

基于能流信息的视频分割

张忠伟,刘贵忠,李宏亮,李永利

(西安交通大学电子与信息学院 信息与通讯工程系,陕西西安 710049)

摘要: 通过把视频序列中图像的像素看作运动的气体分子,本文提出了一个图像温度场的概念,并利用描述气体分子做无规则运动的麦克斯韦-玻尔兹曼分布律,来描述图像序列中像素的运动规律.与传统光流信息相比,由图像温度场估算出的能流信息更加精确地描述了图像间的运动场,将该能流信息进一步应用到视频镜头的自动分割中,仿真实验结果表明,基于能流分析的视频分割方法更加准确地检测出了镜头的切变和渐变.

关键词: 镜头检测;运动估计;图像温度场;能流场;光流场

中图分类号: TN941.1.73 **文献标识码:** A **文章编号:** 0372-2112 (2005) 01-0177-04

Using Energy Flow Information for Video Segmentation

ZHANG Zhong-wei, LIU Gui-zhong, LI Hong-liang, LI Yong-li

(School of Electronic & Information, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China)

Abstract: By looking up pixels of image sequence as those moving gas molecules, a novel concept, image temperature, has been proposed to describe 2-D image motion during video analysis. The idea is motivated by the well-known Maxwell-Boltzman Distribution Law in the gas dynamic theories. The Energy Flow Field (EFF) estimated by the image temperature presents more accurate motion vectors than the traditional Optical Flow Field (OFF). When the information of EFF is applied for video segmentation, the experimental results show that EFF provides better performance.

Key words: shot detection; motion estimation; image temperature field; EFF; OFF

1 引言

视频分割是将一个完整的视频序列分割成多个内容相关的段落或镜头,并用一个关键帧来表示,以便快速有效地实现视频数据的管理、检索和场景分析等,提高视频资源的利用率.视频分割是实现数字视频数据库智能管理的关键技术.

一个完整的视频序列是由多个内容不同的镜头所组成的,镜头的变化通常都伴随着图像的亮度、色度、纹理和物体形状等特征信息的变化,因此,上述信息在视频分割中最早被应用,并开发出了一系列高效的视频分割技术^[1-3].图像的运动信息包含了丰富的视频内容,运动信息的变化反映了视频内容的不同,因此,近年来基于运动信息的视频分割技术的研究得到了快速的发展.图像的运动信息既包括P、B帧中基于块的运动矢量,也包括基于像素的光流信息.又因光流具有更丰富的运动信息和它更接近图像的实际运动场,于是,基于光流分析的视频分割技术受到了更广泛的关注^[4,5].

光流只是三维物体的实际运动在图像平面上的投影,所以它包含了许多虚假的运动信息,使基于光流分析的视频分割方法造成镜头检测的漏检和误检.为此,本文在分析了光流信息的缺点之后,提出了一个全新的能流分析方法来估算图

像间的运动信息,并把该运动信息应用于镜头检测中,显著提高了视频分割的效率.

2 麦克斯韦-玻尔兹曼分布率和图像的温度场

2.1 光流方程(Optical Flow Function, OFE)

光流表达了图像的变化,由于它包含了目标运动信息,因此可被观察者用来确定目标的运动情况,它在视觉运动研究中有非常重要的作用.在1976年,Cafforio与Rocca^[6]首次提出了图像序列运动估计的光流分析方法,该方法是通过图像光流与液体密度流的对比推导出来的.液体密度流方程的前提假设为:液体流动的时候,沿着流线方向上液体密度保持不变.光流方程是液流方程在图像序列运动分析中的一个推广.在利用光流方程进行图像运动估计时,通常假设沿着运动轨迹图像亮度保持不变.假设 $S(x, y, t)$ 表示连续时空的亮度分布,那么根据这个假设可以推导出光流方程

$$\frac{\partial s(x, y, t)}{\partial x} u + \frac{\partial s(x, y, t)}{\partial y} v + \frac{\partial s(x, y, t)}{\partial t} = 0 \quad (1)$$

式中: $\partial s(x, y, t) / \partial x$ 和 $\partial s(x, y, t) / \partial y$ 是图像亮度分别在空间的水平和垂直方向上的梯度, $\partial s(x, y, t) / \partial t$ 是图像亮度在时间上的梯度, $u = dx/dt$ 和 $v = dy/dt$ 分别是水平和垂直方向上

图像的运动速度。

光流方程是一个计算图像运动场的近似方程,运动场是三维物体的实际运动在图像平面上的投影,然而我们所能得到的光流场是图像亮度随时间的变化,所以光流场并不总是对应于运动场^[11]。例如,外部照明每帧都在变化时,即使没有运动,光流也能被观测到;另外,在缺乏足够的空间图像梯度的区域,实际运动往往被观测不到。以上这些问题都是光流方程本身所不能解决的。

2.2 气体动理论和麦克斯韦-玻尔兹曼分布定律

麦克斯韦是最早研究气体分子的速率分布规律的物理学家,并于 1859 年从理论上推导出了理想气体分子在平衡状态下的速率分布函数^[7],即麦克斯韦分布律:

$$f(v) = 4 \left(\frac{m}{2kT} \right)^{3/2} v^2 e^{-mv^2/2kT} \quad (2)$$

式中: m 是气体分子的质量, T 是平衡状态的温度, k 是玻尔兹曼常数, v 是气体分子的运动速率。由式(2)做出的理论曲线与实验曲线是极为接近的,说明麦克斯韦分布律真实地描述了气体分子速率分布的客观规律。

玻尔兹曼在麦克斯韦研究的基础上,进一步研究了不同离散能量状态下微观粒子数的分布规律,即麦克斯韦-玻尔兹曼分布律。该定律的描述是:在一个理想的气体平衡系统中,如果微观粒子只可能具有一系列不连续的能量,其值由小到大排列为 $\epsilon_1, \epsilon_2, \dots, \epsilon_i, \dots$ 。则在能量为 ϵ_i 的状态上分布的粒子数可表示为

$$N_i = C e^{-\epsilon_i/kT} \quad (3)$$

式中: C 是由系统状态所决定的玻尔兹曼常数;在理想平衡状态下,气体分子的能量 ϵ_i 主要是由其动能来决定的,所以能量可写为: $\epsilon_i = mv_i^2/2$ 。由式(2)和式(3)可知:平衡状态下,气体分子的速率分布函数和分子数都是关于速率的“钟”型函数。

麦克斯韦-玻尔兹曼分布律是气体分子动力学的一个基本定律,它定量地描述了气体分子做随机运动时分子数的分布规律。该定律不单在气体分子动力学研究中起了巨大的推动作用,而且在固体物理、激光等近代物理学科中,在研究微观粒子的随机运动规律时,也得到了广泛的应用。

2.3 图像的温度场

由 2.2 的分析可知,麦克斯韦-玻尔兹曼分布律是一个描述微观粒子随机运动属性的广泛定律。视频信号也是随着场景的变化而随机变化的,所以本文把图像序列中的像素看作是随机运动的气体分子,提出了一个“图像温度场”的概念,并用麦克斯韦-玻尔兹曼分布律来描述图像序列中像素的变化规律。

假设视频序列中每帧图像内任何位置都具有一定的“温度”,与气体温度的定义相一致,并且这样的图像“温度”是由该位置处像素的“运动速度”和一个视频单元内具有该“运动速度”的像素总数所共同决定。把像素的亮度信息看作像素的“运动速度” v_i ,并统计一个 GOP 中具有该“运动速度” v_i 的像素总数为 N_i ,由式(3)可得

$$N_i = C e^{-mv_i^2/2kT} \quad (4)$$

对式(4)两端取自然对数,整理后得

$$T = mv_i^2 / 2k(\ln C - \ln N_i) \quad (5)$$

式中: m 是像素的“质量”,本文假设所有的像素都有相同的“质量”值 1; k 和 C 都是归一化常数。

由式(5)可知,图像任何位置处的温度都是由该位置处像素的“运动速度” v_i 与具有该“运动速度” v_i 的像素总数 N_i 共同决定的。利用式(5)可以计算一个视频单元中的图像任何位置处的“温度”,这样就把原始亮度图像序列转换为“温度场”图像序列。

3 图像序列二维运动的能流约束方程

3.1 二维运动的能流约束方程

根据 2.1 推导光流方程的方法,当定义了图像的温度之后,假设沿着运动轨迹上图像的“温度场”保持不变,就可以推导出一个全新的描述图像运动信息的偏微分方程,即

$$\frac{\partial T(x, y, t)}{\partial x} u + \frac{\partial T(x, y, t)}{\partial y} v + \frac{\partial T(x, y, t)}{\partial t} = 0 \quad (6)$$

式中: $\partial T(x, y, t) / \partial x$ 和 $\partial T(x, y, t) / \partial y$ 是图像温度场分别在空间的水平和垂直方向上的梯度, $\partial T(x, y, t) / \partial t$ 是图像温度场在时间上的梯度, $u = dx/dt$ 和 $v = dy/dt$ 表示水平和垂直方向上图像温度场的运动速度。

根据气体分子运动理论,忽略分子之间的势能,气体的温度是气体分子平均动能的度量。所以,视频中像素的温度也是像素能量的度量,沿着运动轨迹上温度场不变,也就是能量保持不变,所以该假设就称为能流约束条件,式(6)即为本文的能流方程。

3.2 由能流方程估算图像的运动场

能流方程式(6)中有两个变量 u 和 v ,而基本等式只有一个,因此只能求出能流场沿梯度方向上的值,而不能同时求出它的两个速度分量 u 和 v 的值。也就是说,从基本等式(6)求解能流场的运动信息是一个不适定问题。光流场的计算最初是由美国学者 Horn 和 Schunck^[8]提出的,他们的方法一直得到广泛的应用。在相邻图像之间的时间间隔很小,且其图像中灰度变化也很小的前提下,他们利用 FIR 高通滤波器近似速度矢量的拉氏算子,得到计算光流场运动信息的 Gauss-Seidel 迭代算法。为了使能流方程与光流方程有一个公平的比较,我们也用文献[8]的方法估算温度图像间的运动场。在计算机的实现中,时间和空间上亮度或温度的梯度都由两个相邻帧亮度或温度的差分来近似,具体的实现方法读者可以参考文献[9]。

3.3 能流运动场与光流运动场的对比

为了检验新算法的有效性,本文分别用光流方程和能流方程估算 News 测试序列的第 488 帧和第 489 帧图像间的运动场,如图 1 所示。图 1(a)和(b)是 News 测试序列的第 488 帧和第 489 帧图像;图 1(c)是两幅图像之间差分的绝对值图像,表示事实上的运动信息,其中较亮的像素是由于运动,相对于前一帧亮度有变化的像素。由图 1(c)可以看出,图 1(a)变化到图 1(b)包含了两个运动:头部的运动和口部的运动。图 1(d)、(e)分别是光流分析方法和能流分析方法估算出的运动场。

从图 1 的比较可以看出,能流方程估算的运动场捕捉到

了绝大多数的实际运动,虽然光流方程估算的运动场也捕捉到了帧间的运动信息,但它同时产生了大量的伪运动信息.分别用光流运动场和能流运动场作为运动补偿信息,由 News488 来重构 News489,并计算两个重构图像与原始 News489 之间差分图像的峰值信噪比 PSNR,PSNR 值越大,说明重构图像与原始 News489 就越接近,运动补偿信息也就越精确,所以 PSNR 可以作为评价运动估算精确性的品质因数.实验计算表明,用光流运动场和能流运动场作为运动补偿信息的两个差分图像的 PSNR 分别为 18.69dB 和 22.18dB.因此,与光流方程相比,用能流方程能够更加精确的估算出图像间的运动信息.

4 用能流信息进行镜头检测

因光流场自身存在的缺点,使利用光流信息的镜头检测方法对那些复杂的镜头渐变无能为力,经常造成镜头的漏检和误检.利用光流信息,Cheong 和 Guo^[4]提出了一种新颖的镜头检测方法,该方法是先计算光流场的平滑度与图像亮度变化梯度的比值,如果该比值大于给定的阈值就认为镜头发生了变化.该方法有效消除了因图像亮度变化产生的虚假运动信息的影响,实验仿真结果也表明,该方法不但成功检测出了所有的镜头切换,也检测出了绝大部分的镜头渐变.该方法的一个不足之处是:为了得到尽可能精确的光流场,必须先消除图像噪声的干扰,并且要求图像在时间上的变化尽可能的平缓.所以在光流场估算之前,Cheong 和 Guo 的方法要对视频序列用高斯滤波器进行空间和时间的滤波,滤波过程不但大大增加了计算量,而且使镜头的边界发生扩散,即因时间滤波导致无法确定镜头发生变化的确切位置,给视频分割造成很大的困难.另外,Cheong 和 Guo 的方法还会因视频序列中图像的突然运动和大目标的运动造成镜头变化的误检.

由第三部分的分析可知,图像温度场估算出的能流场比光流场更加接近图像实际的运动场,所以把能流场应用于镜头检测应该能够得到比用光流信息进行镜头检测更好的结



图 1 光流方程与能流方程估算出的运动场之间的对比

果.与文献[10]相一致我们把能的平滑度与图像温度场空间梯度的比值 D 作为镜头检测的判据.能流平滑度的定义为

$$T_c^2 = \left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial v}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2 \quad (7)$$

式中:矢量 u 和 v 的偏微分由空间梯度算子 Sobel 得到.图像温度场的空间梯度的定义为

$$T_b^2 = \left(\frac{\partial T(x, y, t)}{\partial x} \right)^2 + \left(\frac{\partial T(x, y, t)}{\partial y} \right)^2 \quad (8)$$

则镜头变化的测度 D 定义为 $D = \frac{n_1(i, j)}{n_2(i, j)}$

其中 $n_1(i, j)$ 和 $n_2(i, j)$ 的定义为

$$n_1(i, j) = \begin{cases} 1 & |T_c| > M_c, \text{ 且 } |T_b| > M_i \\ 0 & \text{否则} \end{cases}$$

$$n_2(i, j) = \begin{cases} 1 & |T_b| > M_i \\ 0 & \text{否则} \end{cases} \quad (9)$$

门限 M_c 与 M_i 的值根据实验来选定.

5 实验及其讨论

为了验证能流信息在视频分割中的有效性,我们用 MPEG7 提供的 Eyeexam 作为测试序列.为了实验的方便我们对 Eyeexam 做了进一步的编辑,编辑后的 700 帧图像中包含 7 个镜头的突变(cut)和 13 个镜头的渐变,其中镜头渐变包括 6 个镜头的溶解(dissolve)、4 个镜头的淡入(fade in)和 3 个镜头的淡出(fade out).综合考虑检测结果的“检中率”和检验的“正确率”, D 的阈值分别选为 7.5% 和 10%.如图 2 给出了用光流信息和能流信息进行镜头检测的结果,图中虚线和点线分别为 D 的阈值分别取 10% 和 7.5% 时的阈值线.

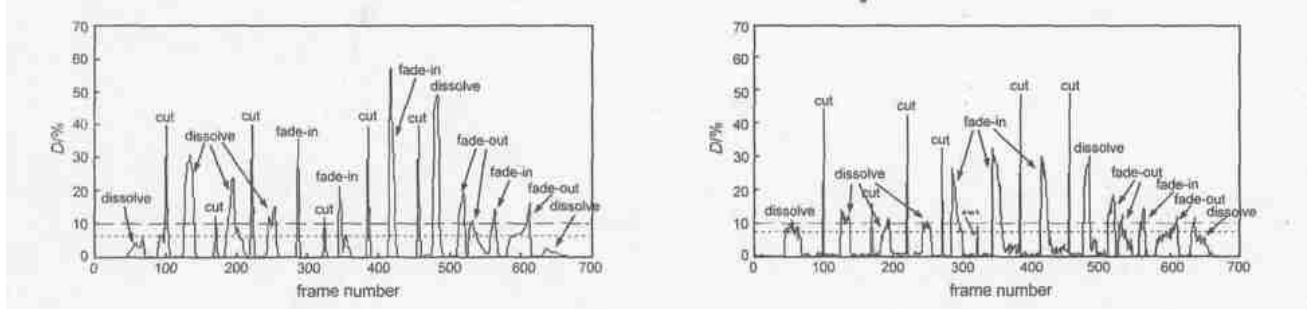


图 2 利用光流信息(左图)和能流信息(右图)进行镜头检测的结果

表 1 用光流信息和能流信息进行镜头检测结果的比较

方法	D 的阈值 (%)	突变		渐变	
		漏检	误检	漏检	误检
光流信息	7.5	0	1	1	2
	10	0	2	2	3
能流信息	7.5	0	0	0	0
	10	0	2	0	2

由图 2 可以看出,利用光流信息不但无法检测到 50 - 70 帧和 640 - 670 帧处的镜头 dissolve,而且镜头边界处也发生了扩散现象,造成无法确定镜头发生变化的准确位置.利用能流信息不但检测出了所有的镜头变化,而且可以精确定位镜头的变化位置.表 1 进一步给出了实验的统计结果,可以看出,利用能流信息进行镜头检测时,大大降低了镜头的漏检和误检.实验结果证明,进行镜头检测时,能流信息比传统光流信

息更具优越性.

6 结论

本文提出了一个图像温度场的概念,利用该温度场估算出的能流信息比用传统的光流信息能够更加精确的捕获图像的实际运动场.与应用光流信息相比,把能流信息应用于视频镜头的自动检测,大大降低了镜头的漏检和误检;因为不需要对视频进行时间和空间上的预滤波,利用能流信息进行镜头检测不但大大减少了镜头检测的计算量,而且可以精确定位镜头的变化位置.本文提出的能流分析方法精确的描绘出了序列图像的运动场,因而,该方法用在序列图像的目标检测和运动目标跟踪中,也是一种极具发展潜力的分析方法.

参考文献:

- [1] N V Patel ,I K Sethi. Video shot detection and characterization for video databases[J]. Pattern Recognition ,Apr. 1997 ,30(4) :583 - 592.
- [2] B L Yeo ,B Liu. Rapid scene analysis on compressed video [J]. IEEE Trans Circuits Systems Video Technol. ,1995 ,5(6) :533 - 544.
- [3] H J Zhang ,C Y Low ,S W Smoliar. Video parsing and browsing using compressed data [J]. Multimedia Tools and Applications ,Mar. 1995 ,1(1) :89 - 111.
- [4] L F Cheong ,H Guo. Shot change detection using scene-based constraint [J]. Multimedia Tools and Applications ,Feb. 2001 ,14(2) :175 - 186.
- [5] P Boutheymy ,F Ganasia. Video partitioning and camera motion characterization for content-based video indexing [A]. IEEE ICIP [C]. Lausanne Switzerland ,1996 ,1 :16 - 19.
- [6] C Cafforio ,F Rocca. Methods for measuring small displacement of television images [J]. IEEE Trans Info Theory ,Sept. 1976 ,22 (5) :573 - 582.
- [7] E R Jones. Contemporary College Physics [M]. Boston :McGraw Hill , 2001.
- [8] B K Horn ,B G Schunck. Determining optical flow [J]. Artificial Intelligence ,Feb. 1981 ,17(2) :185 - 204.
- [9] A M Tekalp. Digital Video Processing [M]. NJ :Prentice Hall ,1995.

作者简介:

张忠伟 1975 年出生,西安交通大学电信学院在读博士生,研究领域主要为图像与视频信号分析与传输控制、基于内容的图像与视频检索、视频数据压缩、小波分析. E-mail :zhweizh@mailst.xjtu.edu.cn

刘贵忠 1962 年生,1989 年 6 月获荷兰 Eindhoven 科技大学博士学位,现为西安交通大学电子与信息工程学院教授,博士生导师,主要从事非平稳信号的分析与处理、音视频数据压缩、模式识别、反演等方面的理论、方法与应用研究.