维普资讯 http://www.cqvip.com

计算机科学 2006Vol. 33№. 6

一种基于SVD协方差加权技术的光流估计算法*)

侯云舒　张艳宁　赵荣椿

（西北工业大学计算机学院　西安 710072）

摘　要　本文以经典的L-K光流方程为出发点，提出了一种高效的基于SVD协方差加权的光流估计算法，并成功应用到柔性目标点跟踪中，有效地解决了传统L-K算法的孔径问题、深度不连续点的估计和长序列视频的漂移问题。基于标准测试序列的试验结果，证明该算法能有效地跟踪较长视频序列中具有2D和1D甚至基本没有纹理的具有退化结构的柔性目标点，同时结果还可以作为半稠密的点对应来解决SFM问题中的一个关键难题correspondence。

关键词　视频跟踪，光流估计，协方差加权，SVD，人脸特征点跟踪

Optical Flow Estimation Based on SVD Covariance Weighted Techniques

HOU Yun-Shu　ZHANG Yan-Ning　ZHAO Rong-Chun

(College of Computer Science, Northwestern Polytechnical University, Xi'an 710072)

Abstract　A novel algorithm namely covariance weighted SVD based optical flow estimation is proposed in this paper, which achieves more robust and precise tracking result comparing to the traditional L-K tracker. Covariance weighted is used to transform the problem from a hyper-ellipse space to a hyper-sphere space. SVD is correspondingly performed to involve the subspace constraints. The proposed algorithm has been evaluated by the standard test sequence. The potential applications vary from articulated automation, structure from motion, computer surveillance to human-computer interaction.

Keywords　Visual tracking, Optical flow estimation, Covariance weighted, SVD, Facial features tracking

1　引言

光流分析技术一直是计算机视觉领域中的研究热点之一，已经形成了很多成熟的算法。H-S算法[1]利用图像的一阶梯度和基于平滑假设的正则化技术计算光流，但对图像边界等不连续点效果不好。L-K方法[2]有简单的数学求解方法，但它只利用了某像素的一个很小邻域内的信息，具有一定的不稳定性。Anandan算法[3]能处理大偏移的光流，但其计算代价较大。从数学上讲，L-K方法实际上是小位移光流条件下Anandan算法的牛顿梯度下降迭代解。上述方法有个通病：当待估计的目标点不具备丰富纹理特征时，往往效果不佳。近来已有一些研究工作者发表了一些新的富有成效的成果[4~6]，对作者的研究颇有启发。

柔性目标跟踪正在得到研究人员的广泛关注[11~14]。一般都是通过预先大量手工标定的样本训练出所跟踪目标的统计模型，然后利用优化技术来估计模型的参数以实现跟踪。数学上可以认为，柔性目标存在于一个低维的线性子空间中。但在已提出的算法里，这个子空间的基需要通过训练大量的标定样本得到，这是极其耗时的工作。本文提出的算法利用了子空间约束，但不用训练而是在算法中推得子空间的基，同时利用了不确定性分析理论，与矩阵分解巧妙结合在一起，使得传统的光流估计算法得以改进，适合跟踪柔性目标中缺乏纹理的目标点。

本文以经典的L-K光流方程为出发点，引入协方差加权技术解决孔径问题，引入基于SVD的子空间约束解决深度不连续点估计和长序列视频的漂移问题。通过有效的融合提出了一种高效的基于SVD协方差加权光流的柔性目标点跟踪算法。标准测试序列的试验结果证明，所提出的算法能有效地跟踪较长视频序列中具有2D和1D甚至基本没有纹理的具有退化结构的柔性目标点。

2　SVD协方差加权技术的引入

2.1　协方差加权技术

如果假定在所考察的小窗口（通常为3×3或5×5）内有一致的光流动(u,v)，我们可以得到经典的L-K光流约束方程[2]：

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} g \\ h \end{bmatrix} \tag{1}$$

$$a=\Sigma(I_x)^2, b=\Sigma(I_x \cdot I_y), c=\Sigma(I_y)^2 \tag{2}$$

$$g=\Sigma(I_x \cdot I_t), h=\Sigma(I_y \cdot I_t) \tag{3}$$

(1)式提供了计算该点光流的两个等式，这是因为假设了在所考察窗口内各点有一致的光流。但仔细观察(1)式左边的系数矩阵，可以发现：当它接近奇异或本身就奇异时，仍然无法从中得出该点光流的可靠估计。通过更本质的分析不难发现，在高斯噪声的假设下，该系数矩阵可以被看作是光流估计f的逆协方差矩阵，且它的两个特征向量给出了该点光流估计的最大、最小不确定性的方向，两个特征值给出了在上述方向上的不确定性的定量描述。这样，如果我们运用协方差加权技术，先将系数矩阵做本征分解，得到变换矩阵Q，然后用Q变换光流后再进行求加号逆运算，即可得到协方差加权意义下鲁棒的光流估计（具体过程见下节），它很好地解决了1D纹理光流估计中的漂流问题。这样做的几何意义是先将

*)该研究得到国防基础研究"多源多目标协同感知及超光谱自主检测技术"和航天创新基金"基于图像分析的作战效果分析与评估"的支持。

侯云舒　博士研究生，主要研究方向为模式识别和计算机视觉，现主要从事人脸的跟踪、识别与理解等方面的研究；**张艳宁**　博士生导师，教授，博士，主要研究方向为信号与信息处理、图像处理、模式识别与计算机视觉；**赵荣椿**　博士生导师，教授，主要研究方向为语音、图像信号处理、模式识别、计算机视觉。

维普资讯 http://www.cqvip.com

点按不同方向可靠性大小做一个坐标变换，在变换空间内具有各向同性的纹理特征，在这一空间中可以直接应用 SVD 在最小均方意义下做子空间约束。

2.2 基于 SVD 的子空间约束

子空间理论是线性空间理论的一部分。当所考察的问题建立数学模型后，由于所加的约束而得到的解空间通常只是整个问题的某个子空间。

以人脸为例，假设人脸可以用几个本征脸之和近似[13]，则在弱透视投影成像模型下，我们可以得到一段视频人脸序列三维外形与运动的数学模型：

$$P_{2F\times N}=M_{2F\times 3K}S_{3K\times N} \tag{4}$$

其中 K 是本征脸的个数；M 给出摄像机与物体之间的相对运动参数和本征脸的加权参数，只与帧相关；S 给出各个本征脸的三维几何信息，只与目标有关。

实际上，运动视觉问题(SFM)的任务就是从视频的点对应矩阵 P 中解出以上各参数。本文的主要目标是如何得到视频的点对应矩阵 P。通过(4)式我们可以发现柔性目标在视频序列中各特征点组成的光流矩阵构成一个低维子空间，即

$$\mathrm{rank}(P)\leqslant 3K \tag{5}$$

对其他柔性目标而言，也有类似的结果成立。对刚性场景而言，只要简单地设 $K=1$ 即可发现光流矩阵 P 存在于一个秩为 3 的子空间中。其他成像模型可以类似得到同样结论。该约束可以有效地替代通常光流估计中的平滑约束，具有几何意义并在深度突变或摄像机运动突变时仍然成立。

2.3 SVD 协方差加权结合使用的原理

第 2.2 节做子空间约束光流的过程实际上是通过 SVD 分解得到的，这自然有着类似于第 2.1 节描述的问题，即矩阵 P 并非各向同性。由于(1)式左边的矩阵可以看作是光流估计的逆协方差，因此用于 SVD 分解的光流矩阵的各分量的不确定性构成一个超椭球。如果能在子空间约束之前先将 P 矩阵变换到各向同性的、具有超球状意义的空间中，那么在此空间中通过 SVD 分解即可得到正确的最小均方意义下的子空间逼近。这正是 SVD 协方差加权结合使用的原理所在。

3 基于 SVD 协方差加权技术的光流估计在柔性目标点跟踪中的应用

为详细介绍算法使用的流程，作者重新定义了如下符号系统。

假设待跟踪的视频为 $V=(I_0, I_1, I_2, \cdots, I_F)$，待跟踪的点在参考帧 I_0 中的坐标为 $P_0(x_1, y_2, \cdots, x_N, y_N)^T$，重心为 $\overline{P_0}$。

作者拟采用(4)式对待考虑的柔性目标进行建模，其他的类似柔性目标直接推广即可。对刚性目标更为简单，直接设 $K=1$ 即可。

现考察第 i 个点在第 j 帧中的光流 $f_{ij}=(u_{ij}\ v_{ij})'$ 的情况。参考(1)式，有

$$y_{2\times 1}=\begin{bmatrix}g_{ij}\\ h_{ij}\end{bmatrix}=\begin{bmatrix}a_i & b_i\\ b_i & c_i\end{bmatrix}\cdot\begin{bmatrix}u_{ij}\\ v_{ij}\end{bmatrix}=X_{2\times 2}\cdot f_{2\times 1} \tag{6}$$

其中，

$$y_{2\times 1}=\int_R (I_0(P_i)-I_j(P_i))\nabla_{P_i}\mathrm{d}P \tag{7}$$

$$X_{2\times 2}=\int_R \nabla_P \nabla_P^T \mathrm{d}P \tag{8}$$

将(6)式代入到有 N 个点、F 为帧的光流场中，可建立类似(4)式紧凑的矩阵形式：

$$Y_{2N\times F}=X_{2N\times 2N}\cdot F_{2N\times F} \tag{9}$$

其中，

$$F_{2N\times F}=\begin{bmatrix}u_{11} & u_{12} & \cdots & u_{1F}\\ v_{11} & v_{12} & \cdots & v_{1F}\\ \vdots & \vdots & \ddots & \vdots\\ u_{N1} & u_{N2} & \cdots & u_{NF}\\ v_{N1} & v_{N2} & \cdots & v_{NF}\end{bmatrix} \tag{10}$$

表示了各点在各帧中的光流估计矩阵。

$$Y_{2N\times F}=\begin{bmatrix}g_{11} & g_{12} & \cdots & g_{1F}\\ h_{11} & h_{12} & \cdots & h_{1F}\\ g_{21} & g_{22} & \cdots & g_{2F}\\ h_{21} & h_{22} & \cdots & h_{2F}\\ \vdots & \vdots & \ddots & \vdots\\ g_{N1} & g_{N2} & \cdots & g_{NF}\\ h_{N1} & h_{N2} & \cdots & h_{NF}\end{bmatrix} \tag{11}$$

反映了各帧各点的时空梯度。当光流估计正确时，在亮度连续性假设的条件下它应趋于零。

$$X_{2N\times 2N}=\begin{bmatrix}a_1 & b_1 & \cdots & 0 & 0\\ b_1 & c_1 & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & a_N & b_N\\ 0 & 0 & \cdots & b_N & c_N\end{bmatrix} \tag{12}$$

在(9)式中，在高斯噪声的假设下，(12)式所表示的系数矩阵 X 可以被看作是光流估计矩阵 F 的逆协方差矩阵，它的特征向量和特征值分别给出了光流估计的最大、最小确定性的方向和相应的确定性的度量。同时，X 可以看作是 Y 的协方差阵。

下面，作者给出 SVD 协方差加权技术在多帧、多点光流估计中的具体应用流程：

1)建立高斯金字塔，对每一层做如下操作；

2)在参考帧中，按(12)式计算 X；

3)对 X 做本征分解，得到 $\Sigma_Y=X=V\Lambda V^T$，令 $R=V^T$，$Q=\Lambda^{1/2}V^T$，$IH=QX^+$；

4)设定起始的光流估计矩阵 F，起始位移矩阵 CT，给出参考帧中参考点的坐标 P_0，并求出其重心 $\overline{P_0}$；

5)根据当前的光流估计，按(7)、(11)式计算 Y；

6)求出当前的位移估计 $T=(Q\cdot(1_{N\times 1}\otimes I_{2\times 2}))\backslash(IH\cdot Y)$，更新 $CT=CT-T$；

7)对当前的估计值做变换，$YY=-IH\cdot Y+R(F\oplus P_0\ominus\overline{P_0}\ominus CT)$；

8)用 SVD 对 YY 做秩为 2×3k 的子空间约束，得到 UV；

9)将 UV 变换回原空间，并做向量转置，用 SVD 对其做秩为 3K 的子空间约束，得到 UUVV；

10)用下式得到当前各点的新的估计：$P=\mathrm{UUVV}\oplus CT\oplus\overline{P_0}$；

11)如果结果收敛则停止，否则转到第(5)步，继续迭代直至收敛。

在上述算法中，需要注意第(9)步的秩约束直接来自于(5)式所得到的 3k，而第(8)步由于对矩阵做了长度为 2 的向量装置，故其秩的上限变为 2×3k。

4 结果分析

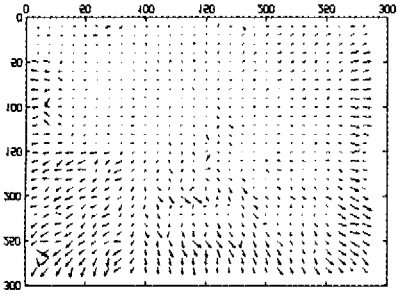
为验证所提出算法的有效性，作者在标准测试序列上分别做了光流估计和点跟踪试验。本文提出的方法在这些数据集上的成功应用，有力地证明了本算法的通用性和推广能力。

4.1 NASA-COKE 序列试验结果

NASA-COKE 序列实际上是在摄像机对准 COKE 瓶子

维普资讯 http://www.cqvip.com

做聚焦，即相当于拉近镜头的过程。仔细观察可以发现，图中存在大量的深度突变的地方，大量的点是没有什么纹理的，用通常的 L-K 算法对它们的光流估计是很混乱的结果。图 1(a)的结果是作者直接应用本文算法得到的，并没有进行任何光滑处理。结果表明，各点的光流不但具有正确的趋势，而且它们的大小也反映了该图像各部分的深度。左下部分的光流明显最大，这与它们离镜头最近是一致的；左右各有两数列的光流也具有比较大的值，因为它们对应与那两个离镜头较近的立杆。



(a) 光流场

(b) 参考帧

图 1　NASA-COKE 序列试验结果

4.2　美国小姐 MISSA 序列试验结果

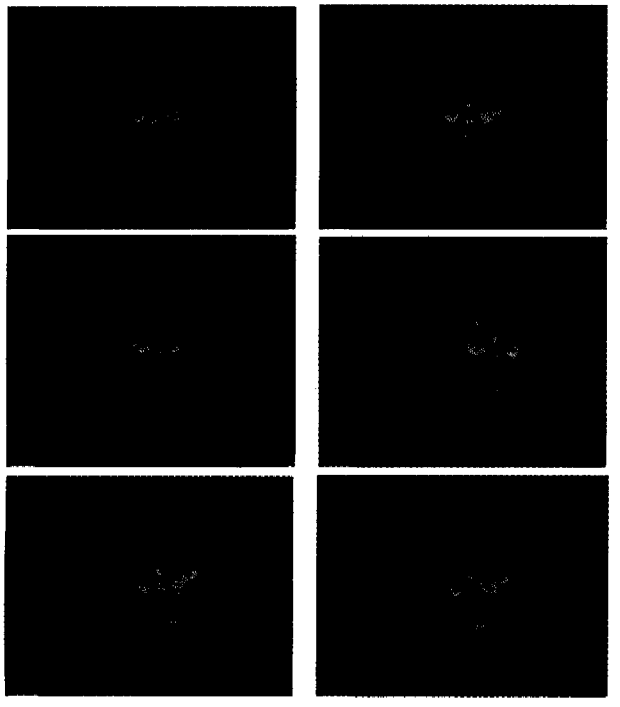


图 2　美国小姐 MIissa 序列试验结果

图 2 为作者应用本文算法在具有较复杂的柔性变形的美国小姐 Missa 序列的试验结果。上述视频中，Missa 的头部有明显的位移，并且眼睛时开时闭，嘴部时张时合，这给精确的点跟踪带来巨大的困难；同时，仔细观察所跟踪的点，它们绝大部分都只具有 1D 的纹理，很容易产生线性漂移的现象。传统的 L-K 方法对下巴轮廓点跟踪会丢失，眼睛的开闭也会有一定问题。

令人兴奋的是，图 2 的跟踪结果表明本方法能有效地在整个视频序列跟踪所有的点。即使当眼睛闭上时，点的估计还是很准确的，并没有乱跑，并随着眼睛的睁开而重新得到。以上各帧包括了脸部的各种柔性变形：眼睛和嘴巴的开与闭、各种表情的变化和面内的旋转等。分别给出的是第 000，006，020，082，118，147 帧的图像点跟踪结果。

结论　在通常的光流估计中，一般只估计角点、拐点等可靠光流，进而用全局优化的技术得到其他所有点的估计。本文引入协方差加权的技术，先将点按不同方向可靠性大小做一个坐标变换，在变换空间内具有各向同性的纹理特征，在这一空间中可以直接应用各种优化技术并保证得到的估计值不会出现在可靠性小的方向出现比较大的流动估计偏差；同时，通过分析真实物体，通过投影成像产生光流这一物理过程，发现这些光流有着流型上的限制，即它们只存在于一个低维线性子空间，有很好的结构特征。这样，在上面提到的变换空间中，我们可以引入低维子空间的秩约束，实际上用光流子空间的基可以组合出所有点的光流。通过 SVD 协方差加权光流，我们可以直接估计具有一维纹理特征甚至某些缺乏纹理特征的点的光流。基于标准测试序列的试验结果证明，本文提出的算法能有效地跟踪较长视频序列中具有 2D 和 1D 甚至基本没有纹理的具有退化结构的柔性目标点。本文提出的算法在诸如人脸表情分析和解决 SFM 中点对应问题等领域有着广泛的应用价值。

参 考 文 献

1　Horn, Schunck. Determining Optical Flow [J]. Artificial Intelligence, 1981,17:185～204

2　Lucas B, Kanade T. An iterative image registration technique with an application to stereo vision [C]. In: Proc. of the Inte Joint Conf. on Artificial Intelligence, 1981

3　Anandan. A computational framework and an algorithm for the measurement of visual motion [J] I. JCV, 1989,2:283～310

4　Shi J, Tomasi C. Good Features to Track [C] . In:Proc CVPR, 1994

5　Brand M. Flexible flow for 3D nonrigid tracking and shape recovery [C]. In: Proc. CVPR, 2001

6　Brand M. Morohable 3D models from video [C]. In: Proc . CVPR, 2001

7　Irani M, Anandan P. Factorization with uncertainty [C]. In: Proc. ECCV, 2000

8　Irani M. Multi-frame optical flow estimation using subspace constraints [C]. In:Proc. ICCV, 1999

9　Bregler C, Hertzmann A, Biermann H. Recovering nonrigid 3D shape from image streams [C]. In: Proc. CVPR, 2000

10　Tomasi C, Kanade T. Shape and motion from image streams under orthography: A factorization method [J]. Inter Journal of Computer Vision, 1992,9(2):137～154

11　Ahlberg J. An active model for facial feature tracking [J]. EURASIP Journal on Applied Signal Processing, 2002(6):566～571

12　Blanz V, Vetter T. A morphable model for the synthesis of 3D faces [C]. In:Proc. SIGGRAPH, 1999

13　Black M J, Jepson A D. EigenTracking: Robust matching and tracking on articulated objects using a view-based representation [C] . In: Proc. ECCV, 1996

14　Hager G D, Belhumeur P N. Efficient region tracking with parametric models of geometryand illumination [J] . IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20 (10):1025～1039