

# 基于不变特征的运动视频序列自动配准算法

李 静 杨 涛 潘 泉 程咏梅

(西北工业大学自动化学院, 西安 710072)

**摘 要** 快速、鲁棒的图像配准是运动视频处理的基础,也是制约后继应用稳定性及可靠性的关键。针对运动视频中存在的图像平移、旋转、尺度及光照变化,提出一种基于不变特征的快速图像配准算法,包括特征点检测、描述和匹配。首先通过多层箱式滤波器构建图像多尺度空间,并同时考虑质量与空间分布检测特征点;然后用主成分分析法对 SIFT (scale invariant feature transform) 特征进行降维,用于特征描述;最后根据描述子主成分的差异设计层叠分类器,加速特征匹配。定量分析实验和对视觉监视系统中球形摄像机和无人机航拍视频的实验结果表明,该算法具有良好的匹配性能,为后继运动载台上的运动目标检测、跟踪、分类等处理提供了坚实基础。

**关键词** 图像配准 不变特征 多尺度空间 主成分分析 层叠分类器

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)02-0335-10

## Invariant Feature Based Automatic Motion Video Registration

LI Jing YANG Tao PAN Quan CHENG Yongmei

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

**Abstract** Fast and robust image registration is an important research problem in motion video processing. In this paper we present a novel invariant feature based automatic image registration method to deal with the large image transformation, rotation, scale and illumination changes. The algorithm includes three parts: invariant feature detection, description and matching. First, the multi-scale space of the image is created via a multi-level box filter, and then the feature points are detected in scale space by considering the quality and special distribution simultaneously. Second, we use the Principle Component Analysis to descend dimension of SIFT (Scale invariant feature transform) for feature description. Finally, based on the principle component of the descriptors, a cascade filtering is designed to speed up the feature matching. Experiments with motion videos captured by the dome camera and the Unmanned Aerial Vehicle demonstrate that the proposed algorithm has satisfied performance, and it can provide a solid foundation for further processing such as moving object detection, tracking and classification from moving platform.

**Keywords** image registration, invariant feature, multi-scale space, PCA, cascade filtering

## 1 引言

运动视频是由搭载在运动平台上的摄像机获取的图像数据。视觉监控系统中,运动平台包括旋转云台、车辆、无人机等。相比静止摄像机采集的视频数据,运动视频能覆盖更大的场景、跟踪拍摄感兴趣的目标。因此,对运动视频的自动处理与分析已成

为智能视觉监控领域的研究热点和前沿课题。

图像配准是运动视频分析的关键技术环节,其目的是恢复图像序列帧间的空间几何关系。考虑到载台运动、摄像机自身旋转、俯仰、缩放等因素,认为针对运动视频,理想的图像配准算法应具有解决下列问题的能力:

(1) 视场旋转 载台运动或摄像机转动造成的图像旋转变化;

基金项目: 国家自然科学基金重点项目 (60634030); 航空科学基金 (2007ZC53037); 高等学校博士学科点专项科研基金 (20060699032)

收稿日期: 2007-06-25 改回日期: 2007-12-07

第一作者简介: 李静 (1979~), 女, 西北工业大学自动化学院控制理论与控制工程专业博士研究生。研究方向为计算机视觉与模式识别、图像配准、智能视频监控。E-mail: jinglinwpu@163.com

- (2) 视场缩放 载台运动或摄像机镜头缩放带来的图像尺度变化;
- (3) 仿射变化 视角变化引起的图像仿射变形;
- (4) 光照变化 不受控环境中的亮度变化;
- (5) 噪声干扰 环境噪声、运动模糊等;
- (6) 实时性 实时性是决定算法是否能够实际用的关键。

尽管图像配准算法众多<sup>[1]</sup>,但早期的序列图像配准算法多假设同一像素在相邻帧间位置变化不大,当运动视频中帧间出现较大位置差异、几何变化时算法性能会降低。

近年来,在图像匹配和目标识别<sup>[2~4]</sup>领域取得巨大成功的不变特征(invariant feature)匹配方法得到学者的广泛关注。所谓“不变特征”是指在图像在几何变化、光照变化、噪声干扰时仍保持稳定性的特征。该类算法的关键步骤包括不变特征检测和特征描述两部分。

Lindeberg率先提出在归一化的高斯拉普拉斯(LoG)尺度空间,通过极值点检测定位出特征所在的位置和尺度<sup>[5]</sup>。为了降低计算量,Low e用双高斯差函数(difference of Gaussian, DoG)对 LoG 进行近似<sup>[2]</sup>。由于 LoG<sup>[5]</sup>和 DoG<sup>[2]</sup>都具有较强的边缘响应,M kolajczyk提出一种 Harris-Laplacian 特征检测算法,用 Harris 检测出图像中的角点,并在拉普拉斯尺度空间定位特征尺度<sup>[6]</sup>。上述检测器均具有良好的旋转、尺度不变性,对图像的仿射变形也具有一定的适应性。但是在构建尺度空间时需要进行高斯卷积,且输出特征点数不可控,不适于对运动视频序列的快速处理。

不变特征描述子主要包括基于分布的描述子、基于时频特性的描述子和基于差分不变量的描述子等<sup>[2,7~9]</sup>。M kolajczyk 针对光照变化、图像旋转、缩放、仿射变形、模糊、压缩等情况,对 11 种典型的描述子进行性能评估<sup>[10]</sup>,结果表明 SIFT 描述子<sup>[2]</sup>的性能最优。

本文针对视觉监视中的运动视频自动配准进行研究。考虑到配准中存在的图像旋转、缩放、光照变化、噪声干扰、实时性等问题,提出一种基于不变特征的快速图像配准算法(见图 1),包括特征检测、描述和匹配。在检测时,首先通过多层箱式滤波器模拟高斯卷积,构建图像多尺度空间,大幅度降低计算量,然后同时考虑质量与空间分布,高效地定位出数

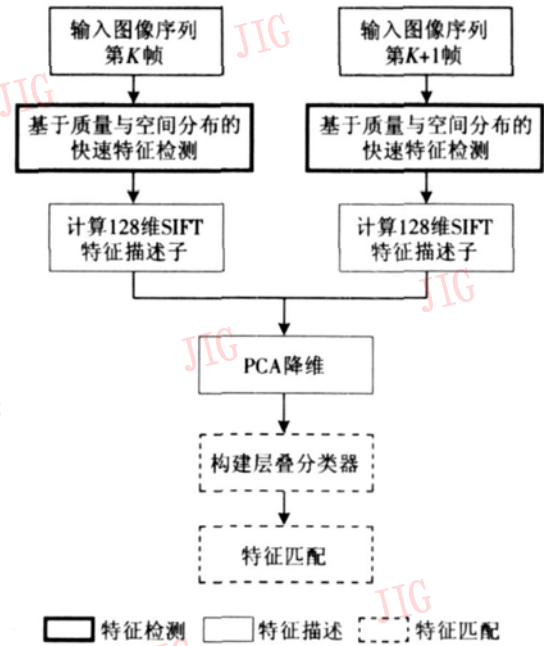


图 1 算法流程图

Fig 1 Flow chat of the algorithm

量可控的特征点;在特征描述时选用 SIFT 描述子。考虑到高维数据对后继匹配速度的影响,采用主成分分析法对 128 维 SIFT 描述子进行降维;在匹配时根据描述子主成分的差异设计层叠分类器,快速滤除大量无关特征点。

定量分析实验结果表明,检测算法在图像旋转、缩放、噪声干扰等条件下均具有良好的可重复性;与直接采用 SIFT 描述子相比,基于 PCA 降维和层叠分类器的描述与匹配算法能在保持匹配性能的同时,大幅度提高匹配速度,更适于对实时性要求较高的运动视频的分析处理。

以本算法为核心,构建了一个自动视频配准与拼接系统,对视频监视中的球形摄像机和无人机航拍获得的运动视频<sup>[11]</sup>的实验表明,算法具有精度高、鲁棒性强等优点。

## 2 基于质量与空间分布的特征点检测

### 2.1 基于多层箱式滤波的尺度空间构建

考虑到摄像机的缩放,运动视频中图像间可能存在较大尺度变化,在特征检测时需要先构建图像的多尺度空间。文献[2]、[5]、[6]在建立尺度空间时都采用高斯核卷积,随着滤波窗宽的增大,计算量会急剧增加。Grabner 提出用窗口内像素灰度的均值来代替高斯卷积<sup>[12]</sup>。但是, Lindeberg 证明高斯卷积核是实现尺度变换的唯一线性变换核<sup>[13]</sup>,而

均值滤波得到的尺度空间与真实值存在较大偏差。基于此, 本文用多层箱式滤波器 (multiple box filter) 模拟高斯卷积, 并通过积分图使得计算量和滤波窗宽大小无关。

假设用高斯函数  $G(x, y, \sigma)$  对输入图像  $I$  进行卷积, 构建尺度空间

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

式中,  $x, y$  为像素点位置,  $\sigma$  是高斯函数的标准差。设计一个  $K$  层阶梯状函数  $MB(x, y, \sigma)$  来模拟高斯  $G(x, y, \sigma)$ , 每一层的权值系数  $P(k)$  由该层覆盖下的环形区域  $C_k$  中高斯核系数的均值得到:

$$P(k) = \frac{1}{Z_k} \sum_{(x, y) \in C_k} G(x, y, \sigma) \quad (2)$$

式中,  $Z_k$  为环形区域  $C_k$  的面积。则尺度空间  $L$  可由下式构建:

$$L(x, y, \sigma) = MB(x, y, \sigma) * I(x, y) = \sum_{k=1}^K \left| P(k) \sum_{(x, y) \in C_k} I(x, y) \right| \quad (3)$$

考虑到  $C_k$  为环形区域, 式 (3) 求和部分不能直接用积分图计算, 因此用两个矩形区域的差求解  $C_k$  内的像素和:

$$\sum_{(x, y) \in C_k} I(x, y) = \sum_{(x, y) \in (C_k, C_{k+1}, \dots, C_K)} I(x, y) -$$

$$\sum_{(x, y) \in (C_{k+1}, \dots, C_K)} I(x, y) \quad (4)$$

将式 (4) 代入式 (3) 并推导得:

$$L(x, y, \sigma) = \sum_{k=1}^K (P(k) - P(k-1)) \sum_{(x, y) \in (C_k, \dots, C_K)} I(x, y) \quad (5)$$

式 (5) 表示尺度空间  $L$  可以由多个不同窗宽的箱式滤波组合得到, 滤波器可用积分图快速计算。系数  $P(k)$  可离线求解, 且有  $P(0) = 0$ 。本文实验均采用 3 层箱式滤波器构建尺度空间, 其滤波窗宽半径分别为  $1\sigma, 2\sigma, 3\sigma$ 。表 1 给出高斯滤波器、均值滤波<sup>[12]</sup>以及本文算法在一个像素点上的计算量比较。可以看出, 本文算法和均值滤波的计算量与窗宽大小无关。图 2 给出了均值滤波和本文算法对高斯滤

表 1 滤波算法计算量比较

Tab 1 Comparison of computational time

滤波方法	加法	乘法
2维高斯滤波	$N^2 - 1$	$N^2$
1维高斯滤波	$2N - 2$	$2N$
均值滤波 <sup>[12]</sup>	3	1
本文算法	$2 + 9$	3

注:  $N$  为滤波器窗宽

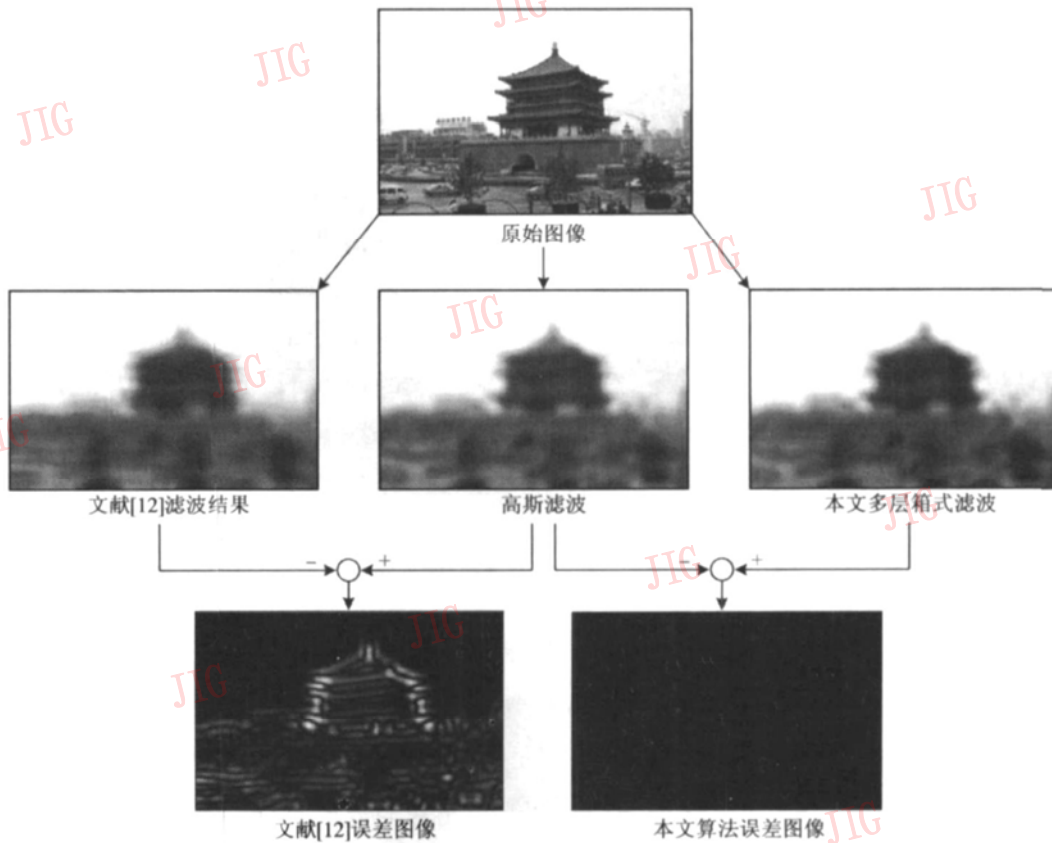


图 2 对高斯滤波模拟精度的比较

Fig 2 Comparison of filtering accuracy

波模拟精度的比较。相比文献 [ 12], 多层箱式滤波能更精确地模拟高斯滤波 (图 2 差异图像), 大幅度降低构建尺度空间的误差。

本文采用多阶多层的思路构造图像多尺度空间  $L$ 。将尺度空间分为  $O$  阶, 每增加一阶, 图像按 2 做降采样,  $O$  由图像尺寸  $(W, H)$  和降采样率 2 求解:

$$O = \log_2(m \text{ in}(W, H)) \quad (6)$$

每阶内将尺度空间细分为  $S$  层。第  $i$  阶第  $j$  层图像由第  $i$  阶第 1 层的图像经多层箱式滤波得到, 滤波器参数由所模拟的高斯函数离线求解。令  $\sigma^{(0)}$  为第 1 层高斯滤波的标准差, 则相应多层箱式滤波器的窗宽  $a^{(0)} = 6 \cdot \sigma^{(0)} + 1$  第  $j$  层图像的滤波窗宽为  $a^{(0)} \cdot 2^{jS}$ 。实验中取  $a^{(0)} = 1.1, S = 5$ 。

### 2.2 基于质量与空间分布的特征点检测

图像多尺度空间建立好以后, 下一步就是如何寻找特征点。目前, 大多数算法<sup>[2,5,6]</sup>在检测特征点时需要遍历整个尺度空间中的所有采样点, 同时根据极值、角点响应等特性进一步确定该点是否为特征点。但这种依次搜索的检测方法存在以下不足:

(1) 该方法将大部分时间消耗在非特征点的判断上, 且搜索时间随图像大小呈非线性上升;

(2) 由于本文采用仿射模型估计运动视频中帧间的几何关系, 只需要较少的匹配点即可精确估计模型参数, 而上述方法输出的特征点数量均无法精确控制。

为此, 提出一种基于质量和空间分布的搜索算法, 从根本上解决上述问题, 而且, 该算法也可以集成到现有的特征点检测算法中。搜索算法包括以下 3 个主要步骤:

#### (1) 构建质量空间

采用相邻尺度图像差的绝对值来构建质量空间, 并用局部极值和 Harris 角点响应判断是否为特征点。由相邻尺度的两个图像相减绝对值得到 DoMB (difference-of-multiple box) 图像:

$$I_{\text{DoMB}}(x, y, \sigma) = |L(x, y, k\sigma) - L(x, y, \sigma)| \quad (7)$$

显然, 某一点的 DoMB 取值越高, 其为局部极值点的概率越大。当给定需要检测的特征点数目后, 如果能先对 DoMB 进行排序, 然后按照取值从高到低搜索特征点, 则能够回避在大量低概率点的搜索, 快速、准确地检测到所需数量的高质量的特征点。由于整个尺度空间中像素点数量巨大, 难以快速对 DoMB 排序。为此, 在生成 DoMB 图像的同时, 根据

每一个像素点的 DoMB 值, 同步建立图像的质量直方图  $Q(n)$  和索引表  $QI$ 。对质量直方图中取值为  $n$  的点, 索引表  $QI$  记录了该点所在的尺度和空间位置。实验中  $n$  的取值范围为  $n \in [0, 60]$ 。

图 3 给出了图 2 原始图像尺度空间的质量直方图。从图中可以看出, 尺度空间中绝大部分像素点都集中在 DoMB 取值较低的部分。这样, 算法根据质量直方图由高到低依次搜索, 仅需要在最可能的点上进行少量的计算即可检测出所需数目的特征点, 大幅度减少了在非特征点上消耗的时间, 从而从根本上加快了检测特征点的速度。大量实验结果表明, 这种方法通常只需要搜索全部采样点的 2% 左右即可以找到足够的特征点。

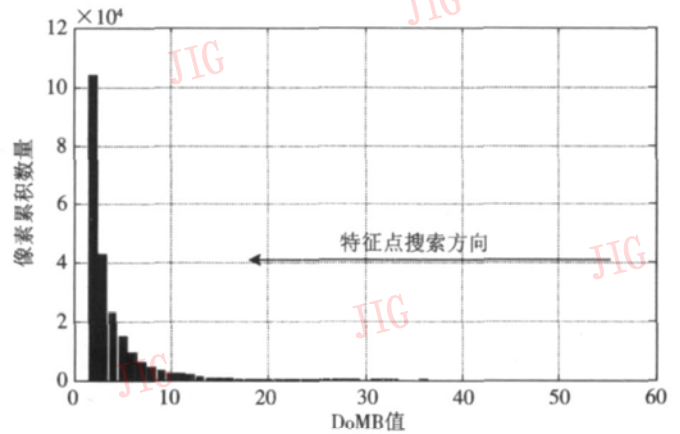


图 3 DoMB 的质量直方图  
Fig 3 Histogram of DoMB

#### (2) 特征点定位

对待检测的采样点, 通过计算其局部极值和 Harris 角点响应进行特征点判断。局部极值检测<sup>[5]</sup>由 Lindeberg 提出, 将采样点分别与上下相邻尺度对应位置的周围 18 个像素以及相同尺度的周围 8 个像素, 共 26 个像素比较, 确保检测到的特征点同时是尺度空间和 2 维图像空间上的局部极值。同时, 为了后继匹配的稳定性, 根据采样点局部自相关矩阵的特征值计算 Harris 角点响应, 去除不稳定边缘点。

#### (3) 基于动态决策图的非极大值抑制

上述基于质量空间搜索的方法具有计算量小, 可以精确控制输出点数的优点, 但是仅按照质量空间搜索的方式可能导致大量特征点集中在图像的某些纹理丰富的区域, 从而降低了特征点的空间分布性。由于载台机动, 运动视频中帧间可能出现仅有小部分重叠区域的情况, 且重叠区域的位置无法事

先得到。为了保证在该情况下仍能对图像正确匹配, 需要检测到的特征点尽可能分布在图像的各个区域。为此, 构造一个和尺度空间等大、初值为 0 的动态决策图  $D(x, y, \sigma)$ 。在按照质量空间从高到低搜索时, 先判断动态决策图中当前点的状态, 如果为 1 则表示当前点的局部邻域内已经存在检测到的特征点, 并终止对该点的进一步判断; 如果为 0 则继续判断该点局部极值和角点响应, 并将检测到的特征点在动态决策图中的邻域像素状态设为 1。其中, 关闭邻域的半径  $\rho$  由特征点所在的尺度图像大小确定, 且有  $\rho = \alpha \times \min(w_s, h_s)$ , 式中  $w_s, h_s$  为特征点所在尺度图像的宽和高, 比例系数  $\alpha$  用于控制关闭范围,  $\alpha$  越大, 特征点的空间分布性越强。这里实验中  $\alpha$  取值为 0.03。

通过结合质量空间和动态决策图, 本文算法可以快速检测到具有较高质量、良好的空间分布、数量精确可控的特征点。

### 3 基于 PCA 和层叠分类器的特征匹配

特征点检测结束后, 下一步就是如何描述和匹配特征点。Mikołajczyk 证明了 SIFT 描述子<sup>[2]</sup>性能最优<sup>[10]</sup>。但由于维数较高, 用 128 维 SIFT 描述子做特征匹配时运算量大, 难以快速处理。为此, 首先用主成分分析法对 128 维 SIFT 描述子降维, 然后根据主成分的误差分布设计层叠分类器, 在搜索匹配时快速滤除大量无关特征点。

#### (1) 构造 SIFT 描述子

为了保证描述子的旋转不变性, 首先要计算特征点的主方向。文献[2]中主方向是指特征点邻域内梯度方向直方图中最大值所对应的方向。此外, 为了保证匹配的稳定性, Lowe 将梯度强度高于最大值 80% 的方向作为辅方向<sup>[2]</sup>, 这样一个特征点可能有多个不同方向的描述子。

构造描述子时, 对任意一个特征点, 以主方向或辅方向为参照, 首先在其所在尺度空间中取出以该点为中心的  $16 \times 16$  大小的邻域, 并将该邻域均分为  $4 \times 4$  个子区域, 每个子区域包含  $4 \times 4$  个像素。然后计算每个子区域的 8 方向梯度方向直方图, 并根据位置将直方图依次排序, 构成一个  $4 \times 4 \times 8 = 128$  维的向量, 即 SIFT 描述子。为了增强描述子对特征点定位误差的鲁棒性, 在匹配前用高斯核对 128 维 SIFT 描述子进行平滑。

#### (2) 基于主成分分析的描述子降维

主成分分析的目标是在误差平方和意义下用低维子空间表示高维数据。用主成分分析法对 128 维 SIFT 描述子降维的过程如下: 首先将两幅待匹配图像中所有  $n$  个特征点的 SIFT 描述子  $x_1, x_2, \dots, x_n$  作为样本, 计算出 128 维均值向量  $\bar{\mu}$  和  $128 \times 128$  的协方差矩阵  $R$ 。

$$\bar{\mu} = [\bar{\mu}_1, \dots, \bar{\mu}_p], \bar{\mu}_i = \frac{1}{n} \sum_{k=1}^n x_k(i) \quad (8)$$

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \dots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix} \quad (9)$$

$$r_{ij} = \frac{\sum_{k=1}^n (x_k(i) - \bar{\mu}_i)(x_k(j) - \bar{\mu}_j)}{\sqrt{\sum_{k=1}^n (x_k(i) - \bar{\mu}_i)^2 \sum_{k=1}^n (x_k(j) - \bar{\mu}_j)^2}} \quad (10)$$

其中, 维数  $p = 128$ ,  $x_k(i)$  表示第  $k$  个 SIFT 描述子的第  $i$  维分量。

然后计算协方差矩阵  $R$  的  $p$  个特征值  $\lambda$  和特征向量  $e$  并将特征值按从大到小顺序排列, 则有  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  和对应的特征向量  $\{e_1, e_2, \dots, e_p\}$ 。选出对应最大  $d$  个特征值的特征向量作为主成分方向,  $d$  的大小由下式给出:

$$v(i) = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k}, i = 1, 2, \dots, p \quad (11)$$

$$d = \arg \min_i (v(i) > T) \quad (12)$$

式中,  $v(i)$  为累计贡献率, 本文实验中累积贡献率的阈值  $T = 0.86$ 。最后构造一个  $p \times d$  的矩阵  $A$ , 它的列由  $d$  个特征向量组成。将原始 128 维 SIFT 描述子按照式 (13) 投影到这个  $d$  维子空间, 得到描述子的主成分表示  $y_1, y_2, \dots, y_n$ 。

$$y_k = x_k \cdot A \quad (13)$$

#### (3) 基于层叠分类器的快速匹配

在匹配时, 对于第 1 幅图像的某一特征点, 需要遍历另一幅图像的所有特征点, 计算每个可能匹配点对描述子的欧式距离。为了快速滤除无关特征点, 进一步提高匹配速度, 用描述子主成分设计一个由  $d$  个判决条件构成的层叠分类器。

为构建层叠分类器, 将两幅图像正确匹配点的描述子主分量相减, 得到匹配点主成分误差图。通

过对不同场景、光照变化、图像缩放、旋转等条件下的大量测试图像的统计结果表明, 匹配点主成分的误差分布近似于一个高斯分布, 且高斯分布的标准差  $\sigma$  取值基本保持恒定, 不随测试图像改变, 有  $3\sigma = 0.15$ 。图 4 (c) 和图 4 (d) 给出一组匹配点主成分

误差曲线和分布的示意图。基于误差分布特性, 在计算欧式距离前, 首先逐一判断该特征点对的主成分差异  $\Delta PC_i$  是否高于阈值  $T_{pc}$  (如图 4 (e) 所示), 为了保证不遗漏真实匹配点, 在实验中  $T_{pc} = 0.25$  略高于  $3\sigma$ 。

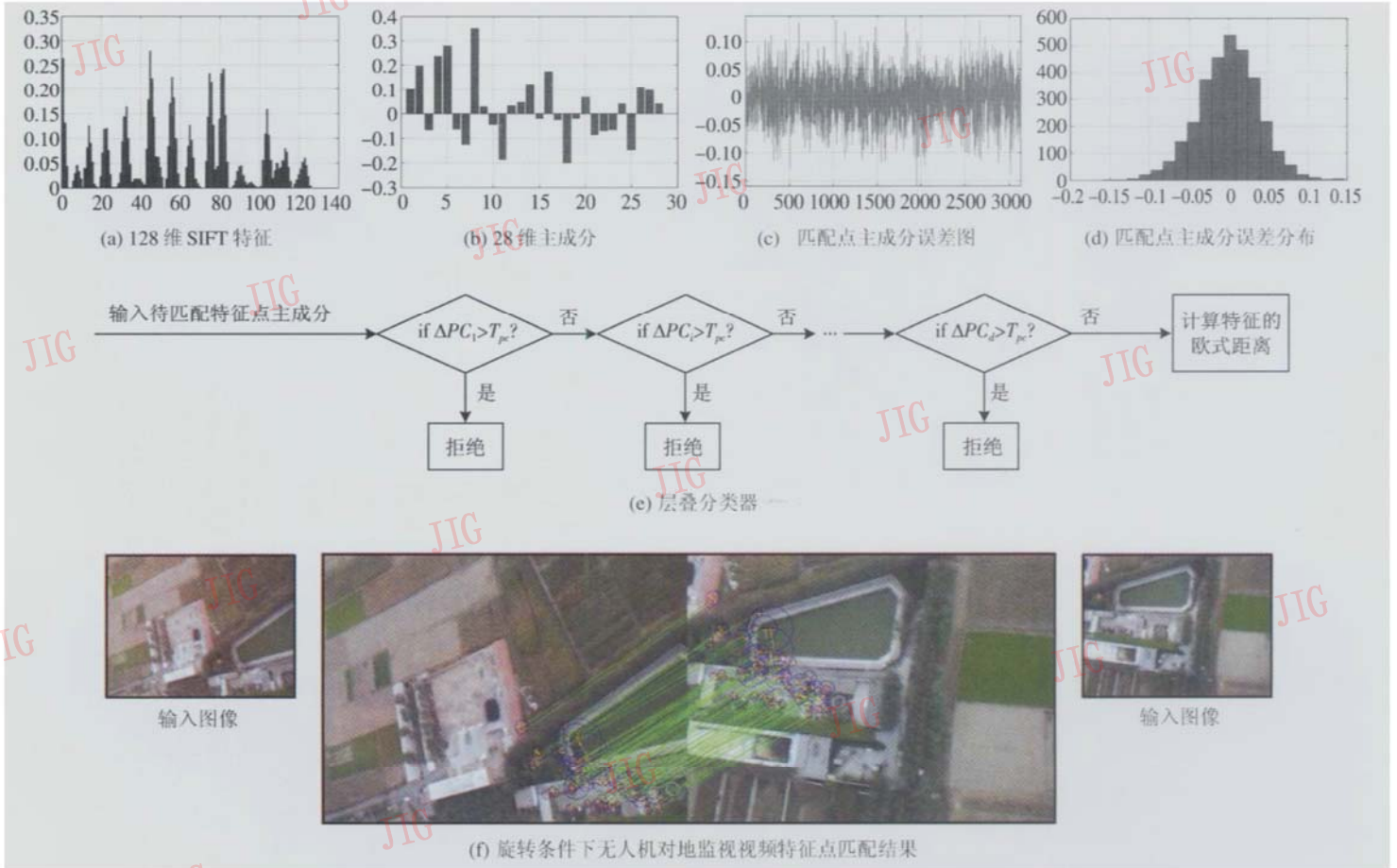


图 4 基于主成分分析和层叠分类器的特征点描述与匹配结果  
Fig 4 PCA and cascade filter based feature description and matching result

在得到可能特征点对的欧式距离后, Lowe 为了保证描述子的独特性, 用所有距离中的最小值和次小值之比做度量, 如果比率小于一定阈值  $T_E$ , 则将取得最小距离的点作为匹配点, 通常  $T_E$  的取值范围为  $0.4 \sim 0.6$ <sup>[2]</sup>。由于一个特征点可以有多个辅方向, 可能得到多个不同的描述子, 因此本文在匹配时对 Lowe 的算法<sup>[2]</sup>进行改进, 要求距离最小值和次小值应对应于不同的特征点。

图 4 给出一个基于主成分分析和层叠分类器的特征点描述与匹配结果。测试图像为自建数据库中无人机对地监视视频, 由于载机机动, 图 4 (f) 左右两幅输入图像存在较大的位置偏差和角度旋转。采用本文第 2 节提出的基于质量与空间分布的快速特征点检测算法, 通过控制输出特征点数, 在两幅输入

图像上分别检测出 1 000 个特征点。用这 2 000 个特征点作样本, 经过主成分分析, 将原始的 128 维 SIFT 描述子 (图 4 (a)) 降为 28 维 (图 4 (b))。直接匹配的方式需要比较 1 000 000 次, 而本文提出的基于主成分差异的层叠分类器 (图 4 (e)) 拒绝了 989 604 次 (占总比较次数的 98.96%) 的无关比较, 最终找到 112 对正确匹配点。图 4 (f) 中用蓝色圆圈标明特征点所在位置和尺度, 从圆心出发的线段标明主方向。

### 4 实验分析

首先对算法性能进行定量分析, 然后给出算法在视频监视系统中的两类典型应用。

(1) 特征检测算法可重复性分析

可重复率 (repeatability rate)是指算法在图像发生几何变换、光照变化以及噪声干扰前后独立检测到相同特征点的比率<sup>[14]</sup>,它的好坏将直接影响后继匹配的性能,是衡量检测算法性能的重要指标。实验采用文献[6]的方法计算可重复率

$$r_r = \frac{C(m_1, m_2)}{mean(m_1, m_2)} \quad (14)$$

假设图像间的变换矩阵  $H$  已知,检测算法在两幅待匹配图像上分别检测出  $m_1$  和  $m_2$  个特征点。式中  $C(m_1, m_2)$ 表示将  $m_1$  个特征点经过变换矩阵  $H$  映射到第 2 幅图像后,找到的可重复特征点总数。这里,判断为相同特征点的标准是位置误差不超过 2 个像素,尺度误差率在 20% 以内。

图 5 给出了图像在旋转、尺度变化和加噪条件下的 4 组可重复性分析实验结果。

图 5(a)中原始图像以  $\pi/16$  为间隔逆时针旋转一周得到 32 幅图像。其可重复率  $r_r$  在图 5(e) 给出。可以看到  $r_r$  在 4 个象限中呈周期性变化,当旋转角度为  $45^\circ$  倍数时降到最低,但仍然达到 0.7 以上,表明本文检测算法具有良好的旋转不变性。

图 5(b)中尺度变化下的图像序列。图 5(f)给出该序列的可重复率。实验结果看出,当缩放比率为 2 时,可重复率仍保持在 0.7 以上。文献 [6] 中给出的 Harris-Laplacian<sup>[6]</sup>和 DoG<sup>[2]</sup>在相同缩放比率下的可重复率分别为 0.67 和 0.5 而 Harris 算法在该尺度变化下仅具有 0.16 的可重复率。当缩放比率进一步增大到 4 时,本文算法的可重复率为 0.3 Harris-Laplacian<sup>[6]</sup>, DoG<sup>[2]</sup> 和 Harris 分别为 0.4, 0.32 和 0.08。可以看出,本文算法在尺度变化下的可重复率同 Harris-Laplacian<sup>[6]</sup>和 Lowe<sup>[2]</sup>性能相当,具有良好的尺度不变性,可以满足运动视频中待配准图像可能出现的尺度变化。

图 5(c)和图 5(d)分别给出图像在椒盐噪声和高斯噪声下的一系列变换图像。图 5(c)中椒盐噪声密度为 0.02~0.2 图 5(d)中高斯噪声的方差从 2 增加到 20 从图 5(g)和图 5(h)可以看出,在本文模拟的噪声强度范围内,可重复性始终高于 0.5 表明算法对噪声也具有一定的稳定性。

(2) 特征描述与匹配算法性能分析

实验选择了 10 组典型图像 (图 6(a)),对基于 PCA 降维和层叠分类器的匹配算法性能进行定量

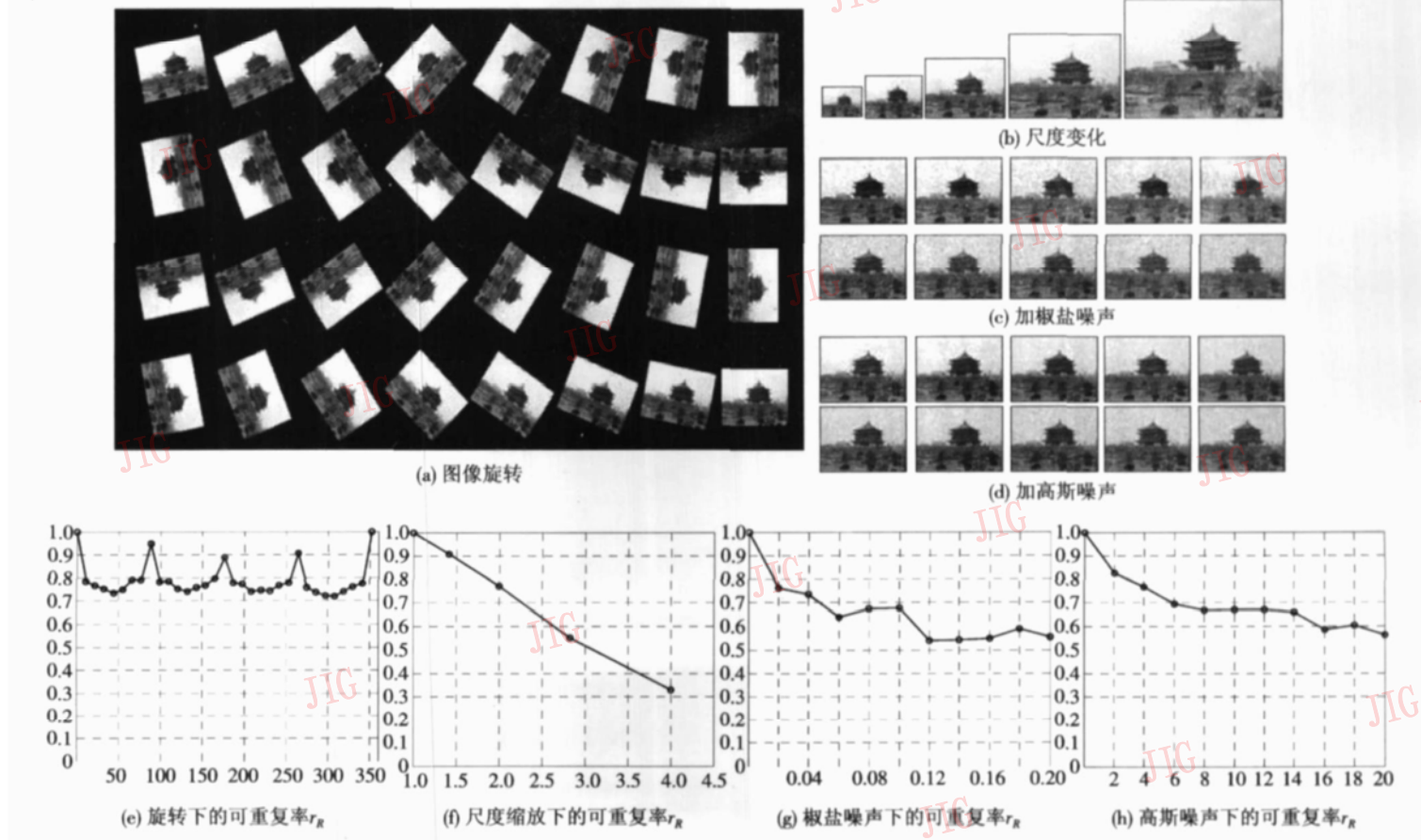


图 5 旋转、缩放、噪声条件下的特征检测算法可重复性

Fig 5 Repeatability of feature detection under rotation, scaling and noises



图 6 不变特征匹配

Fig 6 Invariant feature matching result

分析。测试图像的拍摄平台包括旋转云台 (图 6(a), No 1~ No 4) 和无人机 (图 6(a), No 5~ No 8), 图像间的变化涵盖了旋转 (图 6(a), No 1, No 2, No 6, No 8, No 9) 尺度缩放 (图 6(a), No 1~ No 4, No 6, No 8, No 9), 亮度变化 (图 6(a), No 7), 视角变化 (图 6(a), No 9, No 10) 等情况。

图 6(b) 给出部分匹配结果, 包括旋转云台 (图 6(b) 第 1 行) 和无人机 (图 6(b) 第 2 行) 图像, 图中蓝色圆圈标明特征点所在位置和尺度, 从圆心出发

的黄色线段标明特征主方向。表 2 从描述子维数、特征搜索数量和正确匹配点数 3 个方面给出 PCA 降维和层叠分类器滤波前后的性能比较。从表 2 第 2 列可以看出, 主成分分析使描述子维数从 128 维降至 30 维以下, 平均维数仅相当于降维前的 21.72%。与直接搜索特征点相比, 通过比较两个待匹配点的主成分差异, 基于层叠分类器的滤波算法快速滤除了大量无关特征点, 对 10 组测试图像的平均特征点搜索匹配次数仅为直接搜索的 5.63%

表 2 PCA 降维和层叠分类器滤波前后特征匹配性能比较

Tab 2 Comparison of matching performance with PCA and cascade filtering

图像编号	描述子维数比较			特征点搜索匹配次数比较			正确匹配点数比较		
	SIFT	本文算法	相对比率 (%)	直接搜索	本文算法	相对比率 (%)	SIFT	本文算法	相对比率 (%)
No 1	128	28	21.88	1 452 282	62 065	4.27	78	67	85.90
No 2	128	29	22.66	1 294 164	58 233	4.50	13	12	92.31
No 3	128	28	21.88	1 356 957	73 979	5.45	19	13	68.42
No 4	128	28	21.88	874 752	42 784	4.89	22	21	95.45
No 5	128	27	21.09	840 465	38 533	4.58	89	85	95.51
No 6	128	27	21.09	1 024 985	48 187	4.70	136	139	102.21
No 7	128	27	21.09	1 347 306	73 844	5.48	194	184	94.85
No 8	128	26	20.31	865 242	86 235	9.97	19	25	131.58
No 9	128	29	22.66	4 395 240	284 924	6.48	260	251	96.54
No 10	128	29	22.66	3 577 000	215 270	6.02	21	22	104.76
比率均值			21.72			5.63			96.75

(表 2 第 3 列末尾)。

正确匹配点数是衡量匹配算法性能的重要指标, 从表 2 第 4 列可以看出, 本文算法与直接采用 128 维 SIFT 描述子得到的正确匹配点数基本相当, 为后者的 96.75% (表 2 第 4 列末尾), 对部分测试图像, 本文算法的匹配性能略高于 SIFT (表 2 No 6、No 8、No 10)。

上述结果表明, PCA 降维和层叠分类器能在保持 SIFT 描述子匹配性能的同时大幅度提高匹配速度, 更适于对处理速度要求较高的运动视频分析。

### (3) 景深变化下的图像匹配

图 7 给出算法对不同景深图像的匹配结果。图 7 (a) 中, 由于拍摄目标离镜头比较近, 摄像机转动时, 画面中景深出现较大变化 (图 7 (a), #6), 这些变化会使目标图像发生仿射、透视变化甚至非线性形变。图 7 (b) 和图 7 (c) 分别给出本算法对 #1、#4 和 #1、#6 的匹配结果。从结果中可以看出, 随着景深的逐渐增大, 待匹配图像间发生较大形变, 此时算法的匹配性能会下降。在进一步的研究中拟引入具有仿射或透视不变性的检测与描述方法, 提高算法的稳定性。

### (4) 运动视频自动镶嵌

基于不变特征的图像镶嵌<sup>[15]</sup>能获得大范围、高

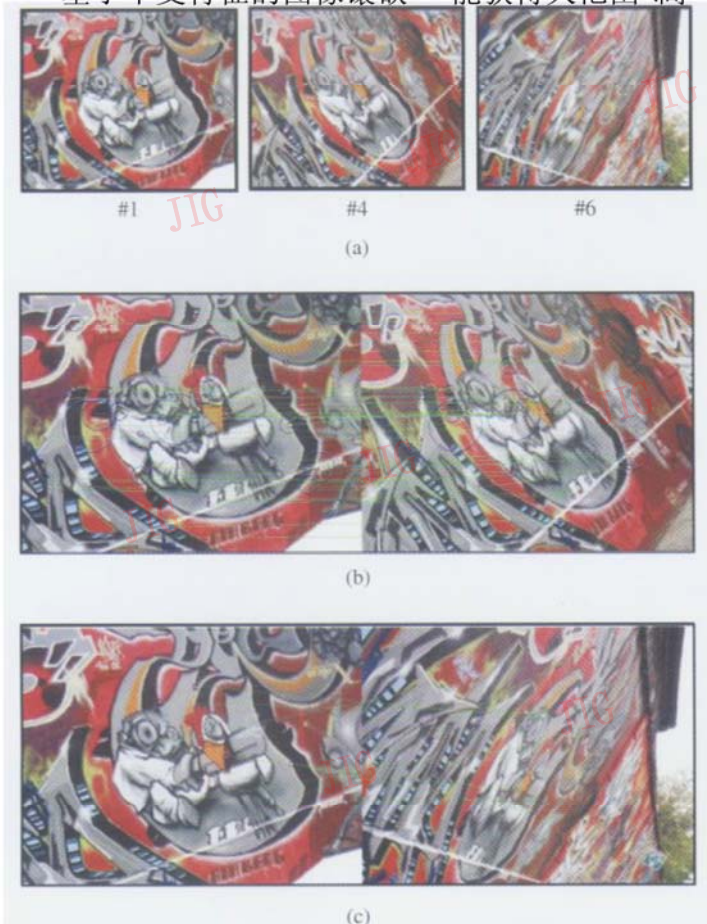


图 7 景深变化下的匹配结果

Fig 7 Matching result under large scene depth changes

分辨率的全景图, 在视频监控中具有广泛应用。以本文算法为核心, 构建了一个自动视频拼接系统。在得到匹配特征点后, 用 RANSAC (random sample consensus)<sup>[16]</sup>去除外点, 并用最小二乘法估计帧间仿射变换参数。

图 8 给出了本文算法对视频监控中的一组运动摄像机视场拼接和定位结果。在现有监控系统中, 当操作人员控制球形摄像机跟踪某个感兴趣目标时, 往往难以定位出当前摄像机视场在整个大范围场景中的位置, 而该参数对于获取目标在整个场景的运动轨迹, 分析操作人员的感兴趣区域等都具有重要意义。图 8 第 1 列为输入视频, 第 2 列中背景为拼接得到的监控场景全景图。可以看出, 通过帧间的自动配准, 本文算法能够在摄像机缩放 (#154、#183)、转动 (#230) 等条件下准确定位出当前视场在全景图中的位置 (白色线框标明区域)。



图 8 运动摄像机自动视场定位

Fig 8 Automatic field of view location for moving camera

图 9 给出了无人机航拍运动视频的自动配准和拼接结果。测试数据来自于美国 DARPA 发布的

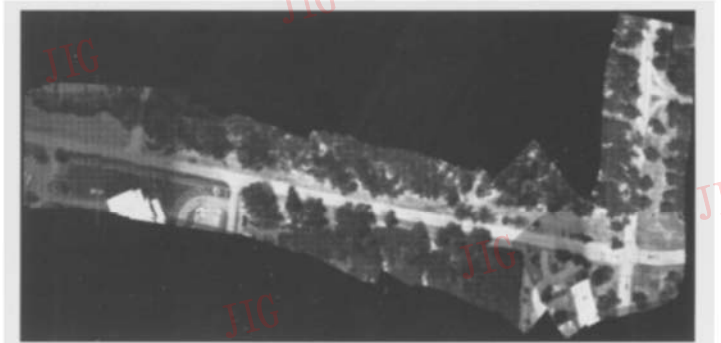


图 9 无人机航拍视频自动拼接结果

Fig 9 Automatic UAV surveillance video mosaicing result

VIVID 数据库<sup>[11]</sup>。实验结果表明,该系统可在无人机转弯机动、光照突变和盘旋等状态下稳定、准确地进行运动视频自动配准和拼接,取得令人满意的效果。

## 5 结 论

提出了一种新的基于不变特征的运动视频自动配准算法,包括基于质量和空间分布的特征点检测、基于主成分分析的描述算法和基于层叠分类特征点匹配算法 3 部分。实验中,对算法性能的定量分析结果表明:(1)检测算法在图像发生旋转、尺度缩放、噪声条件下均具有良好的可重复性,且在保证特征点质量、空间分布的同时,高效检测出数量可控的特征点;(2)与直接采用 128 维 SIFT 描述子的匹配结果相比,本文算法引入 PCA 降维和层叠分类器后,不仅大幅度降低了计算量,而且仍保持了与之相同的匹配性能。

以本文算法为核心,构建了一个自动视频拼接系统,对视频监控中的球形摄像机和无人机航拍获得的运动视频进行了自动配准和拼接。实验结果表明,该算法可以解决运动视频中出现的角度旋转、尺度缩放、光照突变等问题,具有良好的鲁棒性,为后继基于运动载台的运动目标检测、跟踪、识别等处理提供了坚实的基础。

## 参考文献 (References)

- Zitova B, Flusser J. Image registration methods: a survey[J]. *Image and Vision Computing*, 2003, **21**(11): 977~1000
- Lowé D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, **60**(2): 91~110
- Ferrari V, Tuytelaars T, Gool L. Simultaneous object recognition and segmentation from single or multiple model views[J]. *International Journal of Computer Vision*, 2006, **67**(2): 159~188
- Zhang W, Košecák J. Hierarchical building recognition[J]. *Image and Vision Computing*, 2007, **25**(5): 704~716
- Lindeberg T. Feature detection with automatic scale selection[J]. *International Journal of Computer Vision*, 1998, **30**(2): 77~116
- Mikolajczyk K, Schmid C. Indexing based on scale invariant interest points[A]. In: *Proceedings of IEEE International Conference on Computer Vision [C]*, Vancouver, BC, Canada, 2001, **1**: 525~531.
- Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(4): 509~522
- Ke Y, Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors [A]. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition [C]*, Washington, DC, USA, 2004, **2**: 504~513
- Bay H, Tuytelaars T, Gool L V. SURF: speeded up robust features [A]. In: *Proceedings of the 9th European Conference on Computer Vision [C]*, Graz, Austria, Springer LNCS, 2006, **3951**: 404~417.
- Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10): 1615~1630
- Video Verification of Identity, CMU [DB/OL], <http://www.cs.cmu.edu/~hebert/vivid.html>
- Grabner M, Grabner H, Bischof H. Fast approximated SIFT [A]. In: *Proceedings of 7th Asian Conference on Computer Vision [C]*, Hyderabad, India, 2006, 918~927.
- Lindeberg T. Scale-space theory: a basic tool for analyzing structures at different scales[J]. *Journal of Applied Statistics*, 1994, **21**(2): 224~270
- Schmid C, Mohr R, Bauckhage C. Evaluation of interest point detectors[J]. *International Journal of Computer Vision*, 2000, **37**(2): 151~172
- Brown M, Lowe D G. Automatic panoramic image stitching using invariant features [J]. *International Journal of Computer Vision*, 2007, **74**(1): 59~73
- Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, **24**(6): 381~395