

doi: 10.7690/bgzdh.2016.04.012

## 基于 SIFT 特征的视频稳像算法

陈 滨, 杨利斌, 赵建军

(海军航空工程学院兵器科学与技术系, 山东 烟台 264001)

**摘要:** 为将载体上截取的视频进行稳像, 介绍了尺度不变特征变换 (scale invariant feature transform, SIFT) 算法及基于 SIFT 特征的视频稳像算法。基于电子稳像算法框架采用六参数仿射模型表征图像运动, 利用图像的 SIFT 特征估计图像间的模型参数, 通过补偿来消除图像间的运动, 完成了视频稳像任务。最后, 利用峰值信噪比 (peak signal to noise ratio, PSNR) 指标评价稳像效果。实验结果表明: 该算法能很好地消除图像间的运动, 有效提高图像间的 PSNR 值, 完成视频序列的稳像任务。

**关键词:** 视频稳像; SIFT 特征; 运动估计

**中图分类号:** TP391.41 **文献标志码:** A

## Video Stabilization Algorithm Based on SIFT Feature

Chen Bin, Yang Libin, Zhao Jianjun

(Department of Weapon Science & Technology, Naval Aeronautical Engineering Institute, Yantai 264001, China)

**Abstract:** In order to stabilize the video obtained from cameras on kinds of carriers, this paper introduces the scale invariant feature transform (SIFT) algorithm and video stabilization algorithm based on SIFT features. The motions between images are represented with the model of affine model with six parameters, which is estimated by using the SIFT features of images. Then the images are compensated so that the motions would be counteracted and the goal of stabilizing the video would be achieved. At last, peak signal to noise ratio (PSNR) is imported to evaluate the effect of the algorithm. As the experiments show, algorithm proposed in this paper can counteract the motions among images very well and can improve the values of PSNR effectively. Video stabilization of video sequence is also completed with the algorithm.

**Keywords:** video stabilization; SIFT feature; motion estimation

### 0 引言

随着数字化视频设备的发展与普及, 视频数字化存储方式逐渐取代了传统存储方式, 在各个领域得到了广泛的应用。摄像过程中由于载体的抖动, 造成了图像传感器姿态的变化, 进而造成了系统光轴与目标之间产生平移、角运动等无效运动, 从而导致了输出视频图像的不稳定。由于不稳定视频对于图像的后续处理有较大的负面影响, 视频稳像技术成为图像处理领域的热点之一。

视频稳像技术被广泛运用在测绘、摄影和军事领域<sup>[1-2]</sup>。特别是在军事领域, 图像传感器所处的载体通常是机动的, 载体的震动、抖动将会导致图像信息的不稳定, 对于目标自动识别系统则会造成虚警和漏警。目前的稳像算法鲁棒性不强, 当场景中有运动物体或是有遮挡时, 稳像算法表现不佳; 因此, 在各种条件下, 将运动摄像载体上获取的视频进行稳像, 在民用、军用领域有着重要的意义。

### 1 视频稳像技术

稳像系统按照作用机制可以分类 3 类: 机械式

稳像、光学稳像和电子稳像。机械稳像及光学稳像涉及到物理设备, 主要由摄像机生产商和专业摄影者使用<sup>[3]</sup>。

随着数字图像处理的迅猛发展, 利用纯数字图像处理方法来稳定图像成为当前稳像技术发展的趋势。电子(数字)稳像是建立在连续视频图像之间的运动估计的基础上, 通过对视频中每帧图像进行运动滤波和运动补偿处理得到稳定图像。由于电子(数字)稳像相对于机械式、光学稳像来说具有独特的优点, 逐渐成为稳像技术研究的重点研究方向。

#### 1.1 电子稳像算法

电子稳像采用图像处理的方法来完成稳像功能, 通过计算相邻两帧图像之间的全局运动矢量, 滤除高频的抖动分量, 对第二帧图像进行运动补偿, 最后完成图像的稳像处理。综上所述, 稳像过程可以分为 3 个阶段: 全局运动估计、运动补偿和图像生成, 如图 1 所示。

在全局运动估计中, 摄像机的运动模型是一个重要的问题。在设定某种摄像机运动模型的情况下, 图像上所有像素的运动服从相同的规律。目前, 六

收稿日期: 2015-12-26; 修回日期: 2016-02-04

作者简介: 陈 滨(1988—), 男, 广东人, 博士, 从事计算机视觉研究。

参数仿射模型能够精确地描述旋转、平移和缩放运动，表达式为

$$\begin{bmatrix} X_{i+1} \\ Y_{i+1} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \end{bmatrix} + \begin{bmatrix} a_5 \\ a_6 \end{bmatrix} \quad (1)$$

其中  $a_1, a_2, a_3, a_4, a_5, a_6$  为帧间变换参数，利用最小二乘法进行参数估计。

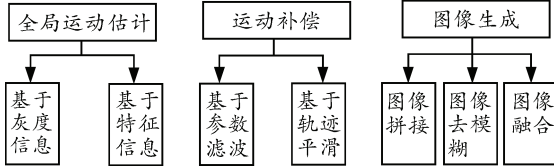


图 1 电子稳像流程

### 1.2 PSNR 评价方法

当所有的运动都被补偿抵消后，理想状况下，两帧图像中每一个像素点相减后得到的结果应该全部为零。但是误差、估计错误以及由于运动模型限制引起的失真都可能得到非零值。为了衡量稳像的效果，采用峰值信噪比 (peak signal to noise ratio, PSNR)<sup>[4]</sup> 可以计算出偏离最佳结果的程度。PSNR 的值越高，表明稳像系统的精确度越高。换句话说，在两帧图像一模一样情况下，PSNR 将取到最大值。

两帧图像之间的 PSNR 定义为：

$$PSNR(I_1, I_0) = 10 \log \frac{255^2}{MSE(I_1, I_0)} \quad (2)$$

其中：灰度图像的灰度级为 255； $MSE(I_1, I_0)$  表示两幅图像之间的均方误差，如下式：

$$MSE(I_1, I_0) = \frac{1}{NM} \times \sum_{n=1}^N \sum_{m=1}^M (I_1(x_n, y_m) - I_0(x_n, y_m))^2 \quad (3)$$

## 2 SIFT 算法

尺度不变特征变换 (scale invariant feature transform, SIFT) 是一种计算机视觉算法，用来检测并描述图像中的局部特征。算法构建尺度空间，并在空间中寻找极值点，并且计算其位置、尺度和旋转不变量<sup>[2]</sup>。SIFT 特征与图像的大小、旋转无关，而是基于图像局部的一些兴趣点，在噪声、视角及光线变化等条件下具有很好的效果。

### 2.1 尺度空间

尺度空间理论的基本思想是：在图像信息处理模型中引入一个被视为尺度的参数，通过连续变化尺度参数获得不同尺度下的视觉处理信息，然后综合这些信息以更加深入地挖掘图像的本质特征。

尺度空间满足视觉不变性：用眼睛观察物体时，一方面当物体所处背景的光照条件变化时，视网膜感知图像的亮度水平和对比度有所不同；因此，要求尺度空间算子对图像的分析不受图像的灰度水平和对比度变化的影响，即满足灰度不变性和对比度不变性。相对于某一固定坐标系，当观察者与物体之间的相对位置变化时，视网膜所感知的图像的位置、大小、角度和形状是不同的；因此，要求尺度空间算子对图像分析和图像的位置、大小、角度以及仿射变换无关，即满足平移不变性、尺度不变性、欧几里德不变性及仿射不变性。

一个图像的尺度空间  $L(x, y, \sigma)$ ，定义为一个变化尺度的高斯函数  $G(x, y, \sigma)$  与原图像  $I(x, y)$  的卷积，如下式所示：

$$L(x, y, \sigma) = G(x, y, \sigma) \times I(x, y) \quad (4)$$

其中：

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp \left[ -\frac{(x-m/2)^2 + (y-n/2)^2}{2\sigma^2} \right] \quad (5)$$

式中： $(x, y)$  代表图像像素位置； $\sigma$  是尺度空间因子，值越小表示图像被平滑得越少，相应尺度也就越小。大尺度对应于图像的外貌特征，小尺度对应于图像的细节特征。

### 2.2 SIFT 流程

SIFT 算法<sup>[5-6]</sup> 主要由 4 个阶段组成：

1) 尺度空间极值检测。

构建尺度空间的第一步就是建立高斯金字塔。

假设原图像大小为  $M_1 \times N_1$ ，塔顶图像大小为  $M_n \times N_n$ ，满足：

$$n = \log_2(\min(N_1, M_1)) - \log_2(\min(N_n, M_n)) \quad (6)$$

构建高斯金字塔模型的流程图如图 2 所示。

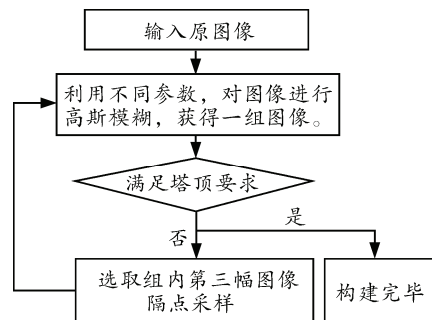


图 2 高斯金字塔构建流程

高斯差分算子能够产生稳定的图像特征，算子  $D(x, y, \sigma)$  如式 (7)；因此，由高斯金字塔生成高

斯差分金字塔，从而产生更加稳定的关键点。实际计算时，将高斯金字塔每组中的相邻两层图像相减，就可得到高斯差分图像；因此，高斯金字塔模型每组含有  $p$  幅图像，高斯差分金字塔仅有  $p-1$  幅图像。

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \times I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (7)$$

得到高斯差分金字塔后，由于极值检测是对像素的图像域及尺度域的 26 个相邻点进行大小比较，金字塔的第一层与最后一层无法检测极值。

## 2) 关键点定位。

由于 1) 中检测到的是离散空间的极值点，为了去除对比度低的关键点和不稳定的边缘响应点，以增强匹配稳定性、提高抗噪声能力，需要通过函数拟合来确定关键点的位置和尺度。

## 3) 方向确定。

在关键点所在的高斯金字塔图像中，对其  $3\sigma$  领域窗口内的像素计算梯度和方向分布特征，使用梯度直方图将  $360^\circ$  分为 36 个区域，每个区域  $10^\circ$ 。使用梯度直方图统计该领域内像素的梯度和方向，并用直方图的峰值作为关键点的主方向，保留峰值大于主方向峰值 80% 的方向作为关键点的副方向。至此，SIFT 特征点可被表示为含有位置、尺度和方向的关键点。

## 4) 关键点描述。

最后，为每一个关键点建立一个描述符，并用一组向量将这个关键点描述出来，令其不受到光照、视角等变化的影响。SIFT 描述子使用关键点为中心的  $4 \times 4$  个窗口，计算每个子窗口内像素点 8 个方向的梯度信息，生成  $4 \times 4 \times 8$  共 128 维向量。这个向量是该区域图像信息的一个抽象，具有唯一性。

## 3 实验及结果

文中的主要工作是基于图像的 SIFT 特征来估计连续图像序列的帧间运动矢量，并完成视频稳像，其基本步骤为：

- 1) 在视频图像序列的每一帧中提取图像的 SIFT 特征；
- 2) 确定帧间特征对应关系；
- 3) 通过特征量计算运动参数，并由此代入图像运动模型，求出全局运动矢量；
- 4) 利用全局运动矢量，得到稳态图像。

### 3.1 实验测试集

实验利用无人机拍摄的视频测试集，选取视频

中连续多帧图像(第 101、102、103、104 帧)，测试算法有效性，并且给出 PSNR 指标评价稳像效果。测试集的差分结果，如图 3 所示。

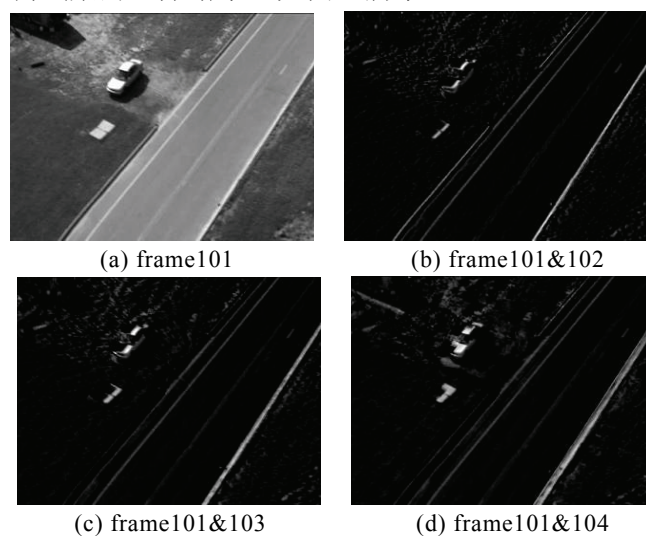


图 3 测集差分图像

测试图像集中，两帧图像间的 PSNR 值如表 1 所示，可见随着时间的推移，摄像机产生某些运动后，后续帧与第 101 帧的距离越来越远，相应的 PSNR 值也逐渐增大。

表 1 测试集 PSNR 值

图像帧数	frame101	frame102	frame103	frame104
frame101	$\infty$	32.85	32.01	31.29
frame102	32.85	$\infty$	32.62	31.63
frame103	32.01	32.62	$\infty$	32.19
frame104	31.29	31.63	32.19	$\infty$

### 3.2 SIFT 特征匹配

利用 SIFT 算法计算相邻两帧图像间的关键点，并且进行匹配。实验分别对测试集内 4 帧连续图像进行关键点检测，并且对相邻两帧图像进行特征点匹配，结果如图 4 所示。由结果可见：SIFT 特征具有较好的稳定性，匹配效果好，可以起到很好的图像局部特征表征作用，利用 SIFT 特征点对应关系计算全局运动矢量具有较强的可行性。

### 3.3 全局运动估计

笔者采用的运动模型为六参数仿射模型。在 SIFT 特征匹配中，得到了匹配点的坐标信息，以此作为输入，利用最小二乘法求解式 (1) 中的参数，对图像的全局运动作出有效估计，为方便计算，将式 (1) 转换成如下形式：

$$\left. \begin{aligned} x_{i+1} &= [x_i \ y_i \ 1] [a_1 \ a_2 \ a_5]^T \\ y_{i+1} &= [x_i \ y_i \ 1] [a_3 \ a_4 \ a_6]^T \end{aligned} \right\} \quad (8)$$



求解出参数后，利用方程组，将后一帧图像中的每一个像素点作为输入，求出运动前其所在位置，得到输出图像。求解得到的参数结果如表 2 所示。



(a) frame101 & 102



(b) frame102 & 103



(c) frame103 & 104

图 4 SIFT 特征匹配结果

表 2 运动模型参数值

参数	frame101 frame102	frame101 frame103	frame101 frame104
$a_1$	0.994 3	1.000 2	0.996 8
$a_2$	0.004 6	0.000 1	-0.009 4
$a_3$	0.010 2	0.002 0	0.009 1
$a_4$	0.994 0	0.996 9	1.012 3
$a_5$	-2.141 2	-5.123 5	-9.283 2
$a_6$	-1.945 7	-1.768 2	-0.895 2

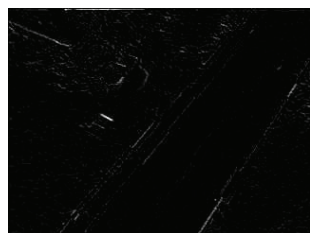
由式 (1) 可知：

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}^{-1} \left( \begin{bmatrix} X_{i+1} \\ Y_{i+1} \end{bmatrix} - \begin{bmatrix} a_5 \\ a_6 \end{bmatrix} \right) \quad (9)$$

式中的参数通过最小二乘估计已经得到，所以可以利用式 (9) 获得后续图像经过稳定后的图像，差分结果如图 5 所示。



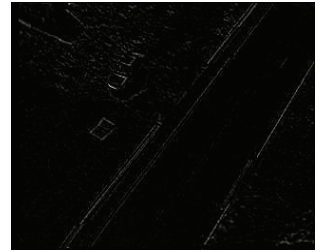
(a) frame101



(b) frame101&102



(c) frame101&103



(d) frame101&104

图 5 稳像后差分图像

稳像后计算图像间的 PSNR 值，结果如表 3。

表 3 稳像后 PSNR 值

图像帧数	frame101	frame102	frame103	frame104
frame101	$\infty$	37.89	38.01	35.69
frame102	37.89	$\infty$	35.23	34.36
frame103	38.01	35.23	$\infty$	34.239 3
frame104	35.69	34.36	34.239 3	$\infty$

由图 3 和图 5 中图像间的差分结果可知，图像间的运动得到了很好的补偿。由表 1 和表 3 对比可知，任意 2 幅图像间的 PSNR 都得到了提高。可见利用 SIFT 特征估计全局矢量，补偿图像间运动，从而稳定视频中的图像序列有较好的效果。

#### 4 结论

笔者利用图像的 SIFT 特征，在航拍视频中选取连续图像进行实验，对连续图像中运动矢量参数进行估计并补偿，从而达到稳定视频图像的目的。实验结果表明：该算法有效地消除了图像间全局运动，降低了 PSNR 值，能够较好地完成视频序列的稳像任务。

#### 参考文献：

- [1] Sachs D, Nasiri S, Goehl D. Image Stabilization Technology Overview[M]. Santa Clara, CA: InvenSense Inc., 2007: 18.
- [2] Chang H C, Lai S H, Lu K R. A Robust Real-time Video Stabilization Algorithm[J]. Journal of Visual Communication and Image Representation, 2006, 17(3): 659-673.
- [3] 陈启立, 宋利, 余松煜. 视频稳像技术综述[J]. 电视技术, 2011, 35(7): 15-17.
- [4] Morimoto C, Chellappa R. Evaluation of image stabilization algorithms[C]. IEEE Int'l Conf. Acoustics, Speech and Signal Processing, 1998: 2789-2792.
- [5] David G. Lowe. Object Recognition from Local Scale-invariant Features[C]. International Conference on Computer Vision, Corfu, Greece: Springer US, 1999: 1150-1157.
- [6] David G. Lowe. Distinctive Image Features from Scale-invariant Keypoints[C]. International Conference on Computer Vision, Springer US, 2004: 91-110.