A hierarchical block matching stereo algorithm based on cepstral analysis for remote sensing

Juan Manuel Ramírez, IEEE, Senior Member, Otto Paz-Luna, Heber Salazar

***Abstract*— This paper presents a stereo vision system based on a hierarchical matching process using cepstral analysis in the disparity detection process. The use of the Cepstrum transform as a measure of matching support allows the automatic decision of the block size during the analysis of the stereo images. A fast Hartley transform algorithm is incorporated for execution time improvement. The automatic determination of local similarity between two data sets or collections of pixels is fundamental in a depth-from-stereo algorithm for numerical evaluation of disparities. This measurement process is embedded within a coarse to fine stereopsis algorithm, providing an initial depth map with the depth information encoded as gray levels. The proposed algorithm has been tested using random dot stereograms and remote sensing imagery stereo pairs.**

***Index Terms*—cepstrum, images, matching, stereo.**

I. INTRODUCTION

The correspondence problem in stereo vision concerns the matching of points or other primitives, such as edges or regions in a stereo pair. Many different stereo systems for range determination using binocular stereo have been developed. The different approaches can be compared by considering their camera modeling, feature acquisition and matching techniques. Image matching is clearly dependent on the choice of feature primitives. When the elements to be matched are low-level and dense, such as the image intensity at each pixel in a neighborhood, the matching strategy is called an area-based process [1-3], while for sparse and usually more abstract and high level features such as edges or zero-crossings, the process is referred as feature-based [4-6]. Several matching techniques have been proposed including correlation in the space domain [3], phase correlation [8], or windowed Fourier phase profiles [7], with some approaches based on dynamic programming [5]. In this work an area-based technique using the Cepstrum transform is embedded in the depth-from-stereo algorithm. A stereo-motion model with a matching technique based on Cepstrum was developed to be used with a sequence of nine images with small disparities between consecutive images [9,10]. The Cepstrum transform was then proved to be very noise tolerant and accurate in the matching procedure. In this work, this concept is extended to be used with only one pair of stereo images. The range image is refined using a median filter to remove possible outliers, and a cubic interpolation as the last step.

II. DISPARITY DETECTION BY CEPSTRUM TRANSFORM

Cepstrum transformation was initially developed in seismic and acoustic signal processing for echo detection. In this work, cepstrum transformation is used for finding disparity between corresponding areas in every step of the depth-from-stereo algorithm described. The power cepstral transformation is defined as:

$$P\{i(x,y)\} = \left| \Im\left\{ \ln \left| \Im\{i(x,y)\} \right|^2 \right\} \right|^2 \qquad (1)$$

where $\Im$ is the notation for the Fourier transform and i(x,y) is the given image function. Let us consider a composed image formed by adding the two corresponding blocks of the stereo images which are supposed to have a small translational difference given by x_0 :

$$i(x,y) = w(x,y) + w(x - x_0, y) \qquad (2)$$

The Fourier transform of i(x,y) is given by:

$$\Im\{i(x,y)\} = W(u,v) + W(u,v)e^{-j2\pi u x_0} \qquad (3)$$

where u and v represent spatial frequencies and W(u,v) is the Fourier transform of w(x,y) ; the Power spectrum is obtained as:

ISBN-13 978-968-7938-03-5

$$\left|\Im\{i(x,y)\}\right|^2 = \left|W(u,v)\right|^2 \left|1+e^{-jux_0}\right|^2$$
$$= \left|W(u,v)\right|^2 \left|1+\cos ux_0 - j\sin ux_0\right|^2$$
$$= \left|W(u,v)\right|^2 \left((1+\cos ux_0)^2 + sin^2 ux_0\right)$$
$$= \left|W(u,v)\right|^2 (2+2\cos ux_0) \quad (4)$$

When the logarithm function is applied to the power spectrum of i(x,y), the multiplicative terms are separated as:

$$\ln\left|\Im\{i(x,y)\}\right|^2 = \ln\left|W(u,v)\right|^2 + \ln(2+2\cos ux_0) \quad (5)$$

Using the logarithm series expansion the second term can be expanded into a convergent infinite series, the application of the power spectrum according to the definition of power spectral transformation yields:

$$\mathrm{P}\{i(x,y)\} = \mathrm{P}\{w(x,y)\} + A\delta(x,y) + B\delta(x \pm x_0, y) + C\delta(x \pm 2x_0, y) + \ldots \quad (6)$$

This derivation considers only horizontal displacements because according to the concept of stereo vision, only horizontal disparities between corresponding blocks are expected. After removing the power cepstrum of w(x,y), the translational difference between the two corresponding windows can be obtained by inspecting the remaining impulse train. In other words, the translational difference corresponds to the distance between the origin and the location of the first peak in the cepstrum plane. A fast recursive algorithm for computing the Hartley transform has been incorporated [12]. Hartley transform improves the excution time by reducing the number of multiplications to $M = \frac{1}{2}N\log_2 N - \frac{3}{2}N + 2$ real operations, providing very good results.

III. HIERARCHICAL SEARCH OF DISPARITIES

In a non-convergent stereo vision model, the depth information is directly derived from the disparity between corresponding points. Random dots stereograms have been widely used to test the performance of stereo vision systems. In a random dot stereogram, the disparities are artificially generated by introducing horizontal displacements in a computer generated image with a pattern of random dots. When a stereoscope is used to analyze these images the three-dimensional effect can be perceived. The goal of any automated depth-from-stereo algorithm, is to find the disparities between corresponding points in the stereo pair using some matching technique. Establishing these corresponding points is, however, the most critical step in the estimation of depth from stereo. In the ideal case, the objective is to find disparities between every individual pixel in both images, however, it is obvious that the intensity value of a single pixel is not enough for finding corresponding points, so a collection of pixels in a neighborhood has to be used to match. The approach used in this work is the technique of successive refinement of parallax based on hierarchical coarse to fine resolution steps [7]. The algorithm starts with a partition of the stereo images in windows of size 32x32. The two dimensional cepstrum is used to obtained disparity between corresponding windows in both images. A second partition by two is performed in each window in order to obtain a correction value of the disparity in the corresponding sub-window. This quadrant subdivision for coarse to fine search of disparities is represented in figure 1.

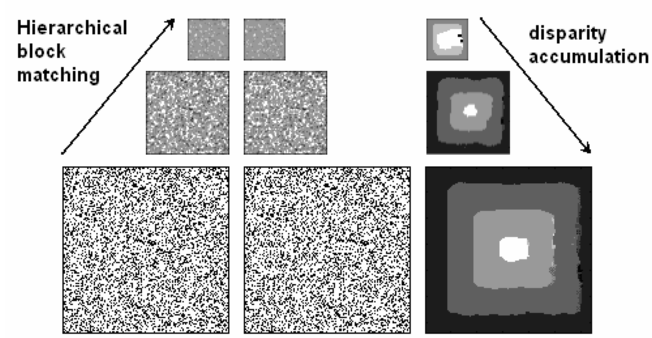


Figure 1. Hierarchical search for disparities.

The coarse disparity assigned initially to the first window is modified in every step to determine the most detailed disparity information. This procedure continues until a window size where no peak can be detected through the cepstrum, is reached. In the implementation described in this work this point was reached usually at a window size of 4x4 pixels. A median filter was applied in order to eliminate outliers, and a cubic interpolation provided a final refining.

IV. RESULTS

A series of synthesized as well as natural stereo images were tested to evaluate the depth-from-stereo algorithm described in this paper. In the first case, a random dot stereogram whose underlying structure is a four layer pyramid (fig.2) was tested yielding good results. If a maximum error of ±1 pixel is tolerated, the percentage of correct values obtained in the case of random dot stereograms was 90% in average. Figure 3 shows the result.

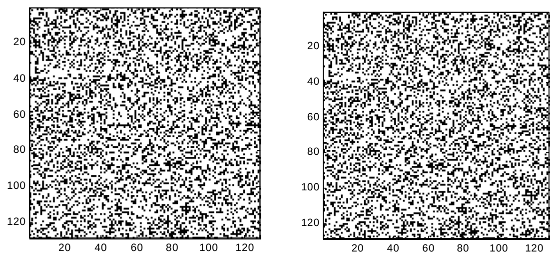


Figure 2. Random dot stereogram; three layer pyramid.

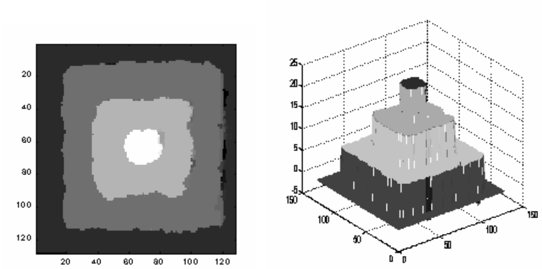


Fig. 3 Range map encoded in gray levels, and isometric view.

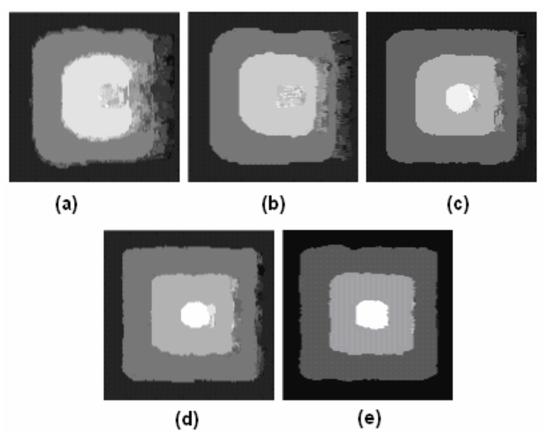


Figure 4. (a-e) Intermediate results in the hierarchical algorithm in five stages.

This number can be estimated in the case of random dot stereograms because they are computer generated and we know in advance the correct values expected. This is different when real stereo imagery was attempted. Figures 4 and 5 show the range map of a surface obtained from a stereo pair of geographical images. A median filter was applied in order to eliminate outliers, and a cubic interpolation provided a final refining.

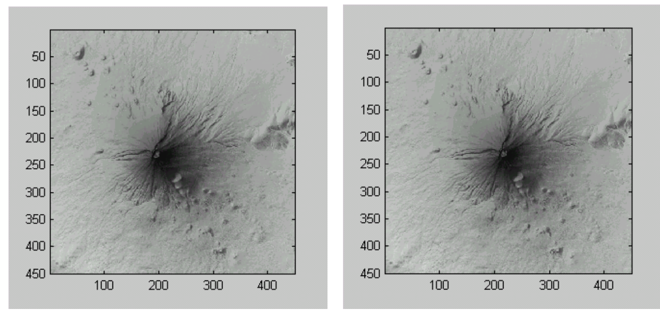
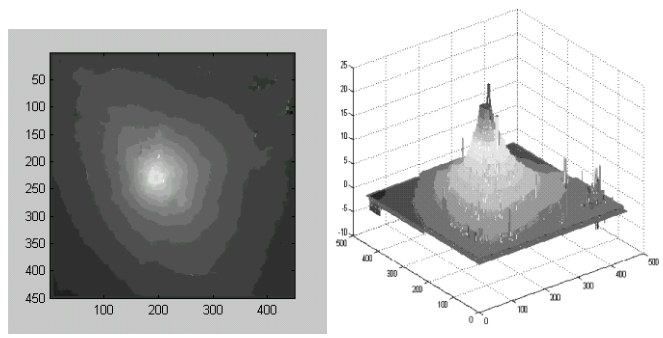


Fig 5. Stereo pair of geographical images



(a) (b)

Fig. 6. (a) Range map and (b) isometric view of the stereo images in fig. 5.

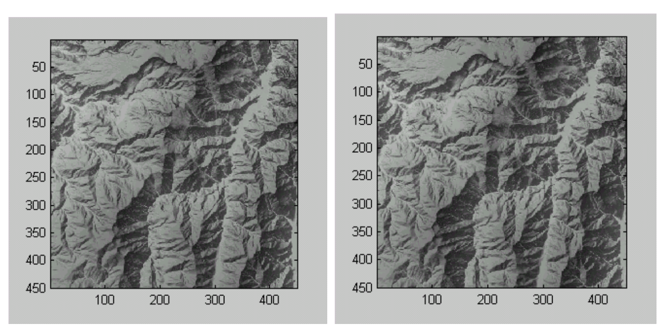
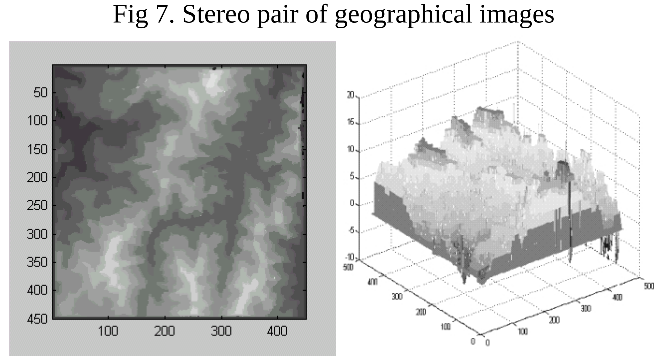


Fig 7. Stereo pair of geographical images



ISBN-13 978-968-7938-03-5

(a) (b)

Fig. 8. (a) Range map and (b) isometric view of the stereo images in fig. 7.

V. CONCLUSIONS

This paper presented a depth-from-stereo algorithm based on disparity detection by cepstrum transform as the required matching procedure. In this context, cepstrum transform represents a good alternative in terms of robustness, accuracy, and execution time, to find disparity between corresponding blocks. This technique is embedded in a coarse to fine search strategy, which provide the control between the various scales at which the operator is to be applied. The performance of the matching procedure was tested using computer generated stereograms formed of random dots as well as natural stereo images with good results. Experimentation with the methodology presented for quantitative analysis of surfaces is currently in progress.

REFERENCES

[1] L.D. Stefano, M. Marchionni, S.Mattoccia, G. Neri, “A fast area-based stereo matching algorithm” 15th International Conference on Vision Interface, Calgary, Canada, May, 2002.

[2] J. Sun, H.Y. Shum, and N.N. Zheng, “Stereo matching using belief propagation”, ECCV 2002, pp. 510-524, 2002.

[3] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts”, Pattern Analysis and Machine Intelligence, vol. 23, No. 11, pp. 1222-1239, 2001.

[4] Qin, I. Li, S. Tian, “A new image matching algorithm with modified greedy for remote sensing imagery”, ISPRS Workshop on Service and Application of Spatial Data Infrastructure, Hangshou, China, October, 2006.

[5] M. Gong, Yee-Hong Yang, “Near real-time reliable stereo matching using programmable graphics hardware”, IEEE Computer Vision and Pattern Recognition Conference”, 2005.

[6] W. Hoff, and N. Ahuja, ¨Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection,¨IEEE Transactions on PAMI, vol. 11, pp. 121-136, February 1989.

[7] J. Weng, ¨A theory of image matching,¨ Proc. Third International Conf. on Computer Vision, pp. 200-209, Osaka, Japan, December 1990.

[8] A.D. Jepson, and M.R.M. Jenkins, ¨The fast computation of disparity from phase differences,¨IEEE Trans. on PAMI, pp. 398-403, 1989.

[9] D.J. Lee, S. Mitra, and T.F. Krile, ¨Dense depth maps from 2-D cepstrum matching of image sequences,¨International Workshop on Robust Computer Vision, Seattle, Washington, October, 1990.

[10] Manuel Ramirez, Pedro Ramos, Pilar Gómez Gil, Heber Salazar “Obtention of Surface Topography using a Stereo System Based on Cepstrum Transform”, 6th World Multiconference on Systemics, Cybernetics and Informatics, Proceedings Vol. III, pp. 135-138, Orlando, Florida, USA, July 2002.

[11] John C. Russ, The Image Processing Handbook, CRC Press Inc., Raleigh, North Carolina, 1995.

[12] Pinneng Yu, Heping Hua, “A new fast recursive algorithm for computing discret Hartley transform and its implementation”, Signal Processing, Vol. 81, No. 10, pp. 2235-2241, October 2001.

ISBN-13 978-968-7938-03-5