A FAST HIERARCHICAL RADIOMETRIC INVARIANT STEREO MATCHING ALGORITHM

Xiaozhou Zhou and Pierre Boulanger

Department of Computing Science
University of Alberta
Edmonton, Alberta, Canada

ABSTRACT

Stereo matching aims at finding the corresponding pixels between two images. Most methods in the literatures are based on pixel intensity comparison. When images are taken at different illumination conditions or using different sensors, it is very unlikely that the corresponding pixels will have the same intensity creating false correspondences if it is only based on intensity matching functions alone. In this paper, a novel hierarchical radiometric invariant method is proposed to solve this problem. Pixels on different disparity planes in a local window are assigned the same label as the central pixel to solve the boundary problem. A more robust mutual information similarity function is used to compare the similarity of windows. A hierarchical structure is proposed to reduce the computational burden caused by using mutual information similarity function. Experimental results demonstrate that both the visual quality and quantitative analysis of the proposed algorithm outperforms most commonly used algorithms in the literatures.

1. INTRODUCTION

Stereo matching looks for corresponding pixels between two images using some similarity functions. By triangulating the corresponding pixels can be converted to a depth map of the scene and used for 3D reconstruction, view synthesis, and free viewpoint video applications. In the past three decades, most stereo matching algorithms [1] have been proposed based on two assumptions: (1) the corresponding pixels have the same intensities; (2) the object surfaces can be represented by a Lambertian model which reflects the light in all directions with the same power. However, the real world is much more challenging. Since most real-world materials are not Lambertian, the perceived illumination from different cameras will be different as it depends on the camera's viewpoint and illumination conditions. Most of the existing stereo matching methods are based on intensity comparison, independently if they use local or global optimization to compute correspondence. Technically, intensity based stereo matching methods cannot work well if corresponding pixels have different intensities and will result in numerous false correspondences.

In this paper, we propose a novel local stereo matching algorithm which could handle radiometric variance. Pixels in a window are marked by 3 indexes according to which disparity plane they are on and the relations with the central pixel. Mutual Information is used to compare two bit strings after the 3-index color conversion. The proposed algorithm runs hierarchically on three levels of resolutions. The hierarchical structure helps to reduce the increased computation burden by using a mutual information function. Since most correspondences can be found at low resolution levels, the running time is reduced by 20 times on average. Holes and occlusions are filled using image inpainting techniques.

The paper is organized as following: Section 2 describes several commonly used radiometric invariant local stereo matching methods. Section 3 explains the proposed algorithm in details and Section 4 compares the results of our algorithm with six methods in the literature. We also prove the quality of the our proposed method by evaluating our results on Middlebury test bed [18]. Finally, we conclude this paper in Section 5.

2. RELATED WORK

In this section, we introduce some existing local and global algorithms which are used to handle radiometry variant stereo matching.

Compared to the general window based stereo matching functions such as: Sum of Squared Differences (SSD) or Sum of Absolute Differences (SAD), Normalized Cross Correlation (NCC) works for linear scale changes, especially if the illuminations in the two images follow a multiple relation:

$$NCC(p,d) = \frac{\sum_{q\in W_p}(I_L(q))(I_R(q-d))}{\sqrt{\sum_{q\in W_p}((I_L(q))^2 \sum_{q\in W_p}(I_R(q-d))^2)}}. \quad (1)$$

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

Zero-mean Normalized Cross Correlation (ZNCC) uses the intensity difference of a pixel I and the mean value I' in a window to replace the intensity q and $q-d$ in (1), thus could handle linear scale illumination changes as well as different offsets. ZNCC could be expressed mathematically as:

$$ZNCC(p,d) = \frac{\sum_{q\in W_p}(I_L(q)-I'_L(W_p))(I_R(q-d)-I'_R(W_{p-d}))}{\sqrt{\sum_{q\in W_p}((I_L(q)-I'_L(W_p))^2\sum_{q\in W_p}(I_R(q-d)-I'_R(W_{p-d}))^2)}}. \quad (2)$$

Rank [2] replaces the intensity of a pixel by its intensity rank within a local window centered at itself. The intensity rank is actually the number of pixels whose intensities are less than the central pixel. Namely,

$$I_{rank(p)} = \sum_{q\in W_p} T[I(q) < I(p)]. \quad (3)$$

The function $T[*]$ returns 1 if the argument is true; otherwise returns 0. Wp is the window centered at pixel p.

Census [2] first converts each pixel to 0 or 1 following Equation (4) and then transforms a window into a bit string where each bit corresponds to a pixel in the window. The similarity of two bit strings are measured by Hamming distance. Namely,

$$R_p(i) = \begin{cases} 1, T[I(i) < I(p)]. \\ 0, otherwise \end{cases} \quad (4)$$

Both *rank* and *census* do no rely on the pixel intensity directly. As long as illumination changes are monotonic, the rank of a pixel in a window stays the same. Thus, these are more reliable to match images under different illumination conditions.

Mutual information[3] was first invented by Shannon to measure the dependence of two random variables X and Y . Mutual information is expressed as:

$$I(X,Y) = H(X) + H(Y) - H(X,Y), \quad (5)$$

where $H(X)$ and $H(Y)$ means the marginal entropies of X and Y, $H(X,Y)$ is the joint entropy. Namely,

$$H(X) = \sum_{x\in X} p(x)\log p(x), \quad (6)$$

$$H(X,Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log p(x,y). \quad (7)$$

Mutual information is one of the most popular methods to measure the similarity of images or signals. Mutual information have been widely used in image registration [4] [5] [6] [7] since 1995. In computer vision, Egnal [8] is the first to introduce mutual information to solve local stereo matching problem.

In the category of global stereo matching, Kim *et al.* [9] regularize mutual information in Markov Random Field (MRF) framework and use a Graph Cut algorithm to minimize the energy function for global optimization. The paper of Heo *et al.* [11] is based on the work of Kim *et al.* [9] and adds SIFT descriptor to find correspondence. Another paper of Heo *et al.* [11] uses Adaptive Normalized Cross Correlation (ANCC) as the data term in energy function. Miled *et al.* [12] estimate the illumination change as a linear model and formulate the stereo matching problem as a convex optimization problem. Disparity map is calculated by minimizing the energy function which is solved by parallel block iterative algorithm.

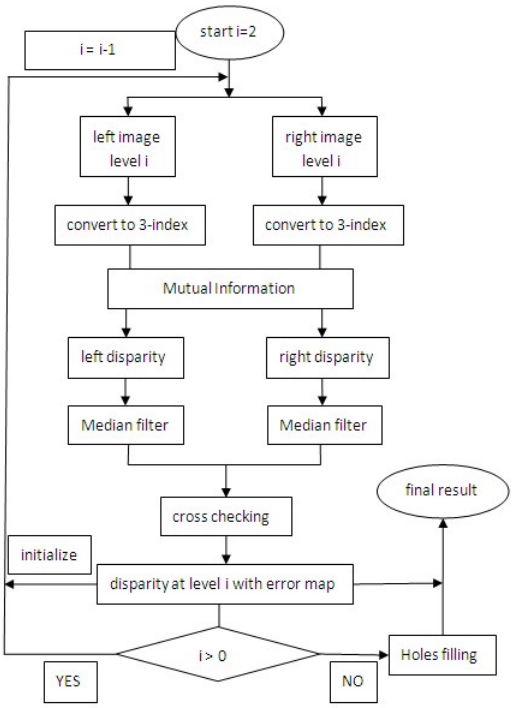


Fig. 1. The flowchart of proposed method

3. PROPOSED METHOD

Figure 1 shows the flowchart of the proposed method. The whole method runs on a hierarchical structure which contains three iterations. The resolutions of the original images are down sampled to half-size and quarter-size using a Gaussian pyramid. The first iteration starts from the lowest resolution. Then, the intermediate scale, and finally, in the third iteration, the full size images are processed.

The processing at each iteration are similar. We summarize the main processing steps here:

Step 1: **Initialization of disparity maps**. (Skip this step in the first iteration.)

Disparity map from previous iteration is used to initialize the disparity maps in the current iteration. Since the size of the image in current iteration is 4 times as large as the previous iteration, the disparity value of each pixel (x,y) is going to be assigned to 4 pixels (i,j), where $\lfloor i/2, j/2\rfloor = (x,y)$. Notice, only those disparity values marked as correct by cross checking are used for initialization. Leave the other pixels blank.

Step 2: **Window selection and comparison**. Only those pixels marked as error will be recalculated. In the first iteration, all pixels in the images are marked as error.

In window based stereo matching, it is straightforward that large window includes more details and leads to better results. However, the large window also includes pixels on different disparity planes (also called bad pixels).

Those bad pixels prevent two windows having high similarity even they are centered at a pair of corresponding pixels. In order to do so we should first detect bad pixels before comparing two windows. In this paper, we propose a 3-index color conversion based on *census*, which could not only get rid of the dependency on intensities but also reduce the impact brought by bad pixels. Zhou *et al.* [13] suggeste Gaussian weighting function and Yoon *et al.* [19] use Gestalt Grouping to classify pixels, but the color difference is a simpler and more direct metric. Color difference in RGB color space is used to measure the distance between a pixel in a local window and the central pixel. The color difference $CD(p,q)$ is calculated by:

$$CD(q,p) = \sqrt{(R_q - R_p)^2 + (G_q - G_p)^2 + (B_q - B_p)^2}. \tag{8}$$

p is the central pixel in a window while q is another pixel in the same window. As described in Equation (9), if the *CD* is equal or lower than the threshold, we assume that this pixel is on the same disparity plane as the central pixel, and then divide them into 0,1,2 according to whether their colors are less, equal, or greater than the central pixel. If the *CD* is greater than the threshold, we assume that this pixel is from another disparity plane. Those pixels are marked as 1 as they have the same color with the central pixel. The function $I(p)$ represents the color of p.

$$pixel(j) = \begin{cases} 0, if I_{ref}(j) > I_{ref}(p) \& CD <= threshold \\ 1, if I_{ref}(j) = I_{ref}(p) \& CD <= threshold \\ 2, if I_{ref}(j) < I_{ref}(p) \& CD <= threshold. \\ 1, if CD > threshold \end{cases} \tag{9}$$

In this classification, the bad pixels are marked the same index as the central pixel as they are on the same disparity plane. Hence, two corresponding local windows on boundaries will look similar because all bad pixels in both windows are indexed by the same number. Moreover, the 3-index color conversion assigns the same index to the corresponding pixels in differently illuminated images as long as the illumination change is monotonic.

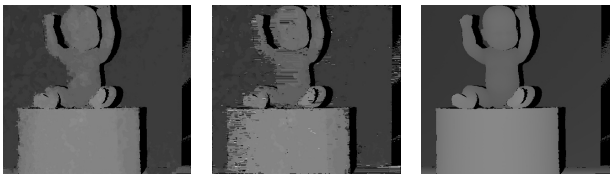
In addition,after the 3-index conversion, *census* transforms two windows into bit strings which are compared by Hamming distance. However, the Hamming distance only simply counts how many pixels are different, we suggest using mutual information which is a more precise measurement of information distance [17] as it is based on the entropy and explore the probability distribution. Corresponding pixels should have the maximal joint entropy. Figure 2 is an example showing the results using Hamming Distance (HD) or Mutual Information (MI) to compare bit strings in *census*.

Step 3: **Post processing**. The disparity maps are both smoothed by using a *median filter* in order to get rid of those too white or too dark noises. Pixels are marked as "correct" or "error" by cross checking [14].

Step 4: **Holes filling**. This step is only executed after the last iteration.

(a) HD (b) MI (c) groundtruth

Fig. 2. Similarity comparison: HD vs. MI.



(a) inpainting (b) interpolation (c) groundtruth

Fig. 3. Holes filling by interpolation and inpainting.

Some view of the 3D world can only be captured by a single camera; therefore, those parts cannot find the corresponding parts in the other image. This is called the occlusion. In global stereo matching, the occlusion parts are assigned disparity values by minimizing an energy function. In local stereo matching, the disparities in the occlusions are usually replaced by the background disparities, or the disparities of the closest pixels, or interpolate from the neighboring pixels. However, without precisely estimated 3D geometry, simple replacement or interpolation must cause lots of artifacts. Besides occlusions, there are also some mismatched pixels which need to be corrected. We find that the disparity map could be thought as the color texture. Each disparity plane has its own texture. Therefore, we use the most popular one - exemplar-based in-painting [16] to fill those holes (occlusions and mismatches). Each pixel in the holes is replaced by the pixel with the most similar surrounding patch. Figure 3 is a comparison of holes filled by interpolating the neighbor pixels and in-painting algorithm. There are obvious streaks in Figure 3(b) caused by interpolation while the disparity map of Figure 3(c) is smoother and continuous.

4. EXPERIMENTAL RESULTS

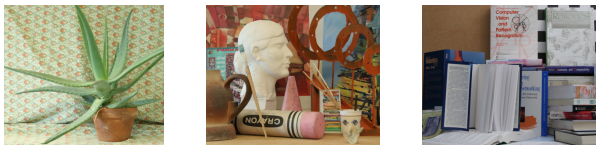
We use the Middlebury database [15] for testing our algorithm and rank its results compared to the best algorithms found in the literature.The results will be compared in three aspects: visual, quantitative, and speed. We will demonstrate that our method produce some of the best results without a huge reduction in processing speed.

4.1. Comparison of Visual Quality

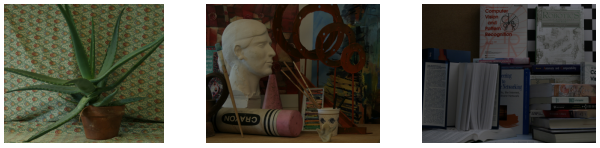
Seven local stereo matching algorithms are compared in this section: SAD, *rank*, *census*, NCC, ZNCC, mutual information [8], and the proposed algorithm. Figure 4 shows three test images in different illumination conditions. The right images are set as reference images. For

Table 1. Quantitative analysis using MAE measure

	aloe	baby	book	wood	art	moebius	cloth	reindeer
SAD	18.98	22.97	25.95	24.54	11.53	11.48	17.87	14.65
Rank	9.67	9.79	13.13	3.25	9.97	15.53	30.07	7.05
Census	13.36	9.81	13.28	6.29	11.59	16.01	23.95	11.26
NCC	4.28	4.19	2.74	3.54	8.60	5.12	1.80	6.31
ZNCC	4.58	4.23	6.35	9.02	8.19	5.49	3.86	5.49
Mutual Information	2.16	3.67	3.70	11.31	7.29	3.75	0.44	9.23
proposed	0.77	0.8	0.98	2.52	1.53	1.24	0.06	1.97



(a) aloe left (b) art left (c) book left



(d) aloe right (e) art right (f) book right

Fig. 4. Test images at different illumination conditions.

all the methods used in the comparison, we set a window size of 25 pixels and use a winner-takes-all to choose the corresponding pixel. The threshold for color distance is set to 5. The results of test images are illustrated in Figure 6, Figure 7, and Figure 8. The SAD algorithm totally relies on intensities and thus does not work at all. It is clear that our method is visually the best for all test images. Among other methods that work well, NCC and ZNCC work better for Art and Book images. Mutual information is the best for the Aloe images.

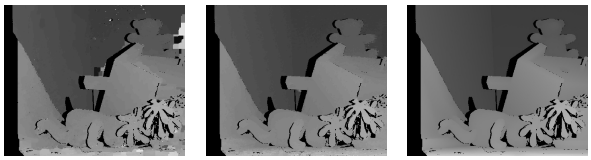
In addition, our local method are also competitive with some of the global illumination invariant stereo matching methods , such as ANCC [10] we mentioned in Section 2. We compare these two algorithms in Figure 5. Ours is visually better and the speed is 40 times faster than ANCC.

4.2. Quantitative Results Analysis

In order to prove the effectiveness of our proposed method, we evaluate more image sets quantitatively by Mean Absolute Error (MAE) and error ratios. MAE measures the mean absolute error between computed disparities and ground truth [12] provided by the Middlebury database:

$$MAE = (\sum_{i,j} |G(i,j) - D(i,j)|)/N. \qquad (10)$$

$G(i,j)$ is the ground truth disparity of (i,j) and $D(i,j)$ is computed disparity value. N is the number of all pixels with computed disparities. The quantitative results are list in Table 1. Table 1 shows the proposed methods have the least error compared with ground truth. Mutual Information is better than NCC and ZNCC except for Wood and



(a) ANCC (b) Proposed (c) groundtruth

Fig. 5. Results comparison of ANCC and proposed method

Reindeer images. NCC and ZNCC have the similar error ratios and rank the third. *Rank* and *census* play the fourth role. SAD does not work for all cases.

Table 2 is the error ratio (the percentage of wrong disparities compared to the ground truth) of the proposed method compares to some of the classic methods on Middlebury test bed [18]. The result of Tsukuba is the best one among four test images.

4.3. Comparison of Running Time

Mutual information is computationally expensive and large window makes it worse. Fortunately, our algorithm starts from quarter-size images, thus the search range is reduced significantly. For example, if the search range is set to 60 in the full size image, it equals to 15 in the quarter-size image. Most correspondences could be found in lower resolution levels. Compare to the proposed without hierarchical structure, the time is reduced on average by 20 times. The codes run on a laptop Thinkpad T400 with Intel(R) Core (TM) 2 Duo CPU, 2.4GHz and 2GB RAM. The running time comparisons are listed in Table 3. Those numbers mean the processing time of a certain algorithm is how many times of the proposed method in average.

5. CONCLUSION

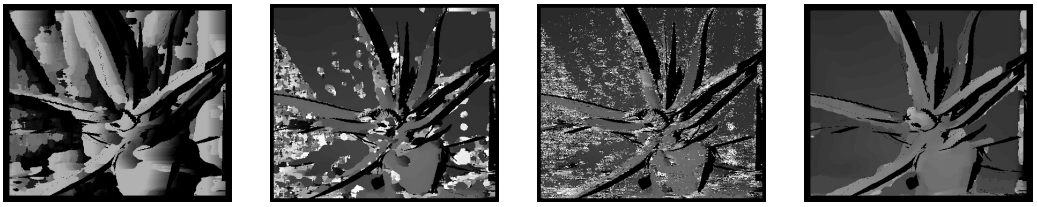
This paper introduces a novel stereo matching method capable of dealing with radiometric variance. Mutual information is employed to find correspondence after the 3-index color conversion. The hierarchical structure guarantees the high matching quality while its computational burden. The experimental results prove that the proposed method produces the best disparity maps and its speed is also the fastest in all test cases.

Table 2. Error Ratio on Middlebury

		Tsukuba			Venus			Teddy			Cones	
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
AdaptingBP	1.11	1.37	5.79	0.10	0.21	1.44	4.22	7.06	11.8	2.48	7.92	7.32
GeoSup	1.45	1.83	7.71	0.14	0.26	1.90	6.88	13.2	16.1	2.94	8.89	8.32
Proposed	0.99	1.13	5.20	3.97	4.55	14.1	5.96	10.4	15.5	4.77	10.2	11.5
AdaptWeight	1.38	1.85	6.90	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26
GraphCut	1.94	4.12	9.39	1.79	3.44	8.75	16.5	25.0	24.9	7.70	18.2	15.3
DP	4.12	5.04	12.0	10.1	11.0	21.0	14.0	21.6	20.6	10.5	19.1	21.1

Table 3. The Comparison of Running Time (times)

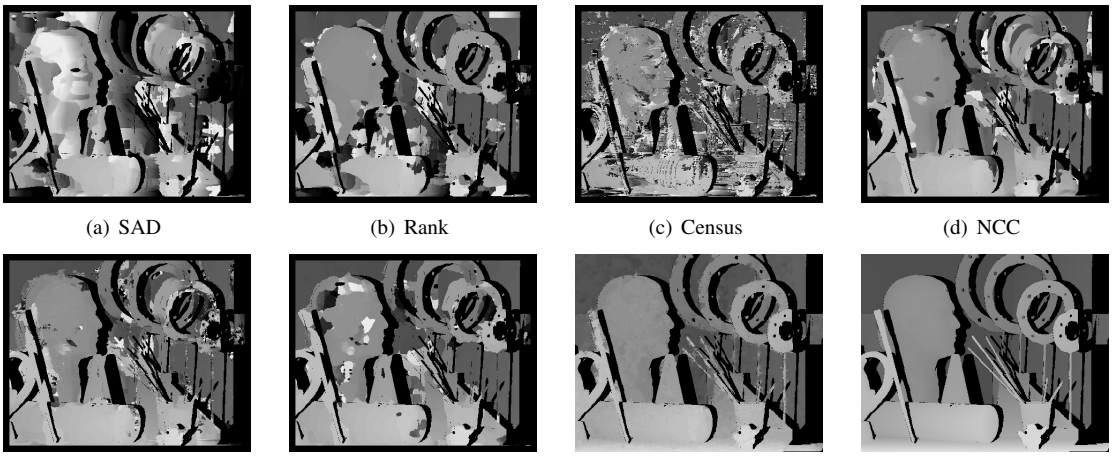
	Rank	Census	NCC	ZNCC	Mutual Information	proposed without hierarchical structure
proposed	1.7	2.7	1.52	1.31	9.61	20.21



(a) SAD (b) Rank (c) Census (d) NCC

(e) ZNCC (f) MI (g) proposed method (h) ground truth

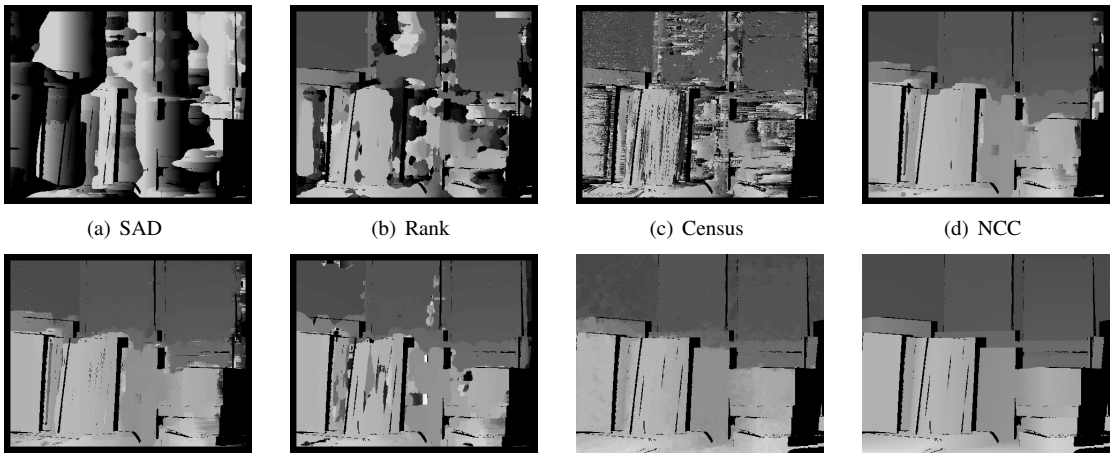
Fig. 6. Result comparison of ”aloe” in different illuminations



(a) SAD (b) Rank (c) Census (d) NCC

(e) ZNCC (f) MI (g) proposed method (h) ground truth

Fig. 7. Result comparison of ”art” in different illuminations



(a) SAD (b) Rank (c) Census (d) NCC

(e) ZNCC (f) MI (g) proposed method (h) ground truth

Fig. 8. Result comparison of "book" in different illuminations

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, 1994.

[3] C.E.Shannon, A mathematical theory of communication. Bell Systmes technical Journal, 27, pp. 379-423, 623-656, 1948.

[4] A.Collignon, F.Maes, D.Delaere, D.Vandermeulen, P.Suetens and G.marchal,Automated multi-modality image registration based on information theory. In Proceedings of International Conference Information in Medical Imaging, France, 1995.

[5] P.Viola and W.M.Wells III, Alignment by maximization of mutual information. In Proceedings of International Conference on Computer Vision, 1995.

[6] A.Rangarajan, H.Chui and J.S.Duncan, Rigid point feature registration using mutual information. Medical Image Analysis, 3(4), pp.425-440, 1999.

[7] B.Likar and F.Pernus, A hierarchical approach to elastic registration based on mutual information. Image Vision Computing, 19(1-2), pp. 33-44, 2001.

[8] G.Egnal, Mutual information as a stereo corresponding measure. Technical Report MS-CIS-00-20, Computer and Information Science, University of Pennsylvania, 2000.

[9] J.Kim, V.Kolmogorov and R.Zabih, Visual Correspondence Using Energy Minimiza- tion and Mutual Information. In Proceedings of International Conference on Computer Vision, France, 2003.

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

[11] 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, 2009.

[12] 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, April, 2009.

[13] X.Zhou and P.Boulanger,Illumination Invariant Stereo Matching Based on Normalized Mutual Information and Census Methods. To appear in Proceedings of Computer Graphics International, 2011.

[14] P.Fua, A parallel stereo algorithm that produces dense depth maps and preserves image features. Machine Vision and Applications, 6(1), pp.35-49, 1993.

[15] http://www.vision.middlebury.edu/stereo/data.

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

[17] M.Li, Information distance and its application. In Proceedings of International Conference of Implementation and Application of Automata,pp.1-9,2006.

[18] http://vision.middlebury.edu/stereo/eval/

[19] K.J.Yoon and S.Kweon, Adaptive support-weight approach for correspondence search. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, issue.4, pp. 650-656, 2006.