Fast stereo matching based on edge energy information

Xiaowei An, Youngjoon Han, Hernsoo Hahn
(*Dept. of Electronic Engineering, Soongsil University, Seoul 156-743, Korea*)

Abstract: A new improvement is proposed for stereo matching which gives a solution to disparity map in terms of edge energy. We decompose the stereo matching into three parts: sparse disparity estimation for image-pairs, edge energy model and final disparity refinement. A three-step procedure is proposed to solve them sequentially. At the first step, we perform an initial disparity model using the ordering constraint and interpolation to obtain a more efficient sparse disparity space. At the second step, we apply the energy function by the edge constraints that exist in both images. The last step is a kind of disparity filling. We determine disparity values in target regions based on global optimization. The proposed three-step simple stereo matching procedure yields excellent quantitative and qualitative results with Middlebury data sets in a fast way.

Key words: stereo matching; disparity space; edge energy model; disparity filling

CLD number: TP391 **Document code**: A
Article ID: 1674-8042(2012)02-0129-04 **doi**: 10.3969/j.issn.1674-8042.2012.02.007

Stereo vision has been one of the most extensively investigated topics in the area of computer vision. As stereo techniques can convert 2D images to 3D model, they have been applied to computer graphics, virtual reality or estimation of relative positions of objects in understanding semantic relationships among different environments. Reliable depth map shows the distance of different objects existing in this world, which has grown in importance in recent years.

Many approaches have been applied to improving the matching quality and efficiency, as Scharstein and Szeliski presented a taxonomy of dense matching methods which gave almost all the classic correspondence algorithms[1]. Usually the main applications focus on two categories: local (area) methods and global (feature) methods. Local methods mainly apply the adaptive window which can get the dense stereo results, but need the expensive computation. Global approaches usually rely on the energy minimization framework with the smoothness constraint to resolve the ill-posed problem of stereo matching. Especially, the adoption of the Markov random field (MRF) model has brought stereo matching research to a new era[2].

It is trivial to balance the tradeoff between dense result and cheap computation time, especially if window containing more pixels may make matching results unclear in the local areas. To smooth this obscure, some relative-costs must be computed in low-texture areas such as normalized cross correlation (NCC), sum of squared difference (SSD) and sum of absolute difference (SAD). Those calculations almost produce coarse results without any edge details. However, methods based on global features are applicable to those issues. The framework based on the graph cuts provided necessary energy minimization techniques in vision[3]. Belief propagation[4] has attracted much attention due to their excellent performances. The dynamic programming for stereo like Baker and Binford paper[5] uses Viterbi algorithm with the occlusion information. They match the standard image pairs with the new corresponding constraints, but do not consider the simple procedure explicitly in the real system, eventually algorithm complex dominates in the matching process.

For those issues, an algorithm for simple stereo matching is developed which produces dense map with edge constraint details for realtime system.

It consists of the following three processing steps: ① Sparse disparity estimation for image-pairs; ② Edge energy model; ③ Final disparity refinement.

In the first step, we focus on improving the quality of disparity estimation in total regions including texture areas. The input image pairs are rectified pictures that could save the cost of computation. In the second step, we fix the simple feature—edge as the corresponding set. This set gives a simple solu-

※ **Received data**: 2011-07-25
Foundation item: MKE (the Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2011-C1090-1121-0010) and the Brain Korea 21 Project in 2010
Corresponding author: Hern-soo Hahn(hanh@ssu.ac.kr)

tion to the matching constraints. In the final step, filling the relative disparities improves the efficiency of the nearest disparities.

Before the matching part, some basic knowledge about epipolar geometry is also presented for facilitating the initial understanding.

1 Epipolar geometry

1.1 Epipolar geometry

If the two cameras are C and C', and a 3D point X is imaged as shown in Figs. 1 and 2.

$$x = PX,\ x' = P'X.$$

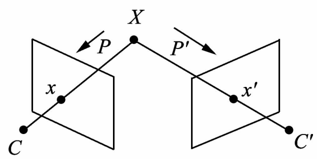

Fig. 1 Camera array

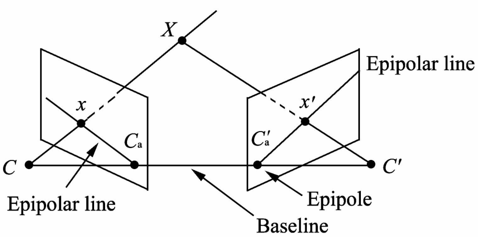

Fig. 2 Epipolar geometry

Once obtaining the baseline—the line connecting C with C', two epipoles C_a and C'_a will be located in both images. Given an image point in one view, we are able to find the corresponding point in the other view[3] (Fig. 3).

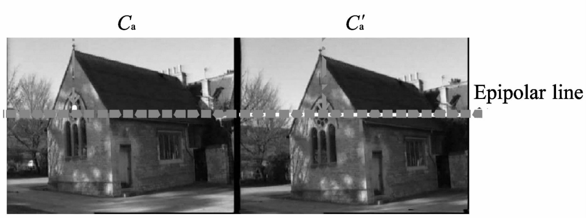

Fig. 3 Epipolar line

Epipolar geometry is a consequence of the coplanarity of the camera centers and scene point (Fig. 4).

In order to determine the detailed geometric configuration of the two parallel cameras, it is necessary to take a calibration before the matching algorithm in terms of system inner geometric parameters (focal length, image center and lens distortion). Calibration refers to the act of evaluating and adjusting the precision and accuracy of measurement equipment in the image pairs processing.

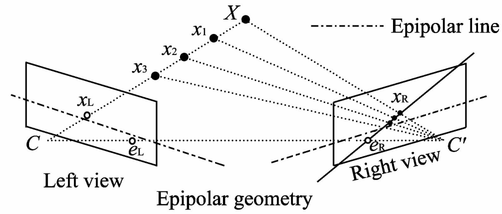

Fig. 4 Rectification

1.2 Triangulation model

After setting the epipolar lines in both images, relationship between pairs are easily determined under the triangulation constraint:

The set of initial matching costs that are fed into a stereo matcher's optimization stage is often called the disparity space image (DSI)[3].

$$DSI_{d_n} = | I_L(x_L + k, d_n) - I_R(x_R, d_n) |,$$

where I_L and I_R are left image and right image respectively; d_n is disparity distance.

Fig. 5 shows that the distance of object (Depth) is inversely proportional to the disparity variance according to the triangulation.

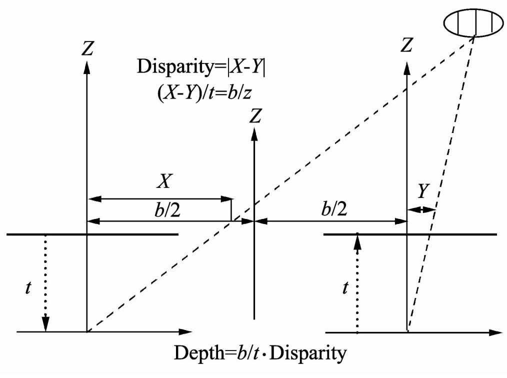

Fig. 5 Depth map

2 Coarse disparity estimation

Common point-wise or fixed size block-wise methods have been used to recover coarse disparity estimation[6]. Since such common methods purely utilize the point-wise intensity subtraction, which cannot present the details of the original image, e.g. edge, texture area and occlusion, our initial matching uses the new method to simulate the coarse disparity.

$$j = 1 : height,$$
$$i = 1 : width,$$

$$V(j,i,1) = \frac{I_{\text{left}}(j,i+1) + I_{\text{left}}(j,i)}{2} - I_{\text{right}}(j,i),$$
$$V(j,i,2) = \frac{I_{\text{left}}(j,i-1) + I_{\text{left}}(j,i)}{2} - I_{\text{right}}(j,i),$$
$$V(j,i,3) = I_{\text{left}}(j,i) - I_{\text{right}}(j,i),$$
$$W(j,i,1) = \frac{I_{\text{right}}(j,i+1) + I_{\text{right}}(j,i)}{2} - I_{\text{left}}(j,i),$$
$$W(j,i,2) = \frac{I_{\text{right}}(j,i-1) + I_{\text{right}}(j,i)}{2} - I_{\text{left}}(j,i),$$
$$W(j,i,3) = I_{\text{right}}(j,i) - I_{\text{left}}(j,i). \qquad (1)$$

Find the i and j position under the condition

$$\min W(j,i,t) * \min V(j,i,t) > 0, \ t = 1,2,3.$$

And then, utilize the obtained i and j in

$$Coarse\ DSI_{d_n} = |\ I_L(j,i+d) - I_R(j,i)\ | \qquad (2)$$

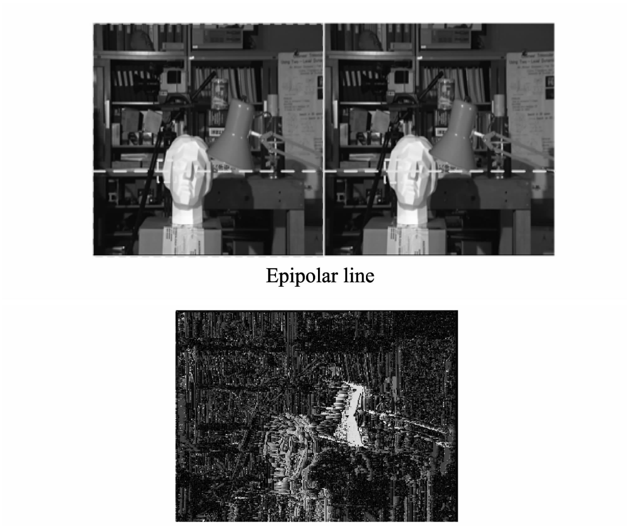

Fig. 6 Coarse disparity space

3 Edge energy model

Edge is the intensity difference between neighbouring pixels. It shows the property of the derivative in the image target areas[7, 8].

To illustrate the edge effect and performance we adopt the edge energy model to accumulate the edge derivative as

$$E(D) = E_{\text{data}}(D) + E_{\text{edgesmooth}}(D), \qquad (3)$$

where $E(D)$ is energy of each disparity, and our task is to find the minimum of the Energy; D is disparity values; $E_{\text{data}}(D)$ is data difference between left and right images.

According to Eq. (2), there is

$$E_{\text{data}}(D) = Coarse\ DSI_{d_n}.$$

In Eq. (3), $E_{\text{edgesoomth}}(D)$ considers point-wise continuous in the scene. It controls disparity derivative, and large variation exists only at depth border. Coarse disparity space is generally ambiguous. Matches can easily have a wrong cost instead of correct ones for the reason of noise.

Here we accept the image model $\exp(-C/T)$. This model facilitates the accuracy of the coarse disparity space.

$$E_{\text{edgesoomth}}(D) = e^{-E_{\text{nerighbouring}}(D)\cdot C/T}. \qquad (4)$$

where C and T are smooth parameters. C stands for weight-balance, which makes $E(D)$ always positive; T is the intensity truncation value, and let $T = 25$.

So the total energy can be transformed as

$$e^{-E(D)\cdot C/T} = e^{-E_{\text{data}}(D)\cdot C/T} * e^{-E_{\text{neighboring}}(D)\cdot C/T}. \qquad (5)$$

If edge exists between x and $x-1$ position:

$$e^{-c\cdot Cost(j,x,d_{x-1})/T} = e^{-c\cdot Sumcost(j,x-1,d_{x-1})/T} \times e^{-c\cdot Csmooth(j,x)\cdot C/T},$$
$$j = 1 : height,\ x = 1 : width,$$

where d_x is the x position disparity, d_{x-1} is the $x-1$ position disparity; *Cost* sepresents the total data difference; *Sumcost* represents the edge accumulation; *Csmooth* is the gradient space of original right (left) image.

$$Csmooth(j,i) = |\ I_L(j,i+1) - I_L(j,i)\ | \text{ or } |\ I_R(j,i+1) - I_R(j,i)\ |.$$

Otherwise:

$$e^{-c\cdot Cost(d_{x-1})/T} = e^{-c\cdot Sumcost(x-1,d_{x-1})/T} \qquad (6)$$

4 Final disparity refinement

The sparse disparity image consists of some ambiguities. Furthermore, there are areas of invalid values need to be recovered, which can be handled by post-processing of the disparity image. For example, one of the most noteworthy features og stereo matching problem is the physical surface in untextured areas[9,10]. e.g. green dotted circles.

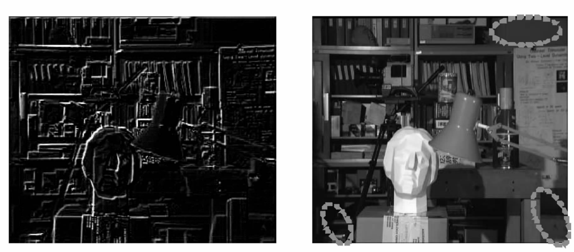

Fig. 7 Gradient smooth space

In order to reduce the ambiguities, we assume:

1) All the untextured areas eepth values similar;
2) Untextured areas own some visible textures;
3) One textured area only own one depth value.

The reasons for those assumptions are that discontinuities occur with texture visible in the disparity space. Otherwise we cannot find the features.

5 Experiment and results

Edge energy model (EEM) has been tested on Middlebury web stereo image pairs. All the experimental activity was supported by the rectified inputs (Fig. 8).

(a) Tsukuba image-sets

(b) Door image-sets

(c) Teddy image-sets

(d) Venus image-sets

Fig. 8 Matching results

Tsukuba 3D display is shown in Fig. 9.

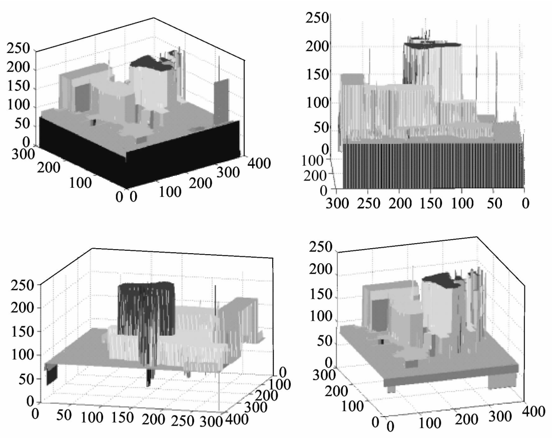

Fig. 9 3D display

Table 1 Time cost

Similar Accuracy	Tsukuda (384 * 284)	
	Adaptive window method	Proposed algorithm
Time cost	2 s	1.1 s
Similar Accuracy	Teddy (384 * 284)	
	Adaptive window method	Proposed algorithm
Time cost	3.5 s	2.5 s
Similar accuracy	venus (434 * 383)	
	Adaptive window method	Proposed algorithm
Time cost	2.8 s	1.6 s
Similar accuracy	Door (311 * 275)	
	Adaptive window method	Proposed algorithm
Time cost	1.6 s	0.9 s

6 Conclusion

A fast edge energy process is presented which has been tested on rectified stereo image pairs. In this model, energy function was formulated with edge properties. Due to its fast computational power, simulating results have verified that the proposed method could get a high accuracy during the short time.

References

[1] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, 2002, 47(1/2/3): 7-42.

[2] Szeliski R, Zabih R, Scharstein D, et al. A comparative study of energy minimization methods for Markov random fields. Proc. of European Conference Computer Vision 2006, 2: 16-29.

[3] Boykov Yuri, Kolmogorov Vladimir. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. on PAMI, 2004, 26 (9): 1124-1137.

[4] Sun J, Zheng N N, Shum H Y. Stereo matching using belief propagation. IEEE Trans. on Pattern Analysis Machine Intelligence, 2003, 25 (7): 787-800.

[5] Baker H H, Binford T O. Depth from edge and intensity based stereo. Proc. of International Joint Conference on Artificial Intelligeit, 1981: 631-636.

[6] Birchfield, C. Tomasi C. A pixel dissimilarity measure that is insensitive to image sampling. IEEE Trans. on PAMI, 1998, 20(4): 401-406.

[7] Mattoccia S. A locally global approach to stereo correspondence. 3DIM 2009.

[8] Cai J. Fast stereo matching: coarser to finer with selective updating. Proc. of Image and Vision Computing, Hamilton, New Zealand, 2007: 266-270.

[9] Barýs Baykant ALAG. Obtaining depth maps from color images by region based stereo matching algorithms. Oncu-Bilim Algorithm and Systems Labs., 2008,8(4).

[10] Szeliski R, Scharstein D. Sampling the disparity space image. Introduction to Robotics, NewYork, 2006.