

基于图割的图像匹配算法

王 年¹, 范益政^{1,2}, 鲍文霞¹, 韦 穗¹, 梁 栋¹

(1. 安徽大学计算智能与信号处理教育部重点实验室, 安徽合肥 230039; 2. 安徽大学数学系, 安徽合肥 230039)

摘 要: 用标号表示视差, 建立能量函数, 把匹配问题转化为能量函数最小化问题; 通过构造网络, 使能量与网络的割的容量相联系; 利用图的网络流理论给出能量函数的最小化, 从而获得图像匹配的视差数据. 与目前已有基于图割的匹配算法相比, 本算法将标号从 1 维向量推广到 2 维向量, 适用于更一般情形下的视觉匹配, 并且在全局上获得能量函数最小. 实验结果表明, 所提的匹配算法准确率较高.

关键词: 匹配; 能量函数; 图割; 网络流; 视差

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2006) 02-0232-05

An Images Matching Algorithm Based on Graph Cuts

WANG Nian¹, FAN Yi-zheng², BAO Wen-xia¹, WEI Sui¹, LIANG Dong¹

(1. Key Laboratory of Intelligent Computing & Signal Processing, Anhui University, Ministry of Education, Hefei, Anhui 230039, China; 2. Department of Mathematics, Anhui University, Hefei, Anhui 230039, China)

Abstract: Label is denoted by disparity and the energy function is established. Then the problem of matching can be transformed into that of energy function minimization. A network is constructed such that the energies can be related to the capacities of the cuts of the network. Finally, the minimal energy is obtained by the network-flows theory, and hence the disparity data are solved. Comparing with some known algorithms based on graph cuts, the algorithm in this paper extends the label from 1 dimension vector to 2 dimension vector, and adapts vision matching of more general conditions; furthermore the algorithm can gain the minimization in global. Experimental results show that the algorithm has a high accuracy.

Key words: matching; energy function; graph cuts; network flow; disparity

1 引言

在众多视觉应用中, 图像匹配是一个关键任务, 即给定同一场景的两幅图像, 寻找同一场景点投影到图像中的像素之间的对应关系. 本文给出立体和运动的匹配算法. 在立体中, 图像被同时从不同视角获得, 匹配给出深度线索. 在运动中, 图像在不同时间获得, 而匹配给出移动线索.

根据视差数据的分布情况, 匹配算法分为稀疏视差匹配和密集视差匹配两大类. 稀疏视差匹配^[1]一般以过边缘轮廓、线段等图像特征为匹配基元, 并仅匹配两幅图像的特征区域. 这些方法通常都比较鲁棒, 但一般情形下获得比较稀疏的结果. 密集视差算法^[2]通常以图像灰度或其它特性作为匹配基元, 由候选点邻域间的相关程度作为判别依据, 常用的方法有: 区域相关法^[3], 动态规划法^[4], 以及基于光流场^[5]和基于图割^[6]的方法等. 其优点在于给出所

有 (或几乎所有) 像素的视差信息, 但会产生一些误匹配.

通过构建能量函数试图获得图像的某些全局性质, 在早期的视觉中得到广泛的应用. 但不可避免的是通常很难获得能量函数的全局最小化. 鉴于此, 很多作者更倾向于寻找局部小的求解. 然而在一般情形下, 局部小可以任意地偏离全局小, 从而不能带来任何的全局性质. 这方面的工作主要有: 梯度下降法^[7]和模拟退火算法^[8], 前者适用于几乎所有连续变量的能量函数, 而后者适用于几乎所有离散变量的能量函数. 此外, 模拟退火算法在理论上虽然能获得能量函数的全局小, 但是它需要指数级的时间. 一个值得注意的工作为 Boykov 等人提出的基于图割理论有效的能量函数优化方法^[6]. 在该方法中, 作者提出标号函数的两种比较大的移动, 扩张移动 (expansion moves) 和交换移动 (swap moves), 并证明了其扩张算法所获得的局部小和全局小相差一个已知的常数, 而交换算法可以处理更一般的能量函数. 文献 [2] 和 [14] 把当前常用的几种优化

收稿日期: 2005-01-17; 修回日期: 2005-05-26

基金项目: 国家自然科学基金 (No. 60473102); 安徽省教育厅自然科学基金 (No. 2005KJ005ZD); 安徽大学 211 工程学术创新团队基金

算法进行了比较,发现:与传统的 8 种算法 (例如模拟退火法、M 估计法等)相比,图割算法不仅总体精度高,而且在不连续区域和低纹理区域的精度也比其它算法都高.不仅如此,即便有些算法 (例如模拟退火法)的精度与图割算法的精度接近,但图割算法在优化过程中收敛更快.

注意到文献 [6] 的图像匹配算法只适应于摄像机沿摄像机坐标系的 X 轴或 Y 轴做平移运动的场合,其标号集中的元素为 1 维向量. (这也包括大多数的匹配算法^[9~13],它们都需要摄像机的某些运动信息).为了适应于更普遍的场合,本文对能量函数进行重新定义,把标号集中的元素扩展到二维向量 (从而适用于一般场合的立体匹配和运动匹配),同时为了获得全局小,建立了不同于文献 [6] 的网络,最后应用网络流理论,获得能量函数的全局最小.算法也体现了区域内部的连续性和边界的不连续性.实验结果表明所提的图像匹配算法精确度比较高.

2 基本理论

首先简要介绍图的网络流理论. 设 $G = (V, E)$ 为一个无向图,其中 V 为点集, E 为边集, $s, t \in V$ 为其固定的两点. 对于连接点集 V 中 x, y 的一条边 e , 可视为有两个不同的方向,即从 x 到 y 和从 y 到 x , 分别记为 (e, x, y) 和 (e, y, x) . 对于 G 的每个边都进行上述操作, 所获得的有向边的集合记为 E . 在 E 上定义容量函数 $c: E \rightarrow \mathbf{R}^+$, 其中 \mathbf{R}^+ 为非负实数. 称上述图 G 及其边集上定义的容量函数构成了一个网络, 记为 $N = (G, s, t, c)$, 其中 s, t 分别称为源点和汇点. 称函数 $f: E \rightarrow \mathbf{R}$ 为网络的一个流, 如果 f 满足如下条件:

- (1) 对所有的 $(e, x, y) \in E, f(e, x, y) = -f(e, y, x)$;
- (2) 对任意的 $x \in V \setminus \{s, t\}, \sum_{(e, x, y) \in E} f(e, x, y) = \sum_{(e, y, x) \in E} f(e, y, x) = 0$;
- (3) 对任意的 $e \in E, f(e) \leq c(e)$.

此时,称 $f(s, V) = \sum_{(e, s, x) \in E} f(e, s, x)$ 为流 f 的流量. 若 $s \in V$, 且 $s \in S, t \in V \setminus S$, 称 $S, V \setminus S$ 之间的边的集合为网络 $N = (G, s, t, c)$ 一个割. 割的容量定义为割中各条边容量之和. 称 $N = (G, s, t, c)$ 中容量最小的割为网络 N 的最小割. 根据 Ford-Fulkerson 的最大流 - 最小割定理 (即一个网络的最大流量等于其最小割的容量), 网络的最小割问题可以通过计算最大流来得到.

在我们的算法中, 首先将匹配问题中的视差看成像素的标号. 对于待匹配的两幅图像, 用 P 表示第一幅图像中所有像素的集合, 匹配的目的是给 P 中的每个像素 p 分配一个标号 $f(p) = \{l_1, l_2, \dots, l_n\}$, 这里标号为像素 p 的可能视差, L 表示所有可能标号的集合. 当摄像机沿摄像机坐标系的 X 轴或 Y 轴做平移运动时, 视差中只有一个分量在改变, 而另一个分量保持不变. 此时标号集 L 中的元素为 1 维向量, 但对摄像机作任意运动, 标号集 L 为二维向量的

集合. 因此, 匹配问题就是: 对第一幅图像中的每个像素 p , 寻找它的恰当标号 $f(p)$, 使得 p 与第二幅图像中的像素 $p + f(p)$ 为一对匹配点, 也即寻找第一幅图像 P 到 L 的一个恰当映射 f .

为了获得像素的恰当标号, 需要构造一个关于标号的能量函数. 对能量函数的构造, 通常有两大约束: 数据约束和平滑约束. 数据约束在图像匹配中描述的是: 第一幅图像中的像素 p 和第二幅图像中匹配的像素 $p + f(p)$ 亮度应该相近; 平滑约束描述的是: 正确的视差映射 f 的变化应该是平滑的, 即一幅图像中邻近的像素 p 和 q 应该有相近的视差 $f(p)$ 和 $f(q)$. 它体现了区域内部的连续性和边界的不连续性. 根据数据和平滑约束, 关于匹配的能量函数可构造如下:

$$E(f) = E_{\text{data}}(f) + E_{\text{smooth}}(f) = \sum_p D_p(f(p)) + \sum_{(p,q) \in N} V_{p,q}(f(p), f(q)) \quad (1)$$

其中, N 表示相邻的像素对构成的集合.

为了最小化上述能量函数 (1), 我们构建一个无向网络 N , 使得标号函数集和网络 N 的 (几乎所有) 的图割构成双射, 并且对于每一个标号 f , 其能量 $E(f)$ 即为其所对应的割的容量. 根据上述映射关系, 能量函数 (1) 的最小化问题等价于 N 的最小割求解问题, 而最小割问题可以通过计算最大流来得到解决.

3 基于图割的匹配算法

3.1 能量函数的构造

设 I_1, I_2 为二灰度图像. 对 I_1 构造标号函数 f , 即 $f(p) = \{l_1, l_2, \dots, l_n\}$ 为像素 p 的标号. 此处 $f(p)$ 为 I_1 中像素 p 在 I_2 中的可能视差. 构造能量函数

$$E(f) = \sum_p [I_{n_1}(p) - I_{n_2}(p + f(p))]^2 + \sum_{(p,q) \in N} [G_{k_1}(p) - G_{k_2}(p + f(p))]^2 \quad (2)$$

其中 $I_{n_k}(p), G_{k_k}(p), (k=1, 2)$ 分别表示像素 p 在第 k 幅图像中的灰度以及梯度向量.

注意到式 (2) 中并没有式 (1) 中的平滑项, 其实, 式 (2) 中关于梯度向量的项 $[G_{k_1}(p) - G_{k_2}(p + f(p))]^2$ 在本质上已经体现了平滑约束, 因为若 $p \in I_1$ 位于某区域的内部 (或边界), 其在 I_2 中实际的匹配点 $p + f(p)$ 也应在匹配区域的内部 (或边界), 此时 $G_{k_1}(p)$ 和 $G_{k_2}(p + f(p))$ 的距离应当很小. 故 $[G_{k_1}(p) - G_{k_2}(p + f(p))]^2$ 体现了区域内部的连续性和边界的不连续性. 我们把相邻像素之间的作用定义为像素的梯度向量之间的距离, 主要是为了有利于如下网络的构造.

3.2 网络的构造

注意到本文中 标号集 L 是一个二维向量集合. 对于 L 中元素, 我们按字典序给它们排列, 即 $(x_1, y_1) < (x_2, y_2)$, 如果 $x_1 < x_2$, 或 $x_1 = x_2$ 但 $y_1 < y_2$. 不妨假设 L 中元素有如下排列:

$$l_1 < l_2 < \dots < l_n$$

构建网络如下:建立三维坐标系 (O; X, Y, Z),把图像 I_1 置于 OXY平面,使得 I_1 的原点, X, Y轴与 OXY平面的原点以及相应的轴重合.在 Z的正半轴上,从原点开始,等距离地放置向量 $l_1, l_2, \dots, l_n, l_1 = 0$.令 $q_0 = l_1 = 0$,对 $i = 1, 2, \dots, n - 1$,在 (l_i, l_{i+1}) 的中点放置点 $q_i, q_n = l_n$.由此,由 OXY平面中像素点 $p = (p_x, p_y)$ 以及 Z的正半轴上的点 q_0, q_1, \dots, q_n 构成了一个立方体网格.可以看出,对 $i = 0, 1, 2, \dots, n - 1$,每个 Z的轴上区间 $[q_i, q_{i+1}]$ 恰含一个 l_{i+1} .记 $(p; q_i) = (p_x, p_y, q_i)$ 为立方体网格上节点, $N(p)$ 为像素点 p 邻域.在立方体网格的上下方添加两个点 s, t (如图 1所示).并在 s 到 l_i 每个像素点之间添加一个边,在 t 到集合 $\{(p_x, p_y, q_n) | (p_x, p_y) \in I_1\}$ 中的每个点,即立方体网格上与 OXY平面相对立的另一个面上节点,添加一个边.由此,获得一个无向图 $G = (V, E)$,其中

$$V = \{s, t\} \cup \{(p; q_i) | p \in I_1, i = 0, 1, \dots, n\},$$

$$E = \{(s, (p; q_0)), (t, (p; q_n)) | p \in I_1\} \cup E_s \cup E_d$$

其中, $E_s = \{(p; q_i), (q; q_i) | q \in N(p), p \in I_1, i = 0, 1, \dots, n - 1\}$ 为光滑边构成的集合, $E_d = \{(p; q_i), (p; q_{i+1}) | p \in I_1, i = 0, 1, \dots, n - 1\}$ 为视差边构成的集合.

设 $H = (V, E \setminus E_s)$ 即 H 为 $G = (V, E)$ 删除光滑边之后的图.定义 H 中各边的容量如下:

(1) $\{(s, (p; q_0)) | p \in I_1\}, \{(t, (p; q_n)) | p \in I_1\}$ 中边的容量为 $+$ 或者充分大的正数;

(2) 视差边的容量为:对任意 $p \in I_1, i = 0, 1, \dots, n - 1$,边 $e_{p, i+1} = ((p; q_i), (p; q_{i+1}))$ 的容量为

$$c(e_{p, i+1}) = [I_{n_1}(p) - I_{n_2}(p + l_{i+1})]^2 + G_{T_1}(p) - G_{T_2}(p + l_{i+1})^2 \quad (3)$$

由此,我们构建了一个网络 $N = (H, s, t, c)$.

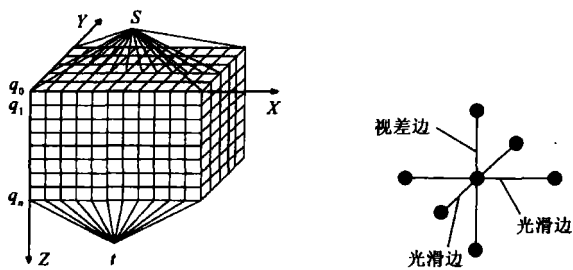


图 1 网络 $N = (H, s, t, c)$

3.3 能量函数的最小化

对于网络 $N = (H, s, t, c)$ 的每一个割 C ,它必分离 s, t .若 C 不含 s 边或 t 边(即与 s 或 t 关联的边),则对每个像素 p ,它至少经过集合 $\{(p; q_i), (p; q_{i+1}) | i = 0, 1, \dots, n - 1\}$ (即由像素 p 出发的视差边)中一条边.由于 s 边或 t 边的容量为 $+$ 或者充分大,显然不可能为最小割.故我们仅考虑如下类型的割:它仅包含每个像素 p 出发的视差边中一条边;称这种类型的割为有效割,记有效割的集合

为 $EC(N)$.

引理 1 存在割 $C \in EC(N)$, C 的容量为网络 $N = (H, s, t, c)$ 的最小割的容量.

证:根据前面讨论, $N = (H, s, t, c)$ 的最小割不可能含 s 边或 t 边.若最小割 C 含有某像素 p 出发的视差边中 k 条边 $e_{p, i_1}, e_{p, i_2}, \dots, e_{p, i_k} (1 \leq i_1 < i_2 < \dots < i_k \leq n, k \geq 2)$,则 $C' = C \setminus \{e_{p, i_1}, \dots, e_{p, i_k}\}$ 依然是 $N = (H, s, t, c)$ 的一个割,且割 C' 的容量不大于 C 的容量.重复上述操作,直到所获得的割中仅含由每个像素出发的视差边中的一条边,即为一个有效割,而且其容量不大于 C 的容量.因此, $N = (H, s, t, c)$ 的最小割在有效割 $EC(N)$ 中达到.

注:若网络 $N = (H, s, t, c)$ 的最小割 C 不在 $EC(N)$ 中,即存在某像素 p ,其出发的视差边中有 k 条边 $e_{p, i_1}, e_{p, i_2}, \dots, e_{p, i_k} (k \geq 2)$ 在 C 中.根据上述讨论,该 k 条边的容量都为 0.此时,从这 k 条边任意删除 $(k - 1)$ 条边,就获得一个新的最小割.对其它和像素 p 有类似性质,可重复操作.最后就获得一个最小割在 $EC(N)$ 中.根据式 (3),上述情形一般不容易出现.

定理 1 对于网络 $N = (H, s, t, c)$,标号函数集和有效割集 $EC(N)$ 存在一个双射,并且每个标号函数 f 的能量 $E(f)$ 等于其所对应的割 $C(f)$ 的容量.

证:对于任给的标号 f ,其对像素 p 的标号 $f(p)$ 为某个 $l_{i+1} = l(p, f) (0 \leq i \leq n - 1)$,而 l_{i+1} 属于唯一的区间 $[q_i, q_{i+1}]$,建立如下映射:

$$f(p) = l(p, f) = l_{i+1} \mapsto e_{p, i+1} = ((p; q_i), (p; q_{i+1}))$$

可以看出,上述映射像的集合构成了 $EC(N)$ 的一个割;并且根据式 (3),割的容量为 $c(e_{p, i+1}) = E(f)$.反之,

对 $EC(N)$ 的任意一个割 C ,对任意的像素 p , C 必经过由 p 出发的视差边 $\{(p; q_i), (p; q_{i+1}) | i = 0, 1, \dots, n - 1\}$ 中唯一的一条边,设为 $((p; q_i), (p; q_{i+1}))$.对 p 的标号为区间 $[q_i, q_{i+1}]$ 中唯一的 $f(p) = l_{i+1}$.因此获得了上述映射的逆映射.

根据定理 1,能量函数 (2)的最小化问题便可转化为有效割集 $EC(N)$ 的最小割问题.根据引理 1, $EC(N)$ 的最小割即为网络 $N = (H, s, t, c)$ 的最小割.根据 Ford-Fulkerson 的最大流 - 最小割定理,网络 $N = (H, s, t, c)$ 的最小割可通过计算最大流来得到.因此,根据引理 1后的注,一旦确定了 $N = (H, s, t, c)$ 的最小割,也就获得了 $EC(N)$ 的最小割.根据定理 1,最小割所对应的标号 f ,满足 $E(f) = \min E(f)$,从而获得了问题的求解.

4 实验及其结果

利用上述算法,分别对立体图像和运动图像进行匹配实验.为了突出水平和垂直视差图的效果,所给出的视差是按一定比例放大的.

图 2 为一对立体图像匹配结果.其中 (a)、(b)分别为

左右图像, (c)、(d)分别为水平和垂直视差图, (e)、(f)分别为左右图像上检测的特征点图像, 特征点数目为 564 个, (g)为根据视差数据得到的特征点匹配图像, (f)为剔除误匹配后的特征点匹配图像, 其匹配特征点数目为 424 个, 图割算法匹配准确率为 75.18%。这里剔除误匹配的方法为: 由左图像上的特征点根据其视差数据如果在右图像上能找到相应的特征点, 则认为其匹配正确, 否则为误匹配。

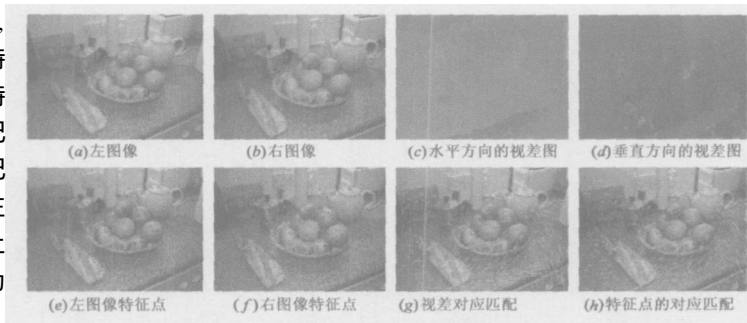


图 2 立体匹配

图 3 为一辆汽车的运动序列图像的匹配结果。其中 (a)、(b)分别为运动序列的第 15 帧和第 20 帧图像, (c)、(d)分别为其水平和垂直视差图, (e)、(f)分别为第 15 帧和第 20 帧图像上检测的特征点图像, 特征点数目为 556 个, (g)为根据视差数据得到的特征点匹配图像, (f)为剔除误匹配后的特征点匹配图像, 其匹配特征点数目为 450 个, 图割算法匹配准确率为 80.9%。这里剔除误区配的准则同上。

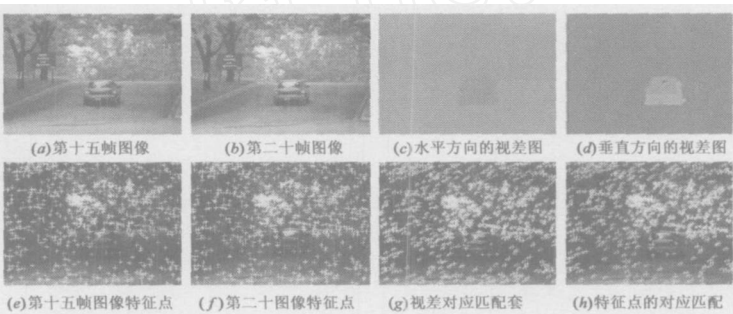


图 3 运动匹配

实验中标号的范围是根据经验来选择的, 如果范围选择过大不仅浪费时间而且没有实际价值, 相反如果范围选择过小则不能正确匹配, 所以标号范围的选取对算法的正确实现尤其重要。

5 结论

本文应用图割理论, 给出了一种匹配算法。该算法不需要摄像机的位置信息 (对立体匹配) 以及运动信息 (对运动匹配)。相对于以往的基于图割的算法, 我们把标号从 1 维推广到 2 维, 从而适用于更一般情形下的匹配问题。此外, 通过用像素的梯度向量的距离来代替平滑项, 更利于网络的构造和算法的实现, 而且我们的算法实现了能量函数的全局最小。真实图像实验反映了我们的算法鲁棒性较好, 匹配准确率达到 75% 以上。

参考文献:

- [1] Y C Hsieh, D M cKeown, F P Perlant Performance evaluation of scene registration and stereo matching for cartographic feature extraction [J]. IEEE Trans, 1992, PAMI-14 (2): 214 - 238.
- [2] D Scharstein, R Szeliski A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. International Journal of Computer Vision, 2002, 47(1): 7 - 42.
- [3] D Gennery. Modelling the environment of an exploring vehicle by means of stereo vision [D]. Stanford: Stanford University, 1980.
- [4] 徐奕, 周军, 周源华. 基于动态规划的相位匹配和遮挡检测 [J]. 电子学报. 2004, 32 (4): 591 - 595.
Xu Yi, Zhou Jun, Zhou Yuan-Hua Phase matching and occlusion detection based on dynamic programming [J].

Acta Electronica Sinica, 2004, 32 (4): 591 - 595.

- [5] M O tte, H -H Nagel Optical flow estimation: advances and comparisons [A]. Eklundh Jan-O lof (Ed) Third European Conference on Computer Vision Volume 1 Series: Lecture Notes in Computer Science, Vol 800 Volume package: Computer Vision-ECCV '94 [C]. Stockholm, Sweden, Proceedings, 1994. 1. 51 - 60.
- [6] Y B oykov, O Veksler, R Zabih. Fast approximate energy minimization via graph cuts [J]. IEEE Trans, 2001, PAMI-23 (11): 1222 - 1239.
- [7] M adLea n W J, Tsotsos J K. Fast pattern recognition using gradient-descent search in an image pyramid [A]. 15 International Conference on Pattern Recognition (ICPR '00) [C]. Barcelona, Spain: IEEE Computer Society, 2000. 2. 2873 - 2877.
- [8] 孙冬梅, 裘正定. 利用薄板样条函数实现非刚性图像匹配算法 [J]. 电子学报, 2002, 30 (8): 1104-1107.
Sun Dong-mei, Qiu Zheng-ding A new non-rigid image matching algorithm using thin-plate spline [J]. Acta Electronica Sinica, 2002, 30 (8): 1104 - 1107.
- [9] Kolmogorov V, Zabih R. Multi-camera scene reconstruction via graph cuts [A]. European Conference on Computer Vision [C]. Copenhagen, 2002. LNCS 2352 (3). 82-96.
- [10] S Roy, I J Cox. A maximum-flow formulation of the N-camera stereo correspondence problem [A]. Proceedings of the Sixth International Conference on Computer Vision [C]. Bombay, India, Narosa Publishing House,

1998. 492-499.
- [11] H Ishikawa, D Geiger Occlusions, discontinuities, and epipolar lines in stereo [A]. Hans Burkhardt, Bernd Neumann (Eds). ECCV 98 [C]. Freiburg, Gemany, Proceedings: Lecture Notes in Computer Science, 1998. 1406 (1) : 232-248.
- [12] V Kolmogorov, R Zabih Computing visual correspondence with occlusions using graph cuts [A]. Proceedings of the Eighth International Conference on Computer Vision (ICCV -01) [C]. Vancouver, British Columbia, Canada: IEEE Computer Society, 2001. 2. 508-515.
- [13] O Veksler Efficient Graph-based Energy Minimization Methods in Computer Vision [D]. New York: Cornell University, 1999.
- [14] R Szeliski, R Zabih An experimental comparison of stereo algorithms [A]. Triggs, A. Zisserman, R. Szeliski

(Eds). Proceedings of the International Workshop on Vision Algorithms: Theory and Practice [C]. Springer-Verlag London, UK. Lecture Notes In Computer Science, 2000. 1883. 1-19.

作者简介:

王 年 男, 1966年生于安徽省和县, 1986年毕业于安徽大学电子工程与信息科学系, 2005年在该校获博士学位, 现为安徽大学副教授, 主要研究领域为计算机视觉、图像处理等, 已发表学术论文 20多篇. Email: wn_xlb@ahu.edu.cn



范益政 男, 1971年生于安徽省无为县, 2001年获中国科学技术大学博士学位, 现为安徽大学教授, 主要研究领域为谱图理论及应用、计算机视觉等, 已发表学术论文 20多篇.