

# 一种基于分割的两步立体匹配算法

张浩峰<sup>1)</sup> 赵春霞<sup>1)</sup> 陈得宝<sup>1),2)</sup>

<sup>1)</sup>(南京理工大学计算机科学与技术学院, 南京 210094) <sup>2)</sup>(淮北煤炭师范学院物理系, 淮北 235000)

**摘要** 提出了一种基于分割的两步立体匹配算法。第1步,利用动态规划和左右一致性约束计算地面控制点,将原始图像分割为许多小区域,假定每一个区域为同一个视差,并通过地面控制点计算部分区域的视差值;第2步将剩余未匹配的区域编码为一条染色体,用遗传算法进行优化,得到最终的稠密视差图。选用了一些标准图像对进行测试,实验结果表明,该算法具有很好的性能。

**关键词** 图像分割 立体匹配 地面控制点 遗传算法

**中图分类号**: TP391.4 **文献标识码**: A **文章编号**: 1006-8961(2007)11-2098-06

## A Segment-based Two-step Stereo Correspondence Algorithm

ZHANG Hao-feng<sup>1)</sup>, ZHAO Chun-xia<sup>1)</sup>, CHEN De-bao<sup>1),2)</sup>

<sup>1)</sup>(Department of Computer Science & Technology, Nanjing University of Science & Technology, Nanjing 210094)

<sup>2)</sup>(Physical Department of Huaibei Coal Industry Teachers College, Anhui 235000)

**Abstract** A segment-based two-step stereo correspondence algorithm is proposed. In the first step, Dynamic Programming and Left-Right Consistency are used to detect ground control points, and the original image is segmented into small areas, where each of them is assumed as a single disparity, meanwhile the detected GCP are used to compute the disparity value of the part of the image; In the second step, the left unmatched areas are coded as a single chromosome, and are optimized with genetic algorithm to generate the dense stereo disparity map. The algorithm is tested on the standard image pairs, and the experimental results show its good performance.

**Keywords** image segmentation, stereo correspondence, ground control point, genetic algorithm

## 1 引言

立体匹配是近年来计算机视觉研究领域的一个重点研究方向,其主要目的就是寻找两幅图像中的对应点。稠密匹配希望在图像对之间找到尽可能多的匹配点,它对3维重建的质量具有十分重要的意义。

国内外许多学者对立体匹配问题进行了相当多的研究,文献[1]、[2]做了详细的描述。基于分割的方法也得到了相当的关注<sup>[3-5]</sup>,因为基于分割的方法在区域内部的视差平滑性约束以及区域之间的视差不连续性约束可以很好地得到实现。Wei等人

用多次迭代的方法逐渐填充视差图像<sup>[3]</sup>;Li等人认为图像是由一系列平滑的连续视差表面所组成<sup>[5]</sup>。

基于遗传算法的立体匹配方法也曾经有人做过一定的研究<sup>[6,7]</sup>,其优势在于能够对图像中的每一块进行编码,并进行不停的迭代优化。Han等人提出的基于遗传算法的立体匹配方法<sup>[6]</sup>,是将图像分成许多的 $9 \times 9$ 的小块,对每一个小块进行编码,并进行交叉变异操作以得到最终的视差图;Gong等人<sup>[7]</sup>提出的算法是将图像用二叉树的结构进行描述,然后对二叉树的每一个叶节点利用遗传算法编码,并进行嫁接变异操作,得到最终视差图。Han等人和Gong等人的算法都是将图像分割成矩形的区

基金项目:国防基础技术研究项目(K1702020302)

收稿日期:2006-06-30;改回日期:2006-08-28

第一作者简介:张浩峰(1983~),男,南京理工大学计算机科学与技术学院模式识别与智能系统专业博士研究生。研究方向为计算机视觉、智能机器人。E-mail:haofeng\_njust@163.com

域进行编码的<sup>[6,7]</sup>,但是图像中同一视差的区域往往不是矩形,这样利用他们的算法将会把同一视差的区域划分成许多小的矩形区域,从而增加了编码的长度和需要的个体数,增加了算法的复杂度。

基于此,本文提出了一种基于分割的两步立体匹配算法,第 1 步利用已知的地面控制点生成半稠密视差图;第 2 步利用遗传算法将为匹配的区域进行优化,得到稠密的视差图。该算法既利用了基于分割方法<sup>[3-5]</sup>的优点,又克服了 Han 和 Gong 等人基于 GA 算法的缺点<sup>[6,7]</sup>。

## 2 问题的定义

### 2.1 视差空间图(DSI)

本文的工作是在已经经过校正的图像对上进行的,设  $I_L$  表示左图像,  $I_R$  表示右图像,那么左视差图  $D_L$  表示在左图像上的任意一个视差为  $d_L$  的点  $I_L(x, y)$ , 和右图像上的点  $I_R(x - d, y)$  为 3 维空间中同一点在两幅图像上的映射;同样,右视差图  $D_R$  表示右图像上点  $I_R(x, y)$  对应于左图像上点  $I_L(x + d, y)$ 。

把  $(p, d)$  叫作一个匹配对,那么形成这样的一个匹配对所需要的能量为  $C(p, d)$ , 其定义为左右图像上相对应的两个点的颜色或者亮度差,

$$C(p, d) = f(I_L(x, y), I_R(x - d, y)) \quad (1)$$

其中,函数  $f$  可以用区域方法,如 SSD(sum of square differences), SAD(sum of absolute differences), NCC(normalized cross correlation) 等等,这里采用的方法是由 Birchfield 等人提出的像素不相似检测方法<sup>[8]</sup>, 因为其对图像的取样噪声具有一定的鲁棒性。文中所使用的图像对为彩色图像,将图像转换到 Lab 空间,图像点之间的匹配能量耗费用像素之间的色差来表示,



图 2 Tsukuba 的 GCP 以及其误匹配点  
Fig. 2 Ground control points and mismatch points of tsukuba

$$\Delta E(I_L, I_R) = \sqrt{((L_L - L_R)^2) + (a_L - a_R)^2 + (b_L - b_R)^2} \quad (2)$$

根据上述说明可以得到视差空间图 DSI (disparity space image)(如图 1 所示)。

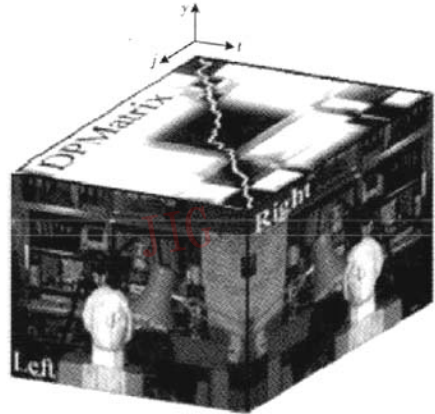


图 1 视差空间图像(DSI)  
Fig. 1 Disparity space image

### 2.2 地面控制点

Bobick 等人提出了地面控制点(ground control points, GCP)<sup>[9]</sup>, 所谓的 GCP 是指视差图中绝对可信的点,即为真实的视差点。本文算法所给出的 GCP 计算方法如下:

(1) 用动态规划的方法分别计算左视差图和右视差图,并利用视差图的左右一致性获得相对较准确的左视差图。

$$D_{LRC}(x, y) = \{d \mid D_L(x, y) = D_R(x - d, y) = d\} \quad (3)$$

(2) 计算上一步得到的视差图上每一个视差点,如果符合下式,则保留,否则删除此点,

$$C(p, d^{GCP}) \leq \lambda_{GCP} * C(p, \hat{d}), \forall \hat{d} \neq d^{GCP} \quad (4)$$

使用测试图像对 tsukuba 做了实验,其中  $\lambda_{GCP} = 0.3$ , 结果图像以及误匹配的视差点如图 2 所示,误



匹配率非常低, 占有所有 GCP 点的 0.3%, 但是这些点并不会影响后来的匹配过程。

### 2.3 视差不变性假设

将左图像  $I_L$  使用 Comaniciu 等人提出的分割方法<sup>[10]</sup>进行分割, 这样可以得到一系列互不交叉的区域, 为了后面算法的可靠性考虑, 先将图像的最小分割区域限制在 80 像素, 经过 2.2 节计算得到的结果, 对每一个分割区域内的真实控制视差进行统计, 如果在同一个区域内有两种或者两种以上不同的视差都超过了一定的比例, 那么将此区域再进行分割。具体分割方法采用 Wei 等人提出的算法<sup>[3]</sup>, 直到不再有可以进一步分割的区域, 假定在同一个区域内的所有像素点的视差相同, 那么算法将变得简单而且有效。

$$d_R = \begin{cases} d_{best} & \left( \frac{\sum_{d \in R_{Matched}^{(i)}} (d = d_{best})}{N_i} > \lambda_{best} \right) \cap \left( \frac{\sum_{d \in R_{Matched}^{(i)}} (d \neq d_{best})}{N_i} < \lambda_{left} \right) \\ \text{Unmatched 其他} & \text{其他} \end{cases} \quad (5)$$

在 2.2 节所计算得到的 GCP 图有部分误匹配点, 但是这些点比较分散, 并且由于图像的最小分割块为 80, 并且还有一部分区域进行多次分割, 所以用式(5)进行计算的时候, 几乎是不可能将这些误匹配点作为区域的最优视差值而赋给整个区域。

### 3.2 基于遗传算法的视差图优化

#### 3.2.1 遗传算法的编码

经过算法的第 1 步处理, 只剩下部分未匹配的区域, 由 2.3 节的视差不变性假设, 将每个未匹配的区域看作一个基因, 图像中所有的未匹配区域看作一条染色体进行编码, 在上一节中, 统计了每个区域的视差直方图, 未匹配的区域主要是因为最优的视差统计达不到一定的比例  $\lambda_{best}$ , 但是即使这样, 这些视差点仍然有较高的可信度, 编码时, 将每个基因值随机赋予视差空间内的任一值, 但是在视差直方图统计内越高的视差被随机赋值的概率越大。设  $hist[i]$  为视差  $i$  在区域内的直方图统计值, 为了保证每一个视差值都有可能被选择到, 将  $hist[i]$  的每一项都加上一个常量  $K$ , 那么视差  $i$  被选择的概率为

$$p(i) = \frac{hist[i] + K}{\sum_{i=0}^{D_{max}} (hist[i] + K)} \quad (6)$$

其中,  $K = \frac{\sum_{i=0}^{D_{max}} hist[i]}{D_{max}}$ ,  $D_{max}$  为允许的最大视差值。

## 3 算法描述

### 3.1 部分视差图的生成

通过第 2 节的计算, 可以得到较为准确的地面控制点视差图像, 且图像已经被分割成许多区域。由视差不变性假设, 在同一个区域内, 视差值相同, 那么就可以假定在同一个区域内, 具有某个视差值  $d_{best}$  的像素达到一定的比例  $\lambda_{best}$ , 而具有其他视差值的像素点总和不超过一定的比例  $\lambda_{left}$ , 则将这个区域内所有像素点的视差都赋值为  $d$ , 否则, 将整个区域标记为未匹配。设第  $i$  个区域共有  $N_i$  个像素, 整个区域  $R_i$  分为已匹配点区域  $R_{Matched}^{(i)}$  和未匹配点区域  $R_{Unmatched}^{(i)}$

### 3.2.2 适应度函数的计算

目前很多算法<sup>[11,12]</sup>对于视差图的优化主要是求取数据项以及平滑项的能量最小时图像的视差值,

$$D = \arg \min_d (E_{data}(d) + E_{smooth}(d)) \quad (7)$$

本文算法将每个个体的适应度函数定义为  $fitness(d_i) = \max_{k \in D_{Unmatched}} E(d_k) - E(d_i)$ , 其中,  $D_{Unmatched}$  表示未匹配区域中可用的视差值。由于在算法的第 1 步已经得到了部分视差图, 所以每个个体的适应度函数只对那些还未匹配的区域进行计算。根据 2.3 节所述的图像分割结果, 将所有未匹配的区域进行编号。

式(7)中数据项  $E_{data}$  只需要计算那些未匹配的区域, 对于区域中所有像素点用下式进行统计:

$$E_{data} = \sum_{R_{Unmatched}} \sum_{p \in R_i} C(p, d) \quad (8)$$

而对于平滑项  $E_{smooth}$  则需要计算相邻的未匹配区域之间的平滑性以及未匹配区域和已匹配区域之间的平滑性(如图 3 所示)。在每一个区域  $R$  边界  $countour(R)$  的 4 邻域的空间  $N = \{(p, q) | p_x - q_x | + |p_y - q_y| = 1\}$  进行搜索, 其中  $countour(R) = \{(p, q) | p \notin R \wedge q \in R \wedge (p, q) \in N\}$ , 求取

$$E_{smooth}(d) = \sum_{(p,q) \in N} S(d_p, d_q) \quad (9)$$

式(9)中  $S(d_p, d_q) = \lambda_{smooth} \cdot f(d_p, d_q)$ , 一般情况下  $f(d_p, d_q) = |d_p - d_q|$ , 但是在实验中实际计算时使

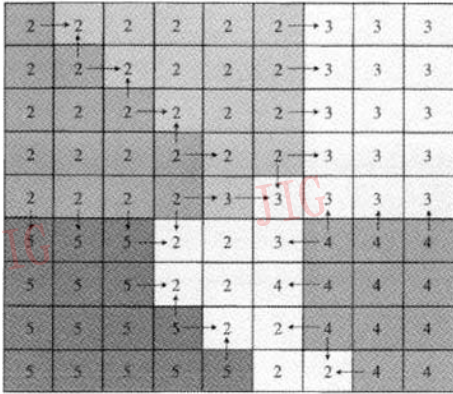


图 3 视差图的边界示意图  
格内的数字表示视差值,暗背景的为未匹配区域,亮背景的为已匹配的区域,箭头表示纳入平滑项计算的边界

图 3 视差图的边界示意图

Fig.3 Illustration of borders of disparity image

用的是 Potts 模型,因为 Potts 模型在实验中取得了更好的效果,

$$f(d_p, d_q) = \begin{cases} 0 & d_p = d_q \\ 1 & d_p \neq d_q \end{cases} \quad (10)$$

### 3.2.3 种群的选择

本文算法采用  $q$  竞争选择规则,设群体共有  $m$  个个体,对其进行交叉变异,以形成个体数目相等的子代,首先计算群体的适应度,从这  $2m$  个个体中选择  $q$  个个体,将已知的  $2m$  个个体按适应度值的大小分别与这  $q$  个个体比较,计算出  $q$  个个体中比已知个体适应度差的个数  $s_i$ ,并将其作为所选个体的得分,待  $q$  个个体都比较完后,按得分高低下降的顺序排列个体,选择得分高的前  $m$  个个体遗传到下一代,在比较过程中,对每个比较都要重新采样,这样在优良个体的基础上遗传和变异,保证最好的个体被置为适应值最大而保存下来。

### 3.2.4 交叉变异

从种群的所有  $m$  个个体中按照概率  $\hat{p}(i)$  (式 (11))选择两个个体进行双交配位交叉,共执行  $m/2$  次操作,然后从生成的  $m$  个子代个体中选择一部分进行变异,每次变异操作在视差范围内,对变异位加上或者减去一个不大于  $\lambda_{\max}$  ( $< 3$ )的整数。

$$\hat{p}(i) = \frac{fitness(X_i)}{\sum_{i=0}^m fitness(X_i)} \quad (11)$$

其中,  $fitness(X_i)$  表示个体  $X_i$  的适应度。

## 4 实验分析

算法的实验是在 Scharstein 和 Szeliski 提供的实验平台上进行的<sup>[1,2]</sup>。测试值  $B$  表示生成的视差图与真实视差图之间误匹配大于  $\delta_d$  个像素的点在全图所占的比例,即

$$B = \frac{1}{N} \sum_{(x,y)} (|d_c(x,y) - d_r(x,y)| > \delta_d) \quad (12)$$

在实验过程中取  $\delta_d = 1.0$ 。

在 2.2 节,用 tsukuba 测试图对得到的地面控制点的结果有 0.3% 的误匹配点,但是从表 1 可以看出,这些误匹配点几乎不影响第 1 步的匹配结果,同样,可以从最终的结果视差图中看出,这些误匹配点对第 2 步的匹配过程影响也很小,因为从图 4 中可以看出,几乎没有整个分割块被赋予错误的视差值(那些误匹配点大多是因为图像分割误差而导致的)。

表 1 本文算法第 1 步结果误匹配率

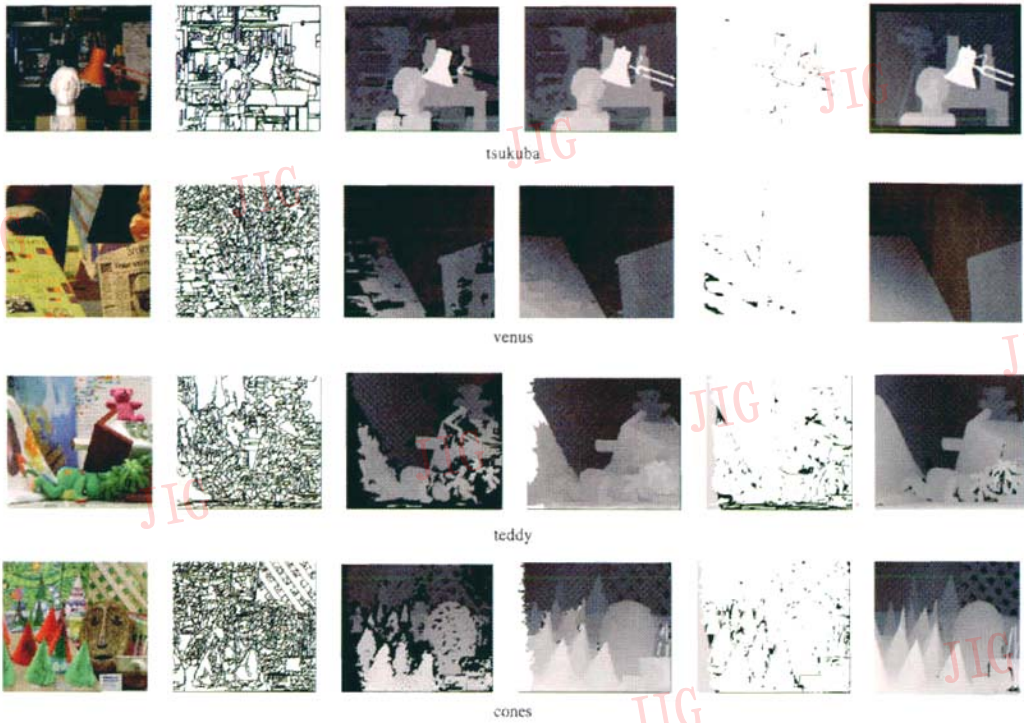
Tab.1 Mismatch rate of first step of our algorithm

	tsukuba	venus	teddy	cones
第 1 步	0.41	0.19	1.40	1.44

表 2,表 3 给出的数据全部为算法第 2 步时,初始化 100 个个体,计算 30 代所得出的结果,同时给出了 DP 算法<sup>[9]</sup>和 GC 算法<sup>[11]</sup>以及其一些改进方法的结果,并与之对比。算法时间的消耗主要是在第 2 步使用遗传算法匹配时较多,第 1 步算法执行时间对于这 4 组图像对平均约 0.86s。整个算法的执行时间与 DP 算法和 GC 算法的时间对比见表 2;从表 3 可以看出,对于 DP 算法及其改进算法,本文算法结果明显要好很多;对于 GC 算法及其改进算法,比原始的 GC 算法要好,和 GC + occ 算法<sup>[14]</sup>相当,比 MulticamGC 算法<sup>[13]</sup>要略差一点。

表 2 本文算法以及 GC 算法、DP 算法的执行时间对比  
Tab.2 Execute time of graph cuts, Dynamic Programming and our algorithm 单位:s

方法	tsukuba	venus	teddy	cones
本文算法	14.79	16.00	57.23	78.31
文献[11]算法	371.0	414.5	623.5	685.2
文献[9]算法	0.35	0.42	1.03	1.06



第 1 列为原始图像队的左图像;第 2 列为左图像的分割结果;第 3 列为算法的第 1 步结果图;第 4 列为最终算法所得视差图;第 5 列为本文算法所得视差图的误匹配点;第 6 列为真实视差图

图 4 本文方法的测试结果

Fig. 4 Experiment results of our algorithm

表 3 本文方法和其他几种方法的定量指标

Tab.3 Comparison of our algorithm and some other algorithms

方法	usukuba			venus			teddy			cones		
	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc
本文算法	1.54	<b>1.83</b>	7.94	2.29	2.65	7.86	<b>9.69</b>	<b>16.0</b>	20.8	7.06	15.3	15.3
文献[11]算法	1.94	4.12	9.39	1.79	3.44	8.75	16.5	25.0	24.9	7.70	18.2	15.3
MulticamGC 算法 <sup>[13]</sup>	1.27	1.99	6.48	2.79	3.13	<b>3.60</b>	12.0	17.6	22.0	<b>4.89</b>	<b>11.8</b>	<b>12.1</b>
GC + occ <sup>[14]</sup>	1.19	2.01	<b>6.24</b>	1.64	2.19	6.75	11.2	17.4	19.8	5.36	12.4	13.0
文献[9]算法	4.12	5.04	12.0	10.1	11.0	21.0	14.0	21.6	<b>20.6</b>	<b>10.5</b>	19.1	21.1
ReliabilityDP 算法 <sup>[15]</sup>	<b>1.36</b>	3.39	7.25	2.35	3.48	12.2	9.82	<b>16.9</b>	<b>19.5</b>	12.9	19.9	19.7
文献[12]算法	1.99	2.84	9.69	<b>1.41</b>	<b>2.10</b>	7.47	15.9	23.9	27.1	10.0	18.3	18.9
SSD + MF 算法	5.23	7.07	24.1	3.74	5.18	11.9	16.5	24.8	32.9	10.6	19.8	26.3

## 5 结论

本文提出了一种基于分割的两步立体匹配算

法,利用彩色图像分割的方法将图像分割成许多不同视差的小区域,用地面控制点填充部分视差图像,用遗传算法来优化能量函数,最终得到稠密视差图。通过标准测试图像对的测试,并与其他一些优化算

法的比较,在综合时间消耗和准确率上显示出本文算法的良好性能。

### 参考文献 (References)

- 1 Scharstein D, Szeliski R. Middlebury stereo vision page [EB/OL]. <http://www.middlebury.edu/stereo/>. 2002-01-10.
- 2 Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. *International Journal of Computer Vision*, 2002, 47(1/2/3): 7 ~ 42.
- 3 Wei Y, Quan L. Region-based progressive stereo matching [A]. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]*, Washington, DC, USA, 2004: 106 ~ 113.
- 4 Wei Y, Lhuillier M, Quan L. Fast segmentation-based dense stereo from quasi dense matching [A]. In: *Proceedings of the 6th Asian Conference on Computer Vision [C]*, Jeju, Korea, 2004: 360 ~ 365.
- 5 Lin M, Tomasi C. Surfaces with occlusions from layered stereo [A]. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]*, Madison, Wisconsin, USA, 2003: 710 ~ 717.
- 6 Han K P, Song K W, Chung E Y, et al. Stereo matching using genetic algorithm with adaptive chromosomes [J]. *Pattern Recognition*, 2001, 34(9): 1729 ~ 1740.
- 7 Gong M L, Yang Y H. Genetic-based stereo algorithm and disparity map evaluation [J]. *International Journal of Computer Vision*, 2002, 47(1/2/3): 63 ~ 77.
- 8 Birchfield S, Tomasi C. A pixel dissimilarity measure that is insensitive to sampling [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(4): 401 ~ 406.
- 9 Bobick A F, Intille S S. Large occlusion stereo [J]. *International Journal of Computer Vision*, 1999, 33(3): 181 ~ 200.
- 10 Comaniciu D and Meer P. Mean shift: A robust approach toward feature space analysis [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 603 ~ 619.
- 11 Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(11): 1222 ~ 1239.
- 12 Veksler O. Stereo correspondence by dynamic programming on a tree [A]. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]*, San Diego, CA, USA, 2005: 384 ~ 390.
- 13 Kolmogorov V, Zabih R. Multi-camera scene reconstruction via graph cuts [A]. In: *Proceedings of Seventh European Conference on Computer Vision [C]*, Copenhagen, Denmark, 2002: 82 ~ 96.
- 14 Kolmogorov V, Zabih R. Computing visual correspondence with occlusions using graph cuts [A]. In: *Proceedings of International Conference on Computer Vision [C]*, Vancouver, Canada, 2001: 508 ~ 515.
- 15 Gong M L, Yang Y H. Near real-time reliable stereo matching using programmable graphics hardware [A]. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]*, San Diego, CA, USA, 2005: 924 ~ 931.