

应用分层 MRF/ GRF 模型的立体图像视差估计及分割

安平,张兆扬,马然

(上海大学通信学院电子信息系,上海 200072)

摘要: 视差估计与分割是立体图像编码及立体视觉匹配的核心问题,本文提出一种基于分层 MRF/ GRF 模型和交叠块匹配 (HMOM) 视差估计算法以及结合主动轮廓模型的视差分割提取算法. 该混合视差估计方法,可得到光滑准确,且具有清晰边缘的视差场;并便于用主动轮廓模型提取感兴趣对象 (OOI) 的视差轮廓. 与通常的变尺寸块匹配 (VSBM) 相比,本算法得到的视差补偿图像的峰值信噪比可提高 2.5dB 左右. 本文得到的视差场及对应的轮廓可进一步用于立体图像编码以及视频对象分割.

关键词: 立体图像; 视差估计; Markov 随机场 (MRF); Gibbs 随机场 (GRF); 交叠块匹配 (OBM); 主动轮廓模型

中图分类号: TN911.1 文献标识码: A 文章编号: 0372-2112 (2003) 04-0597-05

Hierarchical MRF/ GRF Model Based Disparity Estimation and Segmentation for Stereo Images

AN Ping, ZHANG Zhao-yang, MA Ran

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

Abstract: Disparity estimation and segmentation are vital tasks in stereo image coding and stereo vision matching. This paper proposes an algorithm for disparity estimation based on hierarchical MRF/ GRF model and overlapped block matching (HMOM), and for disparity segmentation and extraction using active contour model. By combining the advantages of variable size block matching, overlapped block matching and MRF model, the HMOM scheme estimates a more smoother and accurate disparity field with lower computational complexity. In terms of PSNR of the disparity compensated image, the proposed disparity estimation algorithm achieves about 2.5dB higher PSNR, as compared to conventional variable block matching. A smooth and consistent disparity field with sharp boundary can be obtained by pixel-wise disparity refinement on the HMOM disparity field, and the disparity contour of OOI (objects of interest) can be extracted by the active contour model. The resulting disparity field and corresponding disparity contours can further be used in stereo image coding and video object segmentation.

Key words: stereo image; disparity estimation; Markov random field (MRF); Gibbs random field (GRF); overlapped block matching (OBM); active contour model

1 引言

随着立体显示技术的发展,三维(3D)视频技术的应用正在迅速发展,3D 视频信号将是未来多媒体通信的重要内容.

3D 视频信号的一种表征方法是采用立体图像对,两台摄像机同时从不同位置拍摄同一个场景即可获得立体图像对. 立体图像对间的对应估计称为视差估计. 如果两台摄像机的性能相同,并且光轴被设置成平行排列的,则构成立体平行摄像系统. 通过平行摄像系统采集的是平行图像对,对应点位于相同的扫描线上,基于外极线光学几何原理,二维视差的对应估计就简化为一维,从而降低了视差估计的复杂度.

视差估计 (DE) 或“对应”是立体图像编码和立体视觉分

析的重要环节. 新出台的 MPEG-4 版本 2 也支持对稠密视差场的编码^[1]. 求取视差“对应”问题的方法很多,如图像特征法、像素递归法、视差光流联合估计法及块匹配法等,但这些方法都难以获得光滑而准确的视差场. 特征法以各种图像特征作为匹配的基元,视差估计虽然较准确,但只能生成稀疏的视差场^[2];后几种方法虽然能生成稠密视差场,但视差场的精确度有待提高. 其中,像素递归法用已求出的前面像素的视差矢量预测当前像素视差,便于硬件实现,但其预测因果性限制了视差场的质量^[3];视差光流联合估计法在相互有利的方式下对视差场和光流场进行联合估计^[4],要求摄像系统严格标定,一般很难满足;固定块匹配法 (FSBM) 以固定尺寸的像块作为匹配基元,很难得到好的匹配结果,重建目标图像中块效应严

重^[5];可变尺寸块匹配(VSBM)通过在图像均匀区域采用大块,而在物体边界区域采用小块,虽然能改善匹配效果,但在块尺寸较小时,也会导致不准确的匹配,尤其在物体边界处,误匹配严重^[6].为了保证对应的精确性,在对应匹配过程中,应该利用一些适当的约束条件,如平滑约束和外极线光学几何约束^[6].文献[5]应用平滑约束,提出了一种基于MRF模型的视差估计算法,可以得到较光滑而准确的视差场,然而,由于遮挡效应,常规MRF模型法在物体边界处往往不能求得准确的视差估计.

当前视频对象分割和视频场景分割已分别成为MPEG-4和MPEG-7能否广泛应用的核心问题,但对感兴趣对象(OOI)的分割现大都停留在基于运动信息的分割,如果场景没有运动,往往就显得无能为力.为此,应考虑将其它的信息用于分割,如亮度、边缘以及视差信息等.

立体图像编码的目的是提高编码效率,主要着眼点是在码率和失真之间的折衷,以得到高的压缩比,所以处理视差场的代价不能太高.本文的视差估计研究主要用于立体图像编码,希望得到相对精确且平滑的视差场.鉴于上述现有方法所存在的问题,本文提出一种基于分层MRF/GRF模型和主动轮廓模型的视差估计及分割算法,不仅可以得到光滑而准确的稠密视差场,而且可以从静止场景中分割、提取OOI视差轮廓.整个算法的思想是首先用简单有效的边缘提取算子提取目标图像的亮度边缘,作为初始视差边缘;然后,采用分层MRF/GRF模型进行视差估计,为提高估计的速度,采用我们在文献[7]中给出的视差快速搜索算法,并且将交叠块匹配和分层一阶因果邻域结合在匹配过程中,从而得到光滑而准确的视差场;再结合视差边缘,对视差进行基于MRF模型的逐像素细化,得到既光滑准确又边缘清晰的视差场;最后,用主动轮廓模型提取OOI的视差轮廓,以作进一步的视频对象的分割之用.

2 MRF/GRF模型与视差估计

2.1 用于视差估计的MRF/GRF模型

平滑约束是求对应问题的常用约束条件,而MRF/GRF模型是建立对应平滑模型的理想工具.MRF模型定义在概率域,GRF模型定义在能量域,两者之间有一一映射的关系.因此,可将复杂的MRF最大后验概率(MAP)解化为相对简单的GRF最小能量问题^[8]:

$$f_{MAP} = \arg \max_f P_{MRF \text{ joint}} = \arg \min_f U_{GRF \text{ joint}} \quad (1)$$

其中, $P_{MRF \text{ joint}}$ 指 MRF 联合概率, $U_{GRF \text{ joint}}$ 指 GRF 联合能量.

Geman 等将图像看作由噪声和隐蔽边缘共同作用的随机场^[9].W. Woo 等又将这种随机模型推广到立体视差空间^[5],将MRF模型用到视差估计中.MRF/GRF模型具有严格的数学框架,构造了空间相关变量间相互作用的通用模型,并将间断及遮挡标志结合进能量方程中,从而大大减少了视差误匹配现象.

设 I_1 (参考图像) 和 I_2 (目标图像) 是立体图像对, I_2 相对 I_1 的视差场为 D , 遮挡(即在 I_2 中出现,在 I_1 中未出现的内容)场为 O , 根据 Bayes 理论和 MAP 准则, 视差估计问题可表

示为:

$$(D, O) = \arg \max_{(D, O)} P(D, O | I_1, I_2) \\ = \arg \max_{(D, O)} [U(I_2 | I_1, D, O) + U(D | O) + U(O)] \quad (2)$$

上式右端各项分别对应视差估计中的各种约束,其中第一项表示相似性约束,第二项表示平滑约束,最后一项表示间断约束.对于立体平行摄像系统,图像的扫描线与外极线重合,自然满足了外极线约束^[7,8].

由式(2)表示的视差估计通用模型中,右边第三项因不是视差的直接函数,在视差估计能量函数构造中可忽略.将目标图像 I_2 等间隔地分割为方块,即 $I_2 = \{ b_{2, mn}, (m, n) \}$, 其中 $b_{2, mn}$ 表示第 (m, n) 个块, $\{ \cdot \}$ 表示离散网格,即 $\{ (m, n) | 0 \leq m < N_x, 0 \leq n < N_y \}$, N_x 和 N_y 分别表示垂直和水平方向的块数.对于块匹配,基于MRF模型的视差估计总能量函数可定义为:

$$U(D, O | I_1, I_2) = \sum_{(m, n)} U(d_{mn}, o_{mn} | b_{1, mn}, b_{2, mn}) \\ = \sum_{(m, n)} \{ (1 - o_{mn}) \sum_{(i, j) \in b_{mn}} (f_{1, ij} - f_{2, ij} - d_y)^2 + o_{mn} \sum_{mn} (d_{mn} - d_{mn})^2 (1 - o_{mn}) \} \quad (3)$$

d_{mn} 表示块 $b_{2, mn}$ 的视差矢量, 下标“ mn ”表示沿视差的位移, f_{ij} 表示块内的像素, o_{mn} 为块的遮挡标志,以二值表示,1表示遮挡,0表示非遮挡.在能量函数中结合进遮挡是为了减轻由遮挡造成的匹配误差. mn 表示块 $b_{2, mn}$ 的邻域.邻域遮挡标志 o_{mn} 控制视差 d_{mn} 是否与其邻域视差间断,若某个块被确定为遮挡块,则不对其应用平滑约束,也就是说,平滑约束只对非遮挡块起作用.值得说明的是,视差场的平滑性不仅仅是为了减少噪声的影响,它还能提高视差场的编码效率.常量系数用以控制各约束的作用程度.

2.2 视差估计的分层MRF/GRF模型

由于噪声及遮挡等因素,FSBM会导致不精确的视差估计,并且在重建目标图像中有较严重的块状干扰.若增大块尺寸,会增加相应的估计误差,对于立体图像编码而言,这就使得视差补偿差图像的码率增大;而若减小块尺寸,则表征视差场的码率太大,并且得到的视差往往并非真正的视差,难以作为分割对象的根据.改进FSBM的方法之一就是采用VSBM,将具有较高估计误差的块分割为更小的块.VSBM可以在有效的视差估计/补偿、视差场的表示以及视差补偿差之间取得折衷.可是,在子块较小时,VSBM也会产生不一致的视差估计,导致补偿差图像中沿子块边界处有较多的高频能量.式(3)定义的视差估计的MRF模型利用相似性约束和平滑约束,虽然能改善视差场的准确与光滑性,但由于遮挡效应,在物体边界处往往不能求得准确的视差估计.

为进一步改善视差估计的效率,我们将VSBM和MRF/GRF模型结合起来,建立分层的MRF/GRF模型,以更有效地表示视差场,减少物体边界处的块效应.此外,我们还把交叠块匹配(OBM)引入到分层视差估计中,在匹配时加权了相邻块内容,减少了因局部最小化而造成的视差误匹配^[10].

将 OBM 结合到相似性约束中,并把式(3)改写到分层结构上,我们推导出基于分层 MRF/ GRF 和交叠块匹配(HMOM)视差估计的能量函数:

$$U(D, O | I_1, I_2) = \sum_{(m, n)} [(1 - o_{mn}) \cdot u_{mn} + o_{mn} \cdot su_{mn}] \quad (4)$$

式中,

$$u_{mn} = (1 - \sum_k W \cdot (s_{2, mn}^k - s_{1, mn}^k - d_{mn}^k)^2) + (1 - o_{mn}) (d_{mn}^k - d_{mn}^k)^2$$

$$su_{mn} = \sum_{l+1, k} \{ (1 - \sum_k W \cdot (s_{2, mn}^{l+1, k} - s_{1, mn}^{l+1, k} - d_{mn}^{l+1, k})^2) + [(1 - o_{mn}^{l+1, k}) (d_{mn}^{l+1, k} - d_{mn}^{l+1, k})^2 + (d_{mn}^{l+1, k} - d_{mn}^{l+1, k})^2] \}$$

其中, W 为 OBM 窗函数,本文选用双线性窗。 $s_{2, mn}$ 表示目标图像中块 $b_{2, mn}$ 的扩展块。若第 (m, n) 个块的估计误差较大,则置其遮挡标志 o_{mn} 为 1,须将该块分割为小块,以提高视差估计的精度。这里采用四叉树分割结构, $b_{2, mn}^l, s_{2, mn}^l$ 为第 l 层的块及扩展块,则它们的子块分别为 $b_{2, mn}^{l+1, k}, s_{2, mn}^{l+1, k}, k = \{0, 1, 2, 3\}$ 。 d_{mn}^k 和 o_{mn}^k 分别表示 b_{mn}^k 的视差量及遮挡状态。 u_{mn} 和 su_{mn} 的第一项均表征立体对间的相似性约束,第二项为平滑约束,在视差估计时, u_{mn} 和 su_{mn} 两者选一。首先在最大块尺寸的初层用 u_{mn} 进行基于 MRF/ GRF-OBM 视差估计,如果匹配误差大于预设的阈值,则对此块进行四叉树分割,而若四个子块的误差减量很小的话,就维持原块,不进行分割。重复此分割过程,直

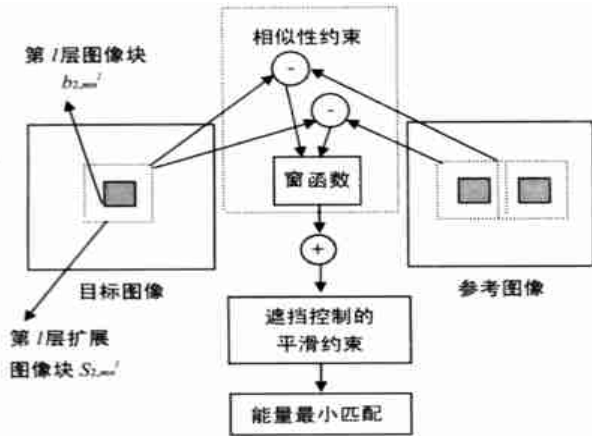


图1 HMOM 视差匹配原理

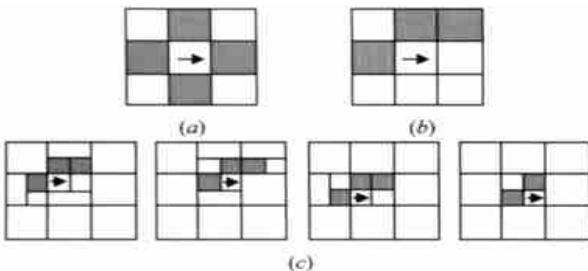


图2 一阶邻域系统,箭头指示区域表示当前像素或块,阴影区域表示邻域。(a)标准邻域;(b)因果邻域;(c)四个子块的分层因果邻域

到当前块的匹配误差小于子块误差的总和或块尺寸足够小为止。用 su_{mn} 对子块进行视差匹配。

式(4)的匹配原理如图1所示,匹配时,在搜索范围内,目标图像中的第 l 层的扩展块 $s_{2, mn}^l$ 与参考图像中的待选扩展块之间的匹配误差经过窗函数 W 的作用,作为匹配能量值的一部分,然后再加上平滑约束能量值,得到总的能量值,按照能量值最小的原则确定匹配值,作为块 $b_{2, mn}^l$ 的匹配值。

常规 MRF 法计算复杂度较高,这是因为需要迭代多次才能得到最优的稠密视差场。用完全搜索策略求解最小能量,并采用一阶邻域系统能够避免迭代引起的大计算量。图2定义了两种一阶邻域系统,即标准邻域和因果邻域,前者用于逐像素处理方式,后者用于逐块分层处理方式。

3 视差估计及分割过程

视差估计与分割的过程如图3所示,以下分别叙述各步骤。

3.1 图像平滑和边缘检测

在视差估计之前,先对图像作自适应低通滤波处理,以减少噪声干扰,并提取目标图像的亮度边缘,用作初始视差边缘。低通滤波过程如下:

$$g_{ij} = \begin{cases} f_{ij} & |f_{ij} - f_{ij}| \geq Th \\ f_{ij} & |f_{ij} - f_{ij}| < Th \end{cases} \quad (5)$$

其中, g_{ij} 为滤波后像素, f_{ij} 为当前像素的邻域均值, Th 为阈值。

采用一些简单有效的边缘提取算子,如 Marr 或沈俊算子,提取亮度边缘。求得的亮度边缘记为 $L = \{l_{ij}\}$,取 1 或 0 值,1 代表边缘像素,0 代表非边缘像素。

3.2 混合视差估计

(1) 初始视差估计

首先应用式(4)定义的 HMOM 法求得光滑的视差场。这里,用我们已在文献[7]中给出的快速搜索取代完全搜索,显著提高了搜索速度,降低了计算量。HMOM 视差场作为视差分割的初始视差场;由于 HMOM 视差场的光滑性和准确性均得到较大提高,对基于视差估计与补偿的立体图像编码是非常有利的,是较理想的用于立体图像编码的视差场。

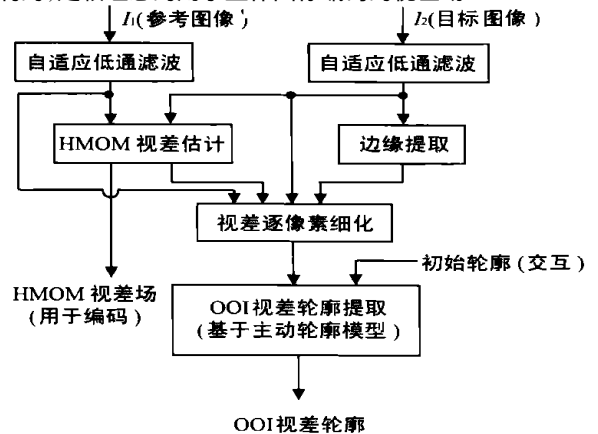


图3 视差估计及分割过程框图

(2) 基于 MRF 模型的视差逐像素细化

在已知初始视差场、边缘信息以及平滑后的立体对 (G₁, G₂) 后,再对视差场进行逐像素的细化处理,以进一步提高视差场的精度,得到光滑且边缘清晰的视差场,以利于分割. 细化处理能量函数根据 MRF/ GRF 模型以及如图 2 所示的一阶标准邻域来定义:

$$U(D, LD | G_1, G_2, L_2) = \sum_{(i,j) \in \Omega_2} \{ (1 - \alpha) (g_{2,ij} - g_{1,ij} - d_{ij})^2 + \sum_{ij} (d_{ij} - d_{ij}')^2 (1 - l_{d,ij}) \} \quad (6)$$

其中 LD 表示视差边缘场, g 和 l_d 分别表示像素的亮度及视差边缘, l_d 控制视差 d 与其邻域视差 d' 之间的间断性. 与式 (3) 和 (4) 类似, α 是控制相似度、光滑度与间断度的权.

3.3 用主动轮廓模型提取视差轮廓

主动轮廓模型 (Snake) 常用来提取物体的光滑闭合边界^[11]. 主动轮廓是指图像网格上的一个有序点集, S = [s₁, s₂, ..., s_K], 这些点位于使其能量函数 E 最小的位置上.

$$E(S) = E_{ext}(S) + (1 - \beta) E_{int}(S) \quad (7)$$

E_{ext} 为外部能量, 其作用是将主动轮廓吸引到图像显著特征

处; E_{int} 为内部能量, 作用主要是保持曲线的光滑性. 外部能量定义为:

$$E_{ext}(S) = - \sum_{i=1}^K h_i^T \nabla G(s_i) \quad (8)$$

其中 ∇G(s_i) 是 s_i 处的 Sobel 亮度梯度, h_i 是主动轮廓在 s_i 处的归一化矢量. 内部能量 E_{int} 定义为:

$$E_{int}(S) = \sum_{i=1}^K (s_i - s_{i-1})^2 + (s_{i-1} - 2s_i + s_{i+1})^2 \quad (9)$$

将主动轮廓模型作用于细化后的视差场, 就能分割、提取出 OOI 的光滑闭合的视差轮廓, 这是将 OOI 从图像中分割出来的重要步骤. 这里, 我们用动态规划法求取主动轮廓的最小能量.

4 实验结果及分析

为验证本文算法的有效性, 我们对多对立体图像对进行了模拟实验, OBM 的扩展窗尺寸在水平及垂直方向均为原图像块的 2 倍, 限于篇幅, 这里仅给出 Room 对的实验结果. 为方便观察视差场, 将视差量映射到 (0~255) 范围, 用亮度来表示视差矢量的绝对值, 亮度大的表示较大的视差位移, 亮度小的



(a) 参考图像

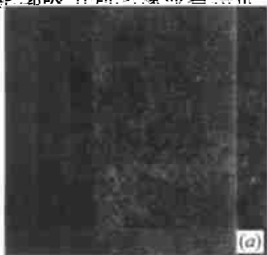


(b) 目标图像

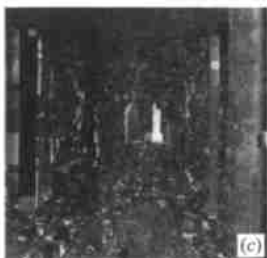


(c) 目标图像的沈俊边缘

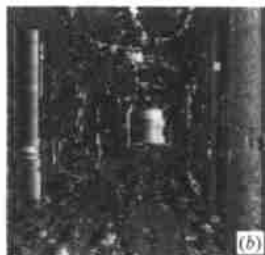
图 4 测试图像对 Room 及其亮度边缘



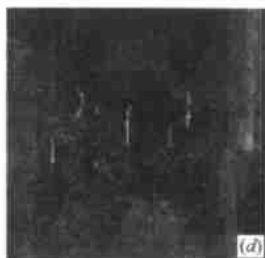
(a)



(c)



(b)



(d)

图 5 各种视差估计算法的视差场比较 (Room). (a)FSBM(块尺寸: 32×32); (b)FSBM(块尺寸: 2×2); (c)VSBM(块尺寸: 32×32 到 2×2); (d)HMOM(块尺寸: 32×32 到 2×2);

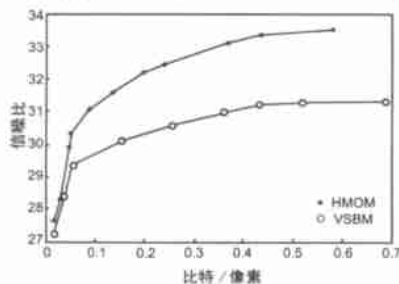


图 6 VSBM 与 HMOM 视差补偿率失真性能比较



(a)



(b)



(c)

图 7 Room 的视差细化及 OOI 视差轮廓提取.

(a) 用 HMOM 法估计的初始视差场;

(b) 细化后的视差场

(c) 用主动轮廓模型提取的 OOI 视差轮廓

表示较小的视差位移.图 4 为实验所用的图像对之一 *Room* 及其目标图像的沈俊边缘,沈俊边缘被用作初始的视差边缘.

图 5 比较了用各种视差估计算法所得到的视差场,搜索窗范围为 $(0, \pm 16)$.其中基于 *FSBM* 的算法,大尺寸块 (32×32) 和小尺寸块 (2×2) 时分别如图 5(a)、(b) 所示,大尺寸块的 *FSBM*,有利于避免局部最小,作为初始视差估计是较合适的,但其视差场缺少细节,会导致视差补偿差的增大;而小尺寸块的 *FSBM* 视差场则充满了噪声,由于匹配仅依据亮度的相似性,这些视差往往没有反映真正的对象深度信息,误匹配严重.正如前面所分析的,*VSBM* 得到的视差场性能优于 *FSBM*,视差场受噪声干扰的程度大为减轻,但当子块变小时,由于噪声、遮挡等干扰,视差场也会出现不一致现象,如图 5(c) 所示.图 5(d) 给出了本文提出的基于 *HMOM* 算法的视差场,视差场的光滑性和精确性得到很大提高.光滑而精确的视差对图像编码是非常有利的,提高了整个立体图像编码的效率.暂不考虑补偿差图像,图 6 比较了 *VSBM* 和 *HMOM* 算法的视差估计/补偿率失真 (*RD*) 性能,在相同的比特率下,*HMOM* 得到的补偿目标图像的信噪比可提高 2.5 *dB* 左右,若再考虑补偿差,*HMOM* 率失真性能还能进一步提高.其中视差场用 *DPCM* 编码,分割二叉树按每个节点 4 比特方式编码.

图 7 给出了 *Room* 的视差估计、细化及 *OOI* 轮廓提取的结果,初始块尺寸选为 32×32 .图 7(a) 是 *HMOM* 视差估计的视差场,图 7(b) 是对 (a) 进行细化后的结果,图 7(c) 示出了用主动轮廓模型提取 (b) 中小球的视差闭合轮廓.实验结果表明,细化后的视差场光滑性较好、边缘清晰,视差比较准确地反映了深度信息.图 7(c) 中标识的 *OOI* 轮廓,可作为 *MPEG-4* 中 *VOP*(视频对象平面)提取的重要依据.

本文提出的基于分层 *MRF/GRF* 和主动轮廓模型的视差估计及分割混合算法,能得到光滑而精确的视差场及 *OOI* 视差轮廓.本算法具有良好的稳健性,适用于各类立体对的视差估计及分割.我们已将 *HMOM* 视差估计算法用于具体的立体视频编码,得到了较好的率失真性能.

参考文献:

- [1] Jens Rainer Ohm. Encoding and Reconstruction of multiview Video Objects [J]. IEEE Signal Processing Magazine, 1999, 16(3): 47 - 54.
- [2] J Liu, R Skerjanc. Stereo and motion correspondence in a sequence of stereo image [J]. Signal Processing: Image Communication, 1993, 5: 305 - 318.
- [3] L Boroczky. Pel-recursive motion estimation [D]. Delft University of

Technology, 1991.

- [4] G Sudhir, S Banerjee, K K Biswas, R Bah. A cooperative integration of stereopsis and optic flow computation [A]. Proc of 12th International Conf. On Pattern Recognition [C]. Jerusalem, Israel: ICOPR, 1994, 356 - 360.
- [5] Wontack Woo, Antonio Ortega. Stereo image compression with disparity compensation using the MRF model [A]. Proc. SPIE VCIP [C]. Orlando, FL: SPIE, 1996, 28 - 41.
- [6] Ebroul Izquierdo M. Disparity/ Segmentation Analysis: Matching with an Adaptive Window and Depth-Driven Segmentation [J]. IEEE Trans. on CSVT, 1999, 9(4): 589 - 607.
- [7] 安平, 张兆扬. 基于立体摄像几何特性的视差匹配快速算法 [J]. 上海大学学报, 2000, 6(5): 399 - 403.
- [8] Andre ' Redert, Emile Hendriks, Jan Biemond. Correspondence Estimation in Image Pairs [J]. IEEE Signal Processing Magazine, 1999, 16(3): 29 - 46.
- [9] S Geman, D Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images [J]. IEEE Trans on PAMI, 1984, 6(6): 721 - 741.
- [10] W Woo, A Ortega. Modified overlapped block matching for stereo image coding [A]. Proceedings of PW-EE-VCIP vol. 3653 [C]. San Jose, CA, USA: VCIP, 1999: 570 - 581.
- [11] M Kass, A Witkin, D Terzopoulos. Snakes: Active contour models [A]. Proc 1st Int. Conf. Computer Vision [C]. London, England: CCC, 1987: 259 - 269.

作者简介:



安 平 女, 1968 年 4 月生于安徽嘉山, 1993 年获合肥工业大学工学硕士学位, 2002 年获上海大学工学博士学位, 现为上海大学通信与信息工程学院副教授, 主要研究领域为视频信号处理、三维图像编码及其应用.



张兆扬 男, 1938 年 9 月生于浙江海宁, 上海大学通信与信息工程学院教授、博士生导师、学科带头人, 主要研究领域为数字视频处理、多媒体通信和计算机视觉.