

基于时空相关性预测的运动估计的优化

张少娴, 俞琼

(南京邮电大学 计算机学院, 江苏 南京 210003)

摘要:运动估计的方法不仅在视频压缩技术中得到广泛的应用,而且也在高质量的视频帧频转换中起到了越来越重要的作用。如何提高运动估计所得到的运动矢量的质量,从而得到更高质量的插入帧图像,是文中研究的目的。在分析比较了现有的几种运动估计算法的基础上,提出了一个优化方法。该方法利用了分水岭算法对视频图像的物体进行正确分割,得到了被识别物体的最优的运动矢量。实验结果表明,该优化方法能够优化所有基于时空相关性预测的运动估计的算法的运动矢量,得到的插入帧图像的质量更高,因而具有较强的实用性。

关键词:运动估计;时空相关性;图像分割;分水岭算法

中图分类号: TN919.81

文献标识码: A

文章编号: 1673-629X(2010)01-0104-04

An Optimization Method for Spatiotemporal Predictive Motion Estimation

ZHANG Shao-xian, YU Qiong

(School of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Not only has method of motion estimation received a wide applications in video compression technology, but also it has played an increasingly important role in high-quality video frame rate conversion. How to improve the quality of the motion vector predicted by the motion estimation algorithm and get the higher-quality images of the insertion frame, is the purpose of this paper. After analyzing and comparing some existing motion estimation algorithms, an optimization method is proposed in this paper, making use of the watershed algorithm to implement video image segmentation, so get the optimum motion vector of the distinguished object. The experiment result show the proposed method can optimize the motion vector predicted by the spatiotemporal predictive motion estimation, thereby the insertion frame of video image's quality can improved, thus has a strong practicability.

Key words: motion estimation; spatiotemporal; image segmentation; watershed algorithm

0 引言

运动估计的方法不仅在视频压缩技术中得到广泛的应用,而且也在高质量的视频帧频转换中起到了越来越重要的作用,比如说由电影胶片的24Hz转换到HDTV的60Hz。同时,随着液晶和等离子电视等新的显示设备在市场上的引入,屏幕尺寸越来越大,对于所显示的视频图像的质量的要求也越来越高。如何提高运动估计所得到的运动矢量的质量,从而得到更高质量的插入帧图像,是文中研究的目的。

1 现有的方法和存在的问题

目前已出现了很多种类的运动估计算法,如:块匹

配法、子波变换法、神经网络法、像素匹配法等。在运动估计的方法中,块匹配运动估计由于其位移跟踪能力强,并且简单和易于实现,是目前最常用的运动估计方法^[1]。在块匹配运动估计的搜索算法中,最基础的是全搜索算法(Full Search, FS)。全搜索算法由于计算强度大,并且在匹配中,图像被简单分割成宏块,独立地逐块进行匹配,相互间没有关联性,造成得到的运动矢量不代表物体的实际运动,即运动矢量缺乏真实性。为了得到代表物体运动的真实的运动矢量,提高运动矢量场的一致性,同时减少运动估计的误差和计算量,提出了基于时间和空间的相关性的预测的运动估计。也就是说,由于视频对象的运动具有连续的运动特征,因此描述该运动特征的运动矢量间也存在时间和空间上的相关性,尤其是在相邻的块之间。运动矢量通常都具有较高的相关性,由此根据这些相邻的块的运动矢量,来预测当前块的运动矢量。例如三维递归搜索算法(3-D Recursive Search, 3DRS)^[2,3]和并

收稿日期:2009-02-04;修回日期:2009-06-25

基金项目:江苏省高校自然科学基金基础研究项目(08KJB620002)

作者简介:张少娴(1972-),女,讲师,硕士,研究方向为数据库、信息安全。

列预测块匹配搜索算法 (Parallel Predictive Block Matching, PPBM)^[4]。

三维递归算法使用了如图 1 所示的两个时空相关预测器。每个时空相关预测器由一个时间预测矢量 $P_{ZA}(P_{ZB})$ 和一个空间预测矢量 $P_{OA}(P_{OB})$ 构成。所谓时间预测矢量,就是前一帧某个位置的块所估计出的运动矢量;空间预测矢量,就是当前帧某一位置的块所估计出的运动矢量。图 1 所示的深黑色的块代表当前的运动估计块,两个箭头分别代表两个时空相关预测器的预测方向。

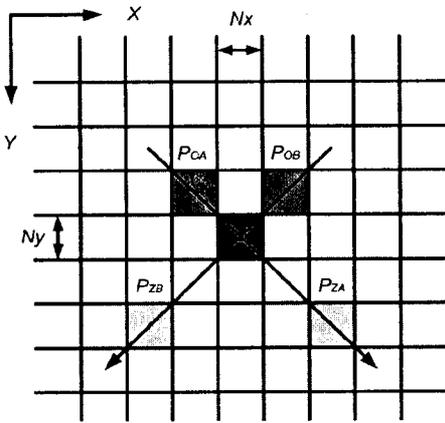


图 1 三维递归算法中时空相关预测器和当前运动估计块的关系

考虑到运动估计中运动物体的速率大小和运动方向的改变,还需要增加一组修正矢量(Update Vector),来确保最终得到的当前块的运动矢量能快速收敛。修正矢量 U 一般随机从当前块周围的块的矢量中选取,并无固定规则。例如可按公式(1)所代表的位置来选取相应的运动矢量作为修正矢量,其中(0,0)代表当前块的位置。

$$U = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \end{pmatrix} \right\} \quad (1)$$

通过时间预测矢量,空间预测矢量和修正矢量的组合,就能得到如公式(2)所示的,时空相关预测器 A 的候选运动矢量组 $C_A(X, t)$,这几个候选运动矢量就可以直接用于当前块的运动估计的匹配计算(SAD)中,由此得到最优的运动矢量。时空相关预测器 B 的候选运动矢量组的构造方法与预测器 A 一样。最后还需要比较时空相关预测器 A 和 B 的最优运动矢量,比较得出的结果就是当前运动估计块的运动矢量。三维递归搜索算法不再需要如全搜索那样在整个搜索窗口里逐块计算,因而在搜索速度上能提高,同时因为其搜索过程中考虑到了运动物体的时间和空间上的相关

性,所以得到的运动矢量能在一定程度上显示出了物体的真实运动。

$$C_A(X, t) = \left\{ \begin{array}{l} 0, \\ V_A(X + P_{OA}, t), \\ V_A(X + P_{OA}, t) + U_A(X, t), \\ V_A(X + P_{ZA}, t - 1) \end{array} \right\} \quad (2)$$

并列预测块匹配搜索算法(PPBM)和三维递归搜索算法在本质上并无区别。主要差别是考虑到了在人的视觉系统中,对水平和垂直运动的物体的辨认敏感度高于斜线运动的物体,从而把两个时空相关预测器的预测方向改为水平的和垂直的。

上述两种基于预测的运动估计得到的运动矢量代表了物体的真实运动,但是只能根据时空相关预测器的预测方向,把整个运动物体的相关性在这个方向上表示出来。也就是说,根据基于预测的运动估计算法得到的运动矢量场,其中反映出来的物体的运动矢量,有很明显的横向纵向(PPBM)或斜向(3DRS)的结构。而作为刚体看待的运动物体,其通过运动估计得出的运动矢量,理论上应该是一致的,也就是说表达该物体运动的运动矢量要一致,不能在物体中有不同的结构或方向性。

2 优化算法的实现

为了解决这个问题,先通过图像分割技术,得到代表物体或其局部的闭合区域,然后根据相对应的矢量场,在闭合区域内部进行的运动矢量的统计,最后把出现最多的运动矢量作为整个区域的运动矢量。

优化的具体的实现步骤是:

1)对当前帧做第一次运动估计,得到相应的运动矢量场。

2)对当前帧进行图像分割,必须要使用能生成闭合区域的分割算法,以便在后续步骤里,能在这个闭合区域内统计相应的运动矢量。这里用的是分水岭算法(Watershed)^[5~8]。

3)通过分水岭算法处理后的图像,是由很多小的闭合区域构成的。这时还需要按照一定的准则,把这些小区域合并到一起。比如可以根据闭合区域内像素的灰度的均值或方差作为合并准则。这里使用的是均值,即当两个相邻的闭合区域各自的灰度和的均值差在一定的范围内,则可以把这两个闭合区域合并为一个大的闭合区域。根据试验,一般要进行 10 次左右的合并过程,才能得到足够大的可辨认出物体的闭合区域。

4)根据步骤 1 得到运动矢量场,统计闭合区域内哪个运动矢量出现得最多,则把这个运动矢量赋给整

个区域,从而得到新的运动矢量场。

5) 把步骤 4 得到的运动矢量场,重新按照运动估计块的大小来划分,并取每块中央的像素点的运动矢量作为分割运动矢量 $V_S(X, t)$ 。

6) 对当前帧做第二次运动估计,同时把步骤 5 得到的分割运动矢量 $V_S(X, t)$ 放入公式(2)中,作为一个新的候选运动矢量。

通过以上步骤,就能得到一个基于物体真实运动的估计误差更小的运动矢量场。

3 结果评测

由于篇幅关系,仅选用了 3 组根据视频对象运动速度区分的、典型的视频序列来给出测试结果。其中视频序列 Wheel 为匀速缓慢转动的车轮和固定的背景图案,视频序列 MatPhone 为正常运动的行人,而序列 Bikechase 为快速运动的骑车人。

对于帧频转换中的运动估计的质量的客观评估,根据文献[9],主要有修正均方误差(Modified Mean Square Error, M2SE, 见公式(3)和(4)和运动矢量场平滑指标(Vector Field Smoothness Indicator, VFSI, 见公式(5)和(6)。其中修正均方误差(M2SE)反映的是当前帧图像 F 与运动补偿图像 F_{mc} 之间的差别。其中 D_x 为水平方向上得到的运动矢量, D_y 为垂直方向上得到的运动矢量。与 MSE 类似,得到的运动矢量的质量越好, M2SE 的数值会越小。

$$M2SE = \frac{1}{N_x \cdot N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \{F(x, y, n) - F_{mc}(x, y, n)\}^2 \quad (3)$$

$$F_{mc}(x, y, n) = \frac{1}{2}(F(x - D_x, y - D_y, n - t) + F(x + D_x, y + D_y, n + t)) \quad (4)$$

运动矢量场平滑指标(VFSI)则反映了运动矢量场的运动矢量之间的一致性程度,也就是运动矢量场的和谐性。该指标需要与 M2SE 一起观察,因为 VFSI 只是个相对值,它的指标越高,只能指出运动矢量场越和谐,越平整,而不能判断运动矢量的质量的好坏。运动矢量场平滑指标(VFSI)见公式(5)和(6),其中 X_B 、 Y_B 分别为宏块的宽和高。

$$VFSI(n) = \sum_x \sum_{k=-1}^{+1} \sum_{l=-1}^{+1} \left\{ \frac{8 \cdot N_B}{|\Delta_x(X, k, l, n)| + |\Delta_y(X, k, l, n)|} \right\} \quad (5)$$

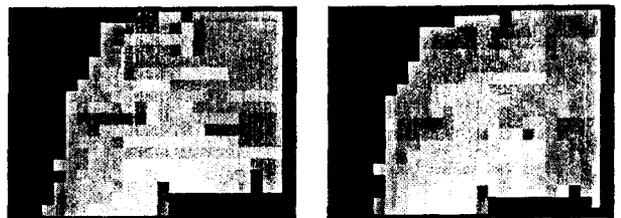
$$\Delta_x(X, k, l, n) = D_x(X, n) - D_x\left(X + \begin{Bmatrix} kX_B \\ lY_B \end{Bmatrix}, n\right)$$

$$\Delta_y(X, k, l, n) = D_y(X, n) - D_y\left(X + \begin{Bmatrix} kX_B \\ lY_B \end{Bmatrix}, n\right)$$

(6)

根据以上两组公式,在表 1 中给出了相应的计算评测结果。由表中可以看到,对于修正均方误差,在优化后,其相对于相应的基于时空相关性预测的运动估计,在数值上均有降低,这说明了优化后的运动矢量的质量得到了提高。而对于全搜索算法(FS),它不能也没必要优化,也是很好理解的,因为对于整个搜索窗口,由该算法计算出的误差值肯定是最底的。但是表中也可以看出,对于整个视频,采用了基于时空相关性预测的运动估计算法,能比全搜索算法得到更低的误差和更好的运动矢量。此外,对于运动矢量场平滑指标,可以看出,在优化后,其相对应的数值均有提高。

由于并列预测块匹配算法(PPBM)比三维递归算法(3DRS)有更好的运动估计结果,限于篇幅关系,只给出了并列预测块匹配算法优化前后的主观(视觉)的比较结果。其中,图 2 是视频序列 MatPhone 其中一帧的运动矢量场的局部截图,并且为了让显示效果更加明显,在图像对比度上进行了增强处理。图中显示出的是一个由左到右运动的人的背部的运动矢量。这里可以注意到,在优化前,其中的运动矢量有很强的横向、纵向的结构。优化之后,其运动矢量显得更加一致与和谐,更加真实地代表了物体的真实运动。图 3 是



(a) 优化前 (PPBM) (b) 优化后

图 2 运动估计矢量场(MatPhone)

表 1 客观评估结果

M2SE	Wheel	MatPhone	Bikechase	均值	VFSI	Wheel	MatPhone	Bikechase	均值
FS	69.5076	114.9148	115.8674	100.0966	FS	0.8194	1.0033	0.9694	0.9307
3DRS	62.1177	76.4911	153.9124	97.5071	3DRS	5.8173	3.7513	3.2953	4.2879
优化后	59.3241	72.8106	139.5144	90.5497	优化后	8.0776	5.0232	4.6319	5.9109
PPBM	56.0236	67.2966	95.8970	73.0724	PPBM	0.8880	0.5828	0.4787	0.6498
优化后	54.8859	65.7952	91.7922	70.8245	优化后	0.9391	0.6282	0.5244	0.6972



(a) 优化前 (PPBM)

(b) 优化后

图 3 插入帧的图像(Wheel)

视频序列 Wheel 其中的一帧,它是经过运动补偿后,重新建立的插入帧的图像。如图所示,在优化前,因为运动估计的错误,在车轮的辐条处有许多明显的缺陷。优化后,这些缺陷就少了很多,这间接说明了优化之后的运动矢量具有更好的质量。

4 结束语

文中分析了基于三维递归搜索算法和并列预测块匹配搜索算法所得到的运动矢量的不足,提出了一种优化方法,利用了分水岭算法对视频图像的物体进行正确分割,得到了被识别物体的最优的运动矢量。实验结果表明,该优化方法能够优化所有基于时空相关性预测的运动估计的算法的运动矢量,从而得到更高质量的插入帧图像,因而具有较强的实用性。

(上接第 4 页)

出现一些干扰和错误。所生成的河网与真实自然水系相比,它所描述的是一种概化了的,会出现很多伪特征。

④分析表明,基于 90m DEM 数据提取的整个珠江三角洲地区流域河网,当阈值取 90(0.729km²)时,生成的流域河网能够较好地反映该地区水系河网。该阈值下计算得到珠江三角洲的平均河网密度大致为 0.86,属于中等偏上。占总整个珠三角面积的 50%左右的坡度小于 3 的地区,主要集中在三江交汇区域和河口地区,平均坡度较低,水量丰富,容易形成支流,其河网密度更是达到了 1.38,属于稠密区。由于珠江三角洲地区河网发达,水系复杂,地势平坦,使得提取的水系河网与实际流域河网难免存在差异,特别是在河口区存在着明显的错误;又由于珠江三角洲地区经济活动剧烈,修建水电站、水库、公路、铁路等人工干预对自然河流的影响相当大,致使所提取的河网在某些地区与实际河流和自然水系不相符合,但是对于整个大区域来说,数字水系能够在一定程度上反映整个珠江三角洲地区流域河网的基本特征。

参考文献:

- [1] 李 淳,马力妮.基于菱形搜索的改进的运动估计算法研究[J].计算机技术与发展,2008,18(11):117-119.
- [2] de Haan G. Progress in motion estimation for video format conversion[J]. IEEE Transactions on Consumer Electronics, 2000,46(3):449-459.
- [3] 王建国,周源华.应用于格式转换的三维递归搜索块匹配算法[J].计算机工程与应用,2003(20):102-104.
- [4] Blume H. Bewegungsschaetzung in Videosignalen mit oertlich-zeitlichen Praediktoren[M]. Dortmund: Tagungsband 5. Dortmunder Fernsehseminar,1993:220-231.
- [5] De Smet P, Pires R. Implementation and analysis of an optimized rainfalling watershed algorithm[C]//Proc of the SPIE, Image and Video Communications and Processing IVCP'2000. [s. l.]:[s. n.],2000:759-766.
- [6] 徐庆伟,李春雷.基于小波变换的图像分割研究刘洲峰[J].计算机应用与软件,2009,26(4):62-64.
- [7] 卢官明.区域生长型分水岭算法及其在图像序列分割中的应用[J].南京邮电学院学报:自然科学版,2000,20(3):51-54.
- [8] 巩 萍,潘冬明.小波分析及其在图像处理中的应用[J].长沙大学学报,2005,19(2):52-54.
- [9] de Haan G. Video Processing for Multimedia Systems[M]. Eindhoven: University Press,2000.

参考文献:

- [1] 曹玲玲,张秋文.基于 SRTM 的数字河网提取及其应用[J].人民长江,2007,38(8):150-152.
- [2] 李 丽,郝振纯.基于 DEM 的流域特征提取综述[J].地球科学进展,2003,18(2):251-256.
- [3] Tarboton D G. A new method for the determination of flow directions and upslope areas in grid digital elevation models[J]. Water Resources Research,1997,33(2):309-319.
- [4] Fairfield J, Leymarie P. Drainage networks from grid digital elevation models[J]. Water Resources Research, 1991, 27(5):709-717.
- [5] 孙友波,宫辉力,赵文吉,等.基于 DEM 的数字河网生成方法的浅议[J].首都师范大学学报:自然科学版,2005,26(2):106-111.
- [6] Band L E. Topographic partition of watersheds with digital elevation models[J]. Water Resources Research, 1986,22(1):15-24.
- [7] 黄诗峰,徐 美,陈德清. GIS 支持下的河网密度提取及其在洪水危险性分析中的应用[J].自然资源学报,2001,10(4):129-132.
- [8] 张行南,叶丽华,井立阳.基于 DEM 的流域流水网对比分析[J].水利水电科技进展,2004,24(2):1-4.