

基于对象的预测递归视差估计

刘佳¹⁾ 安平¹⁾ 张兆扬^{1),2)}

¹⁾(上海大学通信与信息工程学院,上海 200072) ²⁾(上海大学新型显示技术及应用集成教育部重点实验室,上海 200072)

摘要 视差匹配是中间视合成、立体图像/视频编码及物体识别的关键步骤。提出了一种在大基线下基于对象的预测递归视差匹配算法。首先识别前景和背景区域,然后在对前景对象进行匹配的过程中,采用相邻块预测方法,提高了匹配的速度和精度。实验结果表明,本算法能得出良好的视差图。

关键词 视差估计 预测递归 图像分割 基于对象

中图分类号: TP391 **文献标识码:** A **文章编号:** 1006-8961(2005)11-1450-04

Predict Recursion Disparity Match Based on Object

LIU Jia¹⁾, AN Ping¹⁾, ZHANG Zhao-yang^{1),2)}

¹⁾(School of Communication and Information Engineering, Shanghai University, Shanghai 200072)

²⁾(Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai 200072)

Abstract Disparity match is the key technology that realizes intermediate view synthesis, multiocular video coding and object recognition. In this paper, a predict recursion disparity match algorithm based on object is put forward. Firstly, foreground area is distinguished from background, and then the disparity of only object area is estimated with information of neighbor block. Experimental results show that this method can obtain disparity map with high quality.

Keywords disparity estimation, predict recursion, image segmentation, based on object

1 引言

众所周知,估计视差场的过程是中间视合成^[1]、立体图像/视频编码^[2]及物体识别^[3]的关键步骤。根据匹配特征和匹配方式的不同,视差估计基本上可以分为基于区域和基于特征的方法两类,而且前者能够产生稠密的视差场。传统的基于区域的方法,比如基于块的平均绝对差值 MAD (mean absolute difference) 相关匹配算法产生的视差场不连续,视差精度差易受噪声干扰,而且计算速度慢。改进的思路有以下两类:一类是在 MAD 的基础上通过多窗口计算最小匹配代价来克服噪声干扰,并通过分级匹配法来加快计算速度和提高精度,以及利用区域信息进行窗口选择来保证物体边缘处视差场

连续;另一类是采用全局算法,如式(1)所示,在视差分配为 f 的情况下, $E_{\text{data}}(f)$ 是用来衡量图像对的匹配程度,而 $E_{\text{smooth}}(f)$ 则是用来平滑邻域内像素的视差值。全局视差估计算法就是要对某个区域(一行或整幅图)中每个像素分配合理的视差值,以使整个区域匹配代价值 $E(f)$ 最小,典型的方法有动态规划法^[4]和图分割法^[5]。

$$E(f) = E_{\text{smooth}}(f) + E_{\text{data}}(f) \quad (1)$$

本文先通过相邻块预测来减小计算时间;然后再提高精度的基础上,用插值法和平滑因子来克服噪声和视差不连续;最后利用视差检测来纠正补偿不正确的区域,该方法流程如图1所示。采用基于对象的方法,可以确保视差估计准确,由于速度快,因此可达到实时性的要求。

基金项目:上海市科委自然科学基金项目(04ZR14056);上海市教委发展基金项目(04AB56);上海市重点学科建设项目(T0102)

收稿日期:2005-09-04;改回日期:2005-09-15

第一作者简介:刘佳(1980~),男,2003年获上海海事大学学士学位,上海大学信号与信息处理专业硕士研究生。研究方向为3维视觉。E-mail: benliujia@163.com

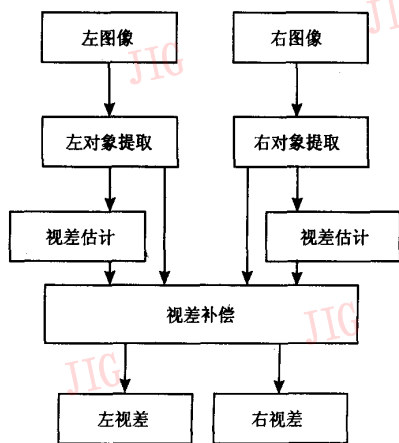


图 1 总体流程图

Fig. 1 Disparity match frame work

2 对象分割

所谓图像分割就是把具有不同意义的区域区分出来,不仅这些区域互不相交,而且每个区域都满足特定区域的一致性。图像分割是基于对象的视频编



图 2 “anna”序列对象分割

Fig. 2 Example of object extract in a “anna” sequence

3 视差匹配算法

3.1 视差估计算法

本文采用 MAD 作为判决准则来寻找左右图像中同名点,其基本思想如下:设 $S_{left}(x, y), S_{right}(x, y)$ 分别表示左右原始图像像素值, $D_{left}(x, y), D_{right}(x, y)$ 分别表示左右图像视差,行坐标为 x ,列坐标为 y ,搜索窗口宽度为 w 。在采用平行像机系统时,由于搜索可以只在水平方向进行,垂直方向的视差可以不予考虑,所以若

$$\nabla F(x, y, \hat{d}) = \min_{d_{min} \leq d \leq d_{max}} [\nabla F(x, y, d)]$$

码和视频分析的关键技术,也是计算机视觉研究中的一个瓶颈。本文假设是在一个背景均匀包含头肩像的场景中(如图 2(a)所示),先用阈值法获得图像的边缘,然后去除孤立点后,再分别从左向右搜索得出左半轮廓,从右向左搜索得出右半轮廓。具体步骤如下:

(1)用阈值获得边缘。设窗口宽度为 w ,计算图像某点 $S(x, y)$ 在 $w \times w$ 大小窗口中的方差值,并将其记作 $\sigma(x, y)$,则图像边缘灰度 $G(x, y)$ 为

$$G(x, y) = \begin{cases} 0 & \sigma(x, y) \leq T_1 \\ 255 & \sigma(x, y) > T_1 \end{cases} \quad (2)$$

其中, T_1 为阈值;

(2)去除孤立点。对于 $G(x, y) = 255$ 的点,计算其邻域非零点的个数 $Sum(x, y)$,即

$$\hat{G}(x, y) = \begin{cases} 0 & Sum(x, y) \leq T_2 \\ 255 & Sum(x, y) > T_2 \end{cases} \quad (3)$$

其中, T_2 为阈值;

(3)搜索轮廓。对每一行,从左向右搜到第 1 个 $\hat{G}(x, y) = 255$ 的点,记下其位置,即为对象的左轮廓位置。从右向左搜索,同理可得右轮廓的位置。

这种方法所得的轮廓线能保证行连续,并可满足对象提取的要求。

则 $D_{left}(x, y) = \hat{d}$, 其中

$$\nabla F(x, y, d) = \sum_{\hat{y}=-w/2}^{w/2} \sum_{\hat{x}=-w/2}^{w/2} |S_{left}(x + \hat{x}, y + \hat{y}) - S_{right}(x + \hat{x}, y + \hat{y} + d)| \quad (4)$$

为了克服噪声,本文将式(4)修改为

$$\nabla F(x, y, d) = k|d - D_{left}(x - 1, y - 1)| + \sum_{\hat{y}=-w/2}^{w/2} \sum_{\hat{x}=-w/2}^{w/2} \min_t^{1/2, 0, -1/2} |S_{left}(x + \hat{x}, y + \hat{y} + t) - S_{right}(x + \hat{x}, y + \hat{y} + t + d)| \quad (5)$$

其中, $k|d - D_{left}(x - 1, y - 1)|$ 是平滑因子, k 是系数, t 是在视差周围的半像素,式(5)中后面部分是通过半像素插值来克服噪声。

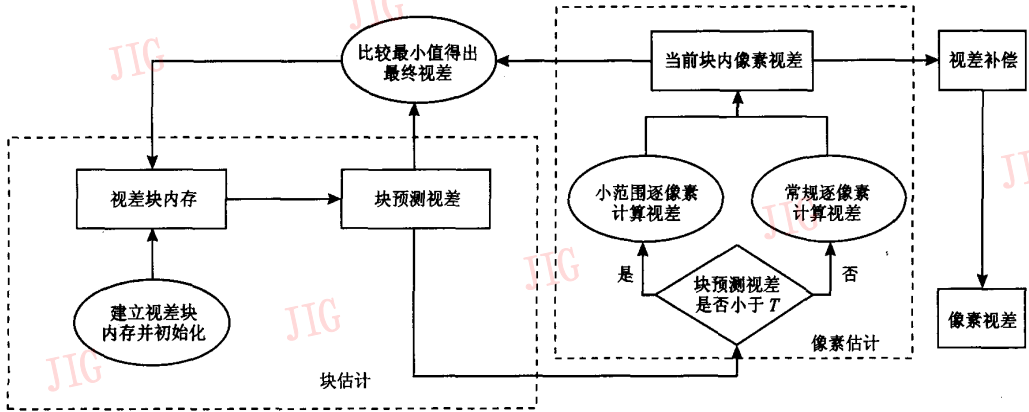


图 3 视差估计算法流程图

Fig. 3 Outline of the disparity estimation algorithm

3.2 基于 Markov 因果系统的预测法

如上所述, MAD 算法虽然结构简单, 但由于对图像中每一个点都要计算 $|d_{\max} - d_{\min}| \cdot \nabla F(x, y, d)$, 然后进行比较, 因此不仅计算量大, 且没有利用到相邻像素的视差相似性。

由于头肩像的视差比较均匀(轮廓处除外), 因此大部分像素都可以通过邻域中的一组像素预测出。以上这种自相关性和局部特性就决定了对于头肩像的视差估计可以采用 Markov/Gibbs 模型。

因此, 本文在 3.1 节描述的基础上采用了一种基于 Markov 因果系统递归预测的视差估计法。该算法流程包含块估计运算和像素估计运算两层运算。前者是根据相邻块来预测块视差, 后者是在像素级上进行估计, 以确定是否需修整, 经模块比较以确定最终视差是采用块预测经像素级修正来得到还是采用常规逐像素估计来得到。

具体方法如下:

(1) 分块及初始化。设 I_{left} 和 I_{right} 是一对立体图像对, 设 I_{left} 为目标图像, I_{right} 为参考图像。将目标图像等间隔分割为方块, 即 $I_{\text{left}} = \{I_{\text{left}}(m, n)\}$, 将对应块的视差值集合记为 $D_{\text{left}} = \{D_{\text{left}}(m, n)\}$, 其中 m, n 分别为分割后离散网格的行坐标和列坐标。

(2) 预测当前块 $I_{\text{current}}^{\text{left}}$ 的视差, 即先确定其邻域, 再确定其因果邻域; 然后在因果邻域中选取最具代表的邻域视差作为 $I_{\text{current}}^{\text{left}}$ 的视差。设邻域为 3×3 , 则因果邻域为 3 块, 依次记为 $I_i^{\text{left}} (i = 1, 2, 3)$, 其因果邻域记为 $I_{i,j}^{\text{left}} (i, j = 1, 2, 3)$ 。设当前块 $I_{\text{current}}^{\text{left}}$ 的预测视差为 $D_{\text{current}}^{\text{left}}$, 邻域 I_i^{left} 的视差记为 D_i^{left} (如图 4(a) 所示), $I_{i,j}^{\text{left}}$ 的视差是 $D_{i,j}^{\text{left}}$ (如图 4(b) 所示)。

若 $C_i = \min(C_i)$, 则 $D_{\text{current}}^{\text{left}} = D_i^{\text{left}}$, 其中

$$C_i = \sum_{j=1}^3 (D_{i,j}^{\text{left}} - D_i^{\text{left}})^2 \quad (6)$$

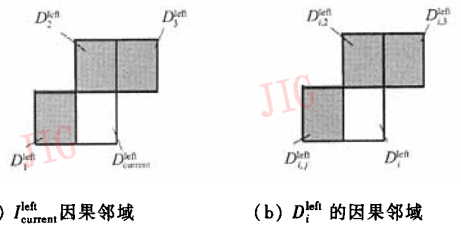


图 4 因果邻域

Fig. 4 Causal neighborhood

(3) 像素级视差估计。如果当前块 $I_{\text{left}}(m, n)$ 的预测值大于阈值 T , 则按照常规算法求视差, 否则, 首先记录该像素位置, 再在右图像相应位置向右移动 $D_{\text{left}}(m, n)$ 个像素位置处设置一个比较小的搜索窗口, 然后利用改进 MAD 算法式(5)获得最优匹配, 并记录其位置, 它和当前像素的距离即为视差值。

(4) 计算当前块的最终视差, 并进行递归调用。当块内每个像素视差值都获得后, 再将其中像素视差的最小值与块视差的预测值进行比较, 然后选取最小的值作为当前块的最终视差值, 存入 $D_{\text{left}}(m, n)$, 供后面使用。重复步骤 2、步骤 3、步骤 4, 直到获得所有像素的视差值。

(5) 视差检测并补偿。

当 $\|D_{\text{left}}(x, y) + D_{\text{right}}(x, y + D_{\text{left}}(x, y))\| < T_0$ (T_0 是阈值) 时, 就认为是正确匹配区域, 否则为误

匹配区域。对于误匹配区域,选用左侧第 1 个准确视差值未对其进行补偿填充,即可得到准确的视差场。对 $D_{right}(x,y)$ 进行处理可类似地得到。

景简单,是典型的视频会议中的头肩序列,其图像尺寸为 384×384 。图 5(b)、图 5(e)是左图像(图 5(a))、右图像(图 5(d))的分割结果。

图 5(c)、图 5(f)为用第 3 节中的算法所获得的左右视差图,视差图中对象亮度越大,表示视差越大,即深度越小。由于针对属于视频对象的像素使用的是基于 Markov 模型的视差估计算法,因此视差图内部比较平滑,运算速度也有较大提高。

4 实验结果

为验证算法效果,本文采用“Man”序列第 5 帧来进行实验。“Man”视频序列为头肩部序列,其背



图 5 “Man”序列视差估计

Fig. 5 Disparity Estimation in a “man” sequence

5 结 论

本文在对象分割的基础上,利用相邻块信息预测,来估计出像素视差,实验结果表明,该方法能获得平滑精确的视差场。而且所提出的算法在 3 维视频会议及需要视差信息的系统中有良好的应用前景。

参考文献 (References)

- 1 McVeigh Jeffrey S, Siegel M W, Jordan A G. Intermediate view synthesis considering occluded and ambiguously referenced image regions[J]. Signal Process: Image Communication, 1996, 9(1): 21 ~ 28.
- 2 Grammalidis N, Strintzis M G. Disparity and occlusion estimation in multicular systems and their coding for the communication of multiview image sequences[J]. IEEE Transactions Circuits System Video Technology, 1998, 8(3):328 ~ 344.
- 3 Naemura T, Kaneko M, Harashima H. 3-D object based coding of multi-view images[A]. In: Picture Coding Symposium '96 [C], Melbourne, Australia, 1996: 459 ~ 464.
- 4 Karathanasis J, Kalivas D, Vlontzos J. Disparity estimation using block matching and dynamic programming[A]. In: Proceedings of the Third IEEE International Conference on Electronics, Circuits, and Systems '96[C], Rodos, Greece, 1996, 2:728 ~ 731.
- 5 Boykov Yuri, Veksler Olga, Zabih Ramin. Fast approximate energy minimization via graph cuts [J]. Pattern Analysis and Machine Intelligence, 2001,23(11):1222 ~ 1239.