Search Space Reduction in the Edge Based Stereo Matching by Context of Disparity Gradient Limit

Payman Moallem, Karim Faez
E.E. Department, Amirkabir University of Technology, Hafez Avenue, Tehran 15914, Iran
Email : {m7523903, kfaez}@cic.aku.ac.ir

Abstract

Traditionally, finding the corresponding points has considered to be the most difficult part of stereo matching algorithms. Usually, the correspondence for a feature point in the first image is obtained by searching in a predefined region of the second image, based on the epipolar line and the maximum disparity. The reduction of the search region can increase the performance of the matching process, in the context of the execution time and the accuracy. We proposed a new matching strategy to reduce the search space for the edge-based stereo correspondence algorithms. Considering the maximum of the disparity gradient in the real scene, we formulated the relation between the maximum search space in the second images with respect to the relative displacement of the continuous edges (as the feature points) in the successive scan lines of the first images. Then we developed some very fast stereo matching algorithms, based on the non-horizontal edges as feature points, and the normalized cross correlation criteria (NCC) with different sizes of the matching block (as the similarity measures). We applied these new algorithms on the Renault stereo image and compared the result with those of a traditional matching algorithm (20 pixels search regions and NCC with size of 15×15). The speed up of these new algorithms is between 2.8 to 13.8 and the percentage of errors is between 0.5 to 5.4.

1. Introduction

Stereo vision refers to the ability to infer information on the 3D structures and the distances of a scene from at least two images (left and right), taken from different viewpoints. The most essential problems in a stereo system is correspondence [1]. The correspondence consists of determining which item in the left image corresponds to which item in the right image. It is usually not a good practice to try to find the corresponding points for all pixels. For example, a point in a uniform region in one image may correspond to many points in the corresponding region in the other image. Feature points or matching primitives are selected so those unambiguous matches could be resulted [2]. Depth in the stereo system is related to the inverse of the disparity, which is the difference between the position of the corresponding points in the left and right images.

The correspondence, which is the most complex stage in stereo algorithms, is classified into three categories: area-based (or correlation based), feature-based, and pixel-based methods [3]. In the area-based methods, the elements to be matched are image windows of fixed or variable sizes, and the similarity criterion can be a measure of the correlation between the windows in the two images. Feature-based methods match feature points in the left image to those in the right image, so these methods restrict the search for the correspondence to a sparse set of feature points. Feature-based methods use a measure of distance between feature descriptors. In these methods, interpolation is a necessity for computing the depth for non-feature points. Pixel-based methods perform the matching at each pixel, using the measurements at a single pixel. In these methods, a dense disparity map can be obtained.

Generally in area-based or feature-based approaches, to find the correspondence of each feature point in one image, the whole of the other image must be examined as the search region. The reduction of the search region can reduce the complexity of matching and increase the accuracy. Most stereo matching algorithms narrow down the number of possible matches for each feature by enforcing suitable constraints on the feasible matches and proper matching strategies, which are discussed in the next subsections.

1.1 Matching constraints

In the stereo correspondence, some matching constraints are generated based on the underlying physical principles of the world imaging and steropsis. Some of the common constraints incorporated in stereo algorithms include:

1- Epipolar constraint: Corresponding points must lie on the corresponding epipolar lines. Epipolar lines are

defined by the intersection of the epipolar planes and the image plane [1].
2- Disparity limit constraint: Regarding to the maximum and minimum of the depth and geometry of stereo system, the maximum disparity range can be estimated [4].
3- Ordering constraint: A left-right ordering relationship between two pixels in the left image should have the same ordering correspondence pixels as in the right image [5].
4- Uniqueness constraint: Correspondence should be unique. In other words, each feature can have at most one match [5].
5- Figural continuity constraint: Disparity along an edge contour changes smoothly, so there should be no disparity discontinuities along a contour [6].
6- Disparity gradient limit: The maximum gradient of the disparity is limited, so the disparity gradient between matched primitives is restricted [7].

The epipolar constraints reduce the search region from the whole of the second image (two-dimensional space) to the one-dimensional (epipolar) line [1]. Moreover, disparity limit narrows down the one-dimensional search from the full search to the restricted space [4]. In some special case, ordering constraint can also reduce the search region [5]. The uniqueness constraintl, the figural continuity constraint [6] and the disparity gradient limit are used to detect false matches and to correct them [8].

1.2. Matching strategy

In addition to the constraints and consistency checks, several control strategies have been proposed by many to further reduce the search region and the ambiguity, and to enhance stereo matching performances. Some of the common and popular ones are:
1- Coarse to fine strategy: In coarse to fine strategy, information obtained at a coarse scale is used to guide and limit the search for the matching of finer scale primitives or feature points. In this approach, the initial matching begins at a coarse scale where the feature density is low due to the scale change. This reduction in feature density reduces the search space, which in turn makes the matching easier and faster, but not necessarily more accurate, because the localization at coarse scale is less accurate. Such multiscale strategy can be used with scale specific primitive representation and can be incorporated into the area-based, the feature based, and the pixel based correspondence techniques. These techniques are classified in two categories: Initial estimation at lowest scale called multiresolution, and initial estimation at finest scale. In the multiresolution strategies, the initial matching begins at the lowest scale and then extended to finer ones [2]. This approach can be applied in both temporal space (graylevel) and frequency space. Usage of the hierarchical Gaussian basis functions3 (in temporal space) and complex multiwavelets [9] (in both temporal and frequency space) are common techniques in stereo vision. In the finest scale method, the disparity estimation of large blocks in the fine level is used as initial estimation of disparity for the entirety of blocks, then the disparity is calculated within each block accuratelyl [5].
2- Structural and hierarchical multi primitives: In hierarchical and structural stereo approaches, semantically rich primitive representations like regions, lines and edge segments are derived from an image and matched. Relational properties are used in addition to the spectral properties in the structural methods to reduce the search space and to disambiguate the stereo matching [2].

Non of these strategies uses the stereo constraints to reduce the search region. But we will introduce the utilization of the disparity gradient limit to do so. In next section, we briefly discuss the disparity gradient concept and its application in some stereo methods. Then we will formulate the reduction of the search region in the stereo matching of the connected non-horizontal edges in successive scan lines (for example edge segments). Then we will develop very fast stereo correspondence strategy based on this new formulization for the connected non-horizontal edges as primitives. This strategy can be used with other primitives and methods.

2. Search space and disparity gradient limit

Fig 1 shows the cameras geometry for stereo vision where the cameras optical axes are parallel to each other and perpendicular to the baseline connecting the two cameras L and R. For a point P(X,Y,Z) in 3D scene, its projections onto the left image and the right image are $p^l(x^l,y^l)$ and $p^r(x^r,y^r)$. Because of this simple camera geometry, $y^l=y^r$ and the disparity d is inversely proportional to the depth Z:

$$d = x^l - x^r = bf / Z \qquad (1)$$

Where f is the focal length of the camera lens and b is the separation of two cameras or baseline.

Given two points $P_1(X_1,Y_1,Z_1)$ and $P_2(X_2,Y_2,Z_2)$, their disparity gradient δd can be defined as the difference in disparities divided by the *cyclopean separation,* where *cyclopean separation* is the average distance between (p^l_1,p^l_2) and (p^r_1,p^r_2). Suppose a virtual camera is placed in the middle of the cameras L and R, i.e., at position of the origin. Therefore, we have:

$$\begin{gathered} d_2 = x_2^l - x_2^r, \quad d_1 = x_1^l - x_1^r \\ p_2^c = \frac{p_2^l + p_2^r}{2}, p_1^c = \frac{p_1^l + p_1^r}{2} \end{gathered} \qquad (2)$$

With these definitions, we can define δd as:

$$\delta d = \left| d_2 - d_1 \right| / \left\| p_2^c - p_1^c \right\| \qquad (3)$$

Where ||.|| denotes the vector norm. Note that from its definition, δd is always a nonnegative number.

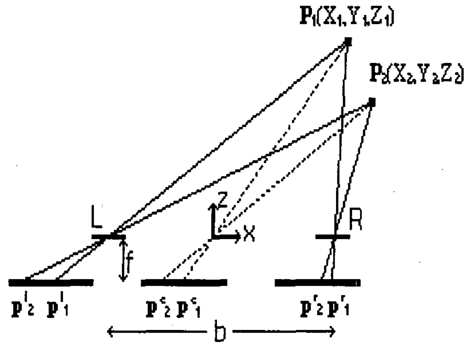


Figure 1. Defining disparity gradient in the stereo system with parallel cameras

The value of δd can be used to define various stereo-matching constraints, which are explained before. A brief summary follows [11]:

δd >2- Violation of non-reversal order constraint.
δd =2- Violation of uniqueness constraint.
δd <1.1 or 1.2- Disparity gradient limit.
δd <<1- Figural continuity constraint.

2.1 Disparity gradient in stereo correspondence

Bult and Julesz [7] provided evidence supporting the claim that, for binocular fusion of random dot stereograms by the human visual system, the disparity gradient must not exceed the unity. Pollard, Mayhew and Frisby [8] suggested that for most natural scene surfaces, including jagged one, the disparity gradient between correct matches is usually less than unity (<1). For discarding ambiguity in the correspondence problem, they impose a disparity gradient limiting constraint among the candidate matches. Pollard et al. [10] pointed out the intrinsic relationship between the disparity gradient, surface orientation, and the depth in 3D scenes. Li and Hu [11] used the disparity gradient as the basis for the unified cooperative stereo matching. They selected some families of neighborhood support functions based on the disparity gradient. In next subsection, we discuss about the restriction of the search region for matching edge segments, based on the disparity gradient limit, and then propose some very fast stereo matching algorithms.

2.2 Relation Between δd and Search Region

Now, we can formulate the relation between δd and the search space in stereo correspondence. If we substitute the relation (2) into equation (3), we have:

$$\delta d = |d_2 - d_1| / \|p_2^c - p_1^c\| = \frac{2|(x_2^l - x_2^r) - (x_1^l - x_1^r)|}{\|(p_2^l + p_2^r) - (p_1^l + p_1^r)\|} = \frac{2|(x_2^l - x_1^l) - (x_2^r - x_1^r)|}{\|(p_2^l - p_1^l) + (p_2^r - p_1^r)\|} \quad (4)$$

We can define Δx_l and Δx_r as the difference between position of p_2 and p_1 in the left and right images. In the other hand, we have:

$$\Delta x_l = x_2^l - x_1^l, \quad \Delta x_r = x_2^r - x_1^r \quad (5)$$

So δd can be changed to:

$$\delta d = \frac{2|\Delta x_l - \Delta x_r|}{\sqrt{(\Delta x_l + \Delta x_r)^2 + (\Delta y_l + \Delta y_r)^2}} \quad (6)$$

Suppose we want to match the continuous non-horizontal edges in successive scan lines; in this case, we have $\Delta y_l = \Delta y_r = 1$, so we have:

$$\delta d = \frac{2|\Delta x_l - \Delta x_r|}{\sqrt{(\Delta x_l + \Delta x_r)^2 + (2)^2}} \quad (7)$$

In a typical stereo system like human vision, the reasonable limit of disparity gradient is about 1.1 or 1.2[11], So we assume that δd<1.2. By substituting in equation (7), we can solve the resulting non-equality of (8).

$$0.64\Delta x_l^2 + 0.64\Delta x_r^2 - 2.72\Delta x_l \Delta x_r - 1.44 \leq 0 \quad (8)$$

Suppose the feature points in the left image are continuous non-horizontal edges in subsequent scan lines. If we know the value of Δx_l in the left image, we can restrict Δx_r in the right image. Five cases are investigated: Δx_l=0, Δx_l=±1 and Δx_l=±2. For example in case of Δx_l = 0 we have, $-1.5 \leq \Delta x_r \leq +1.5$. This means that for two continuous edge points in the subsequent scan line of the left image that their position is the same in x direction, by context of δd <1.2, the maximum allowable range of Δx_r will be 1.5 pixels in the right image.

Table (1) shows the relationship between Δx_l and Δx_r for the above five cases. Since the unit of pixels can not have fractional part, so we rounded Δx_r and then we have extended the range of δd. Third column of the table (1) shows the extended value of δd in both negative and positive limits. In all cases, the condition of $\delta d < 1.2$ is met and even in some cases (negative limit for $\Delta x_l = -1$ and positive limit for $\Delta x_l = +1$), the maximum allowable range of δd for non-reversal reorder constraint is achieved. In the next section, we propose some fast area-based algorithms based on this restriction in the search region. However, emphasis is not on the algorithms themselves, but on the effect of reduction the search region, resulting in lower execution time.

Table 1. The relation between Δx_l, Δx_r and extended range of δd

Value of Δx_l	Rounded range of Δx_r	Extended range of δd
-2	{-9,-8,-7,-6,-5,-4,-3,-2,-1,0}	$1.41 < \delta d < +1.25$
-1	{-5,-4,-3,-2,-1,0,+1}	$-2 < \delta d < +1.26$
0	{-2,-1,0,+1,+2}	$1.41 < \delta d + 1.41$
+1	{-1,0,+1,+2,+3,+4,+5}	$-1.26 < \delta d < +2$
+2	{0,+1,+2,+3,+4,+5,+6,+7,+8,+9}	$1.25 < \delta d < +1.41$

3. Fast stereo algorithms

In the calibrated stereo system with parallel optical axes, the area based or the feature based correspondence consist of two stages: *feature point (or primitives) extraction* and *stereo matching* [1]. Most of fast stereo algorithms use low-level primitives like edges, those that do not require sophisticated semantic analysis in their extraction. Matching the horizontal edges in the stereo system with parallel optical axes is a problem [12], so some authors used non-horizontal edges as feature points [13]. In a correlation based framework, stereo matching for a pixel in a reference image (left) is obtained by searching in a predefined region of the second image (right). Most currently used fast stereo methods belong to the category of linear correlation methods, which include those based on sum of squared differences (SSD) and normalized cross correlation (NCC). In the stereo system that the left and the right cameras are different, NCC is preferable since it is invariant to linear brightness and contrast variations between the perfect matching windows. The value of NCC is between -1 and +1, and a larger value indicates more similarity between windows [14]. The window size in the correlation-based methods is very important. As the window size decreases, the discriminatory power of window based criterion is decreased and some local maximum in NCC could have been found in search region. Moreover, continually increasing the window size causes the performance to degrade because of occlusion regions and smoothing of disparity values across depth boundary [14].

We showed that the search region could be reduced via considering the maximum value of disparity gradient. If the disparity search range could be automatically reduced to an effective range (about 10 pixels) then several local maximum in NCC would stay out of the selection process and therefore, the disparities found would be correct, even if the size of the matching block is small [15]. Therefore, we can use NCC with window size of even 3×3 or 5×5 in restricted search region. Therefore, in the matching stage of connected non-horizontal edges, our algorithms have two phases for each connected edge.

Phase one: At the first point of the connected edges, we use the search region based on maximum value of disparity and NCC with window of 15×15.

Phase two: If we can find the correspondence of that point in the second image, we use the restricted search region based on table 1, and NCC with small size of window for other connected edge points. If we can not find the correspondence, the algorithm finds the first point of the next connected edge and then goes to phase one.

Table 2. The implementation results of our algorithms in Renault stereo image

Algorithm	Threshold on NCC	Matched points (percentage)	Rejected points (percentage)	Error points (percentage)	Speed up
DGRSS	0.8	3706(98.7%)	49(1.4%)	19(0.5%)	2.8
DGRS3	0.7	3642(97.0%)	113(3.0%)	195(5.4%)	13.8
	0.8	3551(94.6%)	204(5.4%)	129(3.6%)	10.2
	0.9	3264(86.9%)	491(13.1%)	74(2.3%)	6
DGRS5	0.7	3640(96.9%)	115(3.1%)	50(1.4%)	10.6
	0.8	3560(94.8%)	195(5.2%)	39(1.1%)	8.8
	0.9	3284(87.5%)	471(12.5%)	18(0.6%)	5.5
DGRSA	0.7	3648(97.2%)	107(2.8%)	51(1.4%)	10.5
	0.8	3581(95.4%)	174(4.6%)	38(1.1%)	8.6
	0.9	3309(88.1%)	446(11.9%)	15(0.5%)	5.3

Here, we propose four algorithms. In the first one, we did not reduce the window size of NCC in phase two (we use 15×15). We called it DGRSS. It means Disparity Gradient based Restricted Search region and NCC with Sufficient size (15×15 in our implementation). In the second one called DGRS3, the window size of NCC in phase two was 3×3. In the third one, the window size of NCC in phase two was 5×5, so we called it DGRS5.

Finally in the last one, the window size of NCC was selected adaptively based on the value of Δx_l (DGRSA). Here the window size is 3×3 for $\Delta x_l = 0$, 5×5 for $|\Delta x_l| = 1$ and 7×7 for $|\Delta x_l| = 2$. For comparing the results of the execution time and accuracy, we choose a popular method as reference. This area-based method which is similar to ours, consists of non-horizontal edge points as primitives. As a similarity criterion, NCC with window size of 15×15 is selected and a search region is considered based on the maximum value of disparity. We implemented and tested the reference algorithm and ours on the Renault stereo images [16]. These images are 256×256 with 256 graylevels, and the feature point extraction stage found 3755 non-horizontal edges. For our algorithms, three value of threshold on NCC criterion is tested (0.7, 0.8 and 0.9). In this test image, the maximum range of the disparity was ±10 pixels. The last column in table 1, is the speed up with respect to the reference algorithm. The results of the matching are shown in table 2, which tabulates the number of the *matched points*, the *rejected points*, the *error points* and finally the *speed up*. The values in error points and speed up columns are with respect to the reference algorithm. Figure 2 shows the disparity maps obtained by DGRSS and the reference algorithm. There is no noticeable difference between these disparity maps.

4. Discussion

Our proposed methods have noticeable speed up and acceptable error rate with respect to the reference algorithm. The threshold value on NCC, the reduction of the search space and the lower size of NCC are the parameters that affect the speed up and the matching performances. In next subsections, these parameters and their effects are investigated in detail.

4.1 Threshold value on NCC

Increasing the value of the threshold on NCC means that the corresponding templates have higher similarity. In other words, increasing this parameter decreases the matching probability and then, the matched points and the error in matching are also decreased. This fact is obvious from table (2). In each of our proposed algorithms, increasing this parameter has noticeable effect on decreasing the speed up. For example in DGRS3, the speed up decreases to 50% when the threshold value is changed from 0.7 to 0.9. For each set of the connected non-horizontal edges in successive scan lines, our proposed algorithms are implemented in two different phases. The search space for the first point of connected set in phase one, is sufficiently large, based on the maximum disparity. But the search space in phase two which is executed for the other points of the connected non-horizontal edges, is reduced based on the table (1). Then the correspondence in phase two is obtained faster than phase one. Increasing the value of the threshold on NCC results in increasing the number of the points that match in the phase one, and so increases the execution time. The phase two of our proposed algorithms is executed subsequently. It means that n-th point, in a connected set of the non-horizontal edges, could be matched in the phase two if all of the previous points 2nd to (n-1)-th are matched in this phase.

4.2 Reduction of the search space

In DGRSS algorithm, the reduced search space based on the table (1) is used while the size of NCC is selected the same as the reference algorithm. So this algorithm can be used to investigate the effect of the reducing the search space on the speed up and on the matching performances, independent from the reducing the size of NCC. Considering the table (1), the rejection of DGRSS is only 1.4% and the error in matching is only 0.5%. These mean that, this search strategy which is using only the reduction of the search space, is as powerful as the reference algorithm. Moreover, DGRSS is executed 2.8 times faster than the reference algorithm.

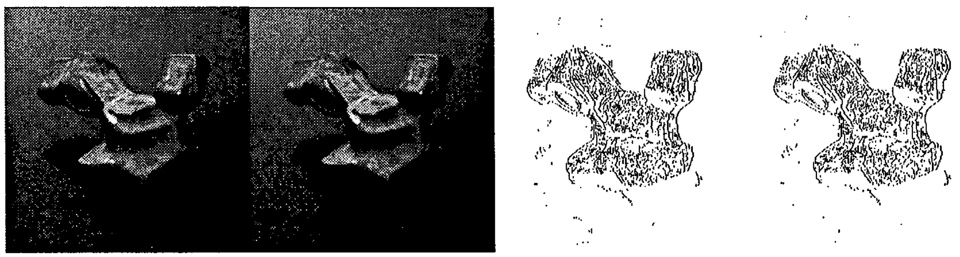
4.3 Reduction of the size of NCC window

In the DGRS3, DGRS5, and DGRSA algorithms, both the search space and the size of NCC are reduced. Considering the threshold value of 0.8 on NCC, the speed up of these algorithms are 10.2, 8.8 and 8.6 respectively. The window size of NCC in DGRS3 is 3×3, which is the smallest size compared with the other proposed algorithms. Then measuring the similarity in DGRS3 is fast and of course its accuracy is a little poor, so this algorithm is executed very fast and of course has a higher error percentage (5.4%) compared to the others (1.1%).

The only difference between DGRS5 and DGRSA is in the case of $\Delta x_l = \pm 2$. The speed up for DGRS5 and for DGRSA is very close together (8.8 and 8.6 respectively), so the case of $\Delta x_l = \pm 2$ occurs rarely. DGRSA has the highest percentage of the matched points (95.4%) and the least percentage of the error in the matching (38 points). However this method has higher execution time than DGRS5, but its performance is always better than those of DGRS5 and DGRS3.

5. Conclusion

In this paper, we have presented the reduction of the search region space in an edge based stereo correspondence by using the context of the maximum disparity gradient. For the reduction of the search region in an area-based stereo correspondence, the continuous non-horizontal edges in the successive scan lines of the first image have been selected as the feature points. Considering the maximum disparity gradient in the real scene, we could formulate the relationship between the search region in the second image and the relative displacement of the feature points in the first image. Then we have proposed four area-based stereo matching algorithms based on this new restriction and the NCC as the criterion similarity measure. The reduction of the search region can result in the reduction of the size of NCC window, so both of the search space and the size of NCC are reduced in our proposed algorithms. Then our algorithms could be executed very fast. Regarding the area-based algorithm with non-restricted search region and NCC with the size of 15×15, our algorithms have 2.8 to 13.8 times faster execution times and the percentages of the errors in matching are only between 0.5 to 5.4. From the point of view of the speed up, DGRS3 is the fastest method but has a higher error rate. If the matching performance is more important with respect to the speed up, DGRSS is the best. This method is faster and more accurate. When both of the speed up and performance are important, DGRSA would be better. For reduction of the search region, our strategy could be used not only in the area-based methods, but also in the other stereo correspondence approaches.



(a) (b) (c) (d)

Figure 1. Renault stereo image and the disparity map obtained by the reference algorithm and DGRSS. a)The left image b)The right image c)The disparity map by the reference algorithm and d)The disparity map by DGRSS. The disparity map was obtained only for the non-horizontal edges. The larger value of the disparity is shown darker.

6. References

[1] 1. E. Trucco, A. Verri, *Introductory Techniques for 3D Computer Vision*, Prentice Hall, 1998.

[2] U. R. Dhond, J. K. Aggarwal, "Structure from stereo - A reviev", *IEEE Trans. on SMC*, vol 19, No 6, Nov-Dec 1989.

[3] G. Q. Wei, W. Brauer, G. Hirzinger, "Intensity- and Gradient-Based Stereo Matching Using Hierarchical Gaussian Basis Functions", *IEEE Trans. On PAMI*, vol 20, No 11, Nov 1998, pp. 1143-1151.

[4] H. Jahn, "Parallel Epipolar Stereo Matching", *IEEE Int. Conf. on Pattern Recognition (ICPR2000)*, 2000, pp. 402-405.

[5] S. B. Marapane, M.M. Trivedi, "Multi-Primitive Hierarchical (MPH) Stereo Analysis", *IEEE Trans. on PAMI*, vol 16, No 3, March 1994.

[6] R. Mohan, G. Medioni, R. Nevatia, "Stereo Error Detecion, Correction and Evaluation", *IEEE Tans. on PAMI*, vol 11, No 2, Feb. 1989.

[7] P. Burt, B. Julesz, "A Disparity Gradient Limit for Binocular Fusion", *Science*, vol 208, 1980, pp. 615-617.

[8] S. B. Pollard, J.E.W. Mayhew, J.P. Frisby, "PMF: A Stereo Correspondence Alogorithm using Disparity Gradient Limit", *Perception*, 14, 1981, pp. 449-470.

[9] J. Magarely, A. Dick, "Multiresolution Stereo Image Matching using Complex Wavelet", *IEEE Int. Conf. on Pattern Recognition (ICPR98)*, vol 1, 1998, pp. 4-7.

[10] S. Pollard, J. Porril, J. Mayhew, J. Frisby, "Disparity gradient, Lipschitz continuity and computing binocular correspondence", in O. D. Faugeras and G. Giralt, Eds. *Proc. 3rd Int. Symp. Robot Research*, 1986, pp. 19-26.

[11] Z. Li, G. Hu, "Analysis of Disparity Gradient Based Cooperative Stereo", *IEEE Trans. on Image Processing*, 5, No 11, Nov. 1996, pp. 1493-1506.

[12] S. Tanaka, A. C. Kak, "Rule-Based Approach to Binocular Stereopsis", *Analysis and Interpretation of Range Image*, R. C Jain, A. K Jain, Springer Verlag, 1990.

[13] Y. Shirai, *Three dimensional Computer Vision*, Springer Verlag, 1987.

[14] D. N. Bhat, S. K. Nayar, "Ordinal Measure for Image Correspondence", *IEEE Trans. on PAMI*, vol 20, No 4, April 1998.

[15] M. Torres, A. Guesalga, "Confidence Factor and Adaptive Disparity Range for Reliable 3D Reconstruction", *Computer Intelligence for Modeling, Control & Automation*, M. Mohammadian, IOS Press, 1999.

[16] CMU stereo Image database at http://www.vasc.ri.cmu.edu/idb/html/stereo/index.html.