Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006 (pp243-247)

Accurately Tuned Search Space in a Feature Based Stereo Matching

PAYMAN MOALLEM[I] MOHSEN ASHOURIAN[II]
BEHZAD MIRZAEIAN[I] MOHAMAD ATAEI[I]
[I] Electronic Engineering Department, Faculty of Engineering
[II] Information Technology Department, Faculty of Engineering
Isfahan University
Hezarjarib Avenue, Isfahan
IRAN

Abstract: Computations in a feature based stereo matching which is basically used for depth extraction are generally very high. These computations essentially include feature extraction and matching which feature matching is usually higher. For a feature-based stereo matching, we accurately tune the search space based on some stereo imaging parameters like the focal length with pixels scale, the displacement of features and the maximum disparity. We show that results of previous matches can be used to narrow down the search space to find current match. We use directional derivative of disparity as a temporary concept to tune the search space accurately. Then we develop a fast feature based stereo algorithm based on the proposed search space tuning and non-horizontal thinned edge points as feature. The execution time of the proposed algorithm is lower than other methods. Moreover the matching rate is also higher.

Key-Words: feature based stereo matching, search spaces tuning

1 Introduction

The execution time is critical in some feature based stereo vision applications, like robot vision [1]. 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 constraints [2]. The tuning of the search region can increase the performance of the matching process, in the context of the execution time and the accuracy. Traditionally, the hierarchical multiresolution techniques using wavelet transform, as the fast methods are used to decrease the search space and therefore increase the processing speed [3].

In this paper, we accurately tune the search space for a typical feature based stereo matching and then obtain a fast feature based stereo matching algorithm based on non-horizontal thinned edge points as feature. We show that the result of previous matching can be used to adjust the search space. We use directional derivate of disparity as a temporary concept to tune the search space.

In the next section, we briefly discuss the concept of directional derivative of disparity [4]. In the section 3, we represent a relationship between directional derivative of disparity range and search space for feature matching based on the pdf[1] of directional derivative of disparity [5]. In the section 4, we tune the search space for feature points in the left image based on the position of the closest matched feature. Then we develop a feature based stereo matching algorithm includes feature extraction and feature matching based on the proposed search space. In the proposed algorithm, the geometric constraints of stereo system like focal length and image resolution and CCD[2] dimension have effects in search space as well as other parameters like distance between features and maximum disparity. Finally, we discuss about the implementation results on some public stereo images with known stereo imaging parameters.

2 Directional Derivative of Disparity

Figure 1 shows a basic stereo system 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 $P^i(X,Y,Z)$ in 3D scene, its projections onto the left image and the right image are $p^i_l(x_l,y_l)$ and $p^i_r(x_r,y_r)$. Considering this

[1] Probability Density Function
[2] Charged Couple Device

Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006 (pp243-247)

camera geometry, it can be shown that $y^i_l = y^i_r = y^i$ and the disparity d is inversely proportional to the depth Z. we have $d=bf/Z$, where f is the focal length of the camera lens and b is the separation of two cameras or baseline [2].

Given two points $\mathbf{P^1}$ and $\mathbf{P^2}$ in 3D scene, directional derivative of disparity or d_α can be defined as the difference in disparities divided by the cyclopean separation, where cyclopean separation is the average distance between (p^1_l, p^1_r) and (p^2_l, p^2_r) [4].

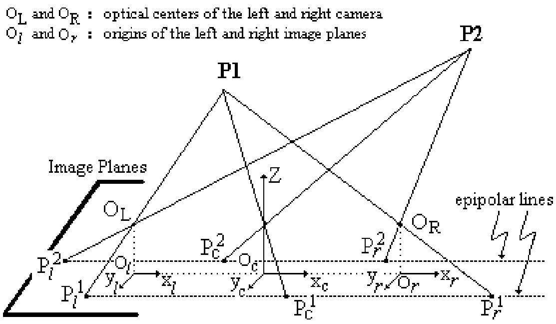


Figure 1 - Basic stereo imaging with parallel optical axes

We define the displacement of P^1 and P^2 in the left and the right images, Δx_l, Δx_r and Δy as:

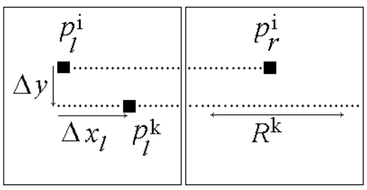
$$\Delta x_l = x_l^2 - x_l^1 \qquad \Delta x_r = x_r^2 - x_r^1$$
$$\Delta y = y^2 - y^1 \tag{1}$$

It has been already shown that d_α between P^1 and P^2 could be computed as [5]:

$$d_\alpha = \frac{\Delta x_l - \Delta x_r}{\sqrt{\left(\frac{\Delta x_l + \Delta x_r}{2}\right)^2 + 1}} \tag{2}$$

3 Search Space and d_α Range

Suppose the point p^i_l in the left image and its correspondence in the right image p^i_r were already known. We want to find the correspondence of a second point p^k_l in the right image. Figure 2 shows a one dimensional restricted search region R^k based on the epipolar line. In this figure, the epipolar lines are shown by a dotted line.



a-Left image b-Right image

Figure 2 - The search region R^k for the correspondence of p^k_l in the right image is shown.

When the correspondence of p^k_l is displaced in R^k (in the right image), Δx_r is changed while Δx_l and Δy are constants and d_α is changed based on the equation (2). Then, $d_\alpha \in Rd_\alpha$ and we have:

$$d_\alpha \in Rd_\alpha : Rd_\alpha = \left\{ d_\alpha \mid d_\alpha^{Lo} < d_\alpha < d_\alpha^{Hi} \right\} \tag{3}$$

The range of d_α or Rd_α is a probabilistic phenomenon and it is dependent on the stereo system parameters and the disparity value. The pdf of d_α for a point in the left image with disparity d can be found by mapping d_α into a tangent direction at [X,Y,Z] in the 3D space of the scene, and using the corresponding relationship to transform pdf of the scene coordinate tangent to the pdf of d_α [4]. Performing these calculations and approximating the results for simplifying results in $P(d_\alpha)$ as:

$$P_{d_a}(d_a) = \frac{f/d}{2(1+(f/d)^2.d_\alpha{}^2)^{3/2}} \tag{4}$$

This pdf depends explicitly on the disparity under consideration (d) and the focal length of the cameras (f). As an example, consider the CIL (Calibrated Imaging Laboratory) images dataset at Carnegie Mellon University [6]. The focal length of the CIL is 57 millimeters and the distance between pixels is 23 μm therefore the focal length is about 2478 pixels. Figure 3 shows the pdf's for the various disparity values as a function of d_α values. These functions have some sharp peaks near zero d_α and become wider when the disparity increases.

Assume the range of d_α is $Rd_\alpha = \{d_\alpha^{Lo} < d_\alpha < d_\alpha^{Hi}\}$, therefore the $P(d_\alpha \in Rd_\alpha)$ can be obtained by integrating the equation (4) over Rd_α as:

$$P(d_a \in Rd_a) = \int_{d_\alpha^L}^{d_\alpha^H} P_{d_a}(d_a).d(d_a) \tag{5}$$

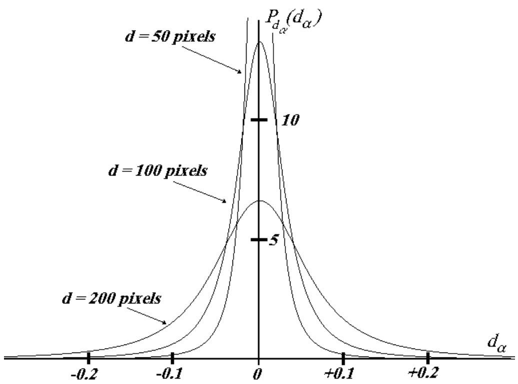


Figure 3 - Graph of $P_{d\alpha}$ vs. d_α for three different disparities d=50, d=100 and d=200 pixels for the CIL images dataset.

Considering the ordering constraint in stereo system [2][7], the maximum range of Rd_α is [-2,+2]. When Rd_α is a subset of [-2,+2], then we have a restricted search region, so we should allow some errors [5]. Assume that this error is less than 0.5% so that $P(d_\alpha \in Rd_\alpha)>0.995$. In a typical stereo algorithm, sometimes up to 10% error in the matching stage can be acceptable, so that the condition $P(d_\alpha \in Rd_\alpha)>0.995$ is suitable [5], since the error due to the search region selection is less than 0.5%.

We are interested in finding the maximum of the disparity (or d_{max}) for which $P(d_\alpha \in Rd_\alpha)>0.995$. For example, suppose $\Delta x_l=2$, $\Delta y=1$ and $\Delta x_r=\{0,1,2,3\}$, so d_α is varied between $d_\alpha^{Lo}=-0.37$ and $d_\alpha^{Hi}=2.0$. We want to find the d_{max} for which $P(d_\alpha \in Rd_\alpha)>0.995$. We increase d from zero and compute $P(d_\alpha \in Rd_\alpha)$ for each value of d considering equations (2), (4) and (5). The maximum value of d that satisfies the condition $P(d_\alpha \in Rd_\alpha)>0.995$ is proposed as d_{max}. For CIL dataset, d_{max} is 130 pixels.

4 Search Space Tuning

In many applications, a maximum disparity is given. It can be assumed that d_{max} is the same as the maximum disparity. In this case, the search space or Δx_r range can be computed when the Δx_l and Δy are known, considering $P(d_\alpha \in Rd_\alpha)>0.995$. In some applications, maximum disparity is large and it is comparable to image dimensions. Consider the maximum disparity or d_{max} less than 130 pixels for CIL stereo dataset, where the image dimensions are 576×384. Considering $P(d_\alpha \in Rd_\alpha)>0.995$, we investigate the search space for two cases: $\Delta y=0$ and $\Delta y=1$. We will see that these two cases are enough for our proposed algorithm. The table (1) shows the search space or Δx_r range when $\Delta y=0$ and Δx_l is varied between 2 to 14 pixels. The search space can be simply computed for $\Delta x_l>14$ pixels.

Table 1 - The relationship between Δx_l and the search region for CIL stereo dataset, when $\Delta y=0$ and Δx_l is between 2 to 14 pixels (refer to text).

Δx_l	2	3	4	5	6	7	8	9	10	11	12	13	14
Min(Δx_r)	0	1	1	2	2	3	3	4	4	5	5	5	6
Max(Δx_r)	3	5	6	8	9	11	13	14	16	17	19	20	22

Table 2 - The relationship between Δx_l and the search region for CIL stereo dataset when $\Delta y=1$ and Δx_l is between -3 to 3 pixels (refer to text).

Δx_l	-3	-2	-1	0	1	2	3
Min(Δxr)	-5	-3	-2	-1	0	0	1
Max(Δxr)	-1	0	0	1	2	3	5

The table (2) shows the search space or Δx_r range when $\Delta y=1$ and Δx_l is varied between -3 to 3 pixels. Both of tables were computed for CIL stereo dataset.

5 Proposed Algorithm

Our algorithm consists of two stages: feature extraction and feature matching. In the feature extraction stage, which is implemented for both of the left and right images, non-horizontal thinned edge extraction is performed. The thinned edge points are classified into two groups: positive and negative, depending on the gray level difference between the two sides of the feature points in the x direction. In the matching stage only similar features can be compared.

To detect non-horizontal edge points, the left and right images are convolved with a proper gradient mask. We use a Sobel 3×3 mask in x direction. A non-horizontal thinned positive edge is localized to a pixel for which the filter response has to exceed a positive threshold and has to obtain a local maximum in x direction. The positive threshold is selected dynamically based on the mean of positive value of the filter response. The extraction of non-horizontal negative thinned edge points is similar to that of the positive ones.

Two parameters are very important in decreasing the execution time of the matching stage: matching criteria and search space. We could reduce the search space to less than 10 pixels when $\Delta y=1$, and also when $\Delta y=0$ and $\Delta x_l<8$. Therefore in this case, if there is only one similar feature in the search space, the candidate point is proposed as the matching result. If there is more than one, the normalized cross correlation (NCC) with window size of 3×3 is chosen as the matching criterion [8]. If the search space is more than 10 pixels, we use the conventional multiresolution technique using Haar wavelet to reduce the search region [3]. Considering other multiresolution computation, such as Gussian pyramids and complex wavelets, implementation of Harr wavelet is simple and its execution is faster than others. We use three levels of Haar wavelet (the original+two lower resolutions) and NCC with the window size of 5×5, 3×3 and 3×3 for the coarse, medium and fine level respectively [7]. In the coarsest level, the search space is selected based on a proposed search region, which can be extended to the disparity range. At the higher resolution levels, the search space is cut around each found maximum correlation location in the previous level.

Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006 (pp243-247)

In the feature matching stage, the systematic scan from the top to bottom of the left image is done and for each scan line, the following steps are executed sequentially:

1- Do systematic scan from the left to right.

2- If the current point is not a feature point, go to step 1.

3- If disparity was already computed for the current point, store its x value as $x0_l$ and go to step 1.

4- Call the x value of the current point as xc_l. If there is not any $x0_l$ then go to step 5, else compute $\Delta x_l = xc_l - x0_l$ and then compute the search space in the right image regarding table 1 (case $\Delta y=0$) and then go to step 6.

5- Compute the search space based on the disparity range.

6- Find the correspondence point of the current point in the right image. The matching criteria and strategy is chosen based on the previous discussion. If there is not any correspondence point, go to step 1, else delete the current point from the left feature image and its correspondence from the right feature image.

7- In the next scan line of the left image; find the similar-close feature point. The absolute difference between the x values of the similar-close feature point and the current point must be less than or equal to 3 pixels. If there is not any similar-close feature point in the next scan line of the left image, go to step 1, else assume this difference as Δx_l.

8- Considering the table 2, compute the search space in the right image (case $\Delta y=1$). Suppose this close-similar feature point as the current feature point and then go to step 6.

6 Implementation Results

We called our algorithm HTS (Hybrid Tuned Search) and compare its implementation results with those of some fast feature based: EO[7], RSHMNE[7] and RS[5]. The thinned non-horizontal edges are selected as the feature points to test various algorithms. The search space in EO (edge only) is selected based on the disparity range but in the other methods the search spaces are reduced. In EO only the similar feature points are tested in the right image but in the other methods, at first only the similar features are examined in the search space and if the correspondence point is not found, other non-feature pixels are examined. As the similarity criterion, NCC with window size of 15×15 is selected in EO.

RSHMNE is similar to our method. In this method, the search space (Δx_r range) is restricted only for the case of $\Delta y=1$ and $\Delta x_l=\{-2,-1,0,1,2\}$, based on the general condition $|d_\alpha|<1.2$. For other feature points, a three level hierarchical multiresolution method is used to reduce the execution time [7]. The search space in hierarchical multiresolution is selected based on the disparity range.

RS method is also similar to RSHMNE, but $d\alpha$ is adjusted based on geometric parameters of stereo imaging system like focal length, CCD dimension and disparity range [5].

These algorithms were tested on two different stereo scenes [6], CIL1 and CIL2, that their disparity is supposed between 0 to 130 pixels. All the scenes used are in 576×384 and gray levels. The codes of the algorithms were written by Watcom C and implemented by a PC under Windows operating system, with an AMD Athlon XP 1700+ processor. The results of the investigated algorithms are shown in table 3 and 4 respectively. Considering the total number of extracted features from an interested area of the left image, the percentage of the matches and failures in the matching stage are shown in the matched (M) and failed (F) columns respectively. Considering the accurate disparity map for the feature points, the error of the matching stage could be computed. The error column (Err) shows the percentage of this error with respect to the total extracted left features. The execution time column (ET) is in millisecond scale (msec). This time includes the execution time of the feature matching. Some results of the algorithms are shown in figures 4 and 5 respectively.

Table 3 - The implementation results of matching on CIL1 scene, where 10046 features are extracted from the interested area of the left image. ET or the execution time is in milliseconds scale

Algorithm	M(%)	F(%)	Err(%)	ET (msec)
EO	41.4%	58.6%	1.6%	269
RSHMNE	86.0%	14.0%	3.1%	40
RS	84.3%	15.7%	2.3%	38
HTS	85.4%	14.6%	3.5%	20

In CIL1, the reduction in execution time of HTS respect to RS is more than CIL2, because the feature points in CIL1 are closer to each other and the reduction in the search space in the case of $\Delta y=0$ is more effective than CIL2.

The execution time of the feature extraction stage is about 24 msec for EO and 30 msec for other methods which includes feature extraction plus three level multiresolution computing of the graylevel and feature images. The execution time of feature extraction is comparable to the matching time of HTS which is only about 20 and 24 msec for the

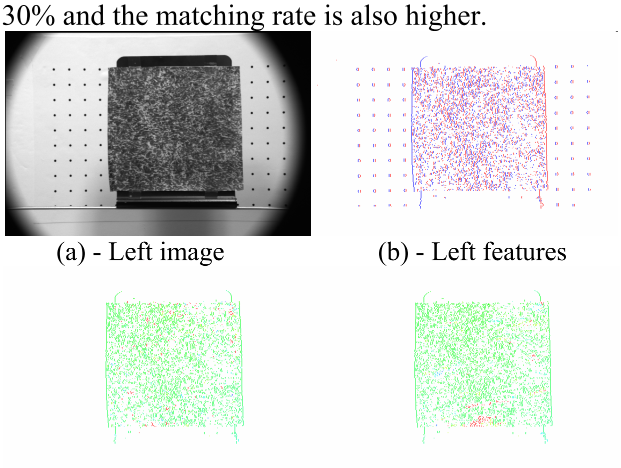
testes stereo images. Therefore in the proposed method, the execution time of the feature extraction is more important than the feature matching.
Considering RS method as the benchmark, HTS stereo matching algorithm is executed 25% to 47% faster and the matching rate is also higher, moreover the error rate of the proposed algorithm is low and it is acceptable for most applications.

Table 4 - The implementation results of matching on CIL2 scene, where 4957 features are extracted from the interested area of the left image

Algorithm	M(%)	F(%)	Err(%)	ET (msec)
EO	70.5%	29.5%	9.2%	77
RSHMNE	92.8%	7.2%	5.5%	35
RS	92.5%	7.5%	5.4%	32
HTS	94.0%	6.0%	1.6%	24

7 Conclusion

To accurately tune the search space in the stereo matching, we could establish the relationships between the search space and the parameters like displacements of features, maximum disparity, focal length and distance between pixels. We combined this new idea and the multiresolution technique using Harr wavelet to reduce the search space in the feature-based stereo matching. We proposed the fast feature based stereo matching algorithm based on the proposed search space tuning. The overall execution time of the proposed algorithm includes feature extraction and matching is decreased 15% to 30% and the matching rate is also higher.

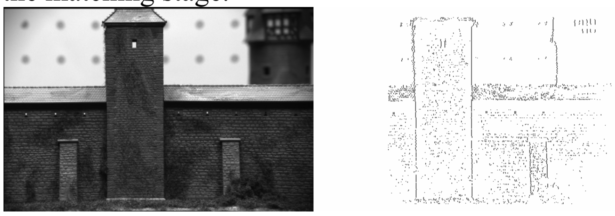


(a) - Left image (b) - Left features

(c) - RS (d) - HTS

Figure 4 - The implementation results on CIL1 stereo scene. The positive and negative feature points are shown together in (b) with magenta and blue color respectively. The disparity map by RS and HTS methods are shown in the figure (c) and (d) respectively, where the red color shows the error points in the matching. The interested area of this scene is selected out of the dotted grid.

The execution time of feature extraction stage, which is implemented separately for the left and right images, is nearly independent of scene complexity. Therefore to decrease the overall execution time, we suggest using two separate processors for implementation of feature extraction in the left and right image. A third processor can be used in pipeline manner to the others to implement the matching stage.



(a) - Left image (b) - HTS

Figure 5 - The implementation results on CIL2 stereo scene. The matched features by HTS method are shown in (b).

References:

[1] W. Ponweiser, M. Ayromlou, M. Vincze, C. Beltran, O. Madsen, A. Gasteratos, "RoboVision, Vision Based Navigation for Mobile Robots", *IEEE International Conf. on* MFI, pp. 109-114, 2001.

[2] M. Bennamoun, G.J. Mamic, Stereo Matching and Reconstruction of a Depth Map, *Object Recognition Fundamentals and Case Study*, Second Section, Springer-Verlag,2002.

[3] S. Brandt, J. Heikkonen, "Multi-Resolution Matching of Uncalibrated Images utilizing Epipolar Geometry and its Uncertainty", *IEEE ICIP*, Vol. 2, pp. 213-216, 2001.

[4] C. Stewart, R. Flatland, K. Bubna, "Geometric Constraints and Stereo Disparity Computation", *Inernational Journal of Computer Vision*, Vol. 20, No 4, pp. 143-168, 1996.

[5] P. Moallem, K. Faez, "Effective Parameters in Search Space Reduction used in a Fast Edge-Based Stereo Matching", *Journal of Circuits, System, and Computers*, Vol 14, No 2, pp. 249-266, April 2005.

[6] CIL Images Dataset at http://www.cs.cmu.edu/~cil/cil-ster.html.

[7] P. Moallem, K. Faez, J. Haddadnia, "Fast Edge-based Stereo Matching Algorithms through Search Space Reduction", *IEICE Transaction Information and Systems*, Vol. E-85, No. 11, pp. 1859-1871, Nov 2002.

[8] M. A. Torres, A. R. Guesalga, "Confidence Factor and Adaptive Disparity Range for Reliable 3D Reconstruction", *Computer Intelligence for Modeling, Control and Automation*, ed. M. Mohammadian, pp. 83-91, IOS Press, 1999.