

文章编号 1004-924X(2020)05-1140-12

采用轮廓特征匹配的红外-可见光视频自动配准

孙兴龙^{1,2}, 韩广良^{1*}, 郭立红¹, 刘培勋¹, 许廷发³

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049;

3. 北京理工大学 光电学院, 北京 100081)

摘要:为了精确地配准近平面场景下的红外-可见光视频序列,本文提出了一种基于轮廓特征匹配的自动配准方法,通过迭代匹配目标轮廓特征来解决异源图像中配准特征的提取和匹配难题。首先,采用运动目标检测技术获取目标轮廓,并由曲率尺度空间(CSS)角点检测算法提取轮廓特征点。此后,建立全局形状上下文描述子和局部边缘方向直方图描述子描述特征,从而实现可靠的特征匹配。来自不同时刻的匹配点对被保存在一个基于高斯距离准则的特征匹配库中。最后,为了克服近平面场景中目标深度变化的影响,本文结合前景样本随机抽样策略计算配准矩阵的损失函数,完成对全局配准矩阵的更新。在LITIV数据库上对方法进行实验验证,结果表明:本文方法的配准精度优于当前先进的对比方法,在9个测试视频上的平均重叠率误差仅为0.194,与对比方法相比下降了18.5%。基本满足了近平面场景下红外-可见光视频序列配准的精度要求,且具有较高的鲁棒性。

关键词:红外-可见光视频序列;图像配准;轮廓特征;特征匹配;全局配准矩阵

中图分类号:TP394.1;TH691.9 **文献标识码:**A **doi:**10.3788/OPE.20202805.1140

Infrared-visible video automatic registration with
contour feature matchingSUN Xing-long^{1,2}, HAN Guang-liang^{1*}, GUO Li-hong¹, LIU Pei-xun¹, XU Ting-fa³(1. Changchun Institute of Optics, Fine Mechanics and Physics,
Chinese Academy of Sciences, Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. School of Optoelectronics, Beijing Institute of Technology, Beijing 100081, China)

* Corresponding author, E-mail: hangl@ciomp.ac.cn

Abstract: To register infrared-visible video sequences precisely in almost-planar scenes, an automatic registration method based on matching the contour features was proposed in this paper. This method could solve the challenging problem regarding extracting and matching features in multimodal images by iteratively matching the contour features of targets. First, this method adopted the technology of moving target detection to identify the contours of targets and extracted the contour feature points with the corner detection algorithm of Curvature Scale Space(CSS). Then, the global shape context

收稿日期:2019-11-11;修订日期:2020-01-29.

基金项目:国家自然科学基金资助项目(No. 61602432)

descriptors and the local histogram of edge orientation descriptors were established to describe the features; these are useful for reliable feature matching. The matched feature pairs from different times were reserved in a reservoir based on the Gaussian distance criterion. Finally, to overcome the influence of target depth variation in almost-planar scenes, the loss function of the registration matrix was calculated by incorporating the strategy of randomly sampling foreground samples, after which the global registration matrix was updated. The method was validated using the LITIV dataset, and the results demonstrate that the proposed method outperforms state-of-the-art methods. The average overlap error of our method on nine test sequences is only 0.194; this value for the compared methods demonstrate a decrease of 18.5%. This essentially satisfies the precise requirement of infrared-visible video registration in almost-planar scenes, and this method is fairly robust.

Key words: infrared-visible video sequence; image registration; contour feature; feature matching; global registration matrix

1 引言

作为计算机视觉领域的重要组成部分,图像融合技术^[1]在近年来得到了深入的研究,尤其是多源图像融合技术^[2-4]。其中,红外图像和可见光图像可以为观测场景提供互补性的光谱信息,融合两者^[2,4]可以显著地提升场景表述能力,因此在人脸识别、安防监控、军事侦察、目标跟踪等领域得到了广泛应用^[5-6]。作为图像融合的关键预处理技术之一,图像配准技术^[7]致力于确定两幅或多幅图像之间的时空变换关系,进而在时间和空间上对齐待融合图像。精确的图像配准可以保证图像融合的有效性和可靠性,具有重要的研究意义。

图像配准的核心工作是在待配准图像间寻找一致性信息,但这对于红外-可见光图像而言是一个十分具有挑战性的问题。红外传感器的成像原理是记录场景的热分布,而可见光传感器是记录场景的反射光分布。成像原理的不同导致两类图像存在强烈的强度差异和纹理差异。例如,红外图像往往遗漏物体的细节信息,但这在可见光图像中却清晰可见。在此情况下,如何有效地为图像配准提取一致性信息是一个值得探究的问题,诸多学者已经对此展开了深入的研究和分析。

目前已经存在的红外-可见光图像配准算法主要基于以下三类信息:区域信息、特征信息及运动信息。区域信息配准算法^[8-10]的基本思想是直接或间接地利用图像强度信息计算一种相似性度量函数,通过最优化度量函数确定配准参数。其

主要包括相关类算法和互信息法。相关类算法^[9]主要通过最大化各个候选图像子窗之间的互相关函数确定匹配窗口对,进而得到一致性信息。互信息方法^[8,10]采用联合信息熵(互信息函数)作为损失函数,通过最优化损失函数确定配准参数。总体而言,区域信息配准算法严重依赖于图像强度信息,难以克服异源图像之间强度差异的影响。

基于特征信息的配准算法首先需要从图像中提取特征,然后通过匹配特征计算两幅图像间的变换关系。经常采用的特征包括点、边缘等。作为最常见的特征,点特征已经被普遍地应用于配准领域,如角点^[11],SIFT特征点^[12]及 SURF^[13]特征点。但由于红外-可见光图像之间的强度差异和纹理差异,利用这些点特征进行配准并不可靠。红外-可见光传感器可以同时捕获场景中的边缘信息,因此采用边缘信息配准是一种可行的方案。Klimaszewski 等人^[14]通过衡量相位相关度匹配红外-可见光边缘,而 Ma 等人^[6]采用复杂的高斯区域准则匹配边缘映射信息。此外,Ming 等人^[15]通过在边缘上提取 CSS(Curvature Scale Space)角点进行图像配准。但是,边缘分布在红外-可见光图像中并不完全一致,且非常容易受到噪声的影响,直接采用边缘信息难以保证配准精度。

由于静态特征的配准精度受限,近年来逐渐出现了基于运动信息的配准方法。此类方法首先采用运动目标检测技术^[16-17]提取目标前景,然后根据前景特征完成配准工作。其中,王洪庆等人^[18-19]通过采用跟踪目标顶点的方式确定目标运动轨迹,并通过轨迹匹配计算配准矩阵。然而

精确地检测目标顶点的难度较大,得到的特征数量也有限。另一类方法^[20-22]是提取目标前景的轮廓曲线,在进行配准时对轮廓曲线或曲线上的特征点进行匹配。此类方法可以达到较高的配准精度,是目前配准效果比较突出的一类配准算法。

然而,基于轮廓特征的配准方法仍然存在一些缺陷。首先,在特征匹配之前并没有对特征进行充分的描述,导致匹配的精确度不足;此外,在近平面场景中物体的深度位置并不完全固定,而这些方法仅仅根据当前时刻的配准效果确定全局配准矩阵,往往导致矩阵陷入局部最优解,无法精确地配准整个视频序列。

为了解决上述问题,本文提出了一种新的配准方法,其根据轮廓曲线的空间分布对特征进行描述,提升了特征匹配的可靠性;在估计全局配准矩阵时,综合考虑了配准参数在多个时刻的配准效果,防止其陷入局部最优解。实验结果表明,本文算法在近平面场景中可以实现高精度的红外-可见光视频配准,其精度优于当前先进的其它配准算法。

2 轮廓特征配准方法的基本原理

基于轮廓特征的配准算法^[20-22]通过利用目标的运动信息克服了红外-可见光图像间强度差异与纹理差异的影响,在红外-可见光图配准领域具有重要的参考价值。这类算法主要包含特征提取、特征匹配及配准参数计算三个步骤。

2.1 轮廓特征提取

为了捕获运动目标的轮廓特征,首先需要采用运动目标检测技术检测目标前景。目前在配准领域被经常使用的运动目标检测算法是 St-Charles 等人提出的 PAWCS 算法^[17]。该算法根据颜色特征信息和二值化特征信息对背景进行统计建模,并对统计模型进行负反馈更新,最后通过背景减除的方式提取运动目标。为了抑制噪声和光线变化等因素的干扰,滤波、腐蚀和膨胀等操作在这一过程也经常被使用。图 1 给出了利用 PAWCS(Pixd-based Adaptive Word Consensus Segmenter)算法提取到的红外-可见光前景图像。

在提取了运动目标以后,利用简单的边缘检测技术即可获得目标轮廓曲线。目前的配准算法通过多种手段从曲线上提取特征。其中,Sonn 等

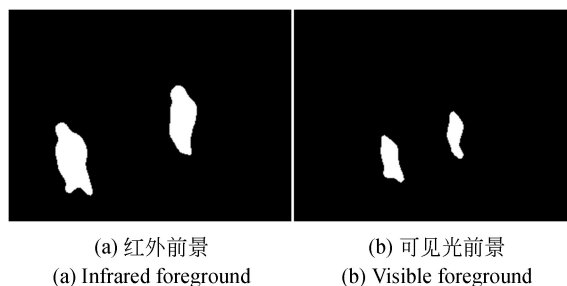


图 1 运动目标检测的结果

Fig. 1 Results of motion target detection

人^[20]采用曲线演化理论获取特征点,而 Charles 等人^[21]直接将所有轮廓点作为特征。另一种方式是利用 CSS 角点检测算法^[15,22]提取特征,主要包含以下步骤:(1)在大尺度空间下计算所有轮廓点的曲率值,选取邻域极大值点作为候选角点;(2)在低尺度空间下精确定位候选角点,将其与 T 型角点进行比较剔除,得到最终的检测结果。该方法具有定位准确、计算量小等优势,且提取的特征具有尺度不变性和旋转不变性。本文采用该算法提取配准特征点,提取到的红外-可见光轮廓特征如图 2 所示。

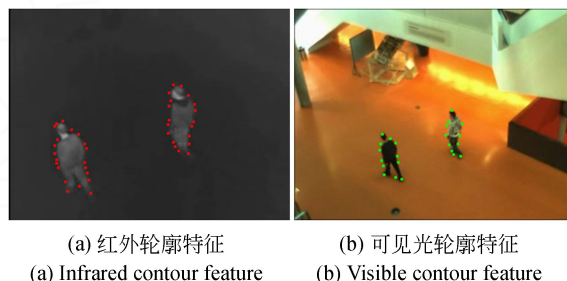


图 2 CSS 角点检测的结果

Fig. 2 Results of CSS corner detection

2.2 轮廓特征匹配

特征匹配是图像配准的核心任务,精确有效的特征匹配是实现高质量图像配准的前提。目前已经出现了多种轮廓特征匹配策略,各种类型的特征描述子也已被采用。Sonn 等人^[20]在曲线演化理论的基础上引入了距离描述子和方向描述子,通过计算描述子之间的距离匹配红外-可见光特征点。Sun 等人^[22]采用归一化位置和边缘方向直方图描述特征点,并以归一化距离作为匹配度量函数。尽管这些描述子在轮廓特征匹配的过程中起到了一定作用,但描述方法过于简单,无法

保证匹配质量。因此本文建立新的特征描述机制,以完成精确地特征匹配。

2.3 配准参数计算

在完成特征点匹配后,可利用匹配点对计算配准参数。目前常用的是随机一致性(RANSAC)算法^[23]。其首先随机选取若干点对计算配准参数,并根据该参数计算各匹配点对之间的距离,方式如式(1)所示:

$$d = \sqrt{(x_i - Z(x_j, \mathbf{M}))^2 + (y_i - Z(y_j, \mathbf{M}))^2}, \quad (1)$$

其中: (x_i, y_i) 和 (x_j, y_j) 分别表示点对中红外和可见光特征点的位置, Z 表示插值变换, $\mathbf{M}$ 表示配准参数。当点对之间的距离小于 T (一般 $T=5$) 时,认为其是矩阵的内点,否则视为外点。多次循环以上步骤,选取内点数量最多的配准参数作为最终结果。

在视频序列配准问题上,仅根据当前时刻的配准参数难以实现精确配准。因此,许多算法^[20-22]采用迭代更新的策略确定一个全局配准矩阵。采用初次得到的配准矩阵初始化全局配准矩阵,并计算当前时刻配准矩阵和全局配准矩阵在当前红外-可见光前景上的重叠率误差,计算方

式如式(2)所示:

$$E = 1 - \frac{Z(\mathbf{S}^I, \mathbf{M}) \cap \mathbf{S}^V}{Z(\mathbf{S}^I, \mathbf{M}) \cup \mathbf{S}^V}, \quad (2)$$

其中: $\mathbf{S}^I$ 表示红外前景图像, $\mathbf{S}^V$ 表示可见光前景图像, $\mathbf{M}$ 表示配准矩阵, Z 表示插值变换。此后,可以根据两个矩阵的重叠率误差判断如何根据当前配准矩阵更新全局配准矩阵,实现最终地配准。

3 改进的轮廓特征匹配式红外-可见光视频配准

3.1 原理框架

图3呈现了本文红外-可见光视频配准方法的基本流程。对于输入的红外-可见光视频序列,首先由PAWCS目标检测算法提取目标前景,并采用CSS角点检测算法在前景轮廓上提取配准特征点。此后,为每个特征点建立两个描述子:根据全局轮廓形状信息建立形状上下文描述子;根据局部轮廓方向信息建立边缘方向直方图描述子。接着,根据描述子之间的相似程度匹配红外-可见光特征点,并用一个基于高斯距离准则的特征匹配库保存不同时刻的特征点对。最后,采用RANSAC(Random Sample Consensus)算法计算

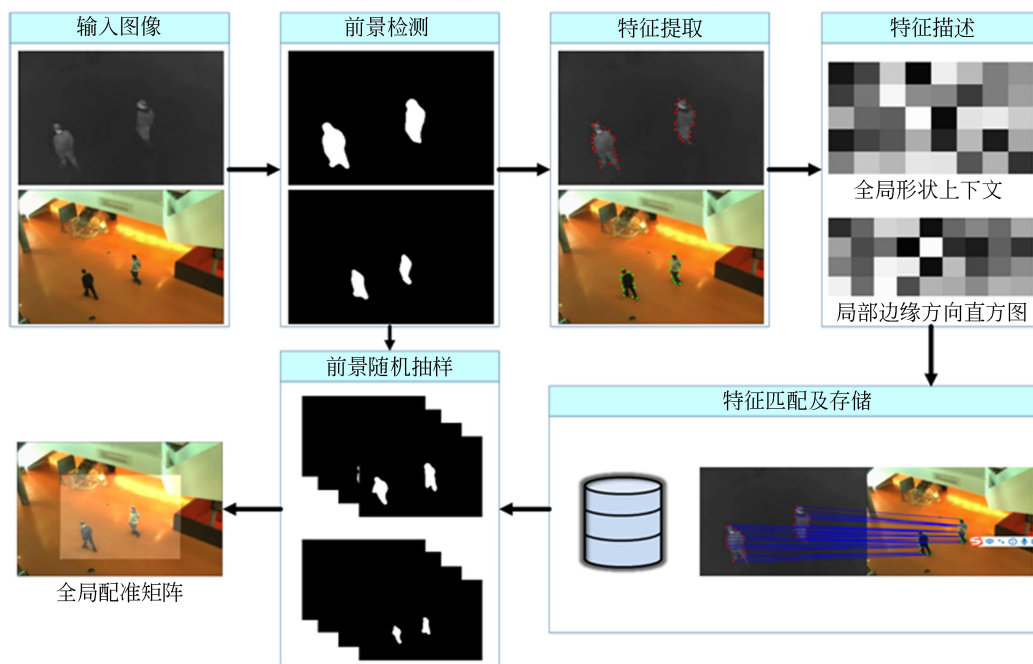


图3 本文配准算法框架示意图

Fig. 3 Framework of proposed registration method

当前时刻的配准矩阵。对不同时刻的前景样本进行随机抽样选取,根据在抽取样本上的前景重叠率误差判断是否更新全局配准矩阵,完成对红外-可见光视频序列的配准。在后续的章节中,会对方法进行详细介绍。

3.2 特征描述

论文首先利用 PAWCS 算法检测目标前景,并从前景轮廓上提取 CSS 特征点。此后,为了保证特征匹配质量,需要进行特征描述。但轮廓特征点来源于二值化前景图像,一般的描述方法并不可靠。为此,论文建立两种新的特征描述子。

3.2.1 全局形状上下文描述子

当红外-可见光相机的拍摄视角差异不是十分明显时,在两种图像中目标的轮廓形状具有相似性。论文采用形状上下文算法统计目标轮廓形状的空间分布,建立全局形状上下文描述子。对于第 k 个 CSS 特征点,首先计算所有轮廓点与该点之间的距离值和角度值,计算公式如下:

$$d_i^k = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2}, \quad \theta_i^k = \arctan((y_k - y_i)/(x_k - x_i)), \quad (3)$$

其中: (x_k, y_k) 和 (x_i, y_i) 分别表示第 k 个特征点和第 i 个轮廓点的位置, d_i^k 和 θ_i^k 分别表示两者之间的距离值与角度值。

在统计了所有轮廓点的空间分布后,需要进行归一化编码处理来保证描述子的尺度不变性和旋转不变性。选取距离当前特征点最远的轮廓点作为归一化编码的参考点,距离归一化编码的公式如式(4)所示:

$$\tilde{d}_i^k = \text{floor}(5d_i^k/d_{\max} - 10^{-6}), \quad (4)$$

其中: $d_{\max}$ 表示参考点到当前特征点的距离, floor 表示向下取整函数, 10^{-6} 用于防止编码越界。 $\tilde{d}_i^k$ 是距离归一化编码的结果,取值为 0, 1, ..., 4。对于角度方向,首先进行归一化处理,公式为:

$$\bar{\theta}_i^k = \begin{cases} \theta_i^k - \theta_{\max} & \theta_i^k \geq \theta_{\max} \\ \theta_i^k - \theta_{\max} + 2\pi & \text{otherwise} \end{cases}, \quad (5)$$

其中: $\theta_{\max}$ 表示参考点相对于当前特征点的角度值, $\bar{\theta}_i^k$ 表示角度归一化的结果。此后对归一化角度值进行量化编码,编码公式如式(6)所示:

$$\check{\theta}_i^k = \text{floor}(4 \cdot \bar{\theta}_i^k / \pi) + 1. \quad (6)$$

在量化编码后,归一化角度编码值只能取 1, 2, ..., 8 等 8 个整数值。最后,对所有轮廓点的距

离编码和角度编码进行直方图统计,为当前特征点建立一个 40 维的形状上下文描述子,表示如下:

$$\mathbf{H}_i = P(i), i = 1, 2, \dots, 40, \quad (7)$$

其中: $i = 8\tilde{d}_i^k + \check{\theta}_i^k$, 表示由距离编码值和方向编码值确定的索引值, $P(i)$ 表示轮廓点具有 i 索引的概率值。

3.2.2 局部边缘方向直方图描述子

运动目标检测的结果存在着一定的误差,导致仅利用轮廓形状的全局空间分布可能会误匹配邻近特征点。因此,论文同时根据特征点邻域的轮廓方向分布描述特征,建立局部边缘方向直方图描述子。为剔除尺度缩放和旋转变换的影响,首先对红外特征的邻域轮廓点 (x, y) 进行重新定位,方式如下:

$$\begin{bmatrix} \tilde{y} \\ \tilde{x} \end{bmatrix} = s \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix}, \quad (8)$$

其中: $s = d_{\max}^{\text{vis}}/d_{\max}^{\text{ir}}$ 表示尺度缩放因子, $d_{\max}^{\text{ir}}$ 和 $d_{\max}^{\text{vis}}$ 分别表示红外特征点之间的最大距离和可见光特征点之间的最大距离。 $\theta = \theta^{\text{vis}} - \theta^{\text{ir}}$ 表示旋转因子, θ^{ir} 和 θ^{vis} 分别表示红外与可见光图像中距离最远的两个特征点之间的角度值。在重定位后,参考梯度方向直方图 (Histogram of Oriented Gradients, HOG) 的构建方式为每个特征点构建边缘方向直方图。但梯度方向直方图统计每个像素点的梯度信息,而边缘方向直方图只利用边缘轮廓点的梯度信息,这可以降低红外-可见光图像之间纹理差异的影响。边缘方向直方图是一个 32 维的描述子,表示如下:

$$\mathbf{E}_i = P(i), i = 1, 2, \dots, 32, \quad (9)$$

其中: i 表示方向索引值, $P(i)$ 表示方向索引值 i 出现的概率。

3.3 特征匹配及存储

3.3.1 特征匹配

论文根据已经建立的特征描述子匹配红外-可见光轮廓特征点。首先对两种描述子进行拼接,为每个特征点建立一个 72 维的综合描述子,表示如式(10)所示:

$$\mathbf{G}_i = G(i), i = 1, 2, \dots, 72. \quad (10)$$

该描述子的前 40 维表示全局形状上下文描述子,后 32 维表示局部边缘方向直方图描述子。此后,采用 χ^2 统计函数计算两个特征点的匹配程

度,公式如式(11)所示:

$$C(\mathbf{G}_I(k), \mathbf{G}_V(k)) = \frac{1}{2} \sum_{k=1}^{72} \frac{[\mathbf{G}_I(k) - \mathbf{G}_V(k)]^2}{\mathbf{G}_I(k) + \mathbf{G}_V(k)}, \quad (11)$$

其中 $\mathbf{G}_I(k)$ 和 $\mathbf{G}_V(k)$ 分别表示红外和可见光特征点的综合描述子。 χ^2 统计函数值越低,两个特征点的匹配程度越高。在匹配过程中,论文采用双向匹配策略。对于每个红外特征点,选取匹配程度最高的可见光特征点作为候选匹配点。同时,根据相同的方式为每个可见光特征点确定红外候选匹配点。若一对红外-可见光特征点互为候选匹配,则将它们视为最终的匹配点对。双向匹配策略可以有效地降低误匹配的概率,特征匹配的结果如图4所示。

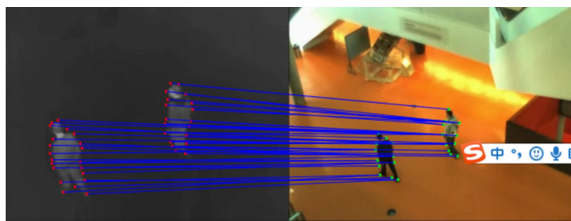


图4 特征匹配的结果

Fig. 4 Results of feature matching

3.3.2 特征存储

在视频序列的配准问题上,仅采用当前时刻的匹配点对难以实现高精度配准。目前,多种特征匹配库^[20-22]已经被用于保存来自不同时刻的特征点对,如先入先出(First in, First out)的匹配库^[20]。为了更有效地剔除外点,保留可靠的匹配点对,论文采用一种基于高斯距离准则的特征匹配库。首先计算每对匹配点之间的高斯距离,公式如式(12)所示:

$$\mathbf{E}_i = \exp\left(-\frac{d^2(Z(\mathbf{x}_I(i)), \mathbf{x}_V(i))}{\sigma^2} - \lambda \cdot C(\mathbf{G}_I(i), \mathbf{G}_V(i))\right), \quad (12)$$

其中: $\mathbf{x}_I(i)$ 和 $\mathbf{x}_V(i)$ 分别表示匹配点对中红外和可见光特征点的位置, Z 表示已经得到的全局配准矩阵(见章节3.4), d 表示欧氏距离函数。 $C(\mathbf{G}_I(i), \mathbf{G}_V(i))$ 表示该点对的匹配程度,计算方式如式(11)。 σ 表示高斯权重因子,一般取值为100。 λ 表示特征点距离和匹配度之间的平衡因子,一般取值为0.3。利用高斯距离准则可以综

合衡量匹配点对的匹配度以及其对当前配准参数的适应性,有利于提升匹配点对的质量。此后,当匹配库中的点对数量超过上限容量时(一般设置为500),将高斯距离最小的1/3视为外点点对。当获取新的匹配点对后,利用其随机替代外点点对,完成匹配库更新。

3.4 计算全局配准矩阵

根据不同时刻的特征点对,采用RANSAC算法即可计算当前时刻的配准参数。此后,可以基于当前时刻的前景重叠率误差更新全局配准矩阵,使全局配准矩阵时刻趋向于配准当前时刻目标。但在近平面场景中,运动目标在不同时刻可能存在深度差异,配准了当前时刻的目标并不意味着精确配准了整个视频序列,全局配准矩阵可能陷入了局部最优解。

为解决该问题,论文提出了前景样本随机抽样策略,核心思想是结合多个时刻的目标前景估计全局配准矩阵,避免其陷入局部最优。该策略首先建立一个前景样本库,保存不同时刻的红外-可见光前景样本对。当库中样本对的数量达到上限时(一般设置为50),根据前一时刻的全局配准矩阵对所有红外前景样本进行插值变换。在变换后,计算红外-可见光前景样本对的重叠率误差,将误差值最大的1/3视为外点样本对。在得到新的前景样本对后,利用其随机替换库中的外点样本对,完成对前景样本库的更新。

在配准矩阵的迭代更新过程,首先从样本库中随机抽取部分前景样本计算当前时刻配准矩阵和全局配准矩阵的重叠率误差,公式如下:

$$E = 1 - \frac{1}{N} \sum_{i=1}^N \frac{Z(\mathbf{F}_I^i, \mathbf{M}) \cap \mathbf{F}_V^i}{Z(\mathbf{F}_I^i, \mathbf{M}) \cup \mathbf{F}_V^i}, \quad (13)$$

其中: Z 表示插值变换, $\mathbf{M}$ 表示当前时刻配准矩阵或全局配准矩阵, N 表示抽取前景样本对的数量(一般设置为10)。 $\mathbf{F}_I^i$ 和 $\mathbf{F}_V^i$ 分别表示红外和可见光前景样本。此后,若当前时刻配准矩阵的重叠率误差小于全局配准矩阵,则对全局配准矩阵进行更新,更新方式如下:

$$\mathbf{M}_p = (1 - \beta) \cdot \mathbf{M}_p + \beta \cdot \mathbf{M}_c, \quad (14)$$

其中: $\mathbf{M}_p$, $\mathbf{M}_c$ 分别表示全局配准矩阵和当前时刻配准矩阵。 β 表示更新步长,一般取值为0.25。通过引入前景样本随机抽样策略,全局配准矩阵可以更准确地配准整个视频序列。

为了清晰地呈现本文方法的主要思想,算法 1 给出了本文方法的基本流程。

算法 1 本文算法的整体配准流程

Algorithm. 1 Integral registration flow of our method

算法 1 配准流程

输入: 红外-可见光视频序列 $\{IR_i, VIS_i\}$ 。

输出: 全局配准矩阵 M_p

步骤:

循环: for $i=1:length(IR_i)$

1. 从当前红外-可见光图像 IR_i, VIS_i 上提取目标前景 F_i^I, F_i^V 。

2. 采用 CSS 角点提取算法从前景 F_i^I, F_i^V 上提取轮廓特征点 $\{K_i^I, K_i^V\}$ 。

3. 为每个特征点建立全局形状上下文描述子 H 和局部边缘梯度方向直方图描述子 E 。

4. 采用 χ^2 统计函数匹配红外-可见光轮廓特征点。

5. 建立特征匹配库,并基于高斯距离准则(公式(10))进行更新。

6. 建立前景样本库,并基于重叠率误差进行更新。

7. 对样本库中的前景样本进行随机采样,根据采样结果计算当前时刻配准矩阵 M_c 和全局配准矩阵 M_p 的重叠率误差 E_c 和 E_p ,更新全局配准矩阵。

循环结束

4 实验测试与结果分析

4.1 实验设置

4.1.1 数据集

为了验证本文方法的精确度和鲁棒性,实验在公开通用的 LITIV 视频配准数据集上进行。该数据集由 Torabi 等人提供^[19],包含 9 组红外-可见光视频序列。视频分辨率为 240×320 ,视频长度在 300~1 300 frame 之间。由于成像传感器与被拍摄场景之间的距离较远,所有的图像均满足近平面假设条件。此外,数据集为每组视频序列提供了手动标定的 Ground-Truth 配准矩阵。在衡量本文方法的配准表现时,可以将该矩阵视为基准参考,从而体现出方法的精确性。

4.1.2 对比算法

实验将本文算法与其他两种先进的配准算法进行比较,进而突出本文算法的优势。第一种算

法是 Charles 等人提出的基于轮廓特征的配准算法^[21]。该算法将所有目标轮廓点作为配准特征,并利用目标形状信息对特征进行描述。第二种算法是王洪庆等人提出的基于目标运动轨迹匹配的配准算法^[18]。该算法通过采用多目标跟踪技术确定每个目标的运动轨迹,并为每条轨迹建立归一化运动距离描述子和归一化运动方向描述子。对比算法皆采用 RANSAC 算法确定当前时刻的配准参数,并根据当前时刻前景上的重叠率误差更新全局配准矩阵。

4.1.3 评价指标

为了量化本文算法和对比算法的配准能力,实验采用多边形重叠率误差作为配准精度的评价指标。首先需要为每对红外-可见光视频序列构建二值化多边形图像对,方法如 Torabi 等人^[19]所述:从红外和可见光图像中人工地选取一些易于识别的关键点并进行手动匹配,按照相同的顺序分别依次连接这些关键点,得到二值化多边形图像。实际上,Charles 等人^[21]已经为每组视频序列建立了二值化多边形图像,第一组视频序列的多边形图像如图 5 所示。实验直接采用这些多边形图像计算配准误差。对于二值化多边形图像的重叠率误差,计算公式同前景图像的重叠率误差,如公式(2)所示。区别是采用二值化多边形图像代替了前景图像,避免了前景检测误差的影响。

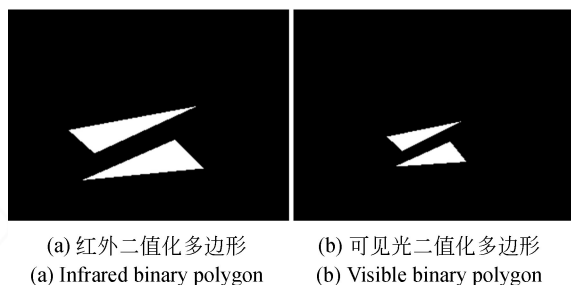


图 5 LITIV-1 序列上的二值化多边形

Fig. 5 Binary polygons on LITIV-1 sequence

4.2 结果分析

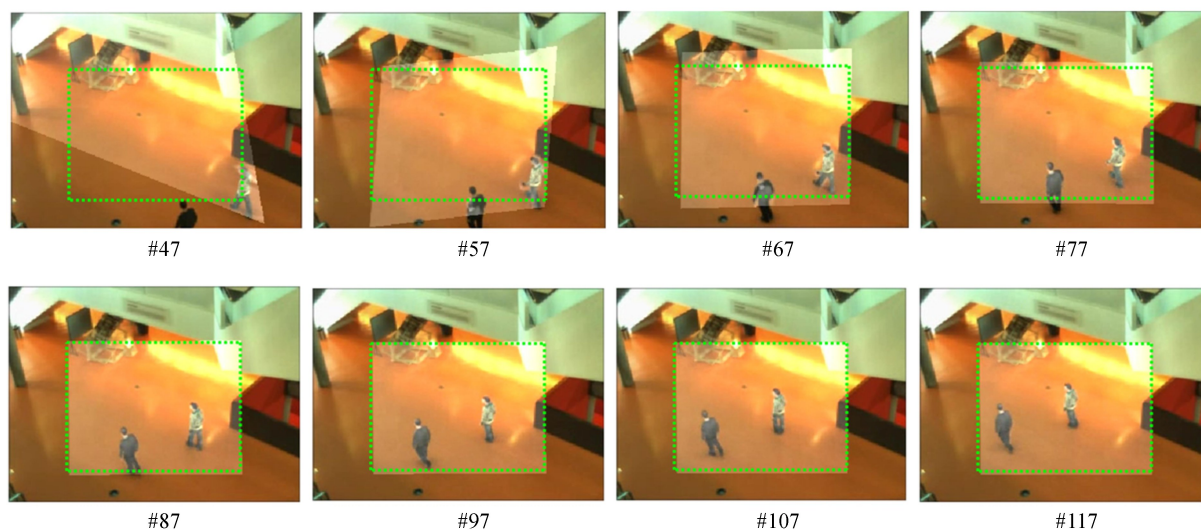
论文方法对 LITIV 数据库中前 3 个视频序列的配准结果如图 6 所示。实验采用全局配准矩阵对红外图像进行插值变换,并将其叠加在可见光图像上,表示最终的配准结果。此外,根据

Ground-Truth 矩阵对红外图像的边界进行重定位,并以绿色矩形框的形式叠加在可见光图像上,表示 Ground-Truth 矩阵的配准结果。

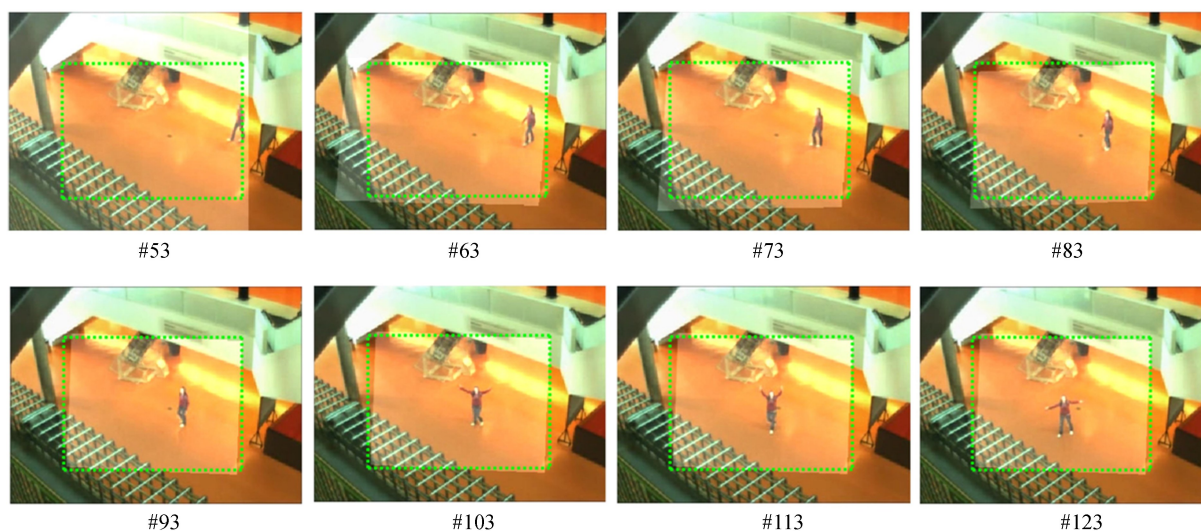
在每个序列的结果图像中,第一幅图像给出了目标初次同时进入红外-可见光视场时的配准结果。可以看出,本文方法并不能立刻精确地配准两幅图像。这是因为此时匹配点对的数量较少,且红外-可见光目标前景存在较大的空间差异。但随着算法的运行,匹配点对的数量会急剧增多,目标的空间分布也会逐渐趋于一致。此时,论文方法输出的配准参数能够迅速收敛,配准结果十分接近于 Ground-Truth 矩阵。这说

明本文方法可以快速有效地配准红外-可见光视频序列。

同时,可以看出本文方法在不同序列上的收敛速度有所差异。对于前两个视频序列,本文提出的方法在目标进入视场后的 40 帧内即可获得准确的全局配准矩阵。但对于第 3 个视频序列,在目标进入视场后需要 80 帧左右的时间才能够计算出可靠的配准参数。引起这一差异的主要原因是目标运动的速度差异和区域差异。在第三个视频序列中,目标的运动速度较慢,且运动区域始终局限在一个较小的空间范围内,导致得到的匹配点对不具有全局代表性。



(a) LITIV-1 视频序列上的配准结果
(a) Registration results on LITIV-1



(b) LITIV-2 视频序列上的配准结果
(b) Registration results on LITIV-2

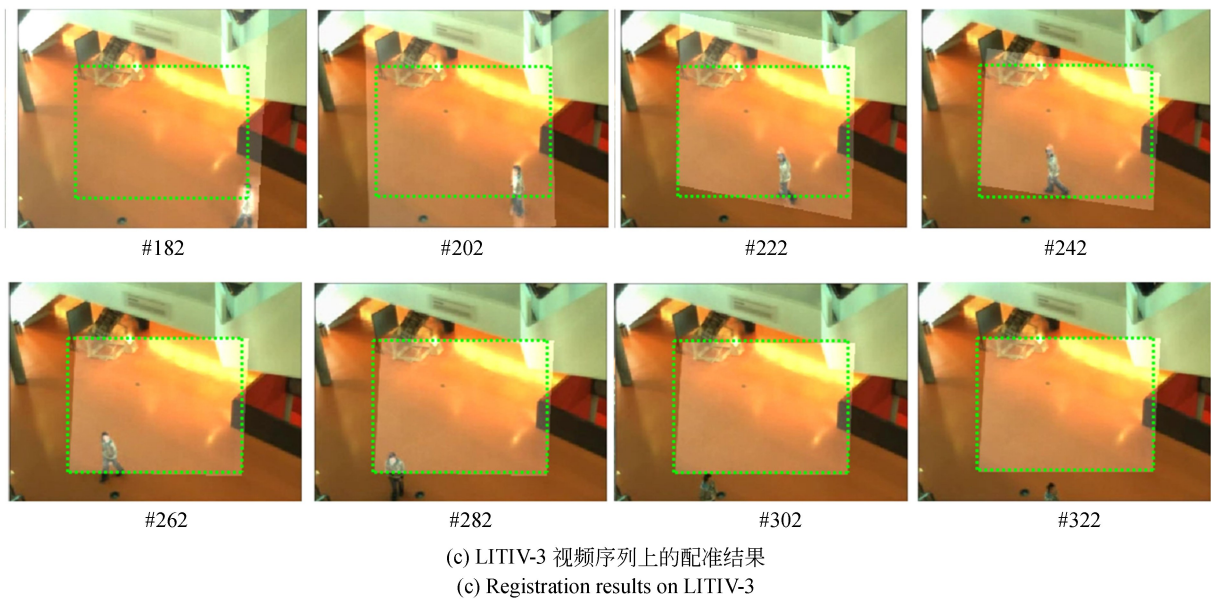


图 6 LITIV 数据集上的配准结果(前 3 个视频对)

Fig. 6 Registration results on LITIV dataset (for the first three video pairs)

为了定量地评价本文方法在各个视频序列上的整体表现,表 1 给出了本文方法和其他对比方法的平均重叠率误差。分析得知,本文方法在 9 个视频序列上的平均重叠率误差仅为 0.194,与对比方法相比下降了 18.5%,充分表明了本文方法的精度优势。

表 1 平均重叠率误差

Tab. 1 Average overlap error

Sequences	Proposed	Wang ^[18]	Charles ^[21]	Ground-truth
LITIV-1	0.248	0.365	0.266	0.149
LITIV-2	0.168	0.325	0.205	0.078
LITIV-3	0.172	0.308	0.193	0.080
LITIV-4	0.217	0.302	0.312	0.221
LITIV-5	0.149	0.229	0.267	0.150
LITIV-6	0.235	0.375	0.413	0.088
LITIV-7	0.118	0.182	0.257	0.136
LITIV-8	0.258	0.318	0.204	0.260
LITIV-9	0.179	0.203	0.185	0.134

下面详细介绍本文方法在各个序列上的配准表现。对于序列 LITIV-4,LITIV-5,LITIV-6 和 LITIV-7,本文方法显著降低了配准误差,下降比例分别为 28.1%,34.9%,37.4%和 35.2%。对

于序列 LITIV-1,LITIV-2,LITIV-3 和 LITIV-9,本文方法也较为明显地提升了配准精度,提升比例分别为 6.8%,18.0%,10.9%和 3.2%。这充分证明了本文方法的有效性:通过建立全局形状上下文描述子和局部边缘方向直方图描述子,有效地提升了特征匹配质量;通过在计算全局配准矩阵的过程中引入前景样本随机抽样策略,有效地避免了全局配准参数陷入局部最优解。与其他视频序列相比,本文方法在序列 LITIV-4,LITIV-5,LITIV-6 和 LITIV-7 上对配准精度的提升更为显著。这主要是因为在这些序列上运动目标的深度变化更为明显,而本文方法在计算配准参数时能够充分降低深度变化带来的影响。

对于序列 LITIV-8,本文方法的配准精度不如 Charles 等人提出的对比方法^[21]。导致此情况的主要原因是该序列的前景检测结果质量较差,严重影响了本文方法在特征提取和匹配时的可靠性。Charles 等人的算法由于直接将所有目标轮廓点视为配准特征,同时仅利用目标形状信息对特征进行描述,所以能够达到更高的配准精度。

本文方法和对比方法在各个视频序列上能够达到的最小重叠率误差如表 2 所示。实际上,实验采用的两种对比算法都是根据当前时刻的前景重叠率误差估计全局配准矩阵,而计算配准误差所用的二值化多边形同样是建立在某一时刻的前

景上。因此在理论上,对比算法应该可以达到更低的最小重叠率误差。但从实验结果可以看出,本文算法在多数视频序列上取得了更低的最小重叠率误差。这一现象充分证明了本文方法的特征匹配质量更高,保证了配准参数的精确性。

表2 最小重叠率误差

Tab. 2 Minimum overlap error

Sequence	Proposed	Wang ^[18]	Charles ^[21]	Ground-truth
LITIV-1	0.158	0.165	0.187	0.149
LITIV-2	0.073	0.100	0.106	0.078
LITIV-3	0.088	0.113	0.108	0.080
LITIV-4	0.121	0.127	0.118	0.221
LITIV-5	0.116	0.103	0.172	0.150
LITIV-6	0.074	0.097	0.609	0.088
LITIV-7	0.054	0.055	0.901	0.136
LITIV-8	0.180	0.197	0.137	0.260
LITIV-9	0.121	0.097	0.095	0.134

根据表1和表2给出的配准误差,可以发现论文方法在某些序列上的精度超过了 Ground-Truth 矩阵。这主要是因为 Ground-Truth 矩阵是通过手动挑选红外-可见光关键点构建的,其存在着不可避免的人为误差。此外,Ground-Truth 矩阵只能够完美地配准平面场景图像,无法克服

近平面场景中目标深度变化的影响。以上因素也导致 Ground-Truth 矩阵在各个序列上的配准误差并不为0。同时,由于本文方法和对比方法都是根据前景重叠率误差迭代更新全局配准参数,往往能够更好地适应目标的深度变化,从而比 Ground-Truth 矩阵取得更高的配准精度。

5 结 论

为了提高近平面场景下红外-可见光视频序列配准的精确度与鲁棒性,本文提出了一种基于轮廓特征匹配的自动配准方法。该方法在特征描述匹配和全局配准矩阵计算两个方面对传统轮廓特征配准算法进行了改进,提升了配准精度。在特征匹配阶段,本文方法为每个轮廓特征点建立了全局形状上下文描述子和局部边缘方向直方图描述子,利用全局轮廓形状信息和局部轮廓方向信息提升特征匹配质量。此后,在计算全局配准矩阵的过程中建立了前景样本随机采样策略,结合多个时刻的红外-可见光前景样本克服目标深度变化的影响,避免全局配准矩阵陷入局部最优解。实验结果显示,本文方法在9个测试视频上的平均重叠率误差仅为0.194,与其他两种先进的配准方法相比下降了18.5%。论文提出的方法基本满足近平面场景下红外-可见光视频序列自动配准的精度要求,且具有较高的鲁棒性。

参考文献:

- [1] YANG Y, NIE Z, HUANG S, *et al.*. Multi-level features convolutional neural network for multi-focus image fusion [J]. *IEEE Transactions on Computational Imaging*, 2019, 5(2): 262-273.
- [2] CHANG NB, BAI K, IMEN S, *et al.*. Multisensor satellite image fusion and networking for all-weather environmental monitoring [J]. *IEEE Systems Journal*, 2018, 12(2): 1341-1357.
- [3] EL-HOSSENY HM, RABAIE ESME, ELRAHMAN WA, *et al.*. Medical image fusion techniques based on combined discrete transform domains [C]. *IEEE 34th National Radio Science Conference*,

Alexandria, Egypt: NRSC, 2017: 471-480.

- [4] MA J, MA Y, LI C. Infrared and visible image fusion methods and applications: A survey [J]. *Information Fusion*, 2018, 45: 153-178.
- [5] KROTOSKY SJ, TRIVEDI MM. Person surveillance using visual and infrared imagery [J]. *IEEE Transactions on Circuits & Systems for Video Technology*, 2008, 18(8): 1096-1105.
- [6] MA J, ZHAO J, MA Y, *et al.*. Non-rigid visible and infrared face registration via regularized Gaussian fields criterion [J]. *Pattern Recognition*, 2015, 48(3): 772-784.
- [7] ZITOVA B, FLUSSER J. Image registration methods: a survey [J]. *Image & Vision Computing*, 2003, 21(11): 977-1000.

- [8] VIOLA P, III W M W. VIOLA PA, WELLS WM. Alignment by maximization of mutual information [J]. *International Journal of Computer Vision*, 1997, 24(2): 137-154.
- [9] ROCHE A, MALANDAIN G, PENNEC X, *et al.*. The correlation ratio as a new similarity measure for multimodal image registration [C]. *Proceedings of the Springer International Conference on Medical Image computing and Computer-Assisted Intervention*, Cambridge, MA, USA, 1998: 1115-1124.
- [10] LEGG P A, RDSIN P L, MARSHALL D, *et al.*. Feature neighbourhood mutual information for multi-modal image registration: an application to eye fundus imaging [J]. *Pattern Recognition*, 2015, 48(6): 1937-1946.
- [11] 车芳, 韩俊刚, 郭志全. SMT-PAAG 下的 Harris 角点检测与匹配算法 [J]. *电子技术应用*, 2017, 43(4): 138-140.
- CHE F, HAN J G, GUO ZH Q. Harris corner detection and matching algorithm at the SMT-PAAG [J]. *Application of Electronic Technique*, 2017, 43(4): 138-140. (in Chinese)
- [12] 耿庆田, 赵浩宇, 王宇婷, 等. 基于改进 SIFT 特征提取的车标识别 [J]. *光学 精密工程*, 2018, 26(5): 269-276.
- GENG Q T, ZHAO H Y, WANG Y T, *et al.*. A vehicle logo recognition algorithm based on the improved SIFT feature [J]. *Opt. Precision Eng.*, 2018, 26(5): 269-276. (in Chinese)
- [13] SONG Z L, ZHANG J. Remote sensing image registration based on retrofitted SURF algorithm and trajectories generated from lissajous figures [J]. *IEEE Geoscience & Remote Sensing Letters*, 2010, 7(3): 491-495.
- [14] KLIMASZEWSKI J, KONDEJ M, KAWECKI M, *et al.*. *Registration of Infrared and Visible Images Based on Edge Extraction and Phase Correlation Approaches* [M]. Berlin: Springer Berlin Heidelberg: Image Processing and Communications Challenges 4, 2013.
- [15] MING C, WONKA P, RAZDAN A, *et al.*. A new image registration scheme based on curvature scale space curve matching [J]. *Visual Computer*, 2007, 23(8): 607-618.
- [16] 于晓明, 李思颖, 史胜楠. 混合高斯融合三帧差的运动目标检测改进算法 [J]. *红外技术*, 2019, 41(3): 256-261.
- YU X M, LI S Y, SHI SH N. An improved algorithm for moving target detection using a gaussian mixture with three-frame difference [J]. *Infrared technology*, 2019, 41(3): 256-261. (in Chinese)
- [17] ST-CHARLES P L, BILODEAU G A, BER-GEVIN R. A self-adjusting approach to change detection based on background word consensus [C]. *IEEE Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA: IW-CACV, 2015: 990-997.
- [18] 王洪庆, 许廷发, 孙兴龙, 等. 目标运动轨迹匹配式的红外-可见光视频自动配准 [J]. *光学 精密工程*, 2018, 26(6): 1533-1541.
- WANG H Q, XU T F, SUN X L, *et al.*. Infrared-visible video registration with matching motion trajectories of targets [J]. *Opt. Precision Eng.*, 2018, 26(6): 1533-1541.
- [19] TORABI A, MASSE G, BILODEAU G A. Feed-back scheme for thermal-visible video registration, sensor fusion, and people tracking [J]. *Journal of Biomedical Materials Research Part A*, 2010, 2(1): 15-22.
- [20] SONN S, BILODEAU GA, GALINIER P. Fast and accurate registration of visible and infrared videos [C]. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Portland, OR, USA: CVPRW, 2013: 308-313.
- [21] ST-CHARLES P L, BILODEAU G A, BER-GEVIN R. Online multimodal video registration based on shape matching [C]. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Boston, MA, USA: CVPRW, 2015: 26-34.
- [22] SUN X, XU T, ZHANG J, *et al.*. A hierarchical framework combining motion and feature information for infrared-visible Video registration [J]. *Sensors*, 2017, 17(2): 384.
- [23] 樊彦国, 柴江龙, 许明明, 等. 基于 ORB 与 RANSAC 融合改进的图像配准 [J]. *光学 精密工程*, 2019, 27(3): 197-212.

FAN Y G, CHAI J L, XU M M, *et al.*. Improved fast Image registration algorithm based on ORB and

RANSAC fusion [J]. *Opt. Precision Eng.*, 2019, 27(3): 197-212. (in Chinese)

作者简介:



孙兴龙(1993—),男,吉林长春人,博士研究生,研究实习员,2018年于北京理工大学获得硕士学位,主要从事异源图像配准融合、目标检测跟踪等方面的研究。E-mail: sunxinglong1123@163.com

导师简介:



郭立红(1964—),女,吉林舒兰人,博士,研究员,博士生导师,1986年于长春光学精密机械学院获得学士学位,1999年和2003年于中科院长春光机所分别获得硕士学位和博士学位,研究方向为光电系统总体集成技术、计算机智能仿真技术等。E-mail: guolh@ciomp.ac.cn

通讯作者:



韩广良(1968—),男,山东嘉祥人,博士,研究员,博士生导师,1990年于内蒙古工业大学获得学士学位,2000和2003年于中科院长春光机所分别获得硕士学位和博士学位,研究方向为图像和视频信息处理、计算机视觉、机器视觉与人工智能等。E-mail: hangl@ciomp.ac.cn

超星