

文章编号:1674-2869(2015)09-0045-05

# SIFT 特征匹配和 VFC 算法的电子稳像技术

姜哲颖,周华兵\*,刘 姣

武汉工程大学计算机科学与工程学院,湖北 武汉 430205

**摘 要:**为了处理较为模糊的抖动视频,提出了一种基于尺度不变特征转换(SIFT)的特征匹配和向量场一致(VFC)优化算法的电子稳像技术.该技术着重于研究视频的运动估计阶段,利用 SIFT 提取高独特性的特征点,根据特征点进行帧间匹配,结合 VFC,通过向量场的学习将外点从内点区分开来,得到两帧图像中特征序列的平移轨迹,然后估算出运动补偿向量并校正每一帧图像的相对位置,从而输出稳定视频.实验表明,该电子稳像技术处理抖动视频时精度高和耗时短.

**关键词:**运动估计;向量场一致性;电子稳像

**中图分类号:**TP391

**文献标识码:**A

**doi:**10.3969/j.issn.1674-2869.2015.09.008

## 0 引 言

近年来,稳像技术得到广泛的应用,无论是航空、车载、舰船上,为了获得清晰不晃动的视频,稳像技术都可以用来消除载体不稳定所引起的视频抖动<sup>[1]</sup>.稳像技术一般分为三种:机械稳像、光学稳像和电子稳像,从各方面的考虑,包括体积小、操作方便、可移植性强、实时性等,电子稳像技术得到更多的普及.电子稳像技术大体分为三步:运动估计,运动滤波和运动补偿.本文大部分的工作集中于运动估计中.处理运动估计有基于灰度和基于特征的方法<sup>[2]</sup>:基于灰度的方法可以得到精确的结果,但要处理所有的像素点而耗时太多;提取图像特征的方法处理速度快,然而精度又得不到保证.对于特征匹配,可以查阅到丰富的文献资料,角点匹配,边缘模型匹配,SIFT 点匹配,MSER 特征匹配,KLT 特征追踪等<sup>[3]</sup>,在处理较为模糊的图片时,SIFT 点匹配的优势就更为明显.在匹配的过程中,消除产生的误匹配,很多用到最小二乘法,然而当数据里有很多外点的情况,该方法处理得并不好.在此,就用到向量场优化算法,向量场一致性是在希尔伯特空间中基于吉洪诺夫正则化,保持向量场的平滑性.在贝叶斯框架下,将每个样点与潜在变量相结合来判断是否为内点,然后构建后验概率问题,通过 EM 算法求解.运动估计过程由局部向

量估计得到全局运动向量,就可以进行运动滤波和运动补偿,从而得到稳定清晰的输出图像<sup>[4]</sup>.

## 1 SIFT 算法的实现

尺度不变特征转换能够从图像中提取高度独特的不变特征,这些特征被用在不同的图像里针对相同对象进行匹配.最初提取这些特征是为了对象识别;将样例图片中检测到的特征与由众多不同视角中提取的特征而组成的数据库进行匹配.

SIFT 算法需要获得全方位的大量特征,主要阶段是对一系列的图像特征的计算:尺度空间极值检测、关键点的精确定位、方向赋值、描述算子的计算<sup>[5]</sup>.

实施的第一步是建立一个尺度空间金字塔,将输入帧进行重复滤波化和下采样.处理过程中,两个连续的高斯分布被一个常量  $k=\sqrt{2}$  分开,每个八度图计算 5 幅图像,将会得到 4 张高斯差分图,因此,就可以在两张不同的图像中找到极值.定位了每一个关键点之后,进行对比,然后通过 Harris 边缘检测.对比阈值和边缘检测算子都会影响最终关键点的个数,成功的关键就是如何去除定位不准的特征或者不稳定的点.

图像处理中,避免不稳定的特征非常重要,运动估计算法需要高度独特的特征,而误匹配往往会影

收稿日期:2015-08-28

基金项目:湖北省教育厅科学技术研究项目(Q20151503)

作者简介:姜哲颖(1991-),男,湖北武汉人,硕士研究生.研究方向:数字图像处理.\*通信联系人.

到向量场一致性算法<sup>[6]</sup>(VFC).

## 2 基于 VFC 的去噪匹配

最小二乘法经常被用来处理去除特征的外点,但是当外点过多时,处理效果不甚明显. VFC 算法有很强的鲁棒性,能够处理含 90% 外点的特征图像.

### 2.1 问题公式化

给定一个输入输出对的集合,  $S = \{(x_n, y_n) \in \mathbf{X} \times \mathbf{Y}\}_{n=1}^N$ , 通过一个随机的向量场, 该向量场包含一些未知的外点, 其中  $\mathbf{X} \subseteq \mathbb{R}^p$ ,  $\mathbf{Y} \subseteq \mathbb{R}^d$  分别是输入空间和输出空间. 目的是将外点从内点中区分开来, 由一个映射  $f: \mathbf{X} \rightarrow \mathbf{Y}$ , 可以得到满意的内点, 其中  $f \in H$ , 假设  $H$  是一个再造核的 Hilbert 空间.

在此做一个假设, 关于内点, 是一个均值为 0, 标准差为  $\sigma$  的高斯噪声; 对于外点, 输出在一个有界的区域  $\mathbb{R}^d$ , 因此外点的分布为平均分布  $\frac{1}{a}$ , 其中  $a$  是一个常量.  $\gamma$  为内点的百分比, 内点和外点的分布模型如下:

$$p(\mathbf{Y}|\mathbf{X}, \theta) = \prod_{n=1}^N p(y_n | x_n, \theta) = \prod_{n=1}^N \left[ \frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|y_n - v_n\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \right] \quad (1)$$

式(1)中  $\theta = \{f, \sigma^2, \gamma\}$  是含未知参数的集合,  $\mathbf{X}_{N \times p} = (x_1, \dots, x_N)^T$ ,  $\mathbf{Y}_{N \times d} = (y_1, \dots, y_N)^T$ ,  $\mathbf{V}_n = f(x_n)$ .

考虑到平滑性的约束, 先验  $f$  可写为:

$$p(f) \propto e^{-\frac{\gamma}{2} \|f\|_H^2} \quad (2)$$

$\lambda > 0$  是正则化参数,  $\|\cdot\|_H^2$  是 RKHS  $H$  的范数.

结合式(1)和式(2), 其后的分布  $p(\theta|\mathbf{X}, \mathbf{Y}) \propto p(\mathbf{X}|\mathbf{Y}, \theta)p(f)$  估算出来.

为了得到最佳的  $\theta$ , 假设一个  $\theta^*$  为

$$\theta^* = \arg \max_{\theta} p(\mathbf{X}|\mathbf{Y}, \theta)p(f) \quad (3)$$

$\theta^*$  对应真正的  $\theta$ , 从而得到  $f$ . 然后运用 EM 方法来讨论  $f$  减少对外点的影响.

### 2.2 EM 方法

EM 算法分为两步, 一是期望步骤, 另一个是最大化步骤.

首先, 将样点  $n$  赋予一个变量  $\mathbf{Z}_n \in \{0, 1\}$ , 当  $\mathbf{Z}_n = 1$ , 代表高斯分布; 当  $\mathbf{Z}_n = 0$ , 代表平均分布. 根据 EM 算法的标准步骤, 并且忽略与  $\theta$  独立的项, 得到:

$$Q(\theta, \theta^{dd}) = \frac{1}{2\sigma^2} \sum_{n=1}^N P(\mathbf{z}_n = 1 | x_n, y_n, \theta^{dd}) \|y_n - v_n\|^2 - \frac{D}{2} \ln \sigma^2 \sum_{n=1}^N P(\mathbf{z}_n = 1 | x_n, y_n, \theta^{dd}) + \ln(1-\gamma) \sum_{n=1}^N P(\mathbf{z}_n = 0 | x_n, y_n, \theta^{dd}) + \ln \gamma \sum_{n=1}^N P(\mathbf{z}_n = 1 | x_n, y_n, \theta^{dd}) - \frac{\gamma}{2} \|f\|_H^2 \quad (4)$$

可以通过把  $\mathbf{Z}_n$  视为混合模型中缺失的数据来实现最大化.

期望步骤:  $\mathbf{P} = \text{diag}(P_1, \dots, P_N)$ , 其中  $P_n = P(\mathbf{z}_n = 1 | x_n, y_n, \theta^{dd})$  可以通过贝叶斯法则来计算:

$$P_n = \frac{\gamma e^{-\frac{\|y_n - v_n\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|y_n - v_n\|^2}{2\sigma^2}} + (1-\gamma) \frac{(2\pi\sigma^2)^{D/2}}{a}} \quad (5)$$

最大化步骤: 定义修正的参数估计为  $\theta^{new}, \theta^{new} = \arg \max_{\theta} Q(\theta, \theta^{dd})$ . 考虑到  $\mathbf{P}$  是对角矩阵, 将  $Q(\theta)$  对  $\sigma^2$  求导并置为零, 得到:

$$\sigma^2 = \frac{\text{tr}[(\mathbf{Y} - \mathbf{V})^T \mathbf{P}(\mathbf{Y} - \mathbf{V})]}{D \cdot \text{tr}(\mathbf{P})}$$

其中  $\text{tr}(\cdot)$  是矩阵的迹. 将  $Q(\theta)$  对  $\gamma$  求导并置为零, 得到:

$$\gamma = \frac{\text{tr}(\mathbf{P})}{N}$$

高斯部分的混合系数  $\gamma$  通过后验概率的均值得到.

EM 算法收敛后, 应该对样点是否为内点做一个判断. 设定阈值  $\tau$ , 可以得到内点集  $T = \{n | P_n > \tau, n = 1, \dots, N\}$ , 这个集合  $T$  就是随机采样一致性中的一致性.

### 2.3 向量场正则化

考虑到式(4)中  $Q$  与  $f$  相关的项, 并把它们乘以 -1, 得到一个能量函数:

$$\varepsilon(f) = \frac{1}{2\sigma^2} \sum_{n=1}^N P_n \|y_n - f(x_n)\|^2 + \frac{\gamma}{2} \|f\|_H^2$$

这个能量函数是一个吉洪诺夫正则化的向量值的推广, 其中第一项可以看成是权值经验误差.

使用向量值表现定理, 对  $f$  的最优化有如下形式:

$$f(x) = \sum_{n=1}^N \Gamma(x, x_n) c_n$$

系数  $c_n$  由一个线性结构决定  $(\tilde{\Gamma} + \lambda \sigma^2 \tilde{P}^{-1})C = \tilde{Y}$ , 其中, 矩阵核  $\tilde{\Gamma}$  和  $\tilde{P}$  是  $N \times N$  的分块矩阵, 每个块是一个  $D \times D$  的标量矩阵.  $\tilde{\Gamma}$  的第  $(i, j)$  个块是  $\tilde{\Gamma}(x_i, x_j)$ .  $\tilde{P} = P \otimes I_{D \times D}$  其中  $\otimes$  表示了矩阵张量积.

$c_n \in R^{D \times 1}$ ,  $C = (c_1^T, \dots, c_N^T)^T$ ,  $\tilde{Y} = (y_1^T, \dots, y_N^T)^T$  是  $D \times N$  维的向量. 另有 VFC 的具体算法可参考文献[6].

### 3 稳像步骤及实验结果

利用 SIFT 算法计算帧间图像的关键点, 以及

匹配点对.

结合 VFC 优化算法, 去除误匹配, 根据准确的帧间匹配关系估算出全局运动向量.

由全局运动向量估算出运动补偿向量并校正每一帧图像的相对位置, 从而输出稳定视频.

从一段较为抖动的视频图像中, 选取了 4 张连续帧图像进行特征匹配实验.

图 1 是平直公路上汽车行驶, 由于后方拍摄工具的不稳定而引起的视频抖动, 从视频中提取前 4 帧图像, 分别对图像帧进行 SIFT 特征提取后和 VFC 的优化算法去除误匹配, 能够得到如图 2 所示两两之间较为准确的匹配关系.



1



2



3



4

图 1 抖动视频中 4 张连帧图

Fig.1 Four successive frames of the shaky video

图 2 中的 4 帧图像里, 分别对 1&2、1&3、1&4 进行了匹配, 从而可以得到帧间运动向量, 即可获知两帧图像之间的相对运动, 然后就可以估计出

运动补偿图像. 图 3 中可以看出, 由局部到整体, 多张运动补偿图像帧即可输出稳定视频图像.

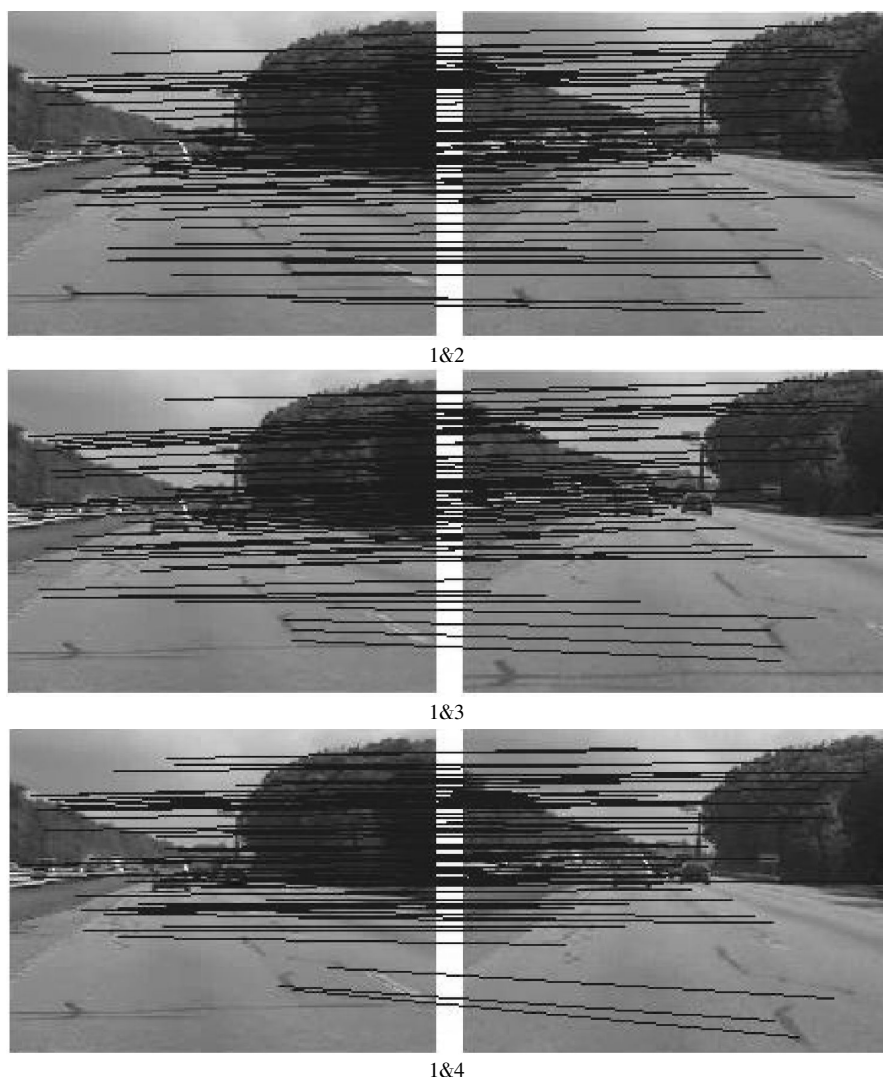


图 2 图像帧之间的匹配关系

Fig.2 The matching relationship between image frames



图 3 处理合成后的稳定帧

Fig.3 Composite steady frame

## 4 结 语

基于 SIFT 特征匹配与 VFC 优化算法的稳像技术,主要针对运动估计过程,将帧间图像的匹配达到高精度而得到清晰的输出视频.在实时性与高精度两者间取得了较为合适的搭配,各取所长、

不失偏颇,也是处理视频抖动研究中一个小的进步.当然,不足之处在于还未与其他方法作直观的对比,和未对实验结果进行更为客观的评价,如帧间保真度、峰值信噪比,并且时间复杂度相对较大,这些问题将在之后作进一步研究.

## 致 谢

感谢湖北省教育厅对本研究的支持!

## 参考文献:

- [1] 张跃飞.车载摄像机数字稳像技术研究[M].成都:电子科技大学,2011.  
ZHANG Yue-fei.The research of the digital image stabilization technology for the In-Car cameras [M]. Chengdu: University of Electronic Science and Technology of China,2011.(in Chinese)
- [2] KUMAR S, AZARTASH H, BISWAS M, et al. Real-time affine global motion estimation using phase corre-

- lation and its application for digital image stabilization [J]. IEEE Trans Image Process, 2011, 20(12): 3406–3418.
- [3] OKADE M, BISWAS P K. Video stabilization using maximally stable extremal region features[J]. Multimedia Tools Appl, 2014, 68(3): 947–968.
- [4] 王海晖, 卢培磊, 吴云韬, 等. 无参考视频平滑度的评价方法[J]. 武汉工程大学学报, 2015, 37(6): 56–62. WANG Hai-hui, LU Pei-lei, WU Yun-tao, et al. Evaluation method of no-reference video smoothness [J]. Journal of Wuhan Institute of Technology, 2015, 37(6): 56–62. (in Chinese)
- [5] BATTIATO S, GALLO G, PUGLISI G, et al. SIFT features tracking for video stabilization[J]. IEEE Computer Society, 2007, 27(5): 825–830.
- [6] ZHAO Ji, MA Jiayi, JIN Wen et al. A robust method for vector field learning with application to mismatch removing[J]. Computer Vision and Pattern Recognition (CVPR) 2011 IEEE Conference on, 2011, 30(6): 2977–2984.

## Electronic Image Stabilization Based on SIFT Matching and VFC

JIANG Zhe-ying, ZHOU Hua-bing, LIU Jiao

School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China

**Abstract:** To process the blurry shaky video, electronic image stabilization based on scale invariant feature transform (SIFT) matching and vector field consensus (VFC) was proposed, which is mainly applied to video motion estimation. First, SIFT was used for extracting highly distinctive invariant feature and setting up the initial feature matching. Then, by the vector field consensus algorithm, we distinguished inliers from outliers and obtained the characteristics of translational trajectory in sequence of two video frames. Finally, we estimated the motion compensation to correct the relative position of every video frame by trajectory. Experiment results show the advantages of the method in precision and efficiency.

**Keywords:** motion estimation; vector field consensus; electronic image stabilization

本文编辑:陈小平