

基于 MAP 的全局运动估计算法

刘征宇, 周孝宽

(北京航空航天大学宇航学院, 北京 100083)

摘要: 常见的全局运动算法存在的问题是仅仅考虑像素的残差, 而忽略了相邻像素和分割信息等因素对算法的影响。为了解决上述问题, 该文提出一个基于最大后验概率(MAP)的全局运动估计算法。通过 MAP 算法定义系统能量, 引入先验能量来描述像素的相邻关系和分割信息, 将系统能量作为目标函数通过 LM 算法估计全局运动参数。实验结果证明了该算法的有效性。

关键词: 全局运动估计; 最大后验概率; Gibbs 能量

Global Motion Estimation Algorithm Based on MAP

LIU Zheng-yu, ZHOU Xiao-kuan

(School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100083)

【Abstract】 One of the main problems related to global motion estimation is that the algorithm only considers the change information in the difference image, but the import information of segmentation and correlation has been ignored. To overcome such drawbacks, a technique of global motion estimation based on MAP is proposed. The system energy is defined through MAP and the potential energy is defined by segmentation mask and the prior information. Using the system energy as object function, the global motion is estimated. Experimental results confirm the effectiveness of the proposed technique.

【Key words】 global motion estimation; MAP; Gibbs energy

1 概述

在大量的实际应用中, 自然背景对象没有自身运动, 由于摄像机运动才造成了图像序列中的背景产生变化, 对这样的背景分析就是全局运动估计。全局运动估计就是要从图像序列中找出背景的运动规律来, 分析的结果广泛地应用于视频编码、索引以及分割等领域, 具有重要的现实意义。

全局运动估计算法首先需要选取背景作为特征点参与计算, 难点主要在于前景和噪声会对全局运动分析产生干扰导致错误的估计结果。因此, 特征点的选取是算法的关键, 目前有如下常见的处理方法: 一般来说前景像素的残差比较大, 所以, 文献[1]为了取得更快的计算速度采用较大的阈值, 仅允许很少的背景参与计算, 这种方法很容易受到前景和噪声的干扰, 必然影响到算法的鲁棒性; 文献[2-3]提出用残差直方图去除固定比例残差最大点的方法来去除前景。MPEG-4 校验模型中采用了这个方法, 该方法由于参与计算的像素多, 因此计算量很大。

现有算法有一个共同的缺点就是, 在处理每个像素时孤立地考虑该像素, 仅仅对该像素的残差进行处理, 而忽略了重要的信息: 和相邻像素之间的关系; 该像素的分割信息: 同样的参差, 属于背景和属于前景对算法的贡献应该是不相同的。本文从概率的角度阐述和理解全局运动估计算法, 通过最大后验概率 MAP 算法定义系统能量, 引入先验能量来描述像素的相邻关系和分割信息, 从而构建基于 MAP 的全局运动估计算法。最后给出了试验结果并论述该方法的合理性和有效性。

2 基于 MAP 的全局运动估计

2.1 基本思路

全局运动估计算法的实现方法虽然各不相同, 但基本的

原理都是用一个数学模型来描述全局运动, 并在此基础上根据相邻帧的相关性估计出模型参数。本文选择可以描述所有的运动八参数模型进行全局运动估计。

全局运动估计问题通常被归类为最小二乘问题, 它通过最小化误差的平方和找到一组数据的最佳函数匹配。换个角度来看, 最小化问题同样可以看作是最大似然估计, 也就是估计参数使得残差在噪声(高斯)分布下的概率最大。最大似然估计本质是一种平均的方法, 在处理每个像素时同样都是处理该像素的参差, 而忽略掉了算法分割信息、相邻关系等先验信息。而最大后验概率估计在最大似然估计的基础上进一步引入先验模型, 由于引入先验统计特性, 理论上 MAP 算法比最大似然估计算法估计得要精确。因此, 本文将在 MAP 模型基础上构建新的全局运动估计算法。

2.2 MAP 模型

在算法中, 每帧图像存在一个用于标记每个像素属于背景还是前景的标记场 z , 一个描述当前帧空间分布信息的事先观察场(亮度场) g 。MAP 用于全局运动估计算法可以描述为: 在已知给定 2 帧的 g_k (当前帧)和 g_{k-1} (搜索帧)和标记场 z 的条件下, 进一步估计未知全局运动估计参数 θ , 使得未知参数 θ 在给定条件下出现的概率最大。

假设给定 2 帧 g_k 和 g_{k-1} , 未知参数 θ 和未知量 z , 根据贝叶斯理论后验概率可以表示为

$$p(\theta, z | g_k, g_{k-1}) = \frac{p(g_k | \theta, z, g_{k-1})p(\theta | z)p(z)}{p(g_k | g_{k-1})} \quad (1)$$

其中, 分母 $p(g_k | g_{k-1})$ 是不变的常数, MAP 估计即求:

作者简介: 刘征宇(1971-), 男, 博士研究生, 主研方向: 模式识别, 控制系统; 周孝宽, 教授、博士生导师

收稿日期: 2008-06-15 **E-mail:** lzyhm@263.net

$$\hat{\theta} = \arg \max p(g_k | \theta, z, g_{k-1}) p(\theta | z) p(z) \quad (2)$$

可以看出 MAP 公式包含 2 个概率模型：在被估计参数给定的情况下被观察图像亮度的条件概率 $p(g_k | \theta, z, g_{k-1})$ ，称为似然模型；被估计参数的先验概率，这里是 θ 和 z 的联合先验概率 $p(\theta, z) = p(\theta | z) p(z)$ ，称为先验模型。

根据 Gibbs 分布，似然模型 $p(g_k | \theta, z, g_{k-1})$ 和先验模型 $p(\theta | z)$ ， $p(z)$ 可以用分别用指数表示，这里定义 U_g 、 U_θ 、 U_z 分别对应以上分布的能量，系统能量函数 U 可以表示为

$$U = U_g + \lambda_\theta U_\theta + \lambda_z U_z \quad (3)$$

其中，参数 λ_θ 、 λ_z 控制各能量之间的比例关系。全局运动的 MAP 估计也等效为

$$\hat{\theta} = \arg \max U \quad (4)$$

也就是说，基于 MAP 的全局运动估计等效于：估计未知全局运动参数 θ ，使得系统能量 U 最小。由于全局运动估计参数 θ 对每个像素都一致，因此 $p(\theta | z)$ 为常数， $U_\theta = 0$ 。系统能量 U 进一步简化为

$$U = U_g + \lambda_z U_z \quad (5)$$

2.3 似然能量

似然模型构建在残差图像的基础上，通常用噪声的概率模型来描述。假设观察噪声是有零均值和方差 σ 的高斯分布，条件概率 $p(g_k | z, g_{k-1})$ 可以表示为

$$p(g_k | z, g_{k-1}) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{r^2}{2\sigma^2}\right\} \quad (6)$$

对应的势能 U_g 为

$$U_g = \frac{\ln(2\pi\sigma^2)}{2} + \frac{r^2}{2\sigma^2} \quad (7)$$

由于和先验能量的比例关系可以通过调整 λ_z 参数加以调整，因此可以忽略掉公式前面的常数，简化定义 U_g 为

$$U_g = \frac{r^2}{2\sigma^2}$$

2.4 先验能量

先验模型就是通过图像的事先信息预测出每个像素先验概率的模型。对于标记的先验概率 $p(z)$ ，常用的量化方法是认为标记场中平滑的区域为背景的可能性高，具有较高的先验概率，而赋予先验能量 U_z 一个较小的值。在文献[4-5]中使用边界像素对来量化区域的光滑程度。所谓边界像素对是指标记场中水平、垂直或对角方向上的 2 个像素，这 2 个像素分别属于前景和背景。光滑度 $S(z)$ 定义如下：

$$S(z) = n_B B + n_C C \quad (8)$$

其中， n_B 表示水平和垂直方向边界像素对的个数； n_C 表示对角方向边界像素对的个数，常数 B 、 C 区别控制不同方向的比例权重。

先验能量 U_z 定义在光滑度的基础上：

$$U_z = d_z (S(z) - 2B - 2C) + 2B + 2C \quad (9)$$

$$d_z = \begin{cases} -1 & z=0(\text{背景}) \\ 1 & z=1(\text{前景}) \end{cases}$$

式(9)中 $2B+2C$ 为八邻域光滑度的中间值，像素为背景时 $U_z = -S(z) + 4B + 4C$ ，为前景时 $U_z = S(z)$ ，该定义使得先验能量 U_z 取值以 $2B+2C$ 为中心按光滑度的大小在 0 、 $(4B+4C)$ 之间分布。在实际的实现中 B 取 2， C 取 1。先验能量 U_z 实际是对相邻像素分割信息的量化，同时也包含了相邻像素之间的相关信息。

3 样实现原理

全局运动估计可以看作是求解如下问题：估计全局运动参数 θ ，使得图像像素的系统能量和有最小值。优化理论中的 LM 法正是解决此类问题。

在 LM 算法中，定义全局运动参数 $\theta^{(m)}$ 为递归实现中的第 m 步的解，下降梯度 $\delta\theta^{(m)}$ (Gauss-Newton 方向)为

$$\delta\theta^{(m)} = -H^{-1}(\theta^{(m)})g(\theta^{(m)}) \quad (10)$$

$$\theta^{(m+1)} = \theta^{(m)} + \delta\theta^{(m)} \quad (11)$$

其中， $H(\theta^{(m)})$ 是目标函数 Hessian 矩阵的近似，忽略了二阶以上导数； $g(\theta^{(m)})$ 是目标函数的梯度向量，两者均由 $\theta^{(m)}$ 定义。代入 U 和 r_i ，可以得到：

$$g_k = \sum_i \frac{\partial U}{\partial r_i} \frac{\partial r_i}{\partial \theta_k} \quad (12)$$

$$H_{kl} = \sum_i \frac{\partial^2 U}{\partial r_i^2} \frac{\partial r_i}{\partial \theta_k} \frac{\partial r_i}{\partial \theta_l} \quad (13)$$

g_k 和 H_{kl} 分别是向量 g 和矩阵 H 在相应位置的单元数据。这样 $\delta\theta$ 可以通过求解如下 k 阶齐次线性方程得到：

$$\sum_l H_{kl} \delta\theta_l = -g_k, k, l = 1, 2, \dots, K \quad (14)$$

在方程式中需要计算的参数为系统能量 U 的一阶和二阶倒数，分别如下：

$$\frac{\partial^2 U_g}{\partial r_i^2} = \frac{1}{\sigma^2}, \quad \frac{\partial U_g}{\partial r_i} = \frac{r_i}{\sigma^2} \quad (15)$$

对于 U_z 用切线来近似导数：

$$\frac{\partial^2 U_z}{\partial r_i^2} = \frac{U_z}{r_i^2}, \quad \frac{\partial U_z}{\partial r_i} = \frac{U_z}{r_i} \quad (16)$$

LM 算法就是不断地递归计算上式，最后所需要估计的参数 θ 。

4 实验结果

全局运动估计算法的复杂度主要在于 LM 算法的递归过程，一帧 CIF 格式的图像往往要经过几十次的递归才能够最终收敛，由于本文提出的系统能量可以更好地反映出像素的先验信息，因此算法可以具有较快的收敛速度。本文用递归次数来衡量算法的速度，递归次数和软件的运行环境无关，相比较运算时间可以更客观地反映算法的复杂度。

92 帧标准测试序列 tennis(352x240)序列在 30 帧附近有比较明显的镜头拉近，相邻 2 帧(28 帧, 29 帧)，本文算法得到的 8 个参数分别为 {1.02467, 2.017e-4, -4.17953, -2.088e-4, 1.02453, -2.9912, 0, 0}；MPEG4 校验模型的计算结果为 {1.02332, 4.034e-4, -4.14182, -3.264 e-4, 1.02334, -2.9836, 0, 0}，从参数结果看两者的差别仍然很小，从图 1、图 2 的视觉效果看，算法有明显的全局运动补偿效果。但得到类似的估计参数 MPEG4 校验模型用了 57 次递归，而本文提供的算法用了 35 次递归。图 3 是全部 92 帧 2 个算法的递归次数对比，MPEG4 校验模型的平均迭代次数为 49 次，本文算法的平均迭代次数为 38 次，本文提出的算法递归次数明显少于 MPEG4 校验模型的递归次数。

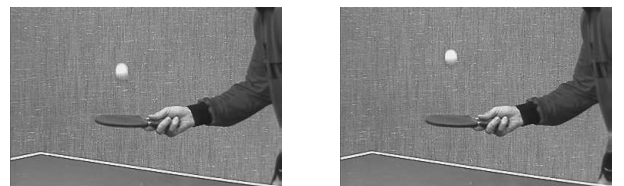
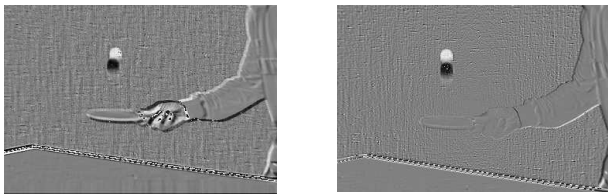


图 1 tennis 序列的相邻 2 帧(28 帧,29 帧)



(a)没有全局运动补偿的帧间差分图像 (b)经过全局运动补偿后的帧间差分图像

图2 帧间差分图像

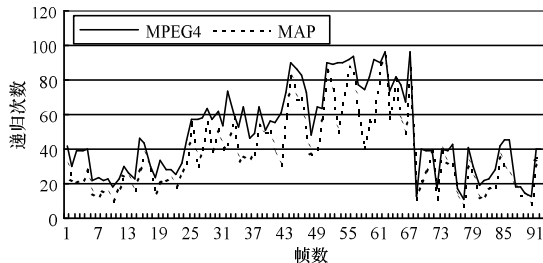


图3 92帧 tennis 序列和 MPEG4 的递归次数比较

表1是一些标准序列用MPEG4校验模型和本文算法计算所用迭代次数的比较,从表的测试结果可以看出,基于MAP的全局运动估计算法有效地降低了算法的复杂度。

表1 MPEG 校验模型和本文算法比较 (%)

序列	MPEG4	MAP	降低
Land(720×576)99 帧	65	52	20.0
Miss America(360×288)150 帧	56	41	26.7
Traffic(352×288)300 帧	57	43	24.6
Tennis(352×240)92 帧	52	38	26.9
Susie(352×240)75 帧	53	37	30.1
Hall Monitor(352×288)200 帧	56	43	23.2

5 结束语

本文将MAP理论引入全局运动估计算法,在传统仅考虑参差的算法基础上进一步引入了相邻信息和分割信息。从对试验结果的分析可以看出,提出的基于MAP的全局运动估计算法有着和MPEG4校验模型相同的精度,但大大地降低了算法的复杂度。和通过选择特征点,简单减少计算像素从而达到提高速度的算法相比,提出的算法有着更好的鲁棒性,同时仍然可以在此基础上进行特征点的选择,在保证精度和鲁棒性的同时降低算法的复杂度是基于MAP全局运动估计算法的优势所在。

参考文献

- [1] Richter H, Smolic A, Stabemack B, et al. Real Time Global Motion Estimation for MPEG-4 Video Encoder[C]//Proc. of Picture Coding Symposium. Seoul, Korea: [s. n.], 2001: 197-200.
- [2] Dufaux F, Konrad J. Efficient, Robust, and Fast Global Motion Estimation for Video Coding[J]. IEEE Transactions on Image Processing, 2000, 9(3): 497-501.
- [3] Konrad J, Dufaux F. Improved Global Motion Estimation for N3[S]. MPEG97/M3096, ISO/IEC/JTC1/SC29/WG11, 1998.
- [4] Derin H, Cole W S. Segmentation of Textured Images Using Gibbs Random Fields[J]. Computer Vision Graph. Image Process, 1986, 35(1): 72-98.
- [5] Aach T, Kaup A, Mester R. Statistical Model-based Change Detection in Moving Video[J]. Signal Processing, 1993, 31(1): 165-180.

(上接第236页)



(a)原始图像 (b)使用 J. Konrad 算法后的二值图像 (c)使用二值函数去除噪声点后的二值图像

图1 CoastGuard 序列



(a)原始图像 (b)使用 J. Konrad 算法后的二值图像 (c)使用二值函数去除噪声点后的二值图像

图2 Stefan 序列

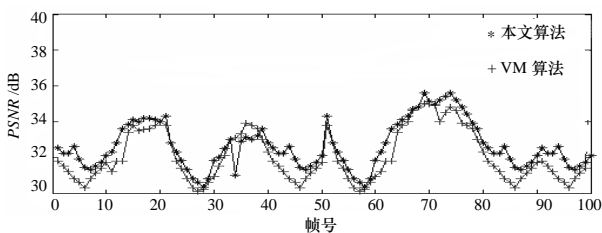


图3 CoastGuard 序列的比较

5 结束语

本文提出一个改进的基于鲁棒M-估计器的全局运动估计方法。实验结果表明,新算法比原算法更快,平均峰值信噪比略有提高。准确有效地去除噪声点和选择对全局运动结果贡献较大的特征点参与计算是加速全局运动估计的关键。因此,怎样有效去除噪声点和快速地选择可靠的特征点,是进一步研究的方向。

参考文献

- [1] Dufaux F, Konrad J. Efficient, Robust and Fast Global Motion Estimation for Video Coding[J]. IEEE Trans. on Image Processing, 2000, 9(3): 497-501.
- [2] Black M J, Anandan P. The Robust Estimation of Multiple Motions: Affine and Piecewise-smooth Flow Fields[J]. Computer Vision and Image Understanding, 1996, 63(2): 75-104.
- [3] Stiller C, Konrad J. Estimating Motion in Image Sequences: A Tutorial on Modeling and Computation of 2D Motion[J]. IEEE Signal Processing Magazine, 1999, 16(4): 70-91.
- [4] Meer P. Robust Techniques for Computer Vision[M]. New York, USA: Prentice Hall, 2004.
- [5] 李董辉, 童小娇, 万中. 数值最优化[M]. 北京: 科学出版社, 2005.
- [6] ISO/IEC JTC1/SC29/WG11. N3312-2000 MPEG-4 Video VM 16.0[S]. 2000-03.

