第 34 卷 第 4 期
2011 年 4 月

计 算 机 学 报
CHINESE JOURNAL OF COMPUTERS

Vol. 34 No. 4
Apr. 2011

一种基于图像分割及邻域限制与放松的立体匹配方法

伍春洪   付国亮

（北京科技大学自动化学院   北京   100083）
（北京科技大学钢铁流程先进控制教育部重点实验室   北京   100083）

摘　要　提出了一种以 K-均值分割为基础的立体匹配方法. 该方法不仅可以根据图像的内容自动调整匹配窗口的形状，还可实现对参与匹配窗口的大小、数目和权重的智能调节. 作者采用 K-均值分割方法精确定位物体边界，保证匹配窗口位于同一物体内部；邻域限制与放松可以进一步根据图像内容灵活地运用匹配窗口周围的环境信息；两种方法的结合有效地提高了匹配过程中窗口选取的智能性. 在国际立体视觉标准平台 Middlebury 网站中测试的结果证实该算法提取的深度图的错误率低于其它局部优化算法，接近全局优化算法，运行效率高于现有的全局优化算法，综合性能是出众的.

关键词　立体匹配；邻域限制与放松；K-均值分割；邻域权重设置；遮挡处理
中图法分类号　TP391　　**DOI 号**：10.3724/SP.J.1016.2011.00755

A Stereo Matching Method Based on *K*-means Segmentation and Neighborhood Constraints Relaxation

WU Chun-Hong   FU Guo-Liang

（*School of Automation and Electrical Engineering，University of Science and Technology Beijing，Beijing   100083*）
（*Key Laboratory for Advanced Control of Iron and Steel Process*（*Ministry of Education*），
University of Science and Technology Beijing，Beijing   100083）

Abstract　The paper proposes a stereo matching method based on *K*-means Segmentation. Within the proposed method，more than one matching windows are involved in one corresponding task. Not only the shape but also the size，number and weight of each matching window can be modified intelligently according to the image content. *K*-means Segmentation was used to detect object edges and kept the matching window within a same object. Neighborhood Constraint and Relaxation Algorithm is further adopted to utilize the environment information. This combination tackles the problem of how to choose an appropriate matching window intelligently. The algorithm is tested using the Middlebury stereo test bed. It was proved that the error percentage of the depth map obtained by the algorithm is lower than other algorithms based on local optimization，approaching to global optimization algorithms. The efficiency of the algorithm is higher than other global optimization algorithms，the overall performance is outstanding.

Keywords　stereo matching；neighborhood constraint relaxation；*K*-means segmentation；neighborhood weight setting；occlusion handling

1 引 言

立体视觉技术在三维测量、虚拟现实以及机器人导航等许多领域都极具应用价值，自 20 世纪 60 年代中期诞生以来就深受关注. 经过多年的研究，立体视觉的发展虽然取得了不少进步，但远未达到成熟阶段[1-2]. 在立体匹配过程中缺少人类的智能来面对

各种不同环境是制约立体视觉技术发展的关键因素. 当前, 立体匹配的主要问题在于如何根据图像的内容灵活地选取匹配窗口的形状和大小, 对缺少纹理的区域如何寻找对应点以及对遮挡区域如何处理这几个问题上[1-4]. 本文针对以上问题, 提出一种以 K-均值分割为基础的立体匹配方法. 该方法采用 K-均值分割方法来精确定位物体边界, 保证匹配窗口位于同一物体内部; 同时采用邻域限制与放松进一步根据图像内容灵活地运用匹配窗口周围的环境信息; 两种方法的有机结合有效地提高了匹配过程中窗口选取的智能性. 不仅可以根据图像的内容自动调整匹配窗口的形状, 还可实现对参与匹配的窗口的大小、数目和权重的智能调节. 在国际立体视觉标准平台 Middlebury 网站中测试的结果证实本算法提取的深度图的错误率低于其它局部优化算法, 接近全局优化算法. 运行效率高于现有的全局优化算法, 综合性能是出众的.

本文第 2 节介绍算法的基本框架; 第 3 节介绍 K-均值分割方法; 第 4 节介绍遮挡区域的检测; 第 5 节介绍邻域权重的设置; 第 6 节用 Middlbebury 标准图像数据库对本算法进行测试, 并和现有算法进行比较; 最后给出结论.

2 算法整体框架

本算法以邻域限制和放松 (Neighhood Constraint and Relaxation, NCR) 的多窗口匹配算法为基本框架[5-7]. NCR 算法的依据是空间一致规则, 即深度空间是分段连续的. 设 $B_{i,j}$ 为待匹配的子窗口, $B_{k,l}$ 为 $B_{i,j}$ 的邻域, 称为邻域子窗口, 如图 1 所示. 在分析区域 $B_{i,j}$ 视差时, 若将相邻区域 $B_{k,l} \in B$ 的视差也考虑进来, 可以取得更加可靠的结果[6-7].
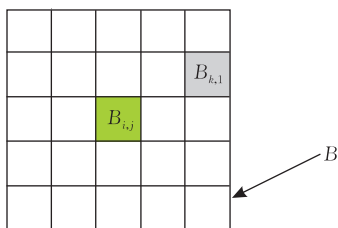

图 1 待匹配块及其邻域

考虑邻域后视差分析的记分函数可表示为

$$score(B_{i,j},d) = \frac{SSD(B_{i,j},d) + \sum_{B_{k,l} \in N(B_{i,j})} w(B_{k,l},B_{i,j}) SSD(B_{k,l},d)}{1 + \sum_{B_{k,l} \in N(B_{i,j})} w(B_{k,l},B_{i,j})}$$

(1)

其中, $w(B_{k,l},B_{i,j})$ 为对应不同邻域块的权重因子. 权重因子可以根据邻域块到待匹配块的距离、邻域块和待匹配块的相似度、邻域块自身的可信度等几个因素来确定. 通过权重因子的设置可以灵活地调整邻域子窗口在匹配过程中对记分函数的贡献. 权重因子的设置将在第 5 节中具体讨论.

考虑到邻域块和待匹配块深度相近但并不一定相同, 邻域放松进一步允许邻域块和待匹配块有一个小的深度变化 δ, 以此来增加匹配的适应性. 考虑邻域放松因素后的代价函数可修正为

$$score(B_{i,j},d) = \left(SSD(B_{i,j},d) + \sum_{B_{k,l} \in N(B_{i,j})} w(B_{k,l},B_{i,j}) \min_{\delta}\{SSD(B_{k,l},d+\delta)\} \right) / \left(1 + \sum_{B_{k,l} \in N(B_{i,j})} w(B_{k,l},B_{i,j}) \right)$$

(2)

式(2)是完整的 NCR 代价函数. 邻域限制是通过考虑周围邻域的记分而实现的, 邻域放松是通过允许邻域的深度和中心块的深度在一定范围内变化实现的. 最后期望的视差值在代价函数取最小值时获得[6-7].

$$d^* = \arg\{\min_{d \in R}\{score(B_{i,j},d)\}\}$$

(3)

以往 NCR 算法的子窗口是固定的正方形, 因此在匹配过程中往往会有部分子窗口跨在不同物体上 (深度不连续区域), 从而使得空间连续性条件不再满足. 为解决这一问题, 我们首先使用 K-均值分割对源图像进行分割, 再将分割后的结果作为传统 NCR 算法中的子窗口.

3 K-均值分割

K 均值聚类的基本思想是: 首先随机选取 K 个点作为初始聚类中心, 计算各个样本到聚类中心的距离, 把样本归属到离它最近的那个聚类中心所在的类; 然后对调整后的新类计算新的聚类中心, 重新聚类, 如此反复的操作, 直到相邻两次的聚类中心的变化在一定范围内时聚类结束[8-11].

本文采用的初始聚类中心是如下设置的: 首先将图像分成等大小正方形区域, 以每个正方形块为初始类, 该类的颜色为每个正方形块内像素的颜色均值, 空间位置为每个正方形块的中心位置.

本文的聚类过程是根据计算每个像素点属于各个类的概率来实现的. 像素点属于某个类的概率通过下式来计算:

$$probability =$$
$$\exp\left(-\left(\frac{color\_distance^2}{\Delta^2}+\lambda\times\frac{spatial\_distance^2}{b^2}\right)\right)$$
$$(4)$$

其中，$color\_distance$ 表示像素点与聚类中心在颜色空间中的距离，$spatial\_distance$ 表示像素点与聚类中心在空间分布上的距离，b 是初始正方形块的边长，Δ 为分割块内的颜色粗糙度，λ 用来控制颜色和空间距离对聚类影响的程度.

图 2 为采用 K-均值分割对 Middlebury 平台的标准图像 Cone 及 Teddy 分割后的结果. 从图中可以看出，没有分割块跨在不同的物体上.

（a）Cone

（b）Teddy

图 2　K-均值分割结果（$b=7,\Delta=5,\lambda=3$）

4　遮挡区域的检测

由于遮挡区域没有相对应的匹配区域，对这些区域进行匹配将没有意义. 而且遮挡区域还会严重影响其邻近非遮挡区域的匹配. 因此，需要将遮挡区域标记出来，去除这部分区域在匹配时的不良影响. 在遮挡区域的检测方面，目前较有效的方法是基于像素点的左右匹配一致性的检测方法[12-14].

基于左右匹配一致性的遮挡检测方法的基本思想是：分别以左、右视图为匹配的基准得到两幅视差图（分别称为左视差图和右视差图）；然后先以左图中区域 A 出发，根据左视差图在右图中找到其对应匹配区域为 B，再以右图中 B 区域为已知量，根据右视差图回到左图中去寻找其对应的区域，若对应区域仍为 A，那么我们称区域 AB 是相互一致的. 相互一致用数学形式可表示为

$$D_L(A)+D_R(B)=0 \qquad (5)$$

这里，$D_L(A)$ 和 $D_R(B)$ 分别表示区域 A 在左视差图和区域 B 在右视差图的视差. 考虑实际匹配过程中存在一定误差，令 $delta=D_L(A)+D_R(B)$. 若 $delta$ 的绝对值小于一定范围，即可认为区域 AB 是相互一致的.

5　邻域权重的设置

采用 K-均值对源图像进行分割以后，可以保证分割后的子窗口位于同一物体之内. 接下来就可以以这些分割后的子窗口为基础，结合相关邻域的信息进行匹配. 其中权重因子需要根据邻域子窗口的内容及和待匹配子窗口的关系来调整. 本节讨论如何根据图像的具体内容设置邻域权重因子.

首先，邻域分割块与中心分割块颜色越接近，其属于同一物体的可能性越大，彼此深度一致的可能性就越大；同样，邻域分割块与中心分割块在空间上越接近，其深度一致的可能性也就越大[15]. 基于这两方面的考虑，我们将邻域分割块的权重设置为颜色距离权重和空间距离权重之积.

$$w(B_{i,j},B_{k,l})=w_c\times w_d=$$
$$\exp\left\{-\left[\frac{color\_dis(B_{k,l},B_{i,j})}{\sigma}+\frac{spatial\_dis(B_{k,l},B_{i,j})}{b\sqrt{n}}\right]\right\}$$
$$(6)$$

这里，$color\_dis(B_{k,l},B_{i,j})$ 是块 $B_{k,l}$ 的平均颜色和块 $B_{i,j}$ 的平均颜色在颜色空间的距离，$spatial\_dis(B_{k,l},B_{i,j})$ 是块 $B_{k,l}$ 的中心和块 $B_{i,j}$ 的中心的欧式距离，n 为邻域的个数，σ,b 为常量.

除了对分割块进行颜色、空间权重设置之外，还需对分割块进行可信度的设置. 可信度的设置是依据匹配块内的颜色粗糙程度和是否为遮挡区域设定的[6-7]. 匹配块内的颜色粗糙程度越高，所含的匹配信息越丰富，匹配结果的可信度就越大. 对于遮挡区域的可信度可以直接设为 0. 设置权重因子后的代价函数如下：

$$score(B_{i,j},\boldsymbol{d})=\left(SSD(B_{i,j},\boldsymbol{d})\times R_{i,j}+\right.$$
$$\left.\sum_{B_{k,l}\in N(B_{i,j})}w(B_{k,l},B_{i,j})\times R_{k,l}\min_{\boldsymbol{\delta}}\{SSD(B_{k,l},\boldsymbol{d}+\boldsymbol{\delta})\}\right)\Big/$$
$$\left(R_{i,j}+\sum_{B_{k,l}\in N(B_{i,j})}w(B_{k,l},B_{i,j})\times R_{k,l}\right) \qquad (7)$$

这里，$R_{k,l}$ 表示邻域分割块 $B_{k,l}$ 的可信度，$R_{k,l}\in[0,1]$.

6　实　　验

我们分别将本文算法用于 Middlebury 平台中的四组标准图像对 Tsukuba，Venus，Cone，Teddy

上,并将结果与同类算法进行了比较.本文所列出的各类算法的性能数据均引用自 Scharsteia 和 Szeliki 的立体平台网站(vision. middlebury. edu/stereo).

6.1 本算法测试结果

图 3 为应用本算法对 Middlebury 平台中的 4 组标准图像对进行深度提取的结果.4 组图像采用的参数相同($b=7,n=48,\lambda=3,\sigma=1$).表 1 给出了采用 Middlebury 标准平台对 4 幅深度图进行定量评估的结果.
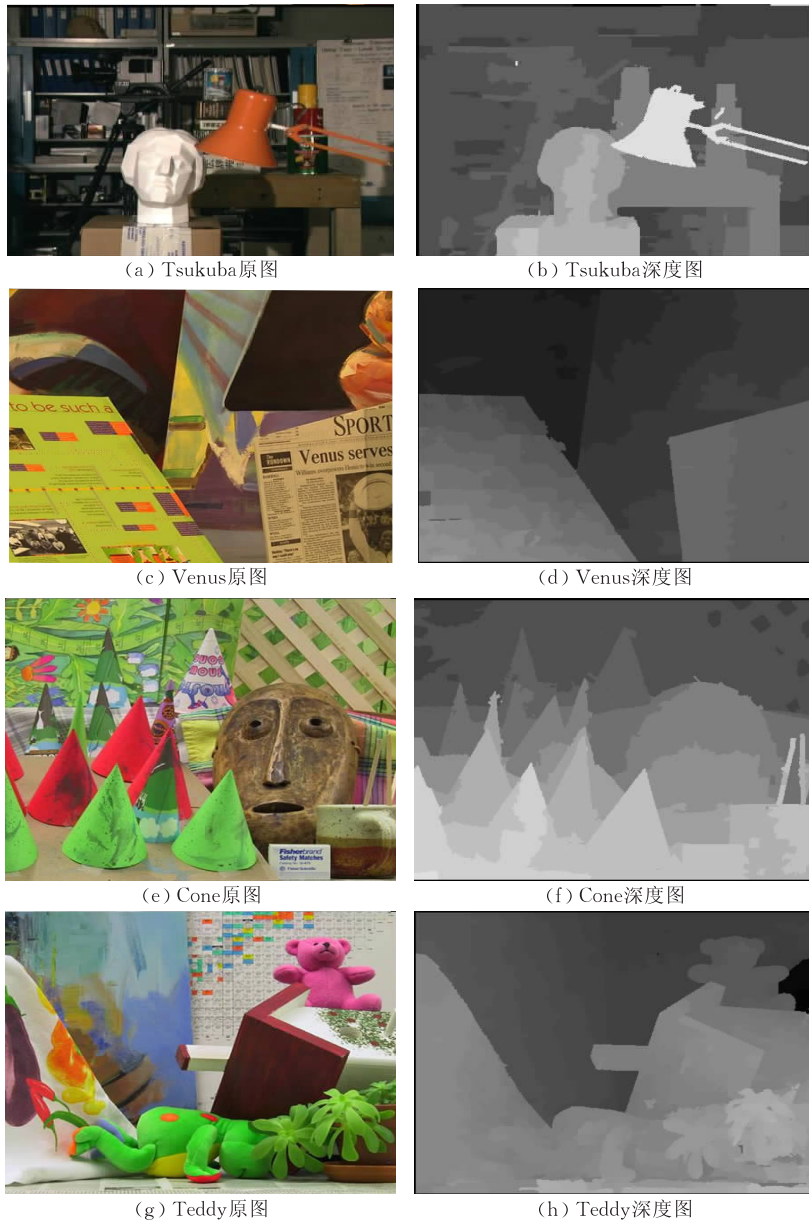
（a）Tsukuba原图 　　　　　　　　（b）Tsukuba深度图

（c）Venus原图 　　　　　　　　（d）Venus深度图

（e）Cone原图 　　　　　　　　（f）Cone深度图

（g）Teddy原图 　　　　　　　　（h）Teddy深度图

图 3 　采用本算法对 Middlebury 平台 4 组标准图像进行深度提取的结果

表 1 　采用 Middlebury 平台对本算法提取结果的定量评估

图像名称	错误百分比/%					均方误差 (RMS)/%
	all	non-occ	untex.	disc.	Occl.	
Tsukuba	1.61	1.39	0.24	8.02	10.23	0.82
Venus	1.03	0.86	0.85	4.18	10.2	0.49
Cone	5.64	3.62	2.26	13.59	63.07	4.45
Teddy	10.22	6.99	12.64	12.04	85.97	5.25

注:all 表示所有的区域,non-occ 表示非遮挡区域,untex. 表示纹理稀少区域,disc. 表示深度不连续区域,occl. 表示遮挡区域. 匹配结果相差 2pixel 以上认为匹配错.

6.2 同类算法的比较

本文算法属于局部优化算法,同类算法中较有代表性的有变换窗口最小值滤波算法(SSD＋MF)[2]、实时相关匹配算法(Realtime correlation)[16]、紧凑窗口算法(Compact windows)[17]、可变窗口算法(Variable windows)[3]. 从表 2 同类算法错误百分比的数据比较中可以看出,本算法的各项错误率均明显低于同类算法.

表 2　同类算法错误百分比比较

	Tsukuba			Venus			Sawtooth		
	all	untex.	disc.	all	untex	disc.	all	untex	disc.
本文算法	1.61	0.24	8.02	1.03	0.86	4.18	0.57	0.15	5.04
Var. Win.	2.35	1.65	12.17	1.23	1.16	13.35	1.28	0.23	7.09
Comp. win.	3.37	3.54	12.91	1.67	2.18	12.33	1.61	0.45	7.87
Realtime	4.25	4.47	15.05	1.53	1.8	12.33	1.32	0.35	9.21
SSD＋MF	5.23	5.8	24.66	3.74	6.28	12.94	2.21	0.72	13.97

6.3　与现有性能最优算法的比较

目前排在平台前几名的算法多为全局优化算法. 它们分别是 Klaus 和 Sormann 提出的 AdaptingBP 算法[18]、Yáng 等人提出的 doubleBP 算法[19]、Zitnick 等提出的 OverSegmentBP 算法[10] 以及 Yoon 的 AdaptWeight 算法[20]. 表 3 给出本文和这几种算法性能比较的结果.

表 3　与目前平台上性能最优算法错误百分比比较

	Tsukuba			Venus			Cone			Teddy		
	non_occ	all	disc.	non_occ	all	disc.	non_occ	all	disc.	non_occ	all	disc.
本文算法	1.39	1.61	8.02	0.86	1.03	4.18	5.60	7.64	13.6	6.99	10.2	12.0
AdaptingBP	1.11	1.37	5.79	0.01	0.21	1.44	4.22	7.06	11.8	2.48	7.92	7.32
DoubleBP	0.88	1.29	4.76	0.14	0.60	2.00	3.55	8.71	9.70	2.90	9.24	7.82
OverSegmBP	1.69	1.97	8.47	0.50	0.68	4.69	6.74	11.9	15.8	3.19	8.81	8.89
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

在 Tsukuba 和 Venus 两幅图像上,本文算法与上述最优算法接近. 但在 Cone 以及 Teddy 两幅图像上,本文算法的错误百分比要略高于平台上的最优算法. 然而,上述基于全局优化的算法在匹配阶段都需要多次迭代,计算复杂度高,计算时间较长. AdaptWeight 算法虽然不需要迭代,但它每次只能计算一个像素点的视差,计算效率相当低. 以上几种算法中只有 Zitnick 的 OverSegmentBP 算法在论文中给出了具体的运行时间和环境参数（CPU 2.8GHz,匹配阶段运行时间 40s）[10]. 本文在同等条件下的匹配过程只需要 5s 左右.

7　结　论

从以上实验数据及分析可以看出,本算法提取的深度图的错误率低于局部优化算法,接近全局优化算法. 但是,运行效率高于现有的全局优化算法. 在综合性能方面是出众的.

参　考　文　献

[1] Trucco E. Introductory Techniques for 3-D Computer Vision. Englewood Cliffs, NJ: Prentice Hall, 1998

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

[3] Kanade T, Okutomi M. A stereo matching algorithm with an adaptive window: Theory and experiment. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(9): 920-932

[4] Zhou Xiu-Zhi, Wen Gong-Jian, Wang Run-Sheng. Fast stereo matching using adaptive window. Chinese Journal of Computers, 2006, 29(3): 473-479(in Chinese)
(周秀芝,文贡坚,王润生. 自适应窗口快速立体匹配. 计算机学报, 2006, 29(3): 473-479)

[5] Dufaux F, Moscheni F. Motion estimation techniques for digital TV: A review and a new contribution. Proceedings of the IEEE, 1995, (6): 857-876

[6] Chen Y, Lin Y, Kung S. A feature tracking algorithm using neighborhood relaxation with multi-candidate pre-screening//Proceedings of the International Conference on Image Processing. Switzerland, 1996: 513-516

[7] Wu C, McCormick M, Aggoun A, Kung S. Depth mapping of integral images through viewpoint image extraction with a hybrid disparity analysis algorithm. Journal of Display Technology, 2008, 4(1): 101-108

[8] Lucchese L, Mitra S. Unsupervised segmentation of color images based on k-means clustering in the chromaticity plane//Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries. Fort Collins, 1999: 74-78

[9] Wagstan K, Cardie C, Rogers S, Schroedl S. Constrained k-means clustering with background knowledge//Proceedings of the 18th International Conference on Machine Learning. San Francisco, 2001: 577-584

[10] Zitnick C, Kang S. Stereo for image-based rendering using image over-segmentation. International Journal of Computer Vision, 2007, 75(1): 49-65

[11] Bleyer M, Gelautz M. A layered stereo algorithm using image segmentation and global visibility constraints//Proceedings of the International Conference on Image Processing. Singapore, 2004: 2997-3000

[12] Intille S, Bobick A. Disparity-space images and large occlusion stereo//Proceedingss of the European Conference on Computer Vision. Stockholm, Sweden, 1994: 179-186

[13] Kang S, Szeliski R, Chai J. Handling occlusions in dense multi-view stereo//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, 2001, 1: 103-110

[14] Sun J, Li Y, Kang S, Shum H. Symmetric stereo matching for occlusion handling//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, 2005: 399-406

[15] Yang Q, Wang L, Yang R, Stewenius H, Nister D. Stereo matching with color-weighted correlation: Hierachical belief propagation and occlusion handling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(3): 492-504

[16] Yoon K, Kweon I. Adaptive support-weight approach for correspondence search. IEEE Transactions on Pattern Analy-

sis and Machine Intelligence, 2006, 28(4): 650-656

[17] Veksler O. Stereo matching by compact windows via minimum ratio cycle//Proceedings of the 8th International Conference on Computer Vision. Vancouver, 2001: 540-547

[18] Hirschmuler H. Improvement in real-time correlation-based stereo vision//Proceedings of the Workshop on Stereo and Multi-Baseline Vision. Hawaii, 2001: 141-148

[19] Yang Q, Yang R, Davis J, Nistér D. Spatial-depth super resolution for range images//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, 2007: 2003-2010

[20] Klaus A, Sormann M, Karner K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure//Proceedings of the 18th International Conference on Pattern Recognition. 2006: 15-18

WU Chun-Hong, born in 1972, Ph. D. , associate professor. Her current research interests include 3D image technology and image analysis.

FU Guo-Liang, born in 1983, M. S. . His research interests include image analysis and computer vision.

Background

Stereo correspondence has traditionally been and continues to be one of the most heavily investigated topics in computer visions. Although much progress has been achieved during the past decades, the corresponding problem is still far from mature. Lack of intelligence as human being is the one of the major problem.

By far, several critical problems in stereo correspondence for dense depth map lies in the matching window size and shapes control, the region lack of texture and the occluded region. This paper attempts to tackle above problems by using a Neighborhood Constraint and Relaxation method base on K-means Segmentation. In our method, more than one matching windows are involved in the matching process. Not only the shape but also the size, number and weight of each matching window can be modified intelligently according to the image content. K-means Segmentation is used to detect object edges and keep the matching window within the same object. Neighborhood Constraint and Relaxation Algorithm is to utilize the environment information. The combination tackles the problem of how to choose an appropriate matching window intelligently. For the region lack of texture, the uncertainty of the corresponding can be reduced by using the information from its neighboring regions. The occluded region is detected and removed from the calculation.

We test our algorithm using Middlebury stereo test bed (http://vision. middlebury. edu/stereo) and compared the results with other algorithms listed on the website. The error percentage of the depth map obtained by our algorithm is

lower than other algorithms based on local optimization and approaching to global optimization algorithms. The efficiency of our algorithm is higher than other global optimization algorithms. The overall performance is outstanding.

The task is one of the subtasks of "Depth extraction from 3D integral images". Integral imaging is a technique capable of displaying 3D images with continuous parallax in full natural color. Extracting depth information from integral image has various applications ranging from remote inspection, robotic vision, medical imaging, virtual reality, content-based image coding and manipulation for integral imaging based 3D TV. The results not only can provide fundamental for data processing of the next generation II-based 3D TV but also can be used as a novel depth measurement methodology.

Below are our representative publications:

1) Depth Mapping of Integral Images Through Viewpoint Image Extraction With a Hybrid Disparity Analysis Algorithm. Journal of Display Technology, 2008

2) Depth measurement from unidirectional integral images using a modified multi-baseline disparity analysis algorithm. Journal of Electronic Imaging, 2005

3) A depth measurement approach based on integral imaging and multiple-baseline stereo matching algorithm. Acta Electronica Sinica, 2006

This current result can be used to further improve the depth measurement result from plane recorded 3D integral images.