

一种基于块匹配的鲁棒运动估计算法

祝远新 徐光佑 俞志和

(清华大学计算机科学与技术系, 北京 100084)

摘要 详细论述了基于块匹配的鲁棒运动估计算法。跟已有的基于块匹配的运动估计算法比较, 首先, 我们引入颜色信息来提高运动估计的准确性; 其次, 在更广泛的意义上运用自适应策略来减少计算量并同时保证算法的鲁棒性; 最后, 提出的基于预测修正的复合查找方法充分利用了物体运动的全局信息, 克服了三步查找算法以及全查找算法的缺点并充分发挥它们二者的优点从而提高查找的效率和匹配精度。实验结果表明基于块匹配的鲁棒运动估计算法具有抗干扰能力强、运动估计准确、计算效率高等优点。

关键词 块匹配, 运动估计, 自适应策略, 运动向量查找, 颜色空间

1 引言

基于计算机视觉的运动估计有着广阔的应用领域, 例如, 军事领域的多目标跟踪、工业过程的动态监控、商业领域的视频数据压缩、视频数据分析、医学领域的心脏运动研究、以及虚拟现实方面的应用等等。基于块匹配的运动估计就是根据物体在上一帧的位置以及匹配准则运用某种查找策略找出物体在当前帧的位置(即估计帧间运动向量), 因此, 基于块匹配运动估计算法的准确性依赖于匹配过程中所使用的匹配准则以及运动向量查找策略。匹配准则一般是以某种匹配度量(误差度量或相似性度量)为基础的, 匹配度量反映两个像素块之间的匹配程度。文献中提及的匹配度量主要有3种: 基于互相关函数的相似性度量、基于均方差的误差度量、以及基于归一化绝对差的误差度量。

不管使用哪种误差度量, 如果采用全查找算法查找运动向量(即计算查找区域内每个像素位置的匹配度量, 然后根据匹配准则得出最终运动向量), 查找运动向量的计算量都很大。很多研究人员都在研究快速块匹配算法, 已经提出的快速块匹配算法有二维对数查找算法^[1]、三步查找算法^[2]、新三步查找算法^[3]、正交查找算法^[4]、交叉查找算法^[5]、四步查找算法^[6]、自适应块匹配算法^[7]、混合自适应查找算

法^[8]等。这些快速块匹配算法绝大多数是为低比特率视频应用服务的, 例如 MPEG1 视频压缩、可视电话等等。因此它们追求的目标是减少计算量并使误差项尽可能小从而减少编码的长度, 但是从跟踪运动物体的角度出发, 不仅要减少计算量更重要的是要保证匹配(即运动估计)的准确性、鲁棒性。在全面了解各种基于块匹配的运动估计算法的基础上, 结合研究背景, 我们提出了基于块匹配的鲁棒运动估计算法。该算法的鲁棒性表现为抗噪声能力强、跟踪准确、以及计算效率高等。我们在 SGI 工作站上开发了一个运动估计系统。整个系统包括 MPEG1 解码子系统和运动估计子系统。MPEG1 解码子系统一次解码一帧图象, 并且自适应地确定是否把当前帧送到运动估计子系统; 运动估计子系统根据物体在上一帧中的位置以及物体运动的全局信息确定物体在当前帧中的位置(最佳位置)并由此给出物体的帧间运动向量(最佳运动向量)。

2 基于块匹配的鲁棒运动估计算法

2.1 颜色空间的选取

基于下述两点理由, 我们选择在 Y'CrCb 颜色空间进行运动估计而不是在 RGB 颜色空间。

(1) 避免重构 RGB 过程所造成的信息丢失。原

始数据经过 MPEG1 解码子系统解码以后,我们首先得到的是 Y'CrCb 颜色空间的图象而不是 RGB 颜色空间的图象。如果再将 Y'CrCb 颜色空间的图象转换到 RGB 颜色空间必将造成进一步的信息丢失。

(2) 减少计算量。因为一方面不必进行颜色空间的转换;另一方面 Cr 以及 Cb 分量矩阵只有 Y 分量矩阵的 1/4 大,而在 RGB 颜色空间里 R、G、B 三个分量矩阵都跟 Y 分量矩阵一样大。

2.2 基于互相关函数的匹配准则

基于块匹配进行运动估计的准确性很大程度上依赖于匹配过程中所用的匹配准则,其中最主要的是如何度量两个像素块之间的相似性。我们定义一

$$S_Y(p, q) = \frac{\sum_{k=1}^M \sum_{l=1}^N Y^{t-1}(k, l) Y^t(k+p, l+q)}{\left(\sum_{k=1}^M \sum_{l=1}^M (Y^{t-1}(k, l))^2 \right)^{1/2} \left(\sum_{k=1}^M \sum_{l=1}^N (Y^t(k+p, l+q))^2 \right)^{1/2}}$$

$$S_{Cr}(p, q) = \frac{\sum_{k=1}^{M/2} \sum_{l=1}^{N/2} Cr^{t-1}(k, l) Cr^t(k+p, l+q)}{\left(\sum_{k=1}^{M/2} \sum_{l=1}^{N/2} (Cr^{t-1}(k, l))^2 \right)^{1/2} \left(\sum_{k=1}^{M/2} \sum_{l=1}^{N/2} (Cr^t(k+p, l+q))^2 \right)^{1/2}}$$

$$S_{Cb}(p, q) = \frac{\sum_{k=1}^{M/2} \sum_{l=1}^{N/2} Cb^{t-1}(k, l) Cb^t(k+p, l+q)}{\left(\sum_{k=1}^{M/2} \sum_{l=1}^{N/2} (Cb^{t-1}(k, l))^2 \right)^{1/2} \left(\sum_{k=1}^{M/2} \sum_{l=1}^{N/2} (Cb^t(k+p, l+q))^2 \right)^{1/2}}$$

如果我们定义 (p^*, q^*) 为最佳运动向量,则式(1)和下面的式(2)结合起来就是我们所定义的基于互相关函数的匹配准则。

$$(p^*, q^*) = \arg \max_{\substack{-W \leq p \leq W \\ -W \leq q \leq W}} (S(p, q)) \quad (2)$$

其中 W 是物体帧间运动的最大位移。如果不考虑物体运动的全局信息,那么物体在水平方向和垂直方向的运动是等可能的,所以一般把查找范围设为正方形窗口(查找窗口的尺度为 $(2W+1) \times (2W+1)$)。

2.3 基于预测修正的复合查找算法

在已有的各种快速块匹配算法中,三步查找算法是最典型,也是最常用的快速块匹配算法,为了便于比较,我们首先简要地介绍三步查找算法。三步查找算法利用两个物体块的亮度值之间的均方差(MSE)作为匹配度量,匹配度量的定义由式(3)给

出。令 M, N 分别是物体块的高度和宽度(以像素为单位), $Y^{t-1}(k, l)$ 是上一帧中物体块的 Y 分量(亮度分量), $Y^t(k+p, l+q)$ 是当前帧内匹配块的 Y 分量,其中 (p, q) 是可能的运动向量(位移向量)。类似地,我们定义: $Cr^{t-1}(k, l), Cr^t(k, l), Cb^{t-1}(k, l)$, 以及 $Cb^t(k, l)$ 。令 $S(p, q)$ 表示两物体块的匹配度量(基于互相关函数的相似度)。类似地, $S_Y(p, q), S_{Cr}(p, q), S_{Cb}(p, q)$ 分别是物体块的 Y 分量、Cr 分量、Cb 分量的匹配度量。

$$S(p, q) = \alpha S_Y(p, q) + \beta S_{Cr}(p, q) + \gamma S_{Cb}(p, q) \quad (1)$$

这里 α, β, γ 分别是 3 个分量的权重,并且

出。

$$MSE(p, q) = \frac{1}{MN} \sum_{k=1}^M \sum_{l=1}^N [I^{t-1}(k, l) - I^t(k+p, l+q)]^2 \quad (3)$$

其中 M, N 的定义同式(1), $I^{t-1}(k, l)$ 是上一帧的物体块, $I^t(k+p, l+q)$ 是当前帧内可能的匹配块, p, q 分别是水平和垂直运动向量。同理,如果定义 (p^*, q^*) 为最佳运动向量,则可用下面的式(4)来描述三步查找算法的匹配准则。其中 $W (=6)$ 个像素间距是帧间最大位移,查找窗口的大小为 13×13 。

$$(p^*, q^*) = \arg \min_{\substack{-W \leq p \leq W \\ -W \leq q \leq W}} (MSE(p, q)) \quad (4)$$

三步查找算法通过假设匹配度量随查找位置偏离最佳位置方向而单调减少,从而减少查找位置的个数,实现快速块匹配。图 1 所示是三步查找算法的示意图(图中每个“·”表示一个像素位置),具体查找过程如下:

第一步,计算距离查找中心(图中用“ \odot ”表示,它是物体块在上一帧的位置)3 个像素间距远的 9

个位置(图中用“①”表示,包括“⊙”)的 MSE,并把 MSE 最小的位置作为第二步的查找中心;

第二步,计算距离查找中心 2 个象素间距远的 9 个位置(图中用“②”表示)的 MSE,并把其中 MSE 最小的位置作为第三步的查找中心;

第三步,计算距离查找中心 1 个象素间距远的 9 个位置(图中用“③”表示)的 MSE,其中使 MSE 最小的位置就是三步查找算法找出的最佳运动向量。

图中找出的运动向量为(4,-6)。图中的箭头从第一步的查找中心出发指向第二步的查找中心,然后从第二步的查找中心出发指向第三步的查找中心,最后从第三步的查找中心指向查找出的最佳位置。

三步查找算法虽然能大幅度地减少查找位置的个数从而提高查找效率,但是它假设误差度量随着查找位置偏离最佳位置方向而单调增加,而误差度量在整个查找区域里可能有多个谷点(极小值点),因此如果在前面的步骤里做出了不合适的选择,那么就会把全局最优的运动向量排除在外。我们参考 Feng 等人提出的自适应块匹配算法对三步查找算法做了两个方面的改进:一方面用上面定义的基于互相关函数的匹配准则取代基于均方差的匹配准则;另一方面引入自适应多候选策略(即自适应地为下一步的查找中心保留一个或多个候选)来减少排除全局最优运动向量的可能性。具体地说,就是按式(5)比较当前步的最大匹配度量跟该步中其它查找位置的匹配度量。式中 MSM_{max} 和分别 MSM_k 是当前步的最大匹配度量和第 k 个查找位置的匹配度量。如果对于所有的 k, G_k 都接近于 1,那么 MSM_{max} 所对应的位置就是下一步的查找中心。如果一个或多个 G_k 接近于 0,那么就需要保留多个候选作为下一步的查找中心,我们称改进后的三步查找算法为自适应多后选 3 步查找算法。

$$G_k = \frac{MSM_k - MSM_{max}}{MSM_k} \quad k = 1, 2, \dots, 9 \quad (5)$$

全查找算法虽然计算量大,但是它的可靠性高;然而,不管是三步查找算法还是全查找算法都没有利用物体运动的全局信息(即物体在过去运动的方向和速率)。物体运动的全局信息指明了物体运动的趋势,因此它可以为运动向量查找提供指导。为了把物体运动的全局信息、自适应多后选三步查找算法、全查找算法等有机地结合起来克服它们各自的缺点,同时提高查找效率和精度,我们提出了基于预测

修正的复合查找算法。

所谓预测修正就是首先利用物体运动的全局信息对查找位置进行更进一步的精确预测,然后用自适应多后选三步查找算法对预测结果进行修正。如图 2 所示,假设物体继续保持原来的运动状态,那么根据物体运动的全局信息就能在查找窗口(图中,窗口中心用“⊙”表示,窗口尺度为 13×13)内定义一个最可能出现的子窗口(图中,子窗口的中心用“+”表示,窗口尺度为 5×5),即子窗口中心的位置由物体在过去运动的方向和速率决定。由于子窗口较小,所以可以在子窗口内运用全查找算法找出子窗口内匹配度量最大的位置对应的运动向量 V_E 。同时,在整个查找窗口内运用自适应多候选三步查找算法找出最佳运动向量 V_F 。然后比较 V_E 和 V_F 对应的匹配度量,选择匹配度量大的运动向量作为最终的运动向量。由于在查找运动向量的过程中,既用到了自适应多候选三步查找算法又用了全查找算法,所以我们称这种查找算法为基于预测修正的复合查找算法。

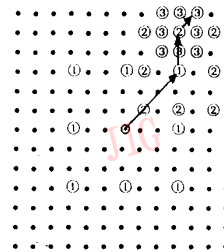


图 1 三步查找算法

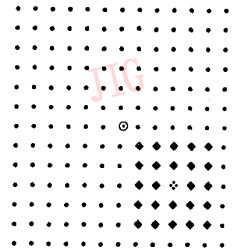


图 2 基于预测修正的复合查找算法。

2.4 运动估计速率自适应控制

物体运动时慢时快,而原始数据的采样速率通常是恒定的(一般为 30 帧/秒)。当物体运动较慢时,显然不必对每一帧都进行运动估计。我们采用自适应控制策略进一步提高运动估计的速率,利用物体运动的全局信息预测物体在未来的运动,并由此确定是否对当前帧进行运动估计。

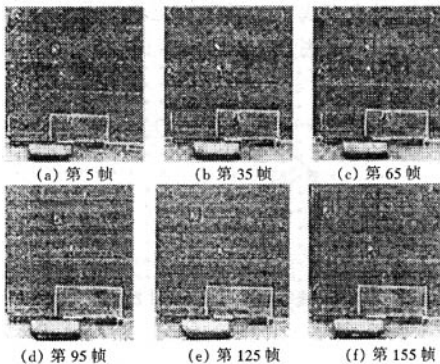
$$M_t = (N_t - N_{t-1}) \times V_{t-1} / W \quad (6)$$

具体地,按式(6)计算 M_t 的值,当 M_t 大于预先设定的阈值时就对当前帧进行运动估计,否则跳过当前帧。式中 N_t 是当前帧的编号, N_{t-1} 是最近一次进行运动估计过的帧的编号, V_{t-1} 是最近一次运动估计得出的运动速率, W 是物体帧间运动的最大位移。

3 实验结果

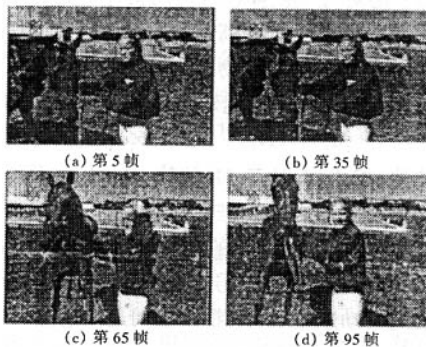
实验数据包括 2 个图象序列。第一个序列是一段足球比赛实况,由 180 帧 384×288 的图象组成。

第二个序列是一段背景非常复杂的电视信号,由120帧 384×288 的图象组成。第一个图象序列(如图3所示)的背景相对简单一些(主要由球场草地组成),但是由于成像距离远,所以关于物体(足球运动员)的信息很少,另外物体本身的运动也很复杂。第二个图象序列(如图4所示)中关于物体(人脸)的信息多一些,但是背景非常复杂,从图中可以看出背景由绿地、运动场、马、电视台标志、树林以及天空等等组成,人脸和天空的灰度级也很接近;更重要的是由于物体运动很快,所以背景变化较大。然而,我们的实验结果表明,无论是对于第一个图象序列还是第二个图象序列,我们提出的基于块匹配的鲁棒运动估计算法对运动向量的估计都很准确,运动估计结果是令人满意的。这充分证明了我们所提出算法的鲁棒性。



位于“□”内的运动员是被跟踪对象。

图3 估计足球运动员运动的实验结果。



位于“□”内的人脸是被跟踪对象。

图4 估计人脸运动的实验结果。

4 结 论

基于块匹配的鲁棒运动估计算法跟过去基于块匹配的运动估计算法比较起来具有可靠性高、计算量小等优点,另外对背景、光照等环境条件约束也很少。我们可以不作任何修改而直接把该算法应用于其它领域,诸如军事领域的多目标跟踪、商业领域的视频压缩、以及虚拟现实方面的应用等等。需要进一步解决的问题包括摄像机标定、遮挡处理等。

参 考 文 献

- 1 Jain J R, Jain A K. Displacement measurement and its application in interframe image coding. IEEE Trans. Comm. COM-29, 1981, (12):1799~1808.
- 2 Koga T, Iinuma K, Hirano A, Iinjima Y, Ishiguro T. Motion-compensated interframe coding for video conferencing, Proc. NTC81, New Orleans, LA, Nov. 1981, C9. 6. 1~9. 6. 5.
- 3 Kappagantula S, Rao K R. Motion Compensated Interframe Image Prediction. IEEE Trans. Commun. 1985, COM-33(9):1011~1015.
- 4 Puri A, Huang H M, D L Schilling. An efficient block-matching algorithm for motion compensated coding. Proc. IEEE ICASSP 1987: 25. 4. 1~25. 4. 4.
- 5 Ghanbari M. The Cross-Search Algorithm for Motion Estimation. IEEE Trans. Comm; 1990, COM-38(7):950~953.
- 6 Po L M, Ma W C. A novel four-step search algorithm for fast block motion estimation. IEEE Trans. Circuit Syst. Video Technol., 1995.
- 7 Feng J, Lo K-T, Mehrpour H, Karbowiak A E. Adaptive block matching motion estimation algorithm for video coding. Electronics Letters, 1995, 31(18):1542~1543.
- 8 Cheung C.-K, Po L W. A hybrid adaptive search algorithm for fast block motion estimation. IEEE Trans. Circuit Syst. Video Technol., 1996.



祝远新 清华大学计算机科学与技术系计算机应用专业博士研究生,1985年毕业于清华大学计算机科学与技术系获学士学位。主要研究方向包括计算机视觉、人机界面(HCI)、图象处理及模式识别。

徐光佑 教授,博士生导师,1963年毕业于清华大学自动控制系,分别于1982~1984年、1993~1994在美国普渡大学、Illinois大学从事计算机视觉和人工智能领域研究。主要研究方向为计算机视觉、多媒体及人机界面(HCI)。已发表专(译)著5部,在国内外重要刊物上发表论文共100余篇。

俞志和 副教授,1968年毕业于清华大学电机工程系,1985年获美国夏威夷大学电气工程硕士,1990年获美国夏威夷大学电气工程博士。主要研究领域有:计算机视觉、水下机器视觉导航、表面检测、图象处理及模式识别。已出版学术专著1部,在国内外重要期刊以及学术会议上发表论文20多篇。

A block matching based robust algorithm for object motion estimation

Zhu Yuanxin Xu Guangyou Yu Zhihe

(Dept. of Computer Science&Engineering, Tsinghua University, Beijing 100084)

Abstract We proposed a robust approach to object motion estimation between frames based on block matching in this paper. Compared with existing methods for object motion estimation between frames, our approach, firstly, employs color information to improve accuracy of block matching; secondly, by adopting a wide range of adaptive scheme, the robustness and high efficiency of the proposed approach can be achieved; finally, prediction-correction based compound search algorithm integrates history information of object motion together with advantages of three-step search algorithm and full search algorithm to yield efficient and accurate motion estimation. Experiment results demonstrate that the proposed approach can track objects moving in a very complex background or that in a low SNR environment, so our approach can achieve more robust motion estimation than existing block matching based methods.

Keywords Block matching, Motion estimation, Adaptive scheme, Motion vector searching, Color space

彩色应用专家联合推出 Colorize.com 网页

美国泰克公司(Tektronix), Kinko's, Adobe及Dynamic Graphics杂志隆重宣布共同推出名为Colorize.com的崭新网页,让商业用户轻松获得彩色应用信息,令商业文件更生动传神。

泰克彩打部市场推广部副总裁Jean Nassar表示:“为方便用户,此次联手行动能将各公司的专业知识融入同一网页。”

这个网址的主页是个完善的信息中心,包括“学习”、“阅读”、“执行”、“提问”、“下载”、“联系”不同部分,每部分提供不同的工具,帮助用户在日常商业文件中增添色彩以达最佳效果。

Colorize.com网页,为用户提供多种广泛内容。“下载”部分可提供多款微软Word,Powerpoint的背景及范例文档;彩色专家Daniel Will-Harris及Roger C. Parker更在每月

“学习”部分撰写彩色使用专栏,探讨如“设计简讯”及“如何运用色彩令信息更生动”等题目。

新网址更是彩色打印机用户交流心得的地方,用户可在“提问”部分发问、寻找答案及投稿交换色彩运用心得。网址的文章皆设有“请教专家”功能,让用户亲自向用者提问。

Colorize.com还为浏览者提供其它资源。“联系”部分引导浏览者进入其它杂志及书籍的网址,从而发掘颜色的重要功能,而“执行”部分备有范例文档、插图、Adobe使用指南及有趣的“贺卡”设计功能,运用色彩自然倍感得心应手。Colorize.com是完备的网上彩色资源中心,帮助用户把缤纷彩色融入商业文件中。

详情参阅Colorize.com网页, <http://www.colorize.com>

(胡青 供稿)