Symmetric segment-based stereo matching of motion blurred images with illumination variations

Wei Wang[1,2,3], Yizhou Wang[3,4], Longshe Huo[3], Qingming Huang[2], Wen Gao[3,4]

[1]Key Lab of Intelligent Information Processing, Chinese Academy of Sciences, Beijing 100080, China
[2]Graduate School, Chinese Academy of Sciences, Beijing 100039, China
[3]Institute of Digital Media, Peking University, Beijing 100871, China
[4]Key Lab of Machine Perception, Peking University, MOE
{wwang, lshuo, qmhuang}@jdl.ac.cn; {Yizhou.Wang, wgao}@pku.edu.cn

Abstract

Most existing methods of stereo matching focus on dealing with clear image pairs. Consequently, there is a lack of approaches capable of handling degraded images captured under challenging real situations, e.g. motion blur is present and an image pair is in different illumination conditions. In this paper we propose a novel approach to handling these challenging situations by formulating the problem into a Maximum a Posteriori (MAP) estimation framework, and adopt a segment-based symmetric stereo matching method to infer a mask of disparity map which indicates whether a disparity is affected by motion blur and estimate the disparity value. The experimental results show that our stereo matching method is able to compute more accurate disparity maps of this type of degraded images.

1. Introduction

Stereo vision has been extensively studied since computer vision area came into being. It has a wide spectrum of important applications, such as 3D scene reconstruction, robot navigation. However, most existing stereo matching methods focus on dealing with image pairs from public data set in which images are taken under control situations, e.g. the Middlebury data set. There is a lack of approaches capable of handling degraded images captured under challenging real situations, e.g. motion blur is present and an image pair is in different illumination conditions, as shown in Fig.1.

In [6], Scharstein and Szeliski provided a comprehensive stereo matching taxonomy and evaluated different stereo matching algorithms. The algorithms can be generally classified into 2 categories: i) local methods compute a disparity at given location depending on its neighboring intensity values and usually make an implicit smoothness assumption by aggregating support. Yoon and Kweon [8] presented a new window-based method using adaptive support-weights based on color similarity and geometric proximity to reduce matching ambiguity. Although it achieved better estimation accuracy than other local methods, it is more computationally expensive. ii) global methods make an explicit smoothness assumption and formulate the stereo matching problem into an energy minimization framework [4][3][2]. Sun *et.al.*[7] proposed a symmetric stereo model to handle occlusion based on visibility constraint. In addition, segment-based methods achieve good performance by assuming that homogeneous color segments are coincident with non-overlapping planes in disparity space.

In this paper, we propose a novel approach to handling a challenging stereo matching problem with degraded images captured under real situations, e.g. motion blur is present and an image pair is in different illumination conditions. In motion blurred regions, objects are blended with background which makes the disparity estimation unreliable in such regions. Hence, in order to estimate the accurate disparity map, we need to infer a *mask of disparity map* (MDM) which indicates whether a disparity is affected by motion blur. However, current methods do not provide a solution to this problem. Besides, illumination difference makes stereo matching more difficult and fails the state-of-the art algorithms.

We formulate this problem into a *Maximum a Posteriori* (MAP) estimation framework to infer the MDM and the disparity map. We also propose a symmetric segment-based stereo matching method to robustly match image pairs under different illumination conditions. In the rest of this paper, we formulate the problem in Section 2. A symmetric segment-based stereo matching method is proposed in Section 3. We present the experimental result in Section 4. Finally, we summarize the paper and offer a discussion in Section 5.

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

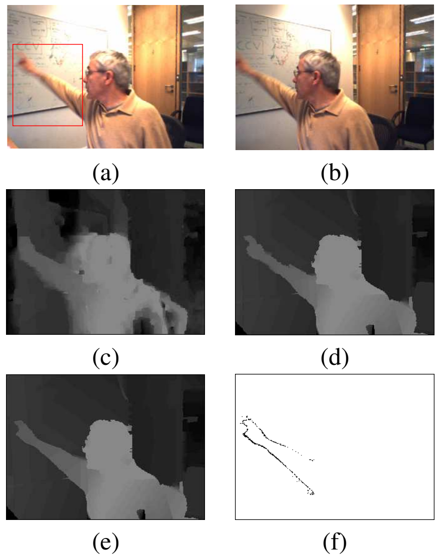

Figure 1: A stereo matching result. (a)(b) an image pair with motion blur and in different illumination conditions. The blurred region is *roughly* marked with a red rectangle in (a). (c) The estimated disparity map by belief propagation. (d) The estimated disparity map using symmetric segment-based stereo matching method. (e) The estimated disparity map combining MDM inference and symmetric segment-based stereo matching method. (f) The inferred MDM. The white region indicates the region where the disparity estimation is not affected by motion blur.

2. Problem formulation

Our goal is to infer the MDM R and the disparity map D simultaneously. $R = \{r_p \in \{0,1\}\}_{p\in\Lambda}$, where Λ denotes disparity map lattice. $r_p = 1$ indicates that p's disparity estimation is not affected by motion blur; $r_p = 0$, otherwise. The two observables are: $\mathbf{I} = (I_L, I_R)$ is the image pair and $\alpha \in [0,1]$ denotes the value in alpha channel of the reference image. We *roughly* mark the blurred region in the images manually and adopt an image matting method[5] to estimate α of that region. In the unblurred regions, we set $\alpha = 1$. It worth noting that even $\alpha_p \neq 1$, r_p can be 1.

Our MAP formulation is as follows:

$$\begin{aligned}(D,R)^* &= \arg\max_{(D,R)} p(D,R|\mathbf{I},\alpha) \\ &= \arg\max_{(D,R)} p(\mathbf{I},\alpha|D,R)p(D,R)\end{aligned}$$

2.1. The likelihood models

We simplify the joint likelihood model by assuming conditional independence:

$$p(\mathbf{I},\alpha|D,R) = p(\mathbf{I}|D)p(\alpha|R)$$

where

$$p(\mathbf{I}|D) \propto \exp\{-\sum_s F(s,d_s,\mathbf{I})\}$$

where $F(\cdot)$ is the matching cost function of pixel s and disparity d_s given images $\mathbf{I}$.

Moreover, we propose a normalized differential image feature to discount the illumination difference between the image pair

$$f_{p,n} = \frac{|I_n - I_p|}{I_p}, \quad n \in \partial p$$

where n is the neighborhood pixel of p. We adopt a cost function similar to Yoon and Kweon's *Adaptive Support Weight* method[8], of which the weights of pixels in the support window are determined by both the color and spatial differences:

$$\omega_{p,n} = \exp\{-(\frac{f_{p,n}}{\beta} + \frac{\delta_{p,n}}{\gamma})\}$$

where β and γ are two parameters determined by the size of the support window and the degree of color similarity, respectively. $\delta_{p,n}$ is the Euclidean distance between pixel p and n.

Finally, the cost function $F(p,d_p,\mathbf{I})$ is defined as:

$$F(p,d_p,\mathbf{I}) = \frac{\sum_{n\in\partial_p, n'\in\partial_{p'}} \omega_{p,n}\omega_{p',n'}|f^{(L)}_{p,n} - f^{(R)}_{p',n'}|}{\sum_{n\in\partial_p, n'\in\partial_{p'}} \omega_{p,n}\omega_{p',n'}}$$

where $p' = p - d_p$ is the corresponding point.

In $p(\alpha|R)$, we assume that when $r_p = 1$, the alpha value is more likely to be close to 1. Then the likelihood term $p(\alpha|R)$ is defined as:

$$p(\alpha|R) \propto \exp\{-\sum_p E(p,r_p,\alpha_p)\}$$

where $E(\cdot)$ is the energy function of pixel p, MDM label r_p and alpha value α_p at p. It is defined as:

$$E(p,r_p,\alpha_p) = \lambda_\alpha|r_p - \alpha_p|$$

where λ_α is a regularization parameter in Table 1.

2.2. The prior models

The joint prior probability is decomposed as follows:

$$
\begin{aligned}
p(D,R) &= p(D|R)p(R) \\
p(D|R) &\propto \exp\{-\lambda_D \sum_p \sum_{n\in\partial_p} \psi(d_p,d_n,r_p,r_n)\}
\end{aligned}
$$

In regions around moving object boundaries, disparity estimation is not reliable or even meaningless due to motion blur. For this reason we introduce a hidden variable R. In region $R_1 = \{p : r_p = 1\}_{p\in\Lambda}$, the regular disparity priors can be applied. Region $R_0 = \{p : r_p = 0\}_{p\in\Lambda}$ is the blur region we pay constant penalty K_D. Thus, the cost function is defined as:

$$
\psi(d_p,d_q,r_p,r_q) = \begin{cases} \min(c|d_p-d_q|,T), & p,q\in R_1 \\ K_D, & p,q\in R_0 \end{cases}
$$

in which the truncated cost function is robust to noise. c and T are regularization and truncation parameters, respectively.

We assume a Potts model for $p(R)$ to enforce label smoothness in MDM.

$$
p(R) \propto \exp\{\beta \sum_{j\in\partial i} \mathbf{1}(x_i = x_j)\}
$$

where $\mathbf{1}(x_i = x_j) = 1$, when $x_i = x_j$; 0, otherwise.

3. Symmetric segment-based stereo matching

Standard stereo matching algorithms are designed for clear image pairs. They generate a lot of matching errors when being applied to challenging image pairs as shown in Fig.1. As there are very few reliable disparities can be used in non-textured regions, we propose a new symmetric segment-based stereo matching method to robustly fit the disparity surfaces. The method uses a cross-checking test to distinguish wrongly matched pixels from occluded pixels and generates two disparity error maps B_L, B_R. After matching these 2 error maps, the disparity estimation result can be refined. Our method is an iterative and symmetric process on both left and right images, which includes 4 sequential steps: an inference step, a cross-checking test and rematching step, a color segmentation and plane fitting step, and a likelihood updating step.

1. **The inference algorithm** We adopt Loopy Belief Propagation[3] to approximate the maximum a posteriori probability of the hidden variables R and D in the initial matching process. At iteration t, we sequentially use the left and right images as the reference image and then compute the left and right disparity maps D_L^t, D_R^t together with the MDM labels R_L^t, R_R^t. These results are used in the next step.

2. **The cross-checking test and rematching** The result of cross-checking test should not be simply considered as occluded pixels. Instead, we divide them into two classes: the occluded pixels and the wrongly matched pixels. The occluded pixels can be derived if there is no pixel matching to p in the reference image from the target image according to the disparity map D^t. The wrongly matched pixels are the remaining ones after removing the occluded pixels from the result of a cross-checking test. These matching errors mainly come from non-textured areas in the image pairs. These disparities can be recovered after a rematching process on B_L^t, B_R^t. We perform a local (window-based) method using *sum of absolute intensity differences* (SAD) to match in B_L^t, B_R^t. After another cross-checking test on the matching error pixels, some matching errors can be corrected and assigned to the reliable matching group, which is used for disparity plane fitting.

3. **The color segmentation and plane fitting** We apply mean shift color segmentation algorithm [1] to both left and right images and fit the disparity plane using weighted least square scheme. The plane fitting is applied only in region $R_1 = \{p : r_p = 1\}_{p\in\Lambda}$ where disparity estimation is not affected by motion blur.

4. **The likelihood updating** We use the output d_{pf}^t of plane fitting at iteration t to update the likelihood:

$$
\begin{aligned}
p_{t+1}(\mathbf{I}|D) &= p_t(\mathbf{I}|D)\cdot\kappa_{t+1}, \\
\kappa_{t+1} &= \exp\{-\sum_{s\in\mathbf{I}}(K_u|d_{pf}^t(s) - d^{t+1}(s)|)\}
\end{aligned}
$$

where $d^{t+1}(s)$ is the disparity of pixel s at iteration $t+1$, K_u is a regularization parameter.

We iterate the above four steps in sequence until convergence. It worth noting that each of the four steps is operated in a symmetric way.

4. Experimental results

We use the stereo image pairs from Microsoft Research Cambridge Lab as our experimental data. This data set has a number of challenging data taken under real situations. Due to limit page, only one image pair "Michel1" is shown in Fig.1.(a) and (b) to demonstrate the effectiveness of our method. There exist motion blurred areas around the man's right arm. The two images are taken under different illumination conditions. It is very challenging to recover the disparity around the man's blurred arm under these situations. In our experiment, the alpha values of the blurred regions are estimated by using the method proposed in [5]. All the parameters used in our approach are listed in Table 1.

Parameters in Proposed Cost Functions		
β	γ	Window size
5	3	3×3

Parameters in Mean-shift Segmentation		
α_{ms}	β_{ms}	γ_{ms}
7	6.5	50

Parameters in MAP Framework						
λ_α	λ_D	K_D	c	T	β	K_u
0.2	1	1	0.1	5	1	5

Table 1: Parameter list.

We compare our result with the one generated by a standard stereo matching method implemented using belief propagation (BP). The estimated disparity map is shown in Fig.1.(c). Fig.1.(d) shows the estimated disparity map using symmetric segment-based stereo matching method with BP to discount illumination variations. And Fig.1.(e) shows the estimated disparity map combining the MDM inference and symmetric segment-based stereo matching method with BP. We did not optimize each implementation so as to offer an fair comparison. We just want to show that by combining the MDM inference and symmetric segment-based stereo matching method, matching result can be greatly improved. From Fig.1.(e), we can see that a clear arm region is recovered from motion blur. Fig.1.(f) shows the inferred MDM.

It is noted that one hidden binary variable r_p is added to infer the MDM at each pixel, compared with the traditional stereo matching formulations, thus it doubles the number of states in belief propagation inference.

5. Summary and discussion

In this paper we propose a novel approach to handling degraded stereo image pairs captured with motion blur and in different illumination conditions. We formulate this challenging stereo matching problem into a MAP framework, and adopt a segment-based symmetric stereo matching method to infer the MDM and estimate the disparity value simultaneously. The experimental results show that our stereo matching method is able to obtain more accurate disparity maps of this type of degraded images.

To further our work, there are several points worth mentioning,

1. This work is not intended to restore the true object boundaries. Even motion deblurring techniques without strong priors will not claim to reliably recovery the object boundaries. In this work, our goal is to infer the regions where disparity estimation can be least affected by motion blurs. In the future, we will consider combining motion deblurring methods to further improve our results.

2. In our method the blurred areas are marked manually by users. In the future, we will develop an automatic method to detect these regions.

Acknowledgments

This work was supported by the China Postdoctoral Science Foundation and the National Hi-Tech Development "863" Program of China under grant No. 2007AA01Z315.

References

[1] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 2002.

[2] A. Criminisi, J. Shotton, A. Blake, and P. Torr. Gaze manipulation for one-to-one teleconferencing. *ICCV*, 2003.

[3] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *CVPR*, 2004.

[4] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions via graph cuts. *ICCV*, 2001.

[5] A. Levin, D. Lischinski, and Y. Weiss. A closed form solution to natural image matting. *CVPR*, 2006.

[6] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 2002.

[7] J. Sun, Y. Li, S. B. Kang, and H.-Y. Shum. Symmetric stereo matching for occlusion handling. *CVPR*, pages 399–406, 2005.

[8] K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *PAMI*, 2006.