ILLUMINATION INVARIANT STEREO MATCHING BASED ON NORMALIZED MUTUAL INFORMATION AND CENSUS METHODS

Xiaozhou Zhou and Pierre Boulanger
Dept. Computing Science, University of Alberta

ABSTRACT

Stereo matching aims at finding corresponding pixels from two or more images where distance information is computed by triangulation. Due to different illuminations, intensities are not reliable to be used to search for corresponding pixels. In this paper, we propose a novel local matching method which is capable of dealing with illumination variations between cameras. This new method can distinguish pixels in the same window but with different disparities allowing for larger window to be used. Moreover, a more precise matching cost function is used to find the correspondence. The proposed method is compared with five other local stereo methods and is proven to be more robust and effective under various illumination conditions.

Keywords **—** computer vision, stereo matching, illumination invariance, normalized mutual information

1. INTRODUCTION

Stereo matching is one of the most challenging research areas in compute vision. Its aims at finding the corresponding pixels from two or more images and at reconstructing distance information for each pixel for applications in 3D games, 3D movies, 3D modeling, and 3D TV. In the last decades, hundreds of methods [1] have been proposed to improve its performance, but most stereo matching methods are based on the assumption that the corresponding pixels have exactly the same intensity. However, even in the same illumination environment, the corresponding pixels may have slight differences, which are caused by different camera settings and non-Lambertian surfaces.

In this paper, we propose a novel local stereo matching method which can deal with illumination variations. The matching cost function is used to compare two local windows one for each image. In stereo matching, the matching cost function should be precise and sensitive. In this paper, we use Normalized Mutual Information (*NMI*) to find the correspondence. Pixels inside a window are divided into three categories 0, 1,or 2 to deal with non-compatible disparities.

The rest of this paper is organized as follows. Section 2 introduces the related literatures on local illumination invariant stereo matching method. Section 3 present our proposed method in detail and Section 4 shows a comparison of our method with five other commonly used methods. We then conclude in Section 5.

2. RELATED WORK

Numerous methods have been proposed to solve the illumination invariance problem. Local optimized methods such as the Rank method [2], the Census method [2], and the Mutual Information [3] method have been used successfully. Other global methods are, such as the work by [5], [6], and [7] built an energy function based on mutual information analysis and then use Graph Cut techniques to minimize the energy function globally. In our on-going project, we want to take advantage of the parallel structure of local optimized methods to achieve real-time performance (30Hz) using Graphics Processing Unit (*GPU*) and global optimization methods are almost impossible to implement in real-time. Therefore, in this paper, we concentrate our attention on methods based on local stereo matching method alone.

The Rank method used the Sum of Squared Differences (*SSD*) or Sum of Absolute Differences (*SAD*) functions to compare two windows around a pixel, but the elements in windows were not intensities [2]. The intensities of pixels are replaced by their intensity ranks within a local window. Comparison depends on the orderings of intensity values instead of intensities themselves. The rank is actually the number of pixels whose intensities are less than the central pixel, and is expressed by:

$$I_{\mathrm{rank(p)}} = \sum_{q\in Wp} T[\mathrm{I(q)} < I(p)]. \qquad (1)$$

The function T [ ] is equal to 1 if the argument is true; otherwise it is 0. The parameter Wp is the window size centered at p.

The Census method transformed a window into a bit string, where each bit corresponds to a pixel in the window [2]. The bit is set to 1 if the intensity of this pixel is less than the central pixel; otherwise, the bit is set to 0.

$$R_p(i) = \begin{cases} 1, \ T[I(i) < I(p)] \\ 0, \ otherwise \end{cases} . \qquad (2)$$

The two bit strings are then compared using Hamming distance. Both the rank and census methods do no rely directly on the pixel intensities. As long as illumination changes are monotonic, the rank of a pixel in a window stays the same. These methods are more reliable to match images under different lighting environments.

Mutual Information method was first introduced by Shannon in his information theory [4] to measure the dependence of two random variables. The greater the Mutual Information function is, the more similar the two variables are. The Mutual information function is expressed as:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \ , \qquad (3)$$

where $H(X)$ and $H(Y)$ are the marginal entropies of X and Y, and $H(X,Y)$ is the joint entropy. Please refer to [3] for more details. For images, the Mutual Information function reaches the maximum if two images are exactly the same. Therefore, Mutual Information function is used as the matching cost in [3] to find the best matched window,

$$C\ (i,j,d) = MI(W_L(i,j), W_R(i-d,j)) \qquad (4)$$

where $W_L(i,j)$ is a local window in left image which centered at (i,j).

The Normalized Cross-Correlation (*NCC*) function is invariant if the illuminations in two images follow a linear relation,

$$NCC(p,d) = \frac{\sum_{q\in W_p}(I_L(q))(I_R(q-d))}{\sqrt[2]{\sum_{q\in W_p}(I_L(q))^2\ \sum_{q\in W_p}(I_R(q-d))^2}} \qquad (5)$$

where $I_L(q)$ is the intensity of pixel q in the left image and W_p is the size a local window centered at p.

3. PROPOSED METHOD

3.1 Normalized Mutual Information (NMI)

Unfortunately, the precision of the Mutual Information function depends on statistics computed from the overlapping regions. Small overlapping regions, such as local windows, reduce the statistical samples and then decrease the precision of function. Thus, we propose to use *NMI* as the matching cost function：

$$NMI(X,Y) = (H(X) + H(Y)) / H(X,Y) \quad . \qquad (6)$$

The maximum value of $NMI(X,Y)$ is 2. *NMI* has been shown to be a better measurement for small overlapping areas than Mutual Information [8].

3.2 Proposed Method

Fig. 1 shows the dataflow of proposed method.

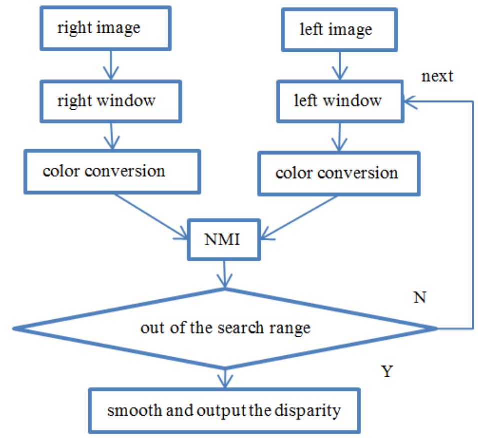


Figure 1: Dataflow of Proposed Method

This algorithm contains four main steps:

Step 1: In our experiments, the right image is set as the reference image. Similar to other local window based stereo algorithms, the local window size is *m* x *n* (25 x 25 in this paper). For each window centered at (s,t) in the reference (right) image, we perform a range search for windows centering on the same scan line in the left image.

Step 2: Before the calculation of *NMI*, the colors are converted to 0, 1 or 2. The following Gaussian weight function is used to divide pixels in a local window into three categories:

$$w(x,p) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(I_{R/L}(x) - I_{R/L}(p))^2}{2\sigma^2}\right) \qquad (7)$$

where σ is the measure of the width of the distribution. p is the central pixel in a window and x is any other pixel in the same window. $I_{R/L}(x)$ is the intensity of pixels in the right or left image.

Generally, in local stereo matching methods, the size of window is expected to be as large as possible, since large windows usually contain more information. However, large windows also include pixels with different disparity planes which should not be included in the matching function. The census method only separates the pixels whose colors are greater or less than the central pixel, which cannot distinguish pixels from different disparity planes. The proposed Gaussian weighting function not only separates those pixels apart to resist the illumination changes, but also looks for the pixels on different disparity planes.

$$pixel\ (j) = \begin{cases} 0,\ if\ I(j) > I(p) \text{and Gaussian weight} \geq \text{threshold} \\ 1,\ if\ I(j) < I(p) \text{and Gaussian weight} \geq \text{threshold} \\ 2, \qquad \text{if Gaussian weight} < \text{threshold} \end{cases} \qquad (8)$$

Table. 1. The error ratio of test images (%)

	Aloe	Baby	Book	Wood	Art	Moebius	Reindeer
SSD	75.28	80.94	77.95	75.43	48.74	57.36	54.47
Rank	31.32	25.32	36.53	8.79	25.42	41.22	18.48
Census	32.51	28.56	35.84	18.66	29.95	41.48	29.40
NCC	15.68	11.52	13.80	13.61	26.92	22.23	25.09
MI	8.96	8.61	12.60	33.63	17.85	14.72	25.89
Proposed	6.32	7.12	8.58	6.52	9.97	9.43	12.99

j is any pixel in a local window and p is the central pixel in the same window. $I(p)$ is the intensity of pixel p. Since the Gaussian weighting function is always greater than 0, the threshold is set as a value higher than 0. For example, the threshold is set to 0.01 in the experiments. If the Gaussian weight of a pixel is lower than the threshold, then this pixel is classified as another disparity planes and marked as 2. Only the pixels from the same disparity plane with the central pixel will be classified as 0 or 1 if its intensity is greater or less than the intensity of the central pixel. In this way, pixels from different disparity planes are marked with different labels and will not disturb the calculation of matching cost function allowing for larger windows.

Step 3: Two windows (right-left) are compared using *NMI*. The window in the left image which has the maximal *NMI* with the window in the reference image is recorded and the central pixels of two windows are considered as corresponding pixels. The disparity of corresponding pixels is the absolute difference of their *X* coordinates.

Step 4: A median filter is used to smooth the disparity map to eliminate cracks and noise and the smoothed map is output.

4. EXPERIMENTS

We use the datasets from Middlebury [9] for testing. The images in Fig. 2(a) and Fig. 2(b) have different illumination conditions and have been rectified. Six methods are compared in Fig. 2(d) to Fig. 2(i): (d) proposed method, (e) *SSD*, (f) Rank, (g) Census, (h) Mutual Information (*MI*), and (i) *NCC*. The intensities of pixels in the results are disparities which could be used to compute distance. The darker the pixel is, the further the pixel is from the viewer. The results are compared with the ground truth image illustrated at Fig. 2(c).

SSD is one of the matching costs functions which perform matching using pixel intensities directly. The results show that when illumination changes, *SSD* fails completely. The rank method just simply replaces the intensity by its rank and the census methods use the relation with the central pixel and are both not robust to noise. Since real illumination conditions are more complicate than a linear model, *NCC* is not a general matching cost function for illumination invariance, although it works better than the rank and census methods. Moreover, pixels in a local window are not sufficient for computing the Mutual Information function. Among the six stereo matching methods, our proposed method got the best results for all test images. We tested more images and listed at Table 1 the error ratios of the six methods for quantitative evaluation. The error ratio measures the percentage of the wrong disparities over all computed disparities. The error ratio of the proposed method is the lowest for all methods. One could also conclude that the proposed method is the most robust because the error ratios do not vary too much on all test images.

5. CONCLUSION

This paper presents a novel local stereo matching for illumination invariance. This method uses NMI instead of Mutual Information for more precise comparison of two windows. In addition, pixels are converted into 3 categories to resist the illumination changes and eliminate the negative effects caused by disparity changes in a window. The experiments demonstrate that the color conversion is a reasonable replacement for colors, and NMI is more efficient than other popular matching costs functions.

6. REFERENCES

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

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

[3] G. Egnal, "Mutual Information As a Stereo Correspondence Measure," Technical Report MS-CIS-00- 20, Computer and Information Science, University of Pennsylvania, USA, 2000.

[4] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1963.

[5] J. Kim, V. Kolmogorov and R. Zabih, "Visual Correspondence using Energy Minimization and Mutual Information",*in Proceedings of* Intern*ational Conference on Computer Vision*, 501-508, 2003.

[6] 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*, 1-8,2008.

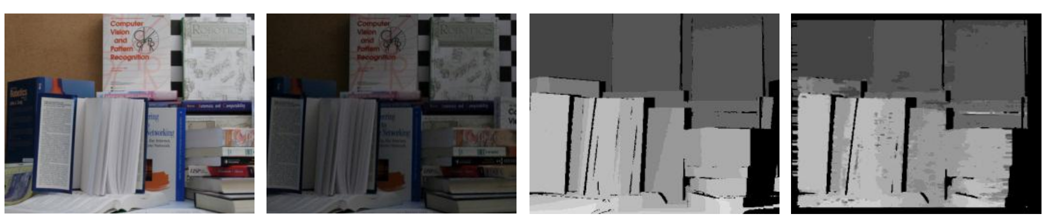
[7] Y. S. Heo, K. M. Lee and S. U. Lee, "Mutual Information-based Stereo Matching Combined with SIFT Descriptor in Log-chromaticity Color Space", *in Proceedings of Computer Vision and Pattern Recognition*, 445-452,2009.

[8] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3-D medical image alignment ", Pattern Recognition, 32(1), pp. 71-86, 1999.

(1) Aloe image:

(2) Art:

(2) Book image:



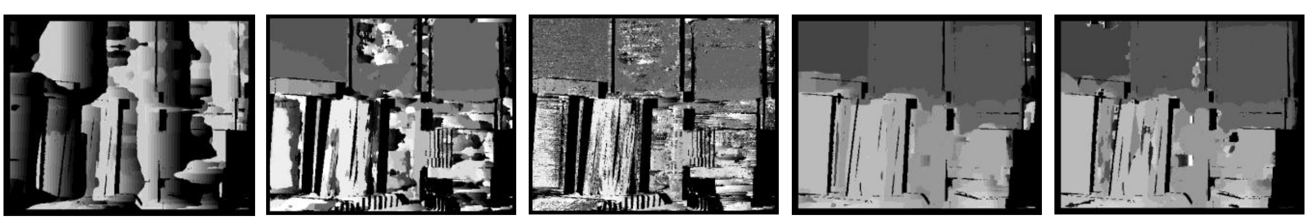


Figure 2: Three Set of Results comparisons under different illuminations: (1) Aloe, (2) Art and (3) Book. The first row in each set is (from left to right): (a) Left Image in Illumination 1 (b) Right Image in Illumination 2 (c) Ground Truth (d) Proposed Method. The second row in each set is (from left to right): (e) SSD (f) Rank (g) Census (h) NCC (i) Mutual Information.