Intensity Invariant Stereo Matching and Disparity Refinement

Junxue He[1,2], Erchao Li[1,2] and Zhanming Li[1,2]

[1]*College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China*
[2]*Key Laboratory of Advanced Control for Industrial Processes, Gansu Province, Lanzhou 730050*
junxue.he@163.com

Abstract

A intensity invariant measure of pixel dissimilarity that compares two pixels is proposed. Horizontal and vertical differences of the image are utilized for dissimilarity measure. The method is intensity invariant for stereo matching. Bad matches arising from illumination change between two cameras are eliminated as much as possible. Moreover, some parameters of belief propagation algorithm for stereo matching can be figure out due to introducing this measure. Disparity gradient constraint is utilized to refine disparity map and extrapolate the disparities into those regions where disparities are not produced properly. Experiments on real images demonstrate that our method is distinctly robust to changes in camera gain and bias or illumination.

Keywords*: Computer vision; Stereo matching; Belief propagation; Dissimilarity measure; Intensity invariant; Disparity refinement*

1. Introduction

Given two rectified images from different viewpoints, the goal of stereo matching is that generates the disparity or depth information of each pixels in the reference image. Stereo matching which is the critical technology on three-dimensional reconstruction is still a problem there is no perfect solution, although there are many high performance algorithms [1-2], because there are many factors would be considered into generating accurate depth map such as image noise, textureless region, depth discontinuity and occlusion. Image noise and illumination change would challenge the robust of stereo matching algorithm. The brightness consistency constraint is invalid in the absence of texture region. Spatial smoothness constraints in the object boundary are invalid. The corresponding match to some of the points in one image does not exist in another image. Moreover, when a point in the world is imaged by a stereo pair of cameras, the intensity values of the corresponding pixels are in general different. Many factors contribute to this difference, such as the fact that the light reflected off the point is not the same in the two directions, the two cameras have different gains and biases. Most stereo matching algorithms compare pixel intensities to minimize a global or local object function. Taking accuracy and computational complexity into account, the results is not ideal because of the factors mentioned above, especially, the two cameras are in harsh environment with changing illumination respectively.

In general, matching algorithms can be classified into local and global methods. Local approaches are utilizing the color or intensity values within a finite window to determine the disparity for each pixel. Global approaches are incorporating explicit smoothness assumptions and are determining all disparities simultaneously by applying energy minimization techniques such as graph cuts, belief propagation, dynamic programming, scanline optimization or simulated annealing. In stereo algorithm whether local or global, dissimilarity measure of pixel is important for stereo matching. When illumination change

between two cameras is distinct, those algorithms that measure dissimilarity with intensity differences simply will cause a large number of bad matches.

Belief propagation (BP) algorithm, have shown significant performance over other stereo matching algorithms [1-3]. However, the standard BP stereo matching has two major challenges for practical applications: high computational cost and varied parameters. Many of iterations are required to ensure convergence of the message values. The matching result is affected by parameters directly. However, proper parameters vary with data which is commonly intensity differences.

There are two contributions in this paper. First, we propose a measure of pixel dissimilarity that compares two pixels using the horizontal and vertical difference of them. Intensity differences are converted to elevation angle difference which is equal to horizontal or vertical difference. This method is distinctly robust to changes in camera gain and bias or illumination. Moreover, proper parameters of belief propagation algorithm for stereo matching can be obtained by adjustment slightly. Our measure is provably insensitive to illumination change between two cameras and is shown to improve the results of BP algorithm on real images. Second, disparity gradient constraint is utilized to optimize disparity map and extrapolate the disparities into those regions which disparities are not produced properly.

2. Previous Work

There are many works such as [2,4-8] involving stereo matching problem. Most of them measure pixel dissimilarity by comparing the pixel intensities (or that surrounding them). The metrics generally are absolute differences, squared differences or linearly interpolated intensity functions [9].

Scharstein and Szeliski [3] pointed out that most of the stereo matching algorithm has four basic steps: matching cost computation; cost (support) aggregation; disparity computation/optimization and disparity refinement. Over the past decade, some local methods such as window-based approach cannot obtain satisfactory results. People have focus attention to the global approach gradually. The global method has global objective function which ensures the result is global optimal. Among them, many kinds of methods that make use of Markov random fields (MRF) model for solving stereo problem show excellent performance.

As a global stereo matching algorithm, belief propagation (BP) [1, 7, 10-11] have shown significant performance over other algorithms [3]. However, the standard BP stereo matching has two major challenges for practical applications: high computational cost and varied parameters. Many of iterations are required to ensure convergence of the message values. the standard implementation of this message passing algorithm on the grid graph runs in $O(nk^2T)$ time, where n is the number of pixels in the image, k is the number of possible labels for each pixel and T is the number of iterations. It takes $O(k^2)$ time to compute each message and there are $O(n)$ messages to be computed in each iteration. Parameters in the matching algorithms need to be estimated automatically to adapt with varying conditions such as lighting, relative motion between the scene and imaging sensors and scene structures.

Felzenszwalb [7] presents three algorithmic techniques to speed up convergence of loopy belief propagation (LBP). One of the techniques reduces the complexity of the inference algorithm to be linear rather than quadratic in the number of possible labels for each pixel, which is important for stereo problems that have a large label set. Another technique speeds up and reduces the memory requirements of belief propagation on grid graphs. A third technique is a multi-grid method that makes it possible to obtain good results with a small fixed number of message passing iterations, independent of the size of the input images. Taken together these techniques speed up the standard algorithm by

several orders of magnitude. Combining the three techniques together yields an algorithm that runs in $O(nk)$ time overall and is very fast in practice. In contrast the standard algorithm summarized above requires $O(nk^2)$ time per iteration and the number of iterations T is generally $O(n^{1/2})$ to allow for information to propagate across the image. Felzenszwalb improved BP algorithm greatly on computational complexity. Klaus [1] proposed a segment-based stereo matching algorithm using BP and a self-adapting dissimilarity measure, which improve the result while increasing computation.

3. Dissimilarity Measure

The most common dissimilarity measures are intensity differences including squared intensity differences (SD), absolute intensity differences (AD), intensity differences from edge detector, or their combination with different proportions. Intensity differences are strictly assuming the constant color constraint. When illumination change between two cameras is distinct, those algorithms that measure dissimilarity with intensity differences will cause a large number of bad matches. Birchfield and Tomasi [9] proposed a measure of pixel dissimilarity that compares two pixels using the linearly interpolated intensity functions surrounding them. Their method are more robust to changes in camera gain and bias.
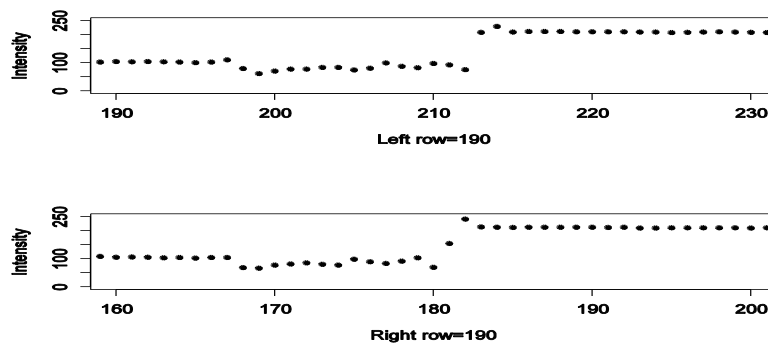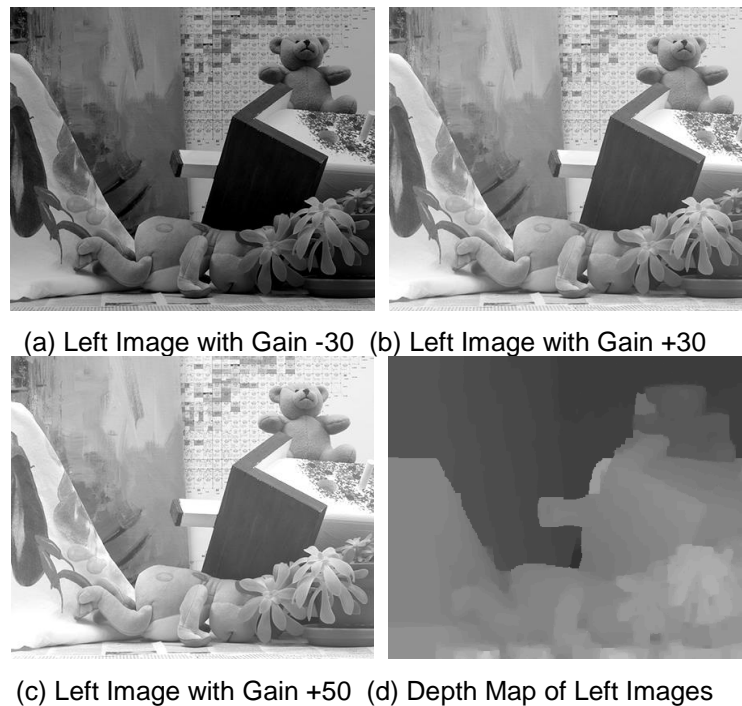

Figure 1. Intensity of Teddy

To know what happened when the corresponding pixels in left and right image are matching, some pixels will be chosen to analyze. Middlebury dataset teddy has wide disparity range. The intensities of pixels at row 190 in left and right image of teddy is show in Figure 1 produced by R [12]. The about location of those pixels are show in Figure 2a. Obviously, lots of bad matches will emerge when we simply make use of intensity difference as dissimilarity measure because of illumination change. We then have the following definition:

Definition 1. The image $H(n, x, y) = H(n-1, x+1, y) - H(n-1, x, y)$ is said to be horizontal n^{th} order difference of $I(x, y)$. The image $V(n, x, y) = V(n-1, x, y+1) - V(n-1, x, y)$ is said to be vertical n^{th} order difference of $I(x, y)$. Where $H(0, x, y)$ and $V(0, x, y)$ are original image $I(x, y)$, $n = 0, 1, 2...N$. N is a constant.

(a) A Interest Region   (b) Matching Schema
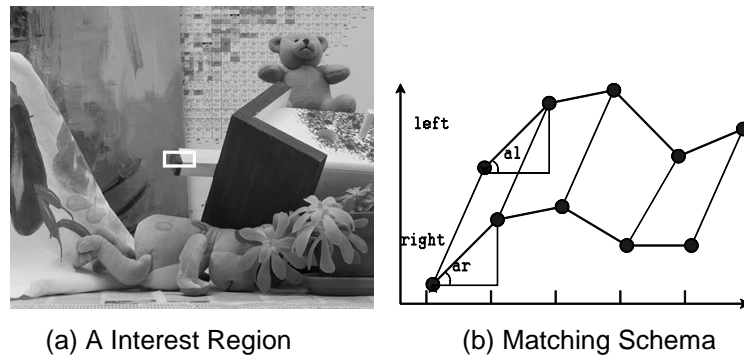
Figure 2. A Interest Region and Matching Schema

As shown in Figure 2b, when two adjacent pixels match correctly, the absolute difference between angle al and ar would reach the global minimum. The elevation angle is equal to horizontal first order difference obviously. This method is distinctly robust to changes in camera gain and bias or illumination. Given a constant right image, the almost same depth map will be obtained from left images with different intensity gain. The result is shown in Figure 3.

(a) Left Image with Gain -30 (b) Left Image with Gain +30

(c) Left Image with Gain +50 (d) Depth Map of Left Images

Figure 3. Same Depth Map from Images with Different Intensity Gains

4. MRF Energy Model

In the Markov random fields (MRF) model of image, the calculation of disparities is regarded as problem of pixel labeling. The general framework for pixel labeling can be defined as follows. Let P be the set of pixels in an image and L be a finite set of labels. The labels correspond to quantities that we want to estimate at each pixel, such as disparities or intensities. A labeling f assigns a label $f_p \in L$ to each pixel $p \in P$. We assume that the labels should vary slowly almost everywhere but may change

dramatically at some places such as pixels along object boundaries. The quality of a labeling is given by an general energy function,

$$E(f) = \sum_{p \in P} D_p(f_p) + \sum_{(p,q) \in N} W(f_p, f_q)$$

(1)

where N are the (undirected) edges in the four-connected image grid graph. $D_p(f_p)$ is the cost of assigning label f_p to pixel p, and is referred to as the data cost. $W(f_p, f_q)$ measures the cost of assigning labels f_p and f_q to two neighboring pixels, and is normally referred to as the discontinuity cost. Finding a labeling that minimizes this energy corresponds to the maximum a posteriori (MAP) estimation problem for an appropriately defined MRF[4] . In stereo problems, the energy equation is of the form,

$$E(d) = \sum_{i \in P} D(d_i) + \sum_{(i,j) \in N} S(d_i, d_j)$$

(2)

where $D(d_i)$ is data term, $S(d_i, d_j)$ is smoothness term, $i = (x_i, y_i)$, $j = (x_j, y_j)$ is pixels and d_i is disparity corresponding to pixel i.

The method is iterative, with messages from all nodes being passed in parallel. Each message is a vector of dimension given by the number of possible labels, k. Let m_{pq}^t be the message that node p sends to a neighboring node q at iteration t. When using negative log probabilities all entries in m_{pq}^0 are initialized to zero, and at each iteration new messages are computed in the following way,

$$\mathbf{M}_{pq}^t(f_q) = \min_{f_p} \left(S(f_p, f_q) + D_p(f_p) + \sum_{s \in N(p) \backslash q} M_{sp}^{t-1}(f_p) \right)$$

(3)

where $N(p) \backslash q$ denotes the neighbors of p other than q. After T iterations a belief vector is computed for each node,

$$\mathbf{b}_q(f_q) = D_q(f_q) + \sum_{p \in N(p)} M_{p \to q}^T(f_q)$$

(4)

In energy function (2), Many functional forms have been proposed for data term $D(d_i)$ and smoothness term $S(f_p, f_q)$, including squared differences, absolute differences, and other robust metrics[3]. According to definition 1, we define the data term $D(d_i)$ of pixel (x, y) as follows,

$$HV(x, y) = H(1, x, y) + V(1, x, y)$$
$$D(d_i) = \lambda * \min\left(| HV_l(x, y) - HV_r(x - d_i, y) |, DATA\_K \right)$$

(5)

where λ is relevant weight. $DATA\_K$ is truncation thresholds . The data cost $D(d_i)$ represents the cost of assigning disparity d_i to pixel i ; the data cost increases as the difference in HV between corresponding pixels increases. The $DATA\_K$ can be eliminated by algorithm 1.

Algorithm 1. Computing $DATA\_K$

$val = |HV_l(x,y) - HV_r(x-d_i,y)|$;

$is = (|H(1,x,y)| < 3.0) \&\&(|H(1,x-d_i,y)| < 3.0)$;

if (is) $DATA\_K$ = 3.0;

else $DATA\_K$ = val/2;

The discontinuity cost function $S(d_i, d_j)$ is based on the magnitude of the difference between labels d_i and d_j. The truncated linear model is a popular choice of top performing stereo algorithms [4, 7] and it has several good properties. The cost increases linearly based on the distance between the labels d_i and d_j up to some level. In order to allow for large discontinuities in the labeling the cost function stops growing after the difference becomes large. We choice the truncated linear model as smoothness term and focus on setting parameters for it,

$$S(d_i, d_j) = min(|d_i - d_j|, DISC\_K)$$

(6)

where $DISC\_K$ controls when the cost stops increasing. The discontinuity cost $S(d_i, d_j)$ represents the cost of assigning disparities d_i and d_j to neighboring pixels; this cost increases as the difference between d_i and d_j increases.

5. Implementation

The standard max-product BP algorithm is speeded up by Felzenszwalb [7]. We implemented the proposed BP algorithm 2 based on Felzenszwalb conjugating with our proposed algorithm 1.

We assume that the range of disparity is $[d_{min}, d_{max}]$, equally, $NR-1$ is equate to the difference of d_{max} and d_{min}, $ITER$ is the iterations of BP algorithm, $LEVELS$ is the levels of data pyramid. The truncation value $DISC\_K$ in the discontinuity cost $S(d_i, d_j)$ are defined in equation(6). Let P be the set of pixels in an image.

Algorithm 2. Proposed BP algorithm

1) Preprocesses. The rectified stereo image pair is loaded and filtered by Gaussian filter with $\sigma = 0.7$ for fading effect of noise and illumination change;

2) According to the formula(5) and algorithm 1, calculate 0 order of the data cost $D_0(d_i)$ for each pixel $i = (u_i, v_i) \in P$;

3) Calculate pyramidal data cost $D_0, D_1, ..., D_{LEVELS}$ of the image;

4) For $levels$ from $LEVELS-1$ to $\mathbf{0}$, perform belief propagation in a coarse-to-fine manner. In the coarsest level $LEVELS-1$, messages are initialized to zero, and otherwise, messages are initialized from values of previous level. Calculate discontinuities cost according to the formula (6). Assume m_d be the message of disparity d, for $iter$ from $\mathbf{0}$ to $ITER$, the message m can be updated by following steps,

4.1) For $d = 0$ to $NR-1$, $m_{d\,min} = min(m_d)$;

4.2) Update message m_d according to the formula (3, 6);

(a) For $d = 1$ to $NR-1$, $m_d = min(m_d, m_{d-1})$;

(b) For $d = NR - 2$ to **0**, $m_d = \min(m_d, m_{d+1})$;

4.3) Truncate message, for $d = 0$ to $NR - 1$, $m_d = \min(m_d, m_{d\min} + DISC\_K)$;

4.4) Normalize message, for $d = 0$ to $NR - 1$, $\sum m_d = 1$;

5) Compute out beliefs according to the formula (4);

6) For $d = 0$ to $NR - 1$, find out the minimum beliefs and generate disparity map;

Sub-pixel accuracy can be obtained by fitting a second degree curve to the correlation coefficients in the neighbourhood of the disparity. The general form of the second degree curve (parabola) is: $f(x) = a + b \cdot x + c \cdot x^2$. The maximum can be found where the slope is zero in the quadratic function. The sub-pixel position can be found at $x = -b/2c$. If only three correlation values at and around the position d are used, *e.g.* values at the points $d - 1$, d, and $d + 1$, the sub-pixel position of the disparity can be calculated using the following formula[13]:

$$x = d + \frac{1}{2} \times \frac{R(d-1) - R(d+1)}{R(d-1) - 2R(d) + R(d+1)}$$

(7)

where $R(d)$ is the correlation value in the 2D matrix at position d, and x is the sub-pixel disparity obtained.

6. Refinement of Disparity Map

In stereo matching algorithm whether local or global, lots of algorithms do not produce disparities in some regions (*e.g.,* borders, occlusion). In this case, refinement or extrapolating the disparities into those regions is required. Disparity is directly related to depth. Disparity changes coincide with depth changes. The disparity gradient limit principle states that the disparity should vary smoothly almost everywhere, and the disparity gradient should not exceed a certain value [14-17]. For two points m_1 and m_2 in the reference image, their disparity are d_1 and d_2 respectively. We define the disparity gradient as:

$$DG = \frac{\|d_2 - d_1\|}{\|m_2 - m_1 + (d_2 - d_1)/2\|} \le K$$

(8)

The disparity gradient limit K can be varied over range $(0.5, 1.9)$. We refine the disparity map utilizing the lemma 1 introduced by Zhang [17].

lemma 1. Given a pair of matched points (m_1, m_1') and a point m_2 in the neighborhood of m_1, the corresponding point m_2' that satisfies the disparity gradient constraint with limit K must be inside a disk centered at $m_2 + d_1$ with radius equal to $\frac{2K}{2-K} D$, which we call the continuity disk, where $D = \|d_2 - d_1\|$ is the distance between m_1 and m_2.

The relation of disparity and gradient limit K is shown in Figure 4. In our implementation, $D = 1$ and $K = 1.8$. This is equivalent to $threshold = 18$ pixels. For all images we have tried, this strategy works well. Algorithm 3 is the pseudo C code for refining a disparity map.
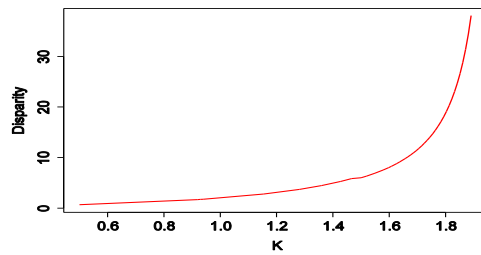
Figure 4. Disparity Gradient Limit

 Algorithm 3. Refining disparity map
input: out; //input initial disparity map
imgmean=mean(out);
threshold =18; //disparity gradient limit
for (int row = 0; row < height; row++) {
   for (int col = 0; col < width; col++) {
   disp0 is the current disparity.
   disp is the disparity in direction left, up, right and down respectively.
   grad is the disparity gradient corresponding to disp.
   if (grad > threshold){
      if (abs(disp0-imgmean) < abs(disp - imgmean))
      disp=disp0;
      else disp0=disp;
      }
   }
}

Table 1. Evaluation of the Proposed Stereo Algorithm on the Middlebury Stereo Benchmark

	nonocc	all	disc
Tsukuba	4.12	6.26	21.5
Venus	2.14	3.09	21.8
Teddy	10.6	19.4	29.1
Cones	6.82	15.3	17.8

7. Results

Results using the Middlebury Stereo datasets [18-19] Tsukuba, Venus, Teddy and Cones are shown in Figure 5, Figure 6, Figure 7 and Figure 8 respectively. In above figures, the images are the reference one, the disparity map by proposed algorithm and the bad pixel map, respectively. Pixels with a disparity error greater than one pixel are displayed in the bad pixel maps, where mismatches in non-occluded areas are indicated in black, in occluded areas in gray color. Table 1 shows evaluation of the proposed stereo algorithms on the Middlebury Stereo benchmark. The error measures the percentage of pixels with wrong disparities. All results are generated on a laptop with a Intel Core 2 Duo CPU running at 2.13 GHz. We used 5 levels and 10 message update iterations per level. In generating all above stereo results we used the following set of parameters: $(DISC\_K, \lambda) = (6.0, 0.2)$.
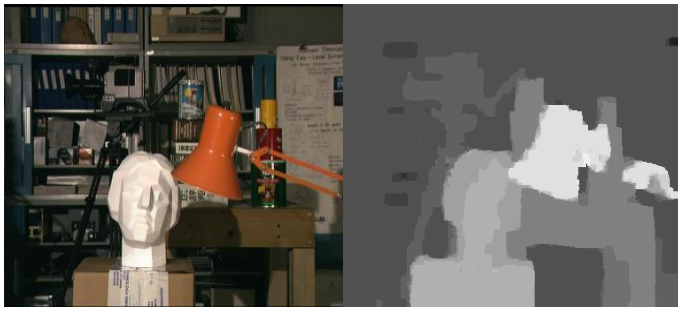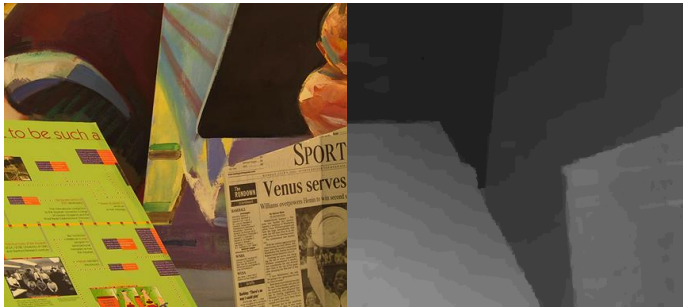
Figure 5. Result for Tsukuba

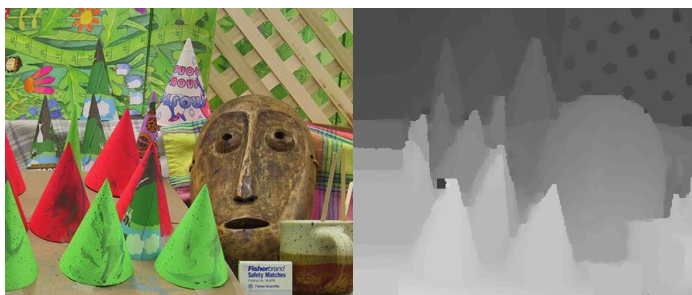**Figure 6. Result for Venus**

Figure 7. Result for Teddy

Figure 8. Result for Cones

8. Conclusions

Two methods are proposed to improve stereo matching. First, a novel dissimilarity measure that compares two pixels has been introduced. Energy equation is constructed by proposed dissimilarity measure and applied to belief propagation stereo matching algorithm. Some parameters of BP can be figure out due to introducing this measure. The measure can be used in most stereo algorithms. Second, disparity gradient constraint is utilized to refine disparity map and extrapolate the disparities into those regions where

disparities are not produced properly. The input initial disparity map can be produced by most global algorithms such as BP and graph-cut. The performance of our methods have evaluated by the Middlebury stereo evaluation test bed. Experiments demonstrate that the methods can yield excellent results.

Acknowledgements

References

[1]   A. Klaus, M. Sormann and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure", Proceedings of the 18th International Conference on Pattern Recognition. ICPR 2006, vol. 3, **(2006)**, pp. 15-18.
[2]   Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, **(2010)**, pp. 1458-1465.
[3]   D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", International journal of computer vision, vol. 47, **(2002)**, pp. 7-42.
[4]   Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, **(2001)**, pp. 1222-1239.
[5]   V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts", Computer Vision-ECCV 2002, **(2002)**, pp. 185-208.
[6]   J. S Yedidia, W. T Freeman, and Y. Weiss, "Generalized belief propagation", Proceedings of the NIPS, vol. 13, **(2001)**.
[7]   P. F Felzenszwalb and D. P Huttenlocher, "Efficient belief propagation for early vision", International journal of computer vision, vol. 70, **(2006)**, pp.41-54.
[8]   Y. Xu, H. Chen, R. Klette, J. Liu and T. Vaudrey, "Belief propagation implementation using CUDA on an NVIDIA GTX 280", AI 2009: Advances in Artificial Intelligence, **(2009)**, pp. 180-189.
[9]   S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, **(2002)**, pp. 401-406.
[10] J. Pearl, "Probabilistic reasoning in intelligent systems: networks of plausible inference", Morgan Kaufmann, **(1988)**.
[11] W. T Freeman, E. C Pasztor and O. T Carmichael, "Learning low-level vision", International journal of computer vision, vol. 40, **(2000)**, pp. 25-47.
[12] "Development Core Team R. R: A Language and Environment for Statistical Computing", Vienna, Austria, **(2010)**.
[13] C. Sun, "Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques", International Journal of Computer Vision, vol. 47, **(2002)**, pp. 99-117.
[14] P. Burt and B. Julesz, "A disparity gradient limit for binocular fusion", Science, vol. 208, **(1980)**, pp. 615-617.
[15] S. B Pollard, J. E Mayhew, and J. P Frisby, "PMF: a stereo correspondence algorithm using a disparity gradient limit", Perception, vol. 14, **(1985)**.
[16] H. P Trivedi and S. A Lloyd, "The role of disparity gradient in stereo vision", Perception, vol. 14, **(1985)**, pp. 685-690.
[17] Z. Zhang and Y. Shan, "A progressive scheme for stereo matching", 3D Structure from Images—SMILE 2000, **(2001)**, pp. 68-85.
[18] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR'07, **(2007)**, pp. 1-8.
[19] D. Scharstein, H. Hirschmuller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang and P, Westling, "High-resolution stereo datasets with subpixel-accurate ground truth", Proceedings of German Conference on Pattern Recognition (GCPR 2014), Germany, **(2014)**.

Authors

Junxue He, he is an associate professor in College of Electrical and Information Engineering, Lanzhou University of Technology, P.R. China. He received his Ph.D. degree from Lanzhou University of Technology in 2011. His research interests include computer vision, image processing, three-dimensional reconstruction, intelligent control and image-based robot navigation. His current research focuses on stereo matching and depth map optimizing. He is the corresponding author of this paper.

Erchao Li, he is an associate professor in College of Electrical and Information Engineering, Lanzhou University of Technology, P.R. China. He received his Ph.D. degree from Lanzhou University of Technology in 2011. His research interests include robot force control, computer vision, image processing and robot control. His current research focuses on robot vision and intelligent control.

Zhanming Li, he is a professor in College of Electrical and Information Engineering, Lanzhou University of Technology, P.R. China. His research interests include control theory and engineering, intelligent information processing and pattern recognition. His current research focuses on computer vision and image processing.