

基于场景复杂度与不变特征的 航拍视频实时配准算法

杨 涛, 张艳宁, 张秀伟, 张新功

(西北工业大学计算机学院, 陕西省语音与图像信息处理重点实验室, 陕西西安 710129)

摘要: 实时、鲁棒的图像配准是航拍视频电子稳像、全景图拼接和地面运动目标自动检测与跟踪的前提和关键技术. 本文以航拍视频序列为处理对象, 提出了一种新的基于场景复杂度与不变特征的实时配准算法, 其主要特点包括: (1) 在对航拍视频配准难点进行详细分析的基础上, 有针对性的提出基于积分图的快速图像尺度空间构建、依据场景复杂度的检测特征点数量在线精确控制、基于描述子误差分布统计特性级的联分类器构造等新方法, 使得算法配准性能不随场景的复杂度发生改变, 能够在各种地貌条件下实时、稳定的进行图像配准; (2) 将多尺度 Harris 角点和 SIFT 描述子相结合, 并通过对帧间变换模型参数进行鲁棒估计, 保证了算法具有良好的旋转、尺度、亮度不变性和配准精度. 实验结果表明, 算法可在场景变化、图像大幅度平移、尺度缩放和任意角度旋转等复杂条件下实时、精确的进行图像配准, 对分辨率为 320×240 的航拍序列的平均处理速度达到 20.7 帧/秒.

关键词: 航拍视频配准; 不变特征检测与匹配; 视频监控

中图分类号: TP391 文献标识码: A 文章编号: 0372-2112 (2010) 05-1069-09

Scene Complexity and Invariant Feature based Real Time Aerial Video Registration Algorithm

YANG Tao, ZHANG Yanning, ZHANG Xiurwei, ZHANG Xingong
(School of Computer Science, Northwestern Polytechnical University, Shaanxi Provincial Key
Laboratory of Speech and Image Information Processing, Xi'an, Shaanxi 710129, China)

Abstract: Real time and robust image registration is the premise and key technology of aerial video stabilization, panorama stitching and ground moving target detection and tracking. This paper presents a novel scene complexity and invariant feature based aerial video registration algorithm. The main characteristics of the proposed method include: (1) Based on analyzing the key difficulties and challenges of aerial video registration, several new methods are presented to realize fast and effective video registration under various real scenes, include integral image based fast image scale space generation, scene complexity based feature number controlling, and statistical error distribution of correspond features based cascade filtering. (2) Through combining the multi scale Harris corner detection, SIFT (Scale Invariant Feature Transform) feature description, and the RANSAC (Random Sample Consensus) based frame geometry transformation parameters estimation, the proposed algorithm achieves satisfied rotation, scaling, brightness invariance and accuracy of registration. Experiment results show that the proposed algorithm carries out real time and precise image registration under complex conditions with change of scene, large image translation, scaling and arbitrary rotation, and the average processing speed for a resolution of 320×240 unmanned aerial video sequences achieves 20.7fps.

Key words: aerial video registration; invariant feature detection and matching; video surveillance

1 引言

随着传感器成像能力的迅速提高和小型化, 航空机载成像设备已经成为无人机对地观测的主要手段, 并在战场侦察、公共安全、灾害分析、环境监测等领域发挥着

重要作用. 目前, 对航拍视频内容进行自动分析和处理已经成为计算机视觉、模式识别和智能视觉监控的研究热点和前沿^[1,2].

实时、鲁棒的图像配准是航拍视频电子稳像、全景图拼接和地面运动目标自动检测与跟踪的前提和关键

收稿日期: 2009-05-16; 修回日期: 2009-10-08

基金项目: 国家自然科学基金(No. 60903126, No. 60872145); 中国博士后基金(No. 20090451397); 国家 863 高技术研究发展计划(No. 2009AA01Z315); 高等学校科技创新工程重大项目培育资金项目(No. 708085)

技术,其目的是精确估计序列帧间的几何变换关系.考虑到载台运动、摄像机平移、旋转、缩放等因素,航拍序列的图像配准需要解决下列问题:

(1) 旋转变换: 载机转弯、光电吊舱转动造成的图像旋转变换.

(2) 尺度变化: 载机飞行高度差异、传感器观测镜头缩放带来的图像尺度变化.

(3) 亮度变化: 光照与大气状况变化,以及摄像机自动白平衡,使得航拍序列中同一地物的灰度发生较大畸变.

(4) 视频抖动: 由于无人机机身重量较轻,其运动特性比较复杂,在飞行过程中容易受到外界环境因素

影响,采集到的视频序列中往往出现抖动;为了从空中一定高度获取地面的高分辨率影像,航空机载成像设备往往采用长焦距镜头,这使得成像平台的微小运动将引起图像序列帧间的较大偏移,严重降低画面的可视性.

(5) 场景变化: 一方面,纹理丰富的复杂场景(如建筑物密集的城区)中往往含有数量众多的图像特征,而过多的特征会影响算法的实时性;另一方面,纹理单一的场景(如田地)却因为缺乏足够特征,引起配准性能下降.

(6) 实时性: 算法的处理速度是决定其能否实际应用的关键.

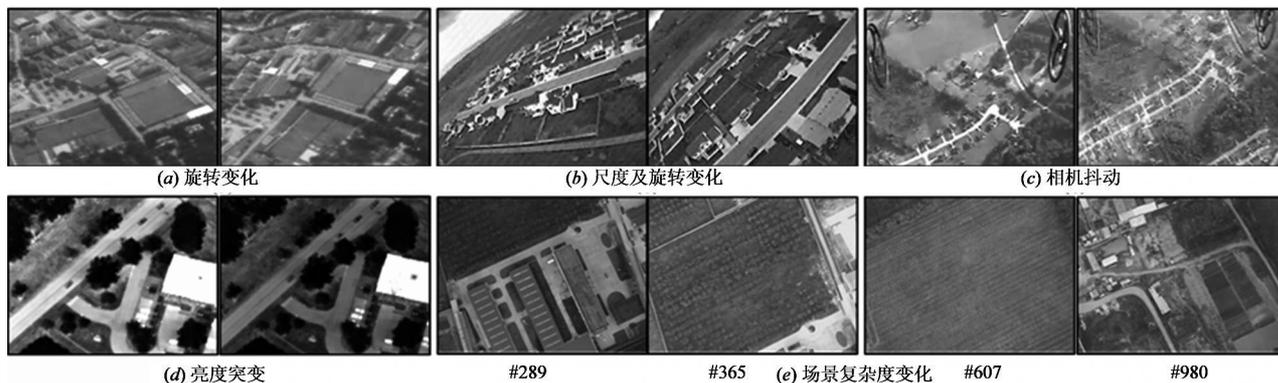


图1 无人机航拍视频序列示例

图1给出航拍视频中经常出现的一些典型问题示例,其中,图1(a)-图1(d)给出帧间可能出现的几何变化和亮度畸变,包括:载机机动造成图像间出现大幅度的平移、旋转、尺度变化(图1(a),(b)),飞行中的载台振动引发的运动模糊(图1(c)),环境亮度突变(图1(d))等;图1(e)给出一组场景复杂度变化下的航拍序列示例,该序列由无人机从低空获取,拍摄的场景中交替出现了纹理和特征丰富的城区建筑(图1(e),#289,#980),以及特征匮乏的农田(图1(e),#365,#607).可以看出,图像的各类畸变和场景复杂度的反差给稳定的特征提取和匹配带来严峻挑战,如何在上述复杂条件下,进行精确、鲁棒、实时的图像配准是机载成像侦查系统设计的关键.

2 相关工作

许多经典的航拍视频配准算法都基于光流约束^[3],但该类算法存在以下不足:首先,算法的推导建立在图像序列的高频采样假设基础上,即以某一点为中心的较小图像窗口在帧间只有微小的平移量.当无人机机动、摄像机镜头摆动造成图像间存在较大位置、角度变化时(如图1(a),(b)),算法性能会急剧降低;其次,光流约束方程是建立在色彩强度不变的假设之上,因此该类算法对环境亮度变化、噪声干扰等引起的图

像灰度畸变(如图1(d))比较敏感.

由于在图像发生几何变化和灰度畸变下表现出良好的稳定性,近年来,基于局部不变特征(Local invariant feature)的图像配准算法^[3-9]受到国内外学者的普遍关注,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图像局部特征点或特征区域,该类算法的处理步骤主要包括不变特征检测和不变特征描述两部分.

不变特征检测的目的是在待匹配图像上独立检测出内容相同的区域.Lindeberg^[4]率先提出在归一化高斯拉普拉斯尺度空间,通过局部极值检测定位特征所在的位置和尺度.Lowe^[5]同时在图像二维空间和双高斯差图像上(Difference of Gaussian, DoG)进行局部极值检测,确定关键点,最终实现算子对尺度和方向的不变性,为了增加后继匹配的稳定性,Lowe在后处理中引入了去除低对比度和边缘点、增加辅方向等机制.2002年,Matas^[6]提出了一种最大稳定极值区域检测算法MSER(Maximally Stable Extremal Regions),用一种类似分水岭算法的思路对相似区域进行分割和定位特征.2006年Grabner^[7]采用箱滤波器快速的得到尺度空间,并提出一种Fast-Hessian检测算法,采用海森矩阵定位特征.Mikolajczyk^[8]通过大量实验比较,指出尺度归一化高斯

拉普拉斯函数的极值是最稳定的图像特征, 可精确定位不变特征位置和尺度. 不变特征描述子是决定匹配性能的关键, 描述子可区分性高低将决定图像匹配的鲁棒性. 典型的局部不变特征描述子可以大致分为以下几类: (1) 基于直方图分布的描述子. 如 SIFT、LBP (Local Binary Pattern)、PCA-SIFT、Spin Images、GLOH (Gradient location and orientation histogram) 等; (2) 基于变换域的描述子, 如傅里叶变换、离散余弦变换、Gabor 变换、小波变换等; (3) 基于微分的描述子, 如微分不变量、方向可调滤波器 (Steerable filter) 等; (4) 其它描述子, 如不变矩等.

做为一种性能优异的不变特征配准算法, SIFT 算法受到国内外学者的广泛关注. Milolajczyk^[8] 在光照变化、旋转、缩放、模糊、压缩等条件下对 11 种典型描述子进行性能评估, 结果表明 SIFT 描述子的性能最优. 我们用 10 组航拍视频序列对 SIFT 算法进行测试, 结果表明该算法可有效解决前述航拍序列配准的问题(1)~(4), 即图像几何变化、亮度畸变和大幅度抖动, 但是, SIFT 算法仍难以克服问题(5)、(6): 一方面, 由于需要用高斯卷积构建图像多尺度空间、加之高维描述子构造与检索等问题, 使得 SIFT 算法计算复杂, 无法满足系统在线实时处理的要求. 另一方面, 该算法的特征检测数量和场景的复杂程度密切相关, 在纹理丰富的场景(如图 1(e), # 298, # 980)中, SIFT 会提取出大量的特征点, 给后继特征描述和匹配带来沉重的计算负担; 而对于纹理简单的场景(如图 1(e), # 365, # 607), 检测和匹配的特征点数急剧减少, 匹配性能会降低.

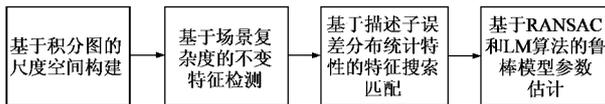


图2 基于场景复杂度与不变特征的实时图像配准算法流程图

针对航拍视频快速、稳健配准面临的上述难点问题, 本文在对不变特征检测和匹配的各个环节进行优化设计的基础上, 提出一种新的基于场景复杂度与不变特征的航拍视频实时配准算法, 包括快速尺度空间构建、基于场景复杂度的不变特征检测、基于描述子误差分布统计特性的特征搜索匹配和鲁棒模型参数估计四部分(图 2). 首先, 针对航拍视频序列中存在的图像几何变化、光照变化、抖动等问题(问题 1~ 问题 4), 算法将多尺度 Harris 角点和 SIFT 描述子相结合, 进行不变特征点提取与匹配, 并用 RANSAC 算法去除误匹配点, 用 LM 算法优化估计帧间变换模型参数, 保证了图像匹配的精度和鲁棒性; 然后, 针对场景复杂度变化下的检测特征数量大幅波动问题(问题 5), 算法通过计算图像的多尺度 Harris 角点响应分布, 动态估计当前场景的复杂度, 并以此精确控制检测的特征点数目; 最后,

针对算法实时性问题(问题 6), 算法主要从以下四个方面加以解决: (1) 用积分图和均值滤波模拟高斯卷积, 快速构建图像尺度空间; (2) 以多尺度 Harris 角点响应为依据, 在保证检测特征点质量的同时, 精确控制检测特征点数, 从而降低后继特征描述与匹配的复杂度; (3) 在特征匹配时, 用 SIFT 描述子误差分布的统计特性构造级联滤波器, 在保证匹配结果的同时, 大幅度缩小特征搜索空间; (4) 在参数估计时, 采用改进的 RANSAC 算法^[13], 用预测的出格点概率动态修正随机采样次数, 减少 RANSAC 迭代次数, 加快参数估计速度.

3 基于场景复杂度与不变特征的实时配准算法

3.1 基于积分图的尺度空间构建

考虑到载机飞行高度差异和传感器观测镜头缩放带来的图像尺度变化, 检测算法需要在多个可能的尺度上独立检测出不变特征点, 因此, 在特征检测前先要构建尺度空间图像. 尺度空间理论是将图像与某变换核进行卷积运算, 得到基于该变换核的图像序列. Koerdtink 和 Lindeberg^[4] 等学者证明具有线性、平移不变性、旋转不变性的高斯卷积核是实现线性尺度空间的唯一变换核. 但是, 随着高斯核窗宽的增大, 高斯卷积的计算量会急剧增加, 无法满足在线处理的要求.

根据中心极限定理, 对一个重要函数族而言, 将函数族中任意一个成员不断与自己卷积, 最终将生成一个高斯函数. 这意味着, 如果我们选择一个新的平滑核函数, 并且重复施加在图像上, 通过卷积的结合律性质, 则最终的结果就像我们使用了高斯函数来平滑该图像一样. 基于此, 本文采用矩形核函数卷积和降采样相结合的方式构造图像多尺度空间, 并用积分图加速计算.

一幅二维图像, 在不同尺度下的尺度空间表示可近似由图像与矩形核函数 M 卷积并亚采样得到:

$$L(x, y, \sigma, i) = \downarrow M(x, y, \sigma) * L(x, y, i-1) \quad (1)$$

其中, L 代表图像的尺度空间, i 代表尺度, (x, y) 代表图像的像素位置, σ 代表均值滤波的窗宽半径, σ 值越小, 表示图像被平滑的越少, 相应的尺度也就越小. 大尺度对应于图像的概貌特征, 小尺度对应于图像的细节特征. 为了加速计算, 在对第 i 级图像 $L(x, y, i)$ 进行卷积前, 首先计算 $L(x, y, i)$ 对应的积分图 $S(x, y, i)$, 从而式(1)可以改写为式(2), 且均值滤波的计算量与窗宽大小无关.

$$L(x, y, \sigma, i) = \downarrow \frac{1}{(2\sigma+1)^2} [S(x-\sigma, y-\sigma) + S(x+\sigma, y+\sigma) - S(x+\sigma, y-\sigma) - S(x-\sigma, y+\sigma)] \quad (2)$$

尺度空间的阶数 O 由图像大小和亚采样率 R 共同决定. 在没有图像缩放比率的先验信息时, 本文实验中采

用固定的亚采样率 R , 根据图像大小计算阶数 O :

$$O = \min_i \{ i | \frac{\min(W, H)}{R^i} \leq s \} - 1 \quad (3)$$

式中 s 为图像亚采样后的最小尺寸. 实验中, 对于原始大小为 320×240 的图像序列, 给定亚采样率 $R = 1.414$, $s = 20$, 则可构造一个 6 阶的多尺度空间. 实验表明, 上述尺度空间构建算法不仅大幅度减少了计算量, 而且精度可以满足后继特征检测和匹配的要求.

3.2 基于场景复杂度的不变特征检测

考虑到拍摄景物、地貌的差异, 航拍视频序列中经常出现场景复杂度交替变化的情况(如图 1e 所示). 由于在设计时缺乏对场景具体特性的考虑, 许多图像配准算法在特征检测时往往采用预先设置的固定参数, 使得该类算法的检测、匹配性能和处理速度会随着场景复杂度大幅度波动, 无法满足无人机系统在线实时处理的要求.

与之不同, 我们认为, 针对航拍视频序列配准设计的特征检测算子应同时具有以下三个特点: (1) 可重复性, (2) 可区分性, (3) 场景自适应性. 可重复性反映的是图像发生几何和灰度变化时, 算子独立检测出相同图像区域的能力, 可重复性越高, 则可能被匹配的点数越大; 可区分性指的是检测的特征点具有独特的纹理, 特征可区分性越强, 则匹配的鲁棒性越高; 场景自适应性指的是算子参数可随着场景复杂度动态调整, 在各种场景中均能精确控制输出特征点的数量和质量, 该特性对整个算法的匹配速度至关重要.

基于此, 本节提出一种基于场景复杂度的不变特征检测算法, 将多尺度 Harris 角点响应作为场景复杂度的依据, 根据响应分布函数在线动态估计最优参数, 在保证算子可重复性和可区分性同时, 精确控制检测的特征点数量, 满足特征描述和匹配对精度和实时性的要求.

作为一种重要的图像局部特征, 角点具有丰富的纹理和良好的可区分性. Harris 算子利用局部自相关函数矩阵 M 的特征值检测角点, 具有良好的旋转和亮度仿射不变性, 不受摄像机姿态及光照的影响. 但该算子对图像的尺度变化比较敏感, 当无人机飞行高度发生变化、或者摄像机镜头缩放时, 不能保持检测特征的可重复性. 因此, 本文在特征检测时, 首先采用积分图快速构建尺度空间图像(见第 3.1 节), 然后采用多尺度 Harris 角点检测算法^[10], 在多个尺度上进行角点检测:

$$M(x, y, i) = \sum_{x, y} w(x, y) \begin{bmatrix} L(x, y, i)_x^2 & L(x, y, i)_x L(x, y, i)_y \\ L(x, y, i)_x L(x, y, i)_y & L(x, y, i)_y^2 \end{bmatrix} \quad (4)$$

式(4)中 $M(x, y, i)$ 为图像在尺度 i 上的局部自相关函数矩阵, $L(x, y, i)_x$ 和 $L(x, y, i)_y$ 代表尺度 i 上 (x, y) 点的水平和垂直梯度, $w(x, y)$ 为高斯窗, 用于降低矩形窗带来的噪声干扰. 矩阵 $M(x, y, i)$ 的两个特征值 λ_1, λ_2 和局部自相关函数的主曲率成正比. 当两个特征值都高时, 局部自相关函数是尖峰状, 窗口强度在任意方向上移动都会增加, 这就表示是角点. 为了回避特征值的计算, Harris 的角点响应 R 采用式(5):

$$R(x, y, i) = \det M(x, y, i) - \lambda \cdot (\text{trace} M(x, y, i))^2 \quad (5)$$

式中 $\det M = \lambda_1 \lambda_2$, $\text{trace} M = \lambda_1 + \lambda_2$, λ 是经验常数, 通常取 0.04~0.06.

在获得多尺度图像的角点响应后, 传统的角点提取算法一般选取角点周围的一个圆形模板, 通过比较中心点和模板其他点的响应值, 提取响应的局部极值点. 由于不考虑图像的自身特点, 对于纹理丰富的场景, 该算法会检测出数量众多的角点, 不仅严重降低了后继特征描述和匹配的速度, 而且使大量的计算浪费在可区分性差的低质量角点上. 与传统算法不同, 本文在非极大值抑制时, 引入一个根据场景图像的复杂度自适应调整的参数 $T(k)$, 并根据下式检测特征点:

$$c(x, y, i) = \begin{cases} 1, & R(x, y, i) - R(x', y', i) > T(k) \\ 0, & R(x, y, i) - R(x', y', i) \leq T(k) \end{cases} \quad (6)$$

式中 $c(x, y, i)$ 表示尺度 i 上检测的角点, $R(x, y, i)$ 和 $R(x', y', i)$ 分别表示模板中心点和非中心点的角点响应值. $T(k)$ 表示序列中第 k 帧的非极大值抑制阈值, 它决定了圆形模板中心点和模板其他点的角点响应差别, $T(k)$ 值越大, 表明特征点的角点响应与周围反差越大, 局部角点响应空间分布呈明显的单峰分布, 相应的, 提取的特征数量减少; 反之, $T(k)$ 值减小, 提取的特征数量会增加. 在线处理时, 参数 $T(k)$ 会根据场景复杂度和拟检测的特征点数 K 动态选取. 综上, 基于场景复杂度的不变特征检测算法计算过程如下:

(1) 对于输入图像, 用式(1)(2)快速构建尺度空间, 并用式(5)(6)计算图像多尺度 Harris 角点响应;

(2) 遍历整个尺度空间, 计算每一个像素的局部角点响应差异 $\Phi(x, y, i)_j$ 的最小值 $\Phi(x, y, i)$:

$$\Phi(x, y, i) = \min(\Phi(x, y, i)_j), j = 1, \dots, m \quad (7)$$

$$\Phi(x, y, i)_j = \begin{cases} R(x, y, i) - R(x'_j, y'_j, i), & \text{if } R(x, y, i) > R(x'_j, y'_j, i) \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

其中 m 表示圆形邻域中非中心点的像素数目, 结合式(6)(7)(8)可知, $\Phi(x, y, i)$ 实质上是将像素 $L(x, y, i)$

检测为角点的最高阈值;

(3) 遍历结束后, 计算离散型随机变量 $\phi(x, y, i)$ 的概率 $p(\phi)$ 和分布函数 $F(\phi)$:

$$p(\phi) = \frac{n_\phi}{N}, \quad F(\phi) = \sum_{t \leq \phi} p(t) \quad (9)$$

式中 n_ϕ 表示角点响应差异为 ϕ 的象素数目, N 是尺度空间图像中象素的总数.

(4) 根据分布函数 $F(\phi)$ 和拟检测的特征点数 K , 计算当前帧最优的特征点检测阈值 $T(k)$:

$$T(k) = \arg \min_{\phi} (F(\phi) > 1 - K/N) \quad (10)$$

(5) 将阈值 $T(k)$ 带入式(6), 重新遍历尺度空间, 检测出角点响应差异最高的前 K 个高质量特征点的位置和对应的特征尺度.

在线配准时, 当给定拟检测的特征点数 K 后, 上述算法可根据当前拍摄场景的复杂度, 动态选择最优阈值 $T(k)$, 在保证检测特征点质量的同时精确控制特征点数, 确保了系统在各种地貌环境下图像匹配的实时性和稳定性.

3.3 基于描述子误差分布统计特性的特征匹配

在采用第 2 节检测算法得到特征点的位置和尺度信息后, 需要对特征点邻域图像进行提取和规范化, 尽量降低图像间可能存在的尺度缩放、旋转的影响, 获得内容一致的子图像. 本文从特征点所在的尺度图像 $L(x, y, i)$ 上选择矩形邻域作为支撑区域, 计算 128 维 SIFT 描述子^[5].

由于维数较高, 直接用 128 维 SIFT 描述子做特征匹配时运算量大, 不适于航拍视频的实时配准. 作者通过大量实验, 发现正确匹配点对的 SIFT 描述子形状具有很强的相似性, 其统一形态特征的误差呈现明显的高斯分布. 依据该特性, 本文提出一种新的搜索匹配策略, 其基本思想是根据离线训练的描述子各维特征的误差分布, 设计一个级联分类器, 在匹配时滤除误匹配点.

为了获得匹配点描述子的误差分布, 同时验证这种误差分布具有良好的稳定性, 我们从 CMU 的捕食者无人机航拍视频数据库和部分自建数据库中, 挑选了多组有代表性的航拍视频序列进行测试. 视频序列中包含了不同地貌环境、不同飞行高度、不同运动模式、不同光照变化等情况(见图 5, No. 1~ No. 6). 在计算时, 先采用本文提出的基于时空相关的自适应特征检测算法, 定位出图像多尺度特征点; 然后用 SIFT 描述子对规范化的支撑区域进行特征描述; 在特征匹配时, 采用 RANSAC 算法进一步去图像中的误匹配点.

用 ϕ_i 和 ϕ'_i 分别代表一对正确匹配点的描述子, 则匹配点描述子误差 ξ 和描述子方差的误差 ξ_{var} 分别为:

$$\xi = \phi_i - \phi'_i \quad \xi_{var} = \text{var}(\phi_i) - \text{var}(\phi'_i) \quad (11)$$

方差表示的是描述子各维特征的离散程度, 它包含了描绘子特征形状的整体信息, 因此, 比较两个描述子方差的差异可以从宏观上判定描述子之间的相似性, 而分析描述子每一维的误差则从微观上刻画描述子特征向量的相似性.

在实验中, 用本文算法对三组捕食者无人机航拍序列进行自动特征检测与匹配, 共得到 6 2117 对正确匹配点, 每一对正确匹配点包含两个 128 维描述子. 分别计算所有匹配点对的匹配点描述子误差 ξ 和描述子方差的误差 ξ_{var} . 其中 ξ 共计 6, 2117 \times 128 个样本, 描述子方差 ξ_{var} 共计 6, 2117 个样本. 图 3 分别给出两种误差的分布曲线, 航拍图像的统计结果表明, 匹配点描述子和描述子方差的相对误差 ξ 和 ξ_{var} 均表现出明显的“单峰性”、“对称性”和“有界性”(见图 3), 误差分布近似于一个高斯分布. 其中 98.9% 的匹配点对描述子方差的误差 ξ_{var} 落在区间 $[-0.5 \times 10^{-4}, 0.5 \times 10^{-4}]$ 中, 99.8% 的匹配点对描述子误差 ξ 落在区间 $[-0.02, 0.02]$ 中. 基于此, 算法在匹配时先计算描述子的形态特征, 并据此快速滤除无关特征点.

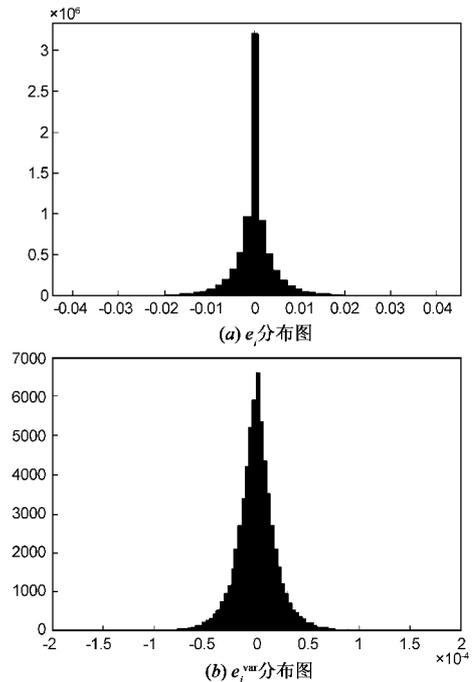


图 3 匹配点描述子误差分布

图 4 给出快速匹配算法流程图. 其中 T_{var} 和 T_e 分别表示描述子方差和每一维的误差阈值, 该参数由正确匹配点的统计分布给出(见图 3). 实验结果表明, 该方法可在在保证匹配性能不变的同时, 快速滤除大量无关误匹配点, 减少搜索空间, 提高特征搜索配速度, 详细实验分析参见第 4 节图 7.

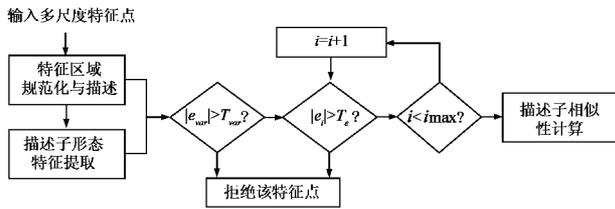


图4 基于描述子误差分布统计特性与级联滤波的快速匹配

3.4 基于RANSAC和LM算法的参数鲁棒估计

模型参数的鲁棒估计是计算机视觉领域的一个核心问题,其关键在于对不符合实际模型的“出格点(outliers)”的处理上.本文采用RANSAC算法^[11]去除误匹配点,并采用实际匹配点与估计匹配点之间的欧式距离作为判决准则,确定内点和出格点,其中变换模型选用仿射模型.在基本的RANSAC算法中,重复抽样次数 K 由外点概率 ϵ ,抽样点对数 S 和 K 次抽样至少有一次全部为内点的概率 P 得到,这种固定抽样次数的方法会严重影响估计算法的处理速度,为此本文采用文献[12]中改进的RANSAC算法,在抽样结束后根据当前内点的个数动态更新出格点概率和采样次数,加快算法速度.

为了进一步提高配准精度,本文在去除误匹配点后,先用最小二乘法估计帧间仿射模型参数,提供较为精确的迭代优化初始值,然后用LM算法^[13]对变换模型参数进行迭代优化.

4 算法实现和结果分析

以本文算法为核心,用VC++ 6.0开发了一个实时航拍视频序列配准、镶嵌与电子稳像系统.系统硬件配置为Intel Core(TM)2 Duo CPU E8400@3.0GHz, 2GDDR2内存.对10组无人机航拍视频序列的测试结果表明,系统可在各类场景中稳定、准确、实时的进行图像配准和增稳,对 320×240 的航拍序列的平均速度达到20.7帧/秒.

本节首先从不同场景下的特征点检测数量、正确匹配点数和计算时间三方面对本文算法进行定量性能评估,并给出算法对航拍视频序列的实时匹配和镶嵌结果;然后用包含不同复杂度场景的航拍视频序列做测试数据,将本文算法与SIFT算法^[5]进行分析比较.

4.1 本文算法性能分析

在实验中选用图5所示的10组无人机航拍视频序列作为测试库,测试数据来自卡耐基梅隆大学的VSAM

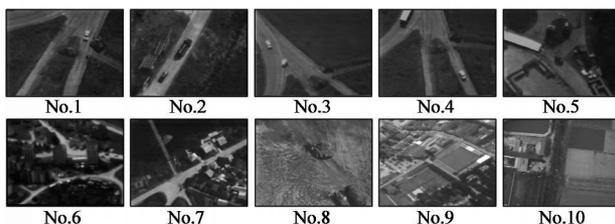


图5 实验采用的无人机航拍视频序列

捕食者无人机数据库(图5, No. 1~No. 7)和自建无人机航拍视频数据库(图5, No. 10).测试序列共计7255帧,分辨率为 320×240 ,拍摄的场景涵盖纹理简单的乡间公路和田野(图5, No. 1~ No. 4, No. 8),纹理丰富的城区建筑物(图5, No. 6~ No. 9, No. 10),拍摄目标有地面运动的车辆(图5, No. 1~ No. 7)、坦克和行人(图5, No. 8),序列中图像间包括大幅度抖动、角度旋转、尺度缩放、亮度突变以及视角变化等情况.

首先选用第六组测试序列(图5, No. 6),分析比较固定阈值和基于场景复杂度的自适应阈值的特征点检测算法性能.序列No. 6是一段美国捕食者无人机对地面车辆的跟踪拍摄视频,序列长度为1518帧,拍摄场景中包括森林、草地、田野、公路、城市建筑等,图像的复杂度始终发生动态变化.试验中,分别选择三组固定的非极大值阈值 T 和本文自适应检测算法进行比较,其中 T 的取值分别为 $T=5$, $T=20$, $T=50$.图6(a)和图6(b)分别给出检测特征点数和抑制非极大值的阈值 T 随时间变化的曲线.可以看出,固定阈值的方法(图6(a))检测的特征点数与场景复杂度密切相关,波动幅度大.对于包含建筑物等纹理丰富场景,特征点数会急剧上升(图6(a),第310帧),增加了后继特征描述和匹配的计算负担;而对于草地、田野等纹理简单的场景,该类算法的检测点数又迅速降低(图6(a),第500帧,第1250帧),无法保证稳定、可靠的匹配.此外,该类算法难以选择一个合适的参数去适应场景的变化.与之不同,本文算法能够场景复杂度实时、动态的选择最优的自适应阈值(图6(b)),当场景复杂时升高阈值,反之则降低,从而在各种场景下检测的特征点数始终保持不变(图6(a)),这一独特优点保证了后继的特征描述和匹配的实时性和鲁棒性.

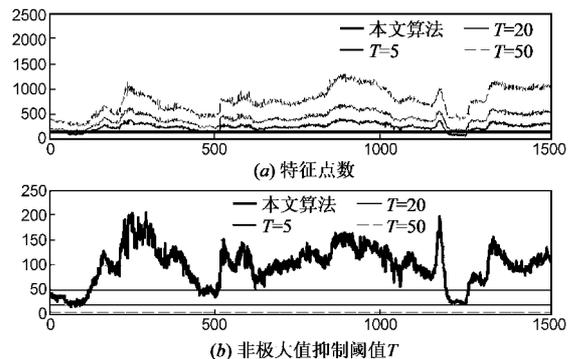


图6 特征点检测结果比较

为了分析基于描述子误差分布的级联滤波算法性能,我们选择了10对航拍图像,分辨率为 320×240 ,并采用本文提出的特征点检测算法,精确控制检测特征点数,分析级联滤波前后匹配性能的变化.实验中检测点数分别设置为200, 400, 800和1600,从图7可以看

出, 级联滤波前后, 平均匹配点数几乎不发生改变(图 7 (a)), 而平均匹配时间明显减少(图 7 (b)). 当检测特征点数为 1600 时, 级联滤波可使匹配时间从 2500ms 降低至 1500ms, 计算时间减少了 40%, 而匹配点数仅有轻微下降. 上述分析结果表明, 基于描述子误差分布特性设计的级联滤波可在保证匹配性能的同时降低匹配时间.

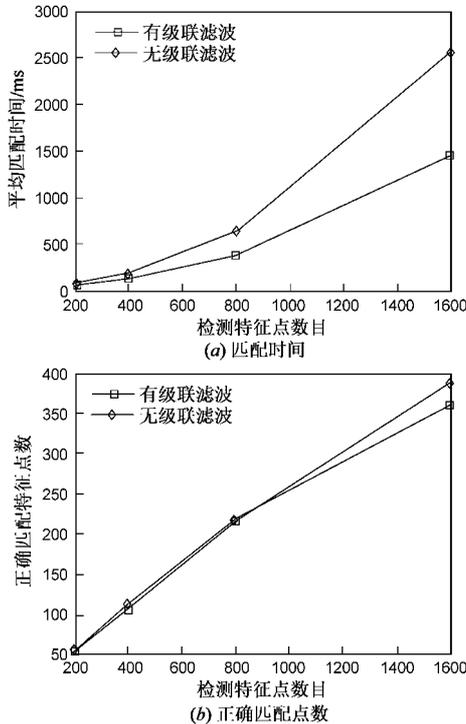


图7 级联滤波前后性能比较

为了进一步获得算法配准性能的统计结果, 用图 5 所示十组航拍序列对算法进行测试, 测试图像序列共计 7255 帧, 分辨率为 320×240 , 拍摄的场景涵盖乡间公路、田野、建筑物, 拍摄目标有车辆、坦克和行人, 实验中检测特征点数设置为 150. 图 8 给出部分特征匹配结果示例, 包括图像的复杂度变化(见图 8 (a), (b))、几何变化(图 8(c)~(f))等情况, 图中用圆圈标明不变特征的支撑区域, 直线标明匹配的特征点对. 表 1 详细给出算法在不同场景下的匹配点数和处理速度. 其中, 定义匹配点比率= 匹配点对/检测特征点数, 该参数值越高, 表明检测的特征点可重复性和可区分性越强. 可以看出, 本文算法在各种场景下均能够找到足够数量的匹配点(表 1, 第四列): 对测试的十组序列, 匹配点数最低为 49.03(表 1, No. 9), 最高为 85.21(表 1, No. 6), 对整个测试序列的 7255 帧图像平均匹配点对为 67.44(表 1, 最后 1 行, 第 4 列), 平均匹配点比率达到 44.83%(表 1, 最后 1 行, 第 5 列), 取得了令人满意的结果. 在处理速度方面, 算法的计算量不随场景的复杂度发生改变, 能够在各种地貌条件下快速、稳定的进行图像配准, 对

7255 帧 320×240 的测试序列平均计算时间为 48.53ms(表 1, 最后 1 行, 第 6 列), 平均处理速度达到 20.7 帧/秒(表 1, 最后 1 行, 第 7 列), 完全能够满足航拍视频实时配准的要求.

表 1 算法对十组航拍序列的匹配点数和处理速度统计结果

序号	帧数	特征点数	匹配点对	匹配点比率	处理时间 (ms)	帧率 (fps)
No. 1	274	150	76.97	50.97%	50.1	19.9
No. 2	971	150	84.27	55.81%	48.9	20.4
No. 3	753	150	10.19	39.86%	49.98	20.0
No. 4	369	150	61.92	41.01%	49.02	20.4
No. 5	951	150	74.23	49.49%	48.02	20.8
No. 6	1518	150	85.21	56.81%	47.57	21.0
No. 7	583	150	71.64	47.76%	41.89	23.9
No. 8	199	150	52.76	35.17%	52.63	19.0
No. 9	32	150	49.03	32.69%	47.19	21.2
No. 10	1630	150	58.19	38.79%	49.98	20.0
平均		150	67.44	44.83%	48.53	20.7

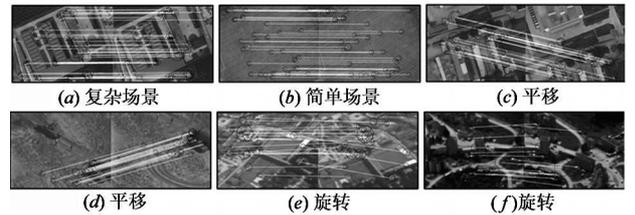


图8 特征匹配结果示例

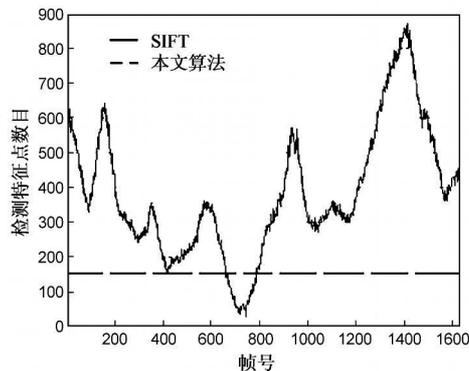
4.2 与 SIFT 算法比较

本节将本文算法与 SIFT 算法用于航拍序列配准, 从检测特征点数、处理速度等方面进行性能评估. 实验中 SIFT 算法采用 Lowe 发布的 C 语言编写的应用程序^[5].

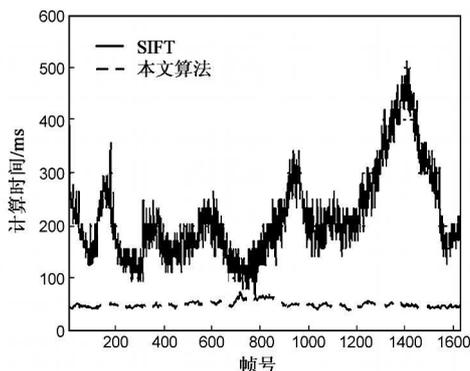
图 9 给出两种算法对第 10 组航拍序列(图 5, No. 10)的处理结果, 图像分辨率为 320×240 , 序列长度 1632 帧, 该序列中交替出现了纹理简单田地和纹理复杂的建筑物. 可以看出, SIFT 算法检测的特征点数始终随场景的复杂程度动态变化(图 9 (a)), 相应的, SIFT 的匹配时间也出现从 78ms 至 509ms 的大幅度波动(图 9 (b)), 这种依赖于场景复杂度的特性使得该算法无法满足在线实时处理的要求. 与之不同, 本文算法的检测特征点数与场景复杂度无关(图 9 (a)), 计算时间稳定在 50ms/帧, 几乎不随场景发生变化(图 9 (b)).

为了比较两种算法在场景变化下的匹配性能, 我们从第 10 组航拍序列中挑选了一组包含建筑物的复杂场景图(图 10 (a), # 1400, # 1401)和一组缺乏特征的农田图像(图 10 (a), # 722, # 723). 图 10 分别给出 SIFT 和本算法对这两组图像的匹配结果比较. 可以看出, SIFT 性能与场景复杂度密切相关, 当场景内出现大量细节丰富的建筑物时, SIFT 算法(图 10 (b), 左图)检测和匹配的特征点数迅速上升, 分别为 864 和 560, 相应的

特征点数的增加导致算法计算量增大, SIFT 算法对该组图像的处理速度仅为 1.96 帧/秒, 而本算法检测特征点数不随场景复杂度改变, 对该组图像获得 59 个匹配点(图 10(b), 右图), 处理速度为 20.83 帧/秒, 是 SIFT 算法的 10.6 倍; 对于缺乏特征的农田图像, SIFT 算法的检测和匹配性能迅速降低, 匹配点数仅为 19 个(图 10(c), 左图), 而本算法通过动态调整特征检测参数, 检测点数保持不变, 匹配点数为 30 个(图 10(c), 右图), 优于 SIFT 算法。

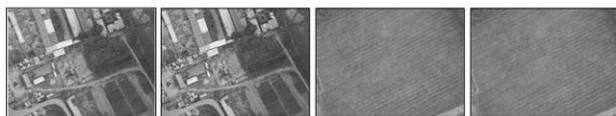


(a) 检测特征点数

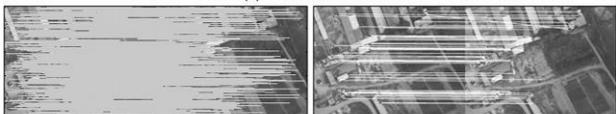


(b) 匹配计算时间

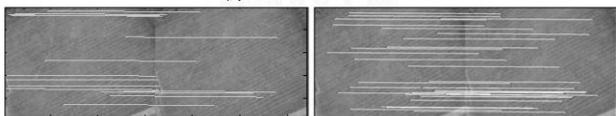
图9 不同场景复杂度下本文算法与SIFT算法的性能比较



(a) 航拍视频示例



(b) 复杂场景匹配结果



(c) 简单场景匹配结果

图10 不同场景复杂度下的匹配结果比较

5 结论

实时、鲁棒的视频配准是航空电子侦察系统的核心和关键技术. 本文在对航拍视频配准难点进行详细分析的基础上, 提出了一种新的基于场景复杂度与不变特征的实时图像配准算法. 在实时性方面, 通过对整个配准算法的各个环节进行优化设计, 包括基于积分图的快速图像尺度空间构建、基于场景复杂度的特征点数控制、基于描述子误差分布统计特性级联分类器设计、采用改进的 RANSAC 算法减少迭代次数等, 使得算法计算量不随场景的复杂度发生改变, 能够在各种地貌条件下实时、稳定的进行图像配准; 在鲁棒性方面, 算法将多尺度 Harris 角点和 SIFT 描述子相结合, 并用改进的 RANSAC 与 LM 算法对帧间变换模型参数进行鲁棒估计, 有力的保证了算法的旋转、尺度、亮度不变性和配准精度. 与相关算法的实验对比表明, 本算法在运行效率、匹配鲁棒性方面均有较大优势.

参考文献:

- [1] Ali S, Reilly V, Shah M. Motion and appearance contexts for tracking and reacquiring targets in aerial videos [A]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Minneapolis, MN, USA, IEEE Computer Society: 2007. 1-6.
- [2] Unmanned Aircraft Systems Roadmap 2005 2030 [R]. Office of the Secretary of Defense, Washington, DC, USA, 2005. 71-77.
- [3] Tomasi C, Kanade T. Detection and tracking of point features [R]. Carnegie Mellon University, Technical Report, 1991. 1-20.
- [4] Lindeberg T. Feature detection with automatic scale selection [J]. International Journal of Computer Vision, 1998, 30(2): 77-116.
- [5] Lowe D G. Distinctive image features from scale invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [6] Matas J, Chum O, Urban M, and Pajdla T. Robust wide baseline stereo from maximally stable extreme regions [A]. In Proceedings of British Machine Vision Conference [C]. Cardiff, UK, 2002. 384-393.
- [7] Grabner M, Grabner H, Bischof H. Fast approximated SIFT [A]. In proceedings of Asian Conference on Computer Vision [C]. India, 2006. 918-927.
- [8] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630.
- [9] 韦燕凤, 赵忠明, 闫冬梅, 曾庆业. 基于特征的遥感图像自动配准算法 [J]. 电子学报, 2005, 33(1): 161-165.

- Wei Yanfeng, Zhao Zhongming, Dongmei Yan, Qingye Zeng. Remote sensing image automatic registration based on feature [J]. Acta Electronica Sinica. 2005, 33(1): 161– 165. (in Chinese)
- [10] Harris C. Geometry from visual motion. Active Vision [M]. Cambridge, MA, USA: MIT Press, 1992. 263– 284
- [11] Fischler MA, Bolles RC. 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.
- [12] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision [M]. Cambridge University Press, 2004. 116– 123.
- [13] Press W H, Flannery B P, Teukolsky S A, Vetterling W T. Numerical Recipes: The Art of Scientific Computing [M]. Cambridge. Cambridge: Cambridge University Press, 1986. 801 – 806.

作者简介:



杨 涛 男, 1979 年生于陕西咸阳, 西北工业大学计算机学院博士后. 研究方向为移动平台智能视频分析、多视阵列成像与应用、多摄像机网络视频目标检测与跟踪、图像配准.

E-mail: yangtao@wpu@163.com



张艳宁 女, 1967 年生于陕西咸阳. 西北工业大学计算学院教授, 博士生导师. 研究方向为计算机视觉、图像处理与模式识别、智能信息处理.

张秀伟 女, 1981 年生于新疆塔城. 西北工业大学计算学院博士生. 研究方向为网络化多源信息协同处理、图像配准.

张新功 男, 1985 年生于陕西咸阳. 西北工业大学计算学院硕士生. 研究方向为景象匹配导航、图像配准.