Advanced Engineering Forum *Online: 2011-12-22*
ISSN: 2234-991X, Vols. 2-3, pp 182-187
doi:10.4028/www.scientific.net/AEF.2-3.182

A New Extension of the Rank Transform for Stereo Matching

Ge Zhao[1, 2,a], Yingkui Du[1,b] and Yandong Tang[1,c]

[1] State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, PR China

[2] Graduate University of the Chinese Academy of Sciences, Beijing, PR China

[a]zhaoge@sia.cn, [b]dyk@sia.cn, [c]ytang@sia.cn

Keywords: Stereo matching, rank transform, Bayesian stereo model

Abstract. Stereo matching methods often use rank transform to deal with image distortions and brightness differences prior to matching but a pixel in the rank transformed image may look more similar to its neighbor, which would cause matching ambiguity. We tackle this problem with two proposals. Firstly, instead of using two values 0 and 1,we increase the discriminative power of the rank transform by using a linear, smooth transition zone between 0 and 1 for intensities that are close together. Secondly, we propose a new Bayesian stereo matching model by not only considering the similarity between left and right image pixels but also considering the ambiguity level of them in their own image independently. We test our algorithm on both intensity and color images with brightnesss differences. Corresponding 2D disparity maps and 3D reconstruction results verify the effectiveness of our method.

Introduction

Stereo matching is one of the most active research areas in computer vision. It consists of determining which pair of pixels, projected on two images, belong to the same physical 3D scene point. For rectified images, two corresponding pixels have the same y coordinate. The difference of their x coordinates is called disparity. In this context, stereo matching is to find the disparity. Since disparity is inversly proportional to 3D depth, stereo matching is often used in 3D reconstruction [1]. Local stereo matcing methods [2], use matching metrics such as sum of absolute differences (SAD) to measure the similarity between two pixels, and thus find the optimal disparity. These algorithms run fast and memory efficicent. However, one problem of them is the sensitivity to radiometric distortions and brightness differences [3]. Some pre-processing methods can be used to remove distortions prior to matching. The rank transform proposed by Zabih [4] is one of them. It replaces the magnitude of a pixel with its order rank in its neighborhood. It has been proven that rank transform is robust against various radiometric distortions [3]. Because rank transform is fast and suitable for hardware implementation, it is widely used by local matching methods. For example, Ambrosch, K embed the rank transform into his FPGA-based matching applications [5]. The disadvantage of rank transform is that the rank transformed image has many pixels with similar appearances. If an image point looks similar to its neighbor, it is very difficult to find its true correspondence in the other image and this problem is referred as matching ambiguity. To resolve this problem, many authors have extended the original rank transform. Banks, J proposed a novel rank constraint to offset the information loss [6]. Wang tackle this problem by using more partitions rather than 0 and 1 to represent the intensity differences [7]. Recently Zheng and Su [8] apply Wang's extension to their own matching algorithm and obtained the top-ranked result on the Middlebury dataset among all of the state-of-the-art local methods. However, Wang's extension fail to wrok if the transform window center is noisy. Besides, it is hard to determine the total number of partitions. Therefore, we propose a new extension of rank transform. Moreover, we also propose a new Bayesian stereo model to further resolve matching ambiguity. The rest of the paper is organized as follows: in section 2, we propose a novel extension of rank transform; in section 3, we propose a new Bayesian stereo matching model to further resolve the matching ambiguity; experimental results are shown in section 4 and we finally conclude our work in section 5.

A Novel Extension of Rank Transform

Rank transform [4] is a widely used non-parametric transform in which the pixel intensities within a window are arranged in increasing (or decreasing) order. The intensity of the center pixel is then replaced by its rank in this window. To further explain the process of the rank transform, let us consider a window W which includes a center pixel. The intensity of the window center is compared to ever other location of the window. Formally, a certain location i of rank window W is defined by

$$W_i = \begin{cases} 1\ (or\ 0), & I_i \geq I_{center} \\ 0\ (or\ 1), & I_i < I_{center} \\ undefined, & i = center \end{cases} \tag{1}$$

Let us denote the number of elements in a set which fulfills the criteria; W_i equals 1, by Card (W_i= '1') (the cardinality of the set) and for the opposite case; W_i equals 0 , by Card (W_i = '0') . The rank transform is then defined by as follows:

$$Rank = Card(W_i = '1')\ or\ Rank = Card(W_i = '0') \tag{2}$$

As discussed in section 1, the rank transform may cause matching ambiguity since the intensity difference is reduced to only 2 grades (0 or 1). Therefore, many extensions of rank transforms are proposed. Currently, the best-performing extension is originally proposed by Kun Wang [7] and furthered by Zheng and Su [8]. In particular, Zheng obtained the top-ranked results on Middlebury stereo dataset among all state-of-the-art window-based methods. Actually, the main difference between Wang's extension and the original one is that Wang define five grades- smallest, smaller, equal, bigger, biggest for every pixel instead of only two grades, namely, 0 and 1. In Wang's method a certain location i of window W is defined by as the following equation:

$$W_i = \begin{cases} -2 & when\ Diff < -v & Smallest \\ -1 & when -v \leq Diff < -u. & Smaller \\ 0 & when -u \leq Diff \leq u & Equal \\ 1 & when\ u < Diff \leq v & Bigger \\ 2 & when\ Diff > v & Biggest \end{cases} \tag{3}$$

where *Diff* indicates the intensity difference between the center and the location i in the rank window W. v and u are two use-specified parameters to define the five grades. If we denote the number of elements in the rank window W as N, then the definition for Wang's rank transform can be simply expressed by

$$Rank = \sum_{i=0}^{N} W_i \tag{4}$$

Though Wang's extension outperforms the original rank transform, it still has some limitations. Firstly, it is quite groundless to say five is the optimal number for grades. Actually, the appropriate number of grades varies in each individual image. Secondly, Wang introduce two user-specified parameters which are hard to configure. Besides, the number of parameters may increase along with the number of grades. Thirdly，Wang's extension is over-reliant on the window center, when the center pixel is a noise then the whole transform would fail, even if most pixels in the rank window are noise-free. Therefore, in order to overcome these limitations while preserving enough useful information like edges, we propose a novel extension of the rank transform by defining a linear, soft transition zone between 0 and 1 for values that are close together:

$$Rank = \sum_{q \in N} \min(1, \max(0, \frac{I(q) - median}{K})) \tag{5}$$

In Eq.5, N indicates the set of pixels in the rank window, median refers to the median intensity value of the rank window. K is a threshold defined by users. As we can see from Eq. 5, our extension depends neither on specific number of grades nor on the center of the transform window, which means easier parameter tuning and more robustness. In addition, our extension can preserve more useful information such as edges than Wang' method and this is verified by Fig.1 as follows:

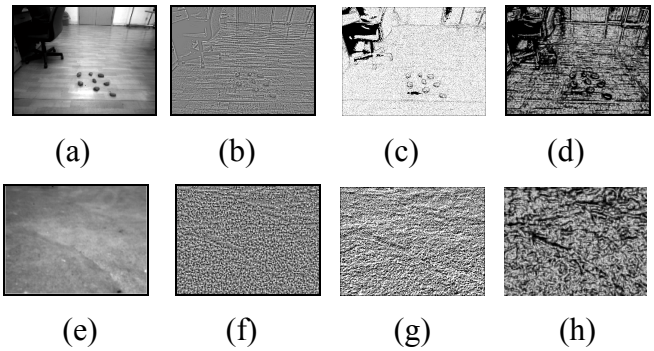


Fig. 1. Different rank transforms on images taken by us. (a), (e) intensity image. (b), (f) original rank transform. (c), (g) Wang's extension. (d),(h) our extension.

As can be seen form Fig.1, our transformed images ((d), (h)) have much richer information content than Zabih's and Wang's ((b) ,(f) and (c), (g)). This advantage is more obviously shown at object boundaries. However, compared with the original intensity images ((a), (e)), we can see all rank transforms would inevitably loose resolution, which is an unavoidable trade-off for robustness. In the following section we will introduce a new Bayesian-based matching model to further deal with this inherent problem associated with the entire rank transform family. Finally, it is worthy of noticing that all our contributions can simply be extended to color images by computing for each color channel separately and then summing the results over all channels.

A New Bayesian Stereo Matching Model

The Derivation of our model . Let p_l and p_r be two pixels in the left and right transformed images respectively and the Bayesian model $M(p_l, p_r)$ models the probability distribution of the event $o_{p_l}^{p_r}$, indicating that $p_l \Leftrightarrow p_r$ is true :

$$M(p_l, p_r) \propto P(O_{p_l}^{p_r} \mid (A_{p_l}, A_{p_r})) \tag{6}$$

where A_{p_l} and A_{p_r} are the local appearances of p_l and p_r . According to the Bayes'theorem, we get

$$P(o_{p_l}^{p_r} \mid (A_{p_l}, A_{p_r}) = \frac{P((A_{p_l}, A_{p_r}) \mid o_{p_l}^{p_r}) P(o_{p_l}^{p_r})}{P(A_{p_l}, A_{p_r})} \tag{7}$$

where $P((A_{p_l}, A_{p_r}) \mid o_{p_l}^{p_r})$ is the data term and $P(o_{p_l}^{p_r})$ is the prior term. In stereo matching, $P(o_{p_l}^{p_r})$ often corresponds to various constraints such as smoothness over disparity space. $P((A_{p_l}, A_{p_r}) \mid o_{p_l}^{p_r})$ corresponds to the similarity between A_{p_l} and A_{p_r} . Most existent Bayesian stereo matching models [9], [10], [11] reduce Eq. 7 to the following equation by omitting $P(A_{p_l}, A_{p_r})$ as :

$$P(o_{p_l}^{p_r} \mid (A_{p_l}, A_{p_r}) \propto P((A_{p_l}, A_{p_r}) \mid o_{p_l}^{p_r}) P(o_{p_l}^{p_r}) \tag{8}$$

Then, intensive research of stereo matching is focused on either improve the robustness of the data term or invent new prior terms to cope with occlusion or slanted surfaces [3]. However, the omitted term $p(A_{p_l}, A_{p_r})$ also includes much useful information which has long been ignored.

Though $p(A_{p_l}, A_{p_r})$ literally means the possibility of the occurrence for A_{p_l} and A_{p_r} in their respective images, it can also be interpreted as the ambiguity levels of them. More specifically, for the left view, if the local neighborhood of p_l looks similar to neighborhoods of its adjacent pixels, then A_{p_l} may be a common pattern in the image, namely, $P(A_{p_l})$ is high, indicating high probability of a false match. In this paper, we use the term $p(A_{p_l}, A_{p_r})$ to derive a new Bayesian model based on it as follows :

$$M(p_l, p_r) \propto \frac{P((A_{p_l}, A_{p_r}) \mid O_{p_l}^{p_r})}{P(A_{p_l}) P(A_{p_r})} \tag{9}$$

Note that we assume A_{p_l} and A_{p_r} are independent so $P(A_{p_l}, A_{p_r}) = P(A_{p_l})P(A_{p_r})$. Besides, because our method is a window-based method which implicitly assume constant disparities within the SAD support window, so we skip the smoothness prior $p(O_{p_l}^{p_r})$.

The implementation of our model. There may be many ways to implement Eq.(9). In this section, we propose a simple but effective implementation. Firstly, we give a specific form of $P(A_{p_l})$ and $P(A_{p_r})$. As discussed in section 3.1, they are in proportion to ambiguity levels of pl and pr . In our implementation, the ambiguity level of one pixel is defined as the multiplicative inverse of the maximum dissimilarity between the pixel and its neighbors as:

$$Dis(p) = \max_{q \in W_p, q \neq p} SAD(p, q) \tag{10}$$

In Eq.(10), *SAD* denotes the sum of absolute differences which is a commonly-used dissimilarity measure in vision and W_p is the *SAD* window which are used in the aggregation of pixel differences. As described at the beginning of section 1, when images are rectified, the two dimensional matching is reduced to one dimensional searching and the searching range is between the maximum and minimum disparity values denoted as d_{max} and d_{min}. By using this heuristic knowledge, we define W_p as a set of points like:

$$W_p = \left\{ p + d \mid d_{\min} - d_{\max} \le d \le d_{\max} - d_{\min} \right\} \tag{11}$$

where $p+d$ is the point with coordinates of p shifted by d in the same image. Now, we can present our implementation by using *Dis* and *SAD* as :

$$P(A_p) \propto \frac{1`}{Dis(p)} \tag{12}$$

$$P((A_{p_l}, A_{p_r}) \mid O_{p_l}^{p_r}) \propto \frac{1}{SAD(p_l, p_r)} \tag{13}$$

and this leads to our implementation :

$$M(p_l, p_r) \propto \frac{P((A_{p_l}, A_{p_r}) \mid O_{p_l}^{p_r})}{P(A_{p_l})P(A_{p_r})} = \frac{Dis(p_l)Dis(p_r)}{SAD(p_l, p_r)} \tag{14}$$

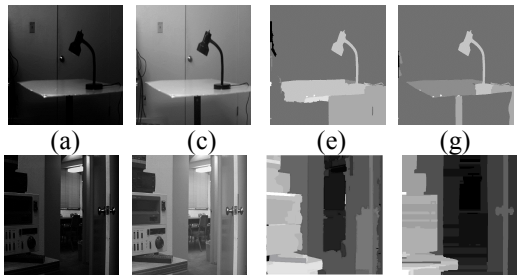
The proposed implementation can easily be generalized for multi-view stereo matching as:

$$M(p_1, p_2, \ldots p_n) = \frac{\prod_{i=1}^{n} Dis(p_i)}{\sum_{i,j \in (1,n) \wedge i \neq j} SAD(p_i, p_j)} \tag{15}$$

Note that in this paper we use a simple greedy search technique called WTA (winner-takes-all) method to pick the disparity having the maximum $M(p_l, p_r)$ within the disparity range and then use the well-known left-right consistency check to eliminate unreliable matches. The LRC check takes the computed disparity value in one image, and re-projects it in the other image. If the difference in the values is bigger than a given threshold $t_{tolerance}$, then this match is marked unreliable.

Experimental Results

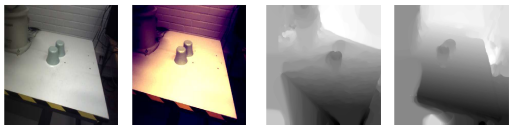
In this section, we evaluate our stereo matching method using both intensity images and color images. In order to reflect the advantages of using rank transform, we deliberately add brightness differences to all input images. We compare our method with Zheng and Su's method [8], which currently the best-performing is matching method based on the rank transform. Fig.2 shows the resultant disparity maps from two intensity stereo pairs. Fig.3 shows the resultant disparity maps from a color stereo pairs.



(a) (c) (e) (g)

(b) (d) (f) (h)

Fig. 2. (a)~(d) input stereo pairs.(e)~(f) Zheng's dispariy maps.(g) ~(h) our disparity maps.



(a) (b) (c) (d)

Fig. 3. Color images. (a),(b) input stereo pairs. (c) Zheng's dispariy maps. (d) our disparity maps.

From Fig.2 we can see that scene objects are more accurately recovered in our disparity maps (Fig.2 (g) and Fig.2 (h)) despite the presence of brightness differences in the input images. For example, the leg of the desk is missing in Zheng's disparity map (Fig.2 (e)) but it is accurately recovered in our disparity map (Fig.2 (g)). Besides, the door knobs are much clearer in our disparity map (Fig.2 (h)) than those in Zheng's disparity map (Fig.2(f)). Fig.4 gives corresponding 3D reconstruction results for our disparity maps and we can see most depth surfaces are accurately reconstructed. Since the stereo pair shown in Fig. 3 has little textures, it is hard to match them. Even though, our method (Fig.3 (d)) still manages to recover both cups on the desk while Zheng's method (Fig.3 (c)) only recovers one of them. This is more obviously illustrated in the following 3D reconstruction results shown in Fig.5. Note that we test Zheng's method with the default parameters in his paper and test our method with parameters shown in Table 1 as follows:

Table 1 :Parameters used by us

d_{min}	d_{max}	k	W_{rank}	W_{SAD}	$t_{tolerance}$
1	22	18	7×7	17×17	3

Fig.4 shows reconstruction results from our disparity maps in Fig.2 and from both Zheng's and our disparity maps in Fig.3. We can see most depth surfaces are recovered correctly in our results.

(a) (b) (c) (d)

Fig. 4. Reconstruction results. (a)~(c) our result. (d) Zheng's result

Conclusion

In this paper, we propose a new extension of rank transform. Compared with existent extensions, the new extension is not only easier to configure but also independent on the center of transform window. Besides, it can preserve more useful information such as edges in the original images. Furthermore,

we extend the Bayesian stereo matching model to cope with the matching ambiguity problem more effectively. The 2D and 3D experimental results show that the proposed methods outperform most rank-based local stereo matching methods with respect to distorted stereo images.

Acknowledgment

This work was supported by Natural Science Foundation of China (No. 60805046 and 60835004).

References

[1] G. Gupta, M.Rawat, "Region growing stereo matching method for 3D building reconstruction", International Journal of Computational Vision and Robotics , vol.2, pp.89-98, 2011.

[2] K. Zhang, J. Lu, "Cross-based local stereo matching using orthogonal integral images", IEEE Trans on Circuits and Systems for Video Technology, vol.19, pp.1073-1079, 2009.

[3] D.Scharstein, R.Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", IJCV, vol.1, pp.1917-1920, 2002.

[4] R. Zabih, J. Woodfill, " Non-parametric local transforms for computing visual correspodence ", ECCV, vol. 8(1), pp. 151 - 158, 1994.

[5] K. Ambrosh, M. Humenberger, "Extending two non-parametric transforms for FPGA-based stereo matching using bayer filtered cameras",CVPR Workshop, vol.1, pp.1-8, 2008.

[6] J. Banks, M. Bennamoun, "Reliability analysis of the rank transform for stereo matching", IEEE Transactions on Systems, Man and Cybernetics, vol.31, pp.870-880, 2001.

[7] K. Wang, "Adaptive stereo matching algorithm based on edge detection", International Conference on Image Processing, vol.1, pp.1-8, 2004.

[8] G. Zheng, X. Su, "Local stereo matching with adaptive support-weight, rank transform and disparity calibration", Pattern Recognition Letters, vol.1, pp.1230-1235, 2008.

[9] C. Li, T. Caelli, "Bayesian stereo matching ", Computer Vision and Image Understanding, vol.5, pp.85-96, 2007.

[10] A. Geiger, R. Martin, R, Urtasun, "Efficient large-scale stereo matching ", Asian Conferencel of Computer Vision, vol.5, pp.25-38, 2010.

[11] Y. Taguchi, B. Wilburn, C. Zitnick, "Stereo reconstruction with mixed pixels using adaptive over-segmentation ", Computer Vision and Pattern Recognition , vol.2, pp.1-8, 2007.