path to the central pixel of interest that crosses no extracted edge are removed. In other words, all pixels that are not reached by a flooding operation started from the central pixel are not taken into account.

Fig. 2a shows a square window V centered around the pixel of interest that has to matched (dark grey). The extracted edges within the area covered by the window are shown on Fig.2b. All pixels that are kept in the adjusted window V' (Fig.2c) are shown in grey.

When height discontinuities correspond with detectable pixel value gradients, this technique guarantees that all pixels used for correlation correspond with points at a similar height as the point represented by the pixel that has to be matched.

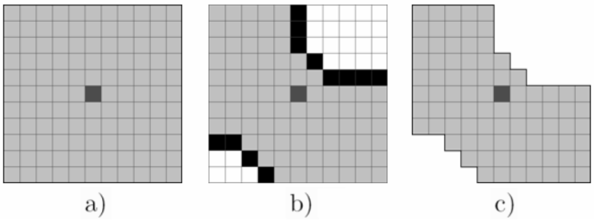


Fig.2: Illustration of adjusted correlation windows using edge information: (a) original window, (b) edge information within window, (c) adjusted window

3. SYSTEMATIC EVALUATION AND CASE STUDIES

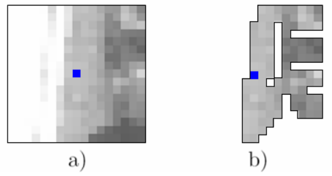
Next to the visual inspection of the different matching results obtained by using both square and adjusted windows based on edge detection, a systematic examination of the causes of this behaviour has been performed.

Four possible situations occur:

1. By using both the square window and the adjusted window the correct match is established. When this situation occurs, both windows seems to differ little. Typically only a few pixels out of the square window are not taken into account to construct the adjusted window.

a) b)

Fig.3: (a) Left image, (b) Right image.
By using a square window the pixel associated with window R is chosen, by using an adjusted window the pixel associated with window R' is chosen.



a) b)

Fig.4: Windows in the left image that are correlated with windows in the right image in order to find a match.
(a) Square window, (b) Adjusted window.

2. The adjusted window gives a correct match, while the square window does not. This typically occurs near sharp height discontinuities (Fig.3 and 4). These figures show the match of a point on the sidewalk near a building. When using a square window for correlation the window R is chosen and an incorrect match is established. When using the adjusted window, the window R' is chosen and a correct match is made because the adjusted window did not contain the part of the square window that covers the building.

3. Both windows give an incorrect correspondence. This occurs in 2 cases:
    - In homogeneous areas matching can be erroneous because of a lack of information in the image.
    - When the square window gives a false correspondence near the edge of building and the detected edge is not continuous (i.e. contains a gap) the constructed adjusted window will be nearly the same as the original square one. This shows the importance of the quality of the extracted edges, which is greatly influenced by the choice of edge detector and parameters.

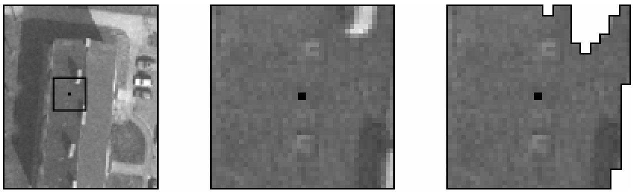


Fig.5: (left) Image with square window and pixel of interest, (center) the square window used for correlation, (right) the adjusted window based on the square window shown in the center. The central pixel of interest is shown.

4. The square window gives a correct correspondence, the adjusted window does not. It was proven that adjusted windows do not always give better matching results. When using adjusted windows, clearly distinguishable features in relatively flat areas tend to be removed from the square windows. Fig.5 shows the left image with the pixel of interest shown together with the square window used for correlation and the adjusted window constructed based on the square window. Fig.6 shows the correlation curves obtained by using both window types. The position of the window in the second image that has the highest correlation value, and thus the highest similarity, is different for each window type. The correct corresponding pixel is determined by the window position obtained by using the square window (Fig.7).

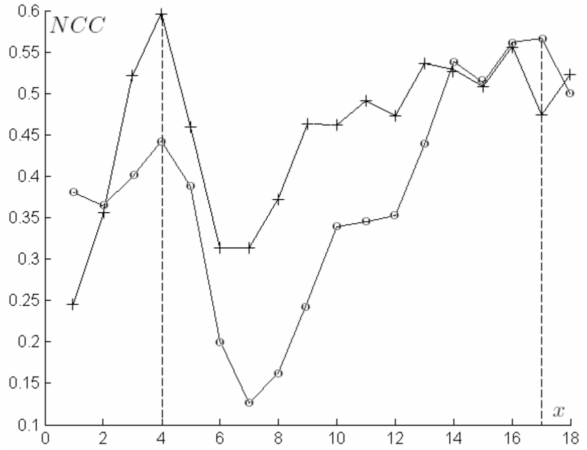


Fig. 6: correlation curves obtained by using a square window (+) and by using an adjusted window (o).
The correlation value (NCC) is shown as a function of the position of the second window in the second image.

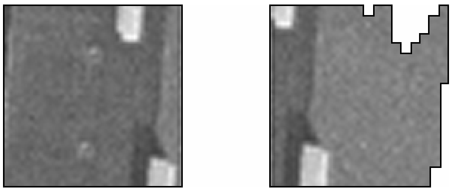


Fig.7: windows in the right image that correspond to window position 4 (left) and 17 (right).
The first window gives the highest correlation when using square windows, the second one when using adjusted windows. Together with Fig.5 (left), it is clear that the first window is correct, the second is not.

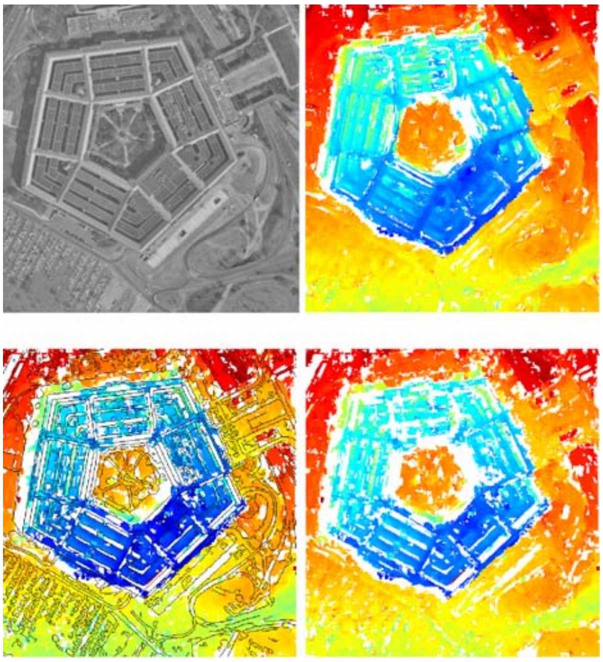


Fig.8: Stereo couple ‘Pentagon’. (Upper left) original image, (upper right) disparity map obtained by using square windows, (lower left) disparity map obtained by using adjusted windows based on square windows with extracted edges shown in black, (lower right) same as lower left without the edges.

Fig.8 shows the matching results of the stereo couple ‘Pentagon’. The upper left figure shows that the building consists of a number of concentric pentagons, each time with sharp height discontinuities. The upper right figure shows the disparity map obtained by using square windows. Every pentagon is made wider in such a way that the sharp height discontinuities disappear.

The lower right figure shows the disparity map obtained by using adjusted windows based on the edges shown in the lower left figure.

It is clear that the sharp height discontinuities are better preserved. Often the height of the pixels between two concentric pentagons is declared invalid: those pixels are almost all part of an occlusion zone.

4. CONCLUSION

In this paper, we have studied the problem of depth discontinuities in correlation based stereo reconstruction. A brief discussion is given on the fundamental problem of fixed size correlation windows and the violation of the assumption of constant depth within the window at object borders which affects the performance. We have given a systematic evualtion of the performance of fixed size and adaptive correlation windows, discussing shortcomings of both techniques. Whereas adaptive windows generally perform better at generating a reliable DSM, the performance depends strongly on the quality of the detected edges. In addition, it has been shown that distinct object features in relatively flat areas can pose problems for adaptive windows.

5. REFERENCES

1. Scharstein, D. and Szeliski, R. (2002). *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. International Journal of Computer Vision, 47(1-3):7–42.

2. Kostkova, J. (2002). *Stereoscopic Matching: Problems and Solutions*. Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University, Prague. PhD Thesis.

3. Cord, M., Paparoditis, N. and Jordan, M. (1998). *Dense, reliable, and depth discontinuity preserving dem computation from hrv urban stereopairs*. International archives of photogrammetry, remote sensing, XXXII(2):57–63.

4. Kanade, T. and Okutomi, M., 1994. *A stereo matching algorithm with an adaptive window: theory and experiment*. IEEE Transactions on Pattern Analysis and Machine Intelligence 16(9), pp. 920–932.

5. Fusiello, A., Roberto, V. and Trucco, E., 2000. *Symmetric stereo with multiple windowing*. International Journal of Pattern Recognition and Artificial Intelligence 14(8), pp. 1053–1066.

6. Tao, H., Sawhney, H. and Kumar, R., 2001. *A global matching framework for stereo computation*. Proceedings International Conference on Computer Vision, Vancouver, Canada, pp. 532–539.