RADIOMETRIC INVARIANT STEREO MATCHING BASED ON RELATIVE GRADIENTS

Xiaozhou Zhou and Pierre Boulanger

Department of Computing Science, University of Alberta
Edmonton, Alberta, Canada
{xzhou3, pierreb}@ualberta.ca

ABSTRACT

Colors images produced by current sensors are affected by many environmental factors resulting in the fact that even for the same illumination conditions, corresponding pixels in stereo pairs cannot be guaranteed to have the same color. In many cases, color based stereo matching is not a good choice to compute a good disparity maps. In this paper, we propose to solve this problem by using a relative gradient algorithm. Boundary and low texture problems are resolved by using a Gaussian weighting function and by limiting the search range. The experimental results show the proposed local method is effective and robust, and even outperforms some of the well-known global methods.

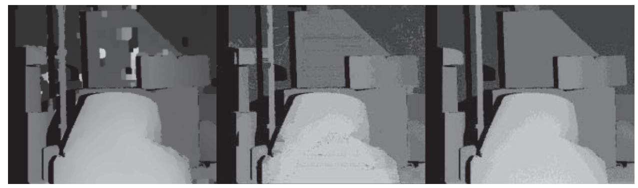
Index Terms-stereo, radiometric invariant, relative gradients

1. INTRODUCTION

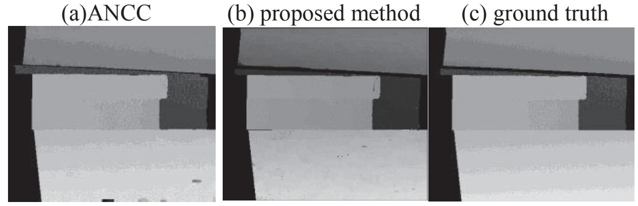
The aim of stereo matching is to find the corresponding pixels from two images. Most stereo matching methods published so far are based on colors matching [1] which assume that the object surface is a Lambertian surface. In the case of a real Lambertian, the color of each 3D point captured from different cameras will be constant. This is due to the fact that Lambertian surface reflects the incident light in all directions with the same power making it viewpoint invariant. However, most real-world objects are not Lambertian and reflect light that is view dependent. View dependent color images vary between multiple cameras and makes simple color matching algorithms unreliable.

In this paper, we propose a novel radiometric invariant stereo matching algorithm based on a relative gradients algorithm. We will demonstrate that relative gradients algorithm eliminates not only the problem of view dependence, but is also robust to color changes caused by illumination variations. Moreover, we also propose a solution to the boundary and low texture problems which exist in all local stereo matching methods. Figure 1 shows two results produced by the proposed method.

The rest of this paper is organized as following: Section 2 introduces the related work about radiometric invariant stereo matching. The propose method is illustrated in Section 3 in details. In Section 4, we list the results of the proposed method and compare them with the state-of-the-arts. Section 5 concludes this paper.



(a)ANCC (b) proposed method (c) ground truth



(a)ANCC (b) proposed method (c) ground truth

Fig.1. The comparison of *ANCC* [3] and proposed method

2. RELATED WORK

One possible solution to handle the radiometric variance problem is to design a color conversion scheme that transforms colors from the standard *RGB* space to another space in which the corresponding pixels will have the same value. The color conversions are usually based on two lighting models. The first lighting model only contains a view independent term, called body reflection:

$$C = eb \,, \tag{1}$$

where e is the illumination energy and b is the surface albedo. The radiometric variance happens only when illuminations energy e changes. Suppose the intensities of the *RGB* channels are: $\begin{cases} C_R = e_R b_R \\ C_G = e_G b_G \\ C_B = e_B b_B \end{cases}$ and if $e_R = e_G = e_B$, then one can define a normalized color space *rgb* as:

$$\begin{cases} r = \frac{C_R}{C_R+C_G+C_B} = \frac{b_R}{b_R+b_G+b_B} \\ g = \frac{b_G}{b_R+b_G+b_B} \\ b = \frac{b_B}{b_R+b_G+b_B} \end{cases} . \tag{2}$$

The normalized rgb space only depends on the surface albedo, making corresponding pixels independent of illumination conditions.

NCC is a more complex similarity function which is defined as:

$$NCC(p,d) = \frac{\sum_{q\in W_p}(I_L(q))(I_R(q-d))}{\sqrt[2]{\sum_{q\in W_p}(I_L(q))^2 \sum_{q\in W_p}(I_R(q-d))^2}}, \quad (3)$$

where $I_L(q)$ is the intensity of pixel q in the left image and W_p is a local window centered at p. *NCC* is well known to be invariant to linear change in illumination.

There are also some global methods based on the assumption (1). The work by Heo *et al.* [3] took the inner working of camera processing into consideration. The proposed transform is:

$$\begin{cases} I'_R(i,j) = \log(I_R(i,j)) - \frac{\log(I_R(i,j))+\log(I_G(i,j))+\log(I_B(i,j))}{3} \\ I'_G(i,j) = \log(I_G(i,j)) - \frac{\log(I_R(i,j))+\log(I_G(i,j))+\log(I_B(i,j))}{3} \\ I'_B(i,j) = \log(I_B(i,j)) - \frac{\log(I_R(i,j))+\log(I_G(i,j))+\log(I_B(i,j))}{3} \end{cases} \quad (4)$$

which linearize many of the non-linear transformation that happens in today's cameras such as: radiosity problems, gamma correction, and CCD sensitivity. In this method an Adaptive *NCC* (*ANCC*) is used as the main matching function and a disparity map is computed by using a Graph Cut algorithm. Another paper of Heo *et al.* [4] utilizes the space information by adding the *Scale Invariant Feature Transform*(SIFT) descriptor to construct the joint probability in a *mutual information* framework. Miled *et al.* [5] regularized the stereo matching problem as a convex optimization problem.

The second lighting model is:

$$C = m_b eb + m_s e. \quad (5)$$

The first term is the same as the first model (1). The second term is the surface reflection, which is view dependent. The term m_b is the weight coefficient for the view independent contribution and m_s is the weight coefficient for the view dependent contribution. Using this model, the Zero mean Normalized Cross-Correlation (*ZNCC)* uses intensity difference in a local window to achieve invariance:

$$ZNCC(p,d) = \frac{\sum_{q\in W_p}\left(\left(I_L(q)-\tilde{I_L}(W_p)\right)\left(I_R(q-d)-\tilde{I_R}(W_{p-d})\right)\right)}{\sqrt[2]{\sum_{q\in W_p}\left(I_L(q)-\tilde{I_L}(W_p)\right)^2 \sum_{q\in W_p}\left(I_R(q-d)-\tilde{I_R}(W_{p-d})\right)^2}}. \quad (6)$$

$\tilde{I}(W_p)$ is the average intensity in the window centered at p.

In addition, there are other methods which are not designed for an explicit lighting model, such as the *rank* and the *census* method. The *Rank* method [2] uses intensity ranks to replace the intensities of pixels. The rank is actually the number of pixels whose intensities are less than the central pixel in a local window. The *Census* method [2] transforms a window into a bit string, where each bit corresponds with a pixel in the window. The bit is set to 1 if the intensity of this pixel is less than the central pixel; otherwise, set the bit to 0. Two bit strings are compared by Hamming distance.

3. PROPOSED METHOD

Color conversion based on the first model cannot handle the view dependent color. Moreover, some outstanding methods, such as *ANCC*, cannot process gray scale images. Therefore, in this paper, we propose a more general radiometric invariant stereo matching method based on the relative gradients [6][7], which takes both the view independent and view dependent colors into consideration, and is able to deal with both color and gray scale images.

The proposed method is based on the following lighting model:

$$C = m_b eb + m_s e + a = \underbrace{m_b eb + a}_{view\ independent} + \underbrace{m_s e}_{view\ dependent}. \quad (7)$$

Equation (7) is the most complete lighting model which is widely used in Computer Graphics. The second term a models diffuse lights reflected from other objects. Two terms in (7) have to be eliminated, $m_s e$ and a. The term $m_s e$ is view dependent. Although a is view independent, it is hard to estimate and should be eliminated as well.

The relative gradient of a pixel (i,j) is expressed as:

$$RG(i,j) = \frac{gradient(i,j)}{maxiaml\ gradient(i,j)+1}. \quad (8)$$

Assuming that the lighting geometries are the same in a 3*3 neighborhood, the subtraction operations in the gradient operation (the numerator of (8)) could counteract the terms $m_s e$ and a so that the pixels only contain view independent radiometric information. Following the gradient, the only difference between corresponding pixels is due to the illumination energy e in the first term.

By normalizing the gradient with the largest gradient value in a 3x3 neighborhood centered at (i,j), defined as:

$$\begin{aligned} &maxiaml\ gradient(i,j) \\ &\qquad = \max\big(gradient(i+m,j+n)\big), -1 \le m,n \le 1, \end{aligned} \quad (9)$$

one can get rid of the influence of the illumination energy after the division operation in (8).The scalar +1 is added to the denominator in case the maximal gradient is close to zero. Equation (8) is the relative gradient function for one channel. For color images, the difference of corresponding pixels in color images is the sum of the difference of relative gradient in each channel. Since the transform of relative gradients is independent to each pixel, stereo matching based on relative gradients can be solved locally or globally. For the speed concern, we implement the proposed method as a window (local) based stereo matching and *Sum of Absolute Differences* (SAD) is used as the matching function, although the relative gradients term could also be integrated as the data term in a global energy function.

However, local stereo matching methods are notorious for the boundary and low-texture areas problems. The proposed method is not an exception. Windows centered at boundaries usually contain multiple disparity planes. The discontinuity of disparities inside a window could interfere with the calculation of the matching cost function. Global methods solve the above two problem by giving heavy penalties to disparity discontinuity. In this paper, we use Gaussian weighing function $G(q)$ to assign a weight to each pixel according to the color distance. This is based on the assumption that if a pixel is far away from the central pixel in color, it probably from other disparity planes. In the reference (left) image, if the color of pixel q is close to the central pixel p, Gaussian weighting function gives a high weight to the matching cost $C(q,d)$; otherwise, pixel q contributes very little to the total matching cost $C(W_p)$:

$$C(W_p) = \sum_{q\in W_p} G(q)C(q,d) = \sum_{q\in W_p} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(I_l(q)-I_L(p))^2}{2\sigma^2}\right) \quad |RG_L(q) - RG_R(q-d)|. \quad (10)$$

In addition, we suggest a way to correct the mismatches caused by low textures. It is based on the observation that the disparities in the same disparity plane change smoothly, thus the disparity of a pixel should be close to its neighbors. Therefore, for each mismatched pixel detected by cross checking [8], we set a new search range and update these pixels in the second round to avoid the black and white noises. The search range is between the previous and next correct disparities on the same epipolar line:

$$search\,range(mismatch) \in [\min(previous(dis), next(dis)), \max(previous(dis), next(dis))]. \quad (11)$$

If matching cost is still too large, the depth of this pixel could be filled by $\min(previous(dis), next(dis))$or by image inpainting technique [11].

4. EXPERIMENTS

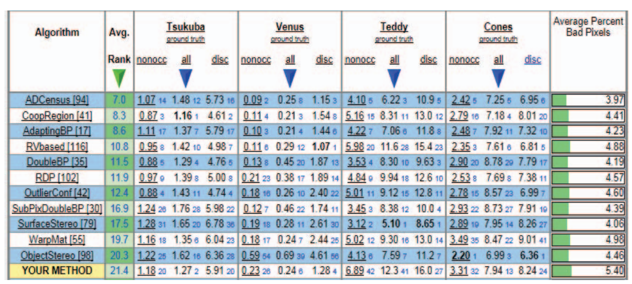
First, we test our proposed method by comparing with the ground truth visually. Disparity maps are gray scale images whose intensities represent the depth information. The darker the pixel is, the further the object is from the viewer. The resulting disparities are compared with the ground truth pixel by pixel. Test images and ground truth images are all from Middlebury dataset. Figure 1 is the comparison between the proposed method and *ANCC*. Two test images are lamp and wood. *ANCC* uses the global optimization to solve the matching problem, which is one of the best so far. Generally, the global methods should be better than the local ones [10]. However, because of the improvements to the boundary and low texture areas, our local method is competitive to *ANCC*. Figure 3 and Figure 4 list the results of commonly used local methods for different illumination conditions. All the local methods use *Winner-Takes-All* (WTA*)* strategy to select the disparity. The window size is 25. One can see that the boundaries in the proposed method are clearer than other methods and the disparities are visually filled well in the low texture areas as well.

Table.1. Error Ratio (%) comparison of local radiometric invariant stereo matching methods (without occlusions)

	art	moebius	wood	rock	book	aloe	baby
rgb	24.83	33.75	73.42	41.06	24.33	12.34	47.65
RANK	38.44	54.42	14.58	7.69	47.05	31.75	32.21
CENSUS	44.81	54.37	25.50	20.91	46.39	43.90	36.02
NCC	27.16	22.75	18.33	13.66	18.01	19.43	12.89
ZNCC	28.02	21.71	21.47	15.61	18.21	25.41	19.14
ANCC	7.15	7.95	3.94	6.46	11.09	3.12	4.75
proposed	6.48	6.40	0.53	2.66	6.40	3.67	3.33

To further demonstrate how well the algorithm works, we compared the results with standard methods with various images and summarize the quantitative analysis in Table.1. The test images all have different illuminations. The error ratio is the percentage of the wrong matches (exclude occlusions) over all pixels. One can see from Table 1, the error ratios of our method are much smaller than others for all the test cases.

Figure 2 is the evaluation on Middlebury test bed [9]. Our proposed method ranks 12th out of 111. Although the rank seems competitive, it cannot fully embody the advantage of our proposed method. Because the proposed method is designed for stereo images suffering from different illuminations, while the four test images Tsukuba, Venus, Teddy, and Cones are at the same illumination.

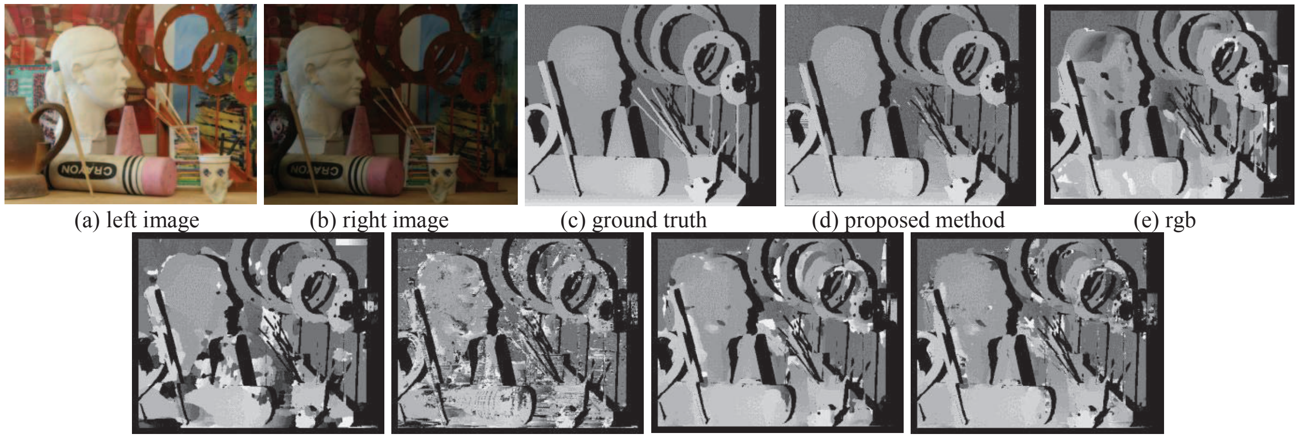


Algorithm	Avg. Rank	Tsukuba (ground truth) nonocc	all	disc	Venus (ground truth) nonocc	all	disc	Teddy (ground truth) nonocc	all	disc	Cones (ground truth) nonocc	all	disc	Average Percent Bad Pixels
ADCensus [94]	7.0	1.07 14	1.48 12	5.73 16	0.09 2	0.25 8	1.15 3	4.10 5	6.22 3	10.9 5	2.42 5	7.25 5	6.95 6	3.97
CoopRegion [41]	8.3	0.87 3	1.16 1	4.61 2	0.11 4	0.21 3	1.54 8	5.16 15	8.31 11	13.0 12	2.79 16	7.18 4	8.01 20	4.41
AdaptingBP [17]	8.6	1.11 17	1.37 7	5.79 17	0.10 3	0.21 4	1.44 6	4.22 7	7.06 6	11.8 8	2.48 7	7.92 11	7.32 10	4.23
RVbased [116]	10.8	0.95 8	1.42 10	4.98 7	0.11 6	0.29 12	1.07 1	5.98 20	11.6 28	15.4 23	2.35 3	7.61 6	6.81 5	4.88
DoubleBP [35]	11.5	0.88 5	1.29 4	4.76 5	0.13 8	0.45 20	1.87 13	3.53 4	8.30 10	9.63 3	2.90 20	8.78 29	7.79 17	4.19
RDP [102]	11.9	0.97 9	1.39 8	5.00 8	0.21 23	0.38 17	1.89 14	4.84 9	9.94 18	12.6 10	2.53 8	7.69 8	7.38 11	4.57
OutlierConf [42]	12.4	0.88 4	1.43 11	4.74 4	0.18 16	0.26 10	2.40 22	5.01 11	9.12 15	12.8 11	2.78 15	8.57 23	6.99 7	4.60
SubPixDoubleBP [30]	16.9	1.24 26	1.76 28	5.98 22	0.12 7	0.46 22	1.74 11	3.45 3	8.38 12	10.0 4	2.93 22	8.73 27	7.91 19	4.39
SurfaceStereo [79]	17.5	1.28 31	1.65 20	6.78 36	0.19 18	0.28 11	2.61 30	3.12 2	5.10 1	8.65 1	2.89 19	7.95 14	8.26 27	4.06
WarpMat [55]	19.7	1.16 18	1.35 6	6.04 23	0.18 17	0.24 7	2.44 25	5.02 12	9.30 16	13.0 14	3.49 35	8.47 22	9.01 41	4.98
ObjectStereo [98]	20.3	1.22 25	1.62 16	6.36 28	0.59 54	0.69 39	4.61 56	4.13 6	7.59 7	11.2 7	2.20 1	6.99 3	6.36 1	4.46
YOUR METHOD	21.4	1.18 20	1.27 2	5.91 20	0.23 26	0.24 6	1.28 4	6.89 42	12.3 41	16.0 27	3.31 32	7.94 13	8.24 24	5.40

Fig.2. The rank of the proposed method on Middlebury

5. CONCLUSION

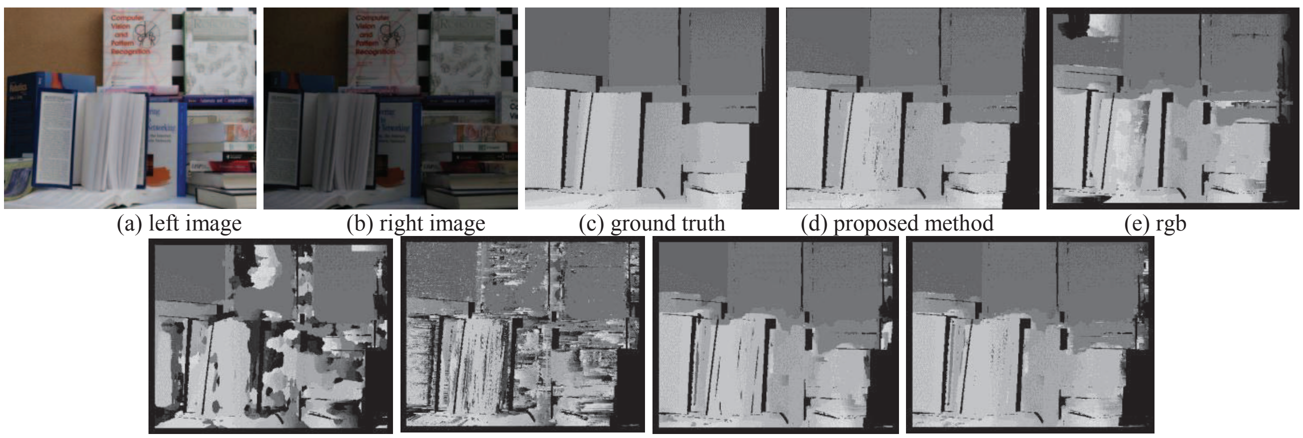
This paper introduces a novel radiometric invariant stereo matching algorithm. It is based on the relative gradients to eliminate the radiometric variance. We also propose to use Gaussian weighting function and search range limitation to improve the disparity estimation at the boundary and low texture areas. Although the proposed method belongs to the category of local stereo matching, the performance compares favorably to global methods. The experiments demonstrate the effectiveness and robust in both visual and quantitative aspects.



(a) left image (b) right image (c) ground truth (d) proposed method (e) rgb

(f)RANK (g) CENSUS (h) NCC (i) ZNCC

Fig. 3. The visual results of "art" in different illuminations



(a) left image (b) right image (c) ground truth (d) proposed method (e) rgb

(f) RANK (g) CENSUS (h) NCC (i) ZNCC

Fig. 4. The visual results of "book" in different illumination

6. REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision, vol.47, pp.7-42,2002.

[2] R. Zabih and J. Woodfil, "Non-parametric local transforms for computing visual correspondence," In Proceedings of European Conference of Computer Vision, pp. 151-158, 1994.

[3] Y.S. Heo, K.M. Lee, and S.U. Lee, "Illumination and Camera Invariant Stereo Matching," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.1-8, 2008.

[4] Y.S. Heo, K.M. Lee, and S.U. Lee, "Mutual Information-based Stereo Matching Combined with SIFT Descriptor in Log-chromaticity Color Space," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 445-452, 2009.

[5] W. Miled, J.C. Pesquet, and M. Parent, "A Convex Optimization Approach for Depth Estimation under Illumination Variation," IEEE Transaction on Image Processing, vol. 18, No.4, pp. 1574-1608, 2009.

[6] S.H. Lai and S.D. Wei, "Reliable image matching based on relative gradients," In Proceedings of International Conference on Pattern Recognition, pp, 802-805, 2002.

[7] Z. Hou and W.Y. Yau, "Relative gradients for image lighting correction," In Proceedings of International Conference on Acoustics, Speech and Signal Proceesing, pp.1374-1377, 2010.

[8] P. Fua, "A parallel stereo algorithm that produces dense depth maps and preserves image features," Machine Vision and Applications, vol. 6, pp.35–49, 1993.

[9] http://www.vision.middlebury.edu/stereo/eval/.

[10] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[11] A. Criminisi, P. Perez, and K. Toyama, "Object removal by exemplarbased inpainting," In Proceedings of IEEE Computer Conference on Computer Vision and Pattern Recognition, vol.2, pp.721-728, 2003.