

## 基于特征点匹配的视频稳定算法

马忠雪, 穆平安, 戴曙光

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘要:** **目的** 解决检测摄像机在拍摄过程中因外界抖动导致视频扭曲失真等问题。**方法** 选取视频第 1 帧作为基准帧, 检测视频每帧的特征点, 并与第 1 帧特征点位置进行对比, 经过仿射变换消除每一帧的扰动。通过 FAST 算法生成每一帧特征点, 利用 BRISK 算法得到特征点的二进制特征描述符, 比较特征点描述符的汉明距离来匹配相邻两帧间的特征点, 利用匹配好的特征点对确定仿射变换模型, 最后利用得到的仿射变换矩阵对每一帧进行校正。**结果** 实验原视频总帧数为 23 帧, 时长为 1 s, 分辨率为 320×240, 视频中除第 1 帧外, 其余 22 帧均得到修正。**结论** FAST 角点检测算法与 BRISK 匹配方法结合可以有效解决视频扭曲等问题。

**关键词:** 图像处理; 视频稳定; 运动校正; 特征点检测; 仿射变换

**中图分类号:** TP391.4 **文献标识码:** A **文章编号:** 1001-3563(2018)21-0229-06

**DOI:** 10.19554/j.cnki.1001-3563.2018.21.039

## Video Stabilization Algorithm Based on Feature Point Matching

MA Zhong-xue, MU Ping-an, DAI Shu-guang

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**ABSTRACT:** The paper aims to solve the problem that the detection camera is distorted due to external jitter during the shooting process. The first frame of the video was selected as the reference frame, and the feature points of each frame of the video were detected. Compared with the position of the feature points of the first frame, and the perturbation of each frame was eliminated by affine transformation. The FAST algorithm was used to generate frame feature point. The BRISK algorithm was used to obtain the binary feature descriptor of the feature point. The Hamming distance of the feature point descriptor was compared to match the feature points between the adjacent two frames; and the affine transformation model were determined with matched feature point pairs. The affine transformation model was finally corrected for each frame with the obtained affine transformation matrix. The total number of frames of the original video was 23 frames, the duration was 1 second, and the resolution was 320×240. All frames except the first frame were corrected. The FAST corner detection algorithm combined with the BRISK matching method can effectively solve problems such as video distortion.

**KEY WORDS:** image processing; video stabilization; motion correction; point-feature detecting; affine transformation

随着市场对产品质量要求的提升, 以及生产成本的不断攀升, 自动化生产技术得到广泛应用。目前大部分产品包装产线也提高了自动化程度, 使用了图像检测技术以检查加工包装的产品是否出现缺陷。由于图像采集设备自身应用环境的影响, 拍摄的视频往往受到不稳定的、无规律抖动的影响, 这对图像采集设备产生了极大的干扰, 降低了图像采集的质量, 同时

降低了检测系统的鲁棒性, 因此需要做进一步的视频稳定处理, 以得到稳定的视频图像, 便于后续检测。

视频稳定是指利用相关算法对视频设备采集到的原始视频进行处理、校正达到去除抖动的效果。视频去抖增稳可作为其他后续处理的预处理措施, 比如缺陷检测和跟踪等。

经典的视频稳定算法一般包括两大步骤: 摄像机

收稿日期: 2018-06-17

作者简介: 马忠雪 (1993—), 男, 回族, 上海理工大学硕士生, 主攻机器视觉检测技术。

通信作者: 穆平安 (1964—), 男, 硕士, 上海理工大学教授, 主要研究方向为智能与虚拟仪器技术、机器视觉检测技术。

运动估计和摄像机运动校正<sup>[1-3]</sup>，也称为全局运动估计与运动补偿。全局运动实际上就是指视频的背景运动，主要为旋转、比例缩放与平移变换，全局运动估计的目的是获得相邻帧间的背景图像的变化量，为运动补偿提供所需参数；运动补偿则是根据全局运动估计获得的转换参数对全局运动进行修正，更确切地说就是对每一帧图像进行平移、旋转等仿射变换从而达到视频顺滑稳定的目的。

视频稳定算法最为核心的部分就是特征点的检测与匹配。特征点检测方法目前有很多种，SIFT<sup>[4-6]</sup>，SURF<sup>[7-8]</sup>提取到的特征点非常优秀（有较强的不变性），但是时间消耗依然很大<sup>[9]</sup>，从系统的整体角度考虑，对目标进行特征提取只是一小部分，此外要进行配准、提纯等后续措施。某一部分的时间消耗增大，将降低系统整体性能。

文中基于 FAST 算法的特征点检测，利用 BRISK<sup>[10]</sup>算法获得特征点的描述符，算法速度较好，校正效果也比较出色。校正之后将所有校正帧导出成视频，从而获得消除抖动的稳定视频。

### 1 全局运动估计

完整的视频稳定流程见图 1，其中全局运动估计作为下面所有步骤的预处理流程。全局运动是由于视频设备因受到外界扰动，视频背景图像出现相应的抖动。全局运动估计就是对这些背景图像的抖动进行估计。

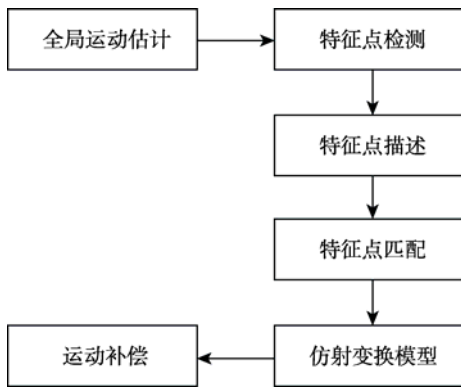


图 1 视频稳定流程  
Fig.1 Video stabilization process

#### 1.1 基于 FAST 角点检测算法的特征点获取

FAST 角点的定义为：若某像素点与其周围邻域内足够多的像素点相差较大，则该像素可能是角点<sup>[11]</sup>。角点也即特征点。

如图 2 所示，以某一点  $p$  为圆心，以 3 个像素点距离为半径，得到像素点  $p_1, p_2 \dots p_{16}$ 。根据情况设定一个阈值，计算  $p_1, p_9$  与中心  $p$  的像素值差，绝对值都小于阈值，则  $p$  点不可能是特征点；否则，做为候选点，等待下一步计算；若  $p$  是候选点，则计算  $p_1, p_9$ ,

$p_5, p_{13}$  与中心  $p$  的像素值差，若绝对值有至少 3 个超过阈值，则做为候选点，再进行下一步计算；否则，舍弃；若  $p$  是候选点，计算  $p_1 \dots p_{16}$  这 16 个点与  $p$  中心的像素值差，若至少 9 个超过阈值，则为特征点；否则，舍弃<sup>[12]</sup>。

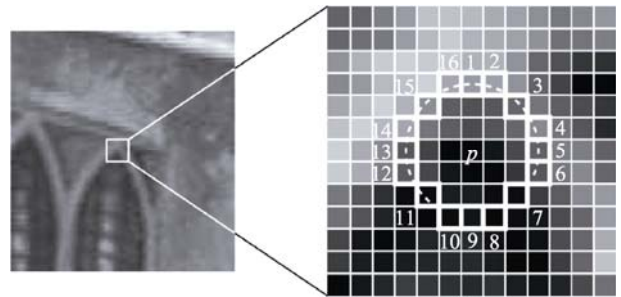


图 2 FAST 特征点  
Fig.2 FAST feature points

之后对图像进行非极大值抑制，对特征点的 FAST 得分值（即 score 值，也即  $s$  值）进行计算，在特征点  $p$  为中心的一个邻域（ $3 \times 3, 5 \times 5$  等）内，计算每个特征点的  $s$  值（16 个点与中心差值的绝对值总和）。如果  $p$  的响应值在该邻域内的特征点中最大，保留该特征点；否则，舍弃。若邻域内只有一个特征点（角点），则保留。得分计算公式为（ $v$  为得分， $p_v$  为像素点灰度值）：

$$v = \max \sum |p_v - p| \tag{1}$$

根据经验，阈值一般设置为 50。也可根据具体情况进行加减，加减幅度应根据结果进行更改。

视频中截取的相邻 2 帧见图 3，首先对进行灰度处理，因为算法不需要颜色，针对 2 帧图像利用 FAST 算法得到图像的特征点，其中阈值设置为 50。

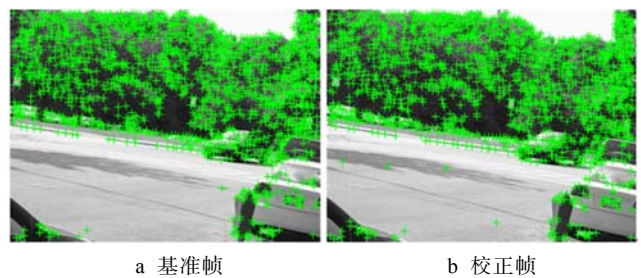


图 3 特征点提取  
Fig.3 Extraction of feature point

在图 3a 为第 1 帧且作为基准帧，图 3b 为第 2 帧并将基于第 1 帧的特征点位置进行校正。

#### 1.2 基于 BRISK 算法的特征点匹配

利用 BRISK 算法来进行特征点的描述，具有较好的旋转不变性、尺度不变性、鲁棒性等<sup>[13]</sup>。

首先利用图像金字塔对原图像进行多尺度表达（见图 4）构造  $n$  个 octave 层（用  $c_i$  表示）和  $n$  个

intra-octave 层 (用  $d_i$  表示),  $n=4, i=\{0,1\dots n-1\}$ 。假设有图像  $I$ , 其中, octave 层的产生:  $c_0$  层为原图像,  $c_1$  层通过对  $c_0$  层的 2 倍下采样生成,  $c_2$  层通过对  $c_1$  层的 2 倍下采样生成, 以此类推。

intra-octave 层:  $d_0$  层通过对  $I$  的 1.5 倍下采样生成,  $d_1$  层通过对  $d_0$  层的 2 倍下采样 (即  $I$  的  $2 \times 1.5$  倍下采样) 生成,  $d_2$  层通过对  $d_1$  层的 2 倍下采样生成, 以此类推<sup>[14-15]</sup>。

对每层图像进行空间上的非极大值抑制: 特征点在位置空间 (8 邻域点) 和尺度空间 (上下层  $2 \times 9$  个点), 共 26 个邻域点的 FAST 得分值要最大, 否则舍弃特征点。

对极值点的所在层和相邻上下层之中对应的像素点的 FAST 得分值作二维二次函数插值, 得到极值点的坐标位置; 之后在尺度方向上进行一维插值, 即得到极值点所对应的尺度以此作为特征尺度。

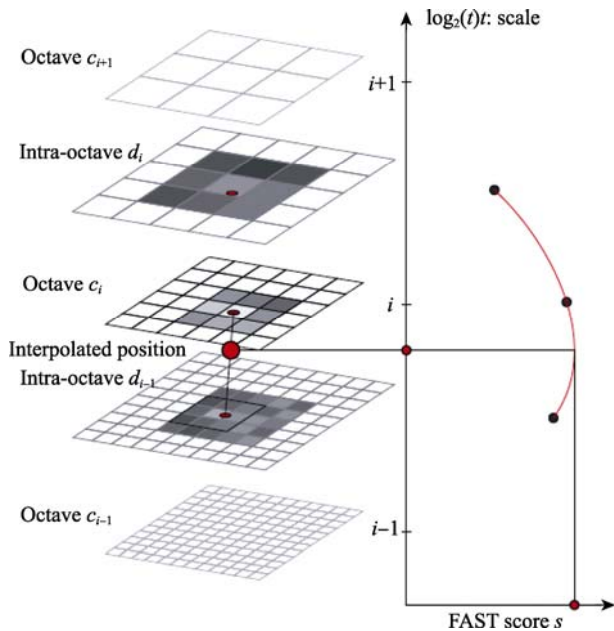


图 4 尺度空间特征点检测  
Fig.4 Feature point detection in scale space

以特征点为圆心, 做半径不同的同心圆, 在每个圆上获得一定数目等间隔采样点, 并对同心圆上采样点进行高斯滤波, 高斯滤波的中心为采样点, 方差设置为  $\delta$ , 滤波半径与高斯方差成正比, 最终使用的  $N$  个 ( $N$  值一般设为 60) 采样点是经过高斯平滑后的采样点。采样点两两一对, 所有组合方式通过下面的集合表示:

$$A = \left\{ (p_i, p_j) \in \mathbf{R}^2 \times \mathbf{R}^2 \mid i < N \wedge j < i \wedge i, j \in \mathbf{N} \right\} \quad (2)$$

式中:  $I(p_i, \sigma_i), I(p_j, \sigma_j)$  为像素,  $\delta$  为尺度。用  $g(p_i, p_j)$  表示局部梯度集合, 则有:

$$g(p_i, p_j) = (p_i - p_j) \cdot \frac{I(p_j, \sigma_j) - I(p_i, \sigma_i)}{\|p_j - p_i\|^2} \quad (3)$$

定义短距离点对和长距离点对:

$$S = \left\{ (p_i, p_j) \in A \mid \|p_j - p_i\| < \sigma_{\max} \right\} \subseteq A \quad (4)$$

$$L = \left\{ (p_i, p_j) \in A \mid \|p_j - p_i\| > \sigma_{\min} \right\} \subseteq A \quad (5)$$

接下来就可以按式 (6) 获得特征点的主方向:

$$g = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \frac{1}{L} \cdot \sum_{(p_i, p_j) \in L} g(p_i, p_j) \quad (6)$$

$$\delta = \arctan 2(g_y, g_x) \quad (7)$$

旋转特征点周围的采样区域至主方向, 对新的采样区域进行采样, 模式同上。BRISK 算法获得的特征描述子为二进制描述子, 上述过程得到的长、短距离子集, 考虑其中短距离子集的 512 个短距离点对, 进行二进制编码:

$$b = \begin{cases} 1, & I(p_j^\delta, \sigma_j) \forall (p_i^\delta, p_j^\delta) \in S \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

式中上标为经过  $\alpha$  度旋转之后得到的新采样点, 这样就得到了 512Bit 的二进制编码, 这个二进制编码就是特征点的描述符。

特征点配对利用汉明距离进行: 2 个特征编码对应比特位上相同元素的个数小于 128, 一定不配对; 一帧上特征点与另一帧上特征的编码按对应比特位上相同元素的个数最多的配成一对。

第 1 帧与第 2 帧 2 帧的复合图像见图 5,  $\circ$  为第 1 帧中的特征点, 十字为第 2 帧中的特征点。

图 5 显示了相邻 2 帧中特征点的对应关系。这样就获得了特征点对。由于此时尚未对第 2 帧进行校正, 因此复合图像中来自第 1 帧与第 2 帧的特征点的中心 (圆心和十字中心) 并没有重合。



图 5 特征点匹配  
Fig.5 Matching of feature point

### 1.3 利用特征点对建立仿射变换模型

仿射空间是数学中的几何结构。在仿射空间中, 点与点之间做差可以得到向量, 点与向量做加法将得到另一个点, 但是点与点之间不可以做加法。

仿射变换又称仿射映射, 是指在几何中, 一个向量空间进行一次线性变换并接上一个平移, 变换为另

一个向量空间。

一个向量  $\vec{x}$  按  $\vec{b}$  平移, 并按  $A$  进行旋转放大或缩小的仿射映射为:

$$\vec{y} = A\vec{x} + \vec{b} \quad (9)$$

在齐次坐标上可得到式 (10)。

$$\begin{bmatrix} y \\ 1 \end{bmatrix} = \begin{bmatrix} A & \vec{b} \\ 0, \dots, 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} \quad (10)$$

在摄像机拍摄过程中, 摄像机的光轴与平面垂直, 摄像机的平移和旋转会引起图像的平移和扭曲。它与该平面的距离的变换则会使视频帧与帧之间景物的比例发生变换。如果假定参考帧的中点  $(x, y)^T$  运动到下一帧的位置  $(x', y')^T$ , 则仿射变换模型为:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_3 \\ a_6 \end{bmatrix} \quad (11)$$

式中: 参数  $a_1, a_5$  为比例变换 (即缩放);  $a_2, a_4$  为旋转, 其余表示位移。

使用 Matlab 中的 EstimateGeometricTransform 函数来获得仿射变换矩阵, 这个函数在给定一组点对应关系时, 将搜索有效的内部对应关系。然后从中得出仿射变换, 使得第 1 组点的内点与第 2 点的内点最接近匹配。这个仿射变换将是一个  $3 \times 3$  矩阵的形式:

$$\begin{bmatrix} a_1 & a_2 & t_x \\ a_3 & a_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

参数  $a$  定义了变换的比例变换、旋转变换, 而参数  $t$  为平移参数。

## 2 根据参数矩阵进行运动校正

### 2.1 单帧校正结果

根据上述步骤得到的仿射变换模型与变换参数矩阵对图像进行校正:

$$T_c = T_p \cdot H_p \quad (13)$$

式中:  $T_p$  为原始视频帧;  $H_p$  为仿射变换矩阵;  $T_c$  为校正帧。

将第 2 帧根据第 1 帧校正之后与第 1 帧进行复合, 见图 6。第 2 帧图像经过仿射变换后, 与第 1 帧按点对点复合 (以第 1 帧的特征点位置为基准), 边缘出现缺失现象 (右侧与底侧的黑边)。可以看到校正后的复合图像里圆心与十字的中心重合。

### 2.2 多帧校正结果

给定 1 组视频, 视频设备在固定拍摄时受到了不稳定的扰动, 以致图像出现抖动, 第  $i$  帧  $T_i, i=1, 2, \dots$  与  $i+1$  帧  $T_{i+1}$  的仿射变换矩阵为  $H_i$ , 第  $i$  帧相对于第 1 帧的累积变换矩阵为:

$$H_{\text{cumulative}, i} = \prod_{n=1}^{i-1} H_n, i > 1 \quad (14)$$



图 6 校正后的复合图像  
Fig.6 Corrected composite image

第  $i$  帧  $T_i$  的校正帧  $T_{ci}$  为:

$$T_{ci} = T_i \cdot H_{\text{cumulative}, i} \quad (15)$$

原视频时长为 1 s, 帧率为 23, 分辨率为  $320 \times 240$ , 原视频的所有帧取平均值, 见图 7a, 校正后视频的所有帧取平均值见图 7b。这些图像在下面并列显示。在图 7a 中通过观察可以直观看出由于镜头的抖动, 原视频中存在大量的失真, 而通过图 7b 的校正帧的平均值图像可以直观看出视频的核心内容几乎没有丢失, 虽然出现了边缘丢失和前景目标模糊 (这是由于车辆持续移动的原因)。这也显示了稳定算法的功效。



图 7 稳定前后的复合图像  
Fig.7 Composite image before and after stabilization

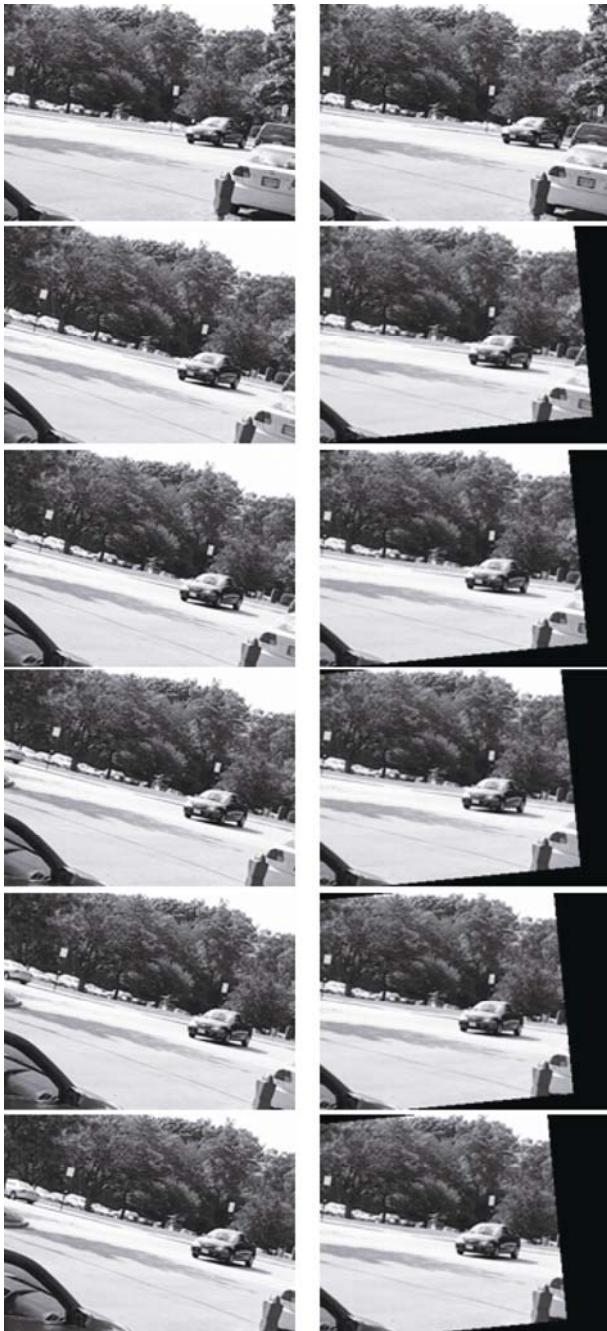
### 2.3 突变帧处理

文中算法以第 1 帧为参考帧, 继而对后续帧进行校正。当后续帧场景与第 1 帧场景完全不同时, 此时前后帧之间匹配好的特征点对数将会减少, 若低于上一次匹配的特征点对数的一半, 则把突变帧设定为基准帧, 对其后续帧进行校正, 校正方法仍然同上。

## 3 实验分析

### 3.1 实验方法

图 8 中原视频总帧数为 23 帧, 时长为 1 s, 分辨率为  $320 \times 240$ 。FAST 算法阈值为 60。在对原始帧进行校正时以第 1 帧为基准帧校正, 可以观察到原视频中由于外界的干扰使图像整体向右倾斜, 而校正后的帧每一帧的背景目标的位置与第 1 帧相比几乎无变化。



a 原始视频帧                      b 校正后的图像

图 8 校正结果比较

Fig.8 Comparison of calibration results

定量衡量稳定效果，采用相邻帧间 PSNR 作为指标，其反映了视频各帧经过处理后与原视频像素匹配程度。PSNR 越大代表效果越好。定义为：

$$MSE(I_{K+1}, I_K) = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N [I_{K+1}(x, y) - I_K(x, y)]^2 \quad (16)$$

$$PSNR = 10 \log_2 \left( \frac{255^2}{MSE} \right) \quad (17)$$

式中： $I_K(x, y)$  和  $I_{K+1}(x, y)$  分别为第  $K$  帧与

第  $K+1$  图像在  $(x, y)$  处的灰度值； $M, N$  为图像的长和宽。文中算法处理后与原视频 PSNR 对比见图 9（图中横坐标为帧序号，纵坐标为 PSNR 值）。

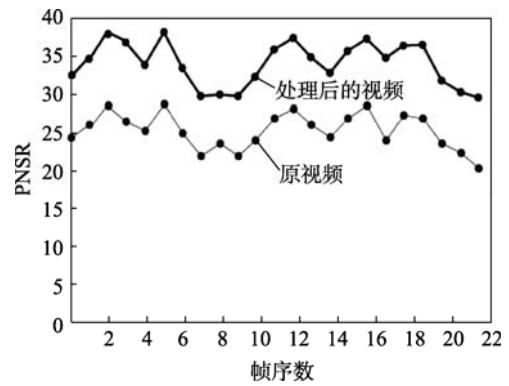


图 9 视频处理前后 PSNR 对比

Fig.9 PSNR comparison before and after video processing

#### 4 结语

通过 FAST 算法检测到特征点之后，再利用 BRISK 算法对特征点进行匹配，接下来利用仿射变换模型对视频的每一帧进行仿射变换，实验证明该算法能够较好地视频进行去抖动达到视频稳定的效果。仿射变换对于三维场景中的帧与帧之间的失真不太合适的。变换之后的视频没有对边缘进行补偿，因此校正后视频帧存在边缘缺失，有待深入研究。

#### 参考文献：

- [1] 李棱钦. 基于块匹配和局部子空间的视频稳像方法研究[D]. 大连: 大连理工大学, 2015.  
LI Ling-yi. Study on the Methods of Video Stabilization Based on Block Matching and Local Subspace[D]. Dalian University of Technology, 2015.
- [2] WANG Wei-xing, CAO Ting, LIU Sheng, et al. Remote Sensing Image Automatic Registration on Multi-scale Harrislaplacian[J]. Journal of the Indian Society of Remote Sensing, 2015, 43(3): 501—511.
- [3] GHOSH D, KAABOUC N. A Survey on Image Mosaicing Techniques[J]. Journal of Visual Communication & Image Representation, 2015, 34(C): 1—11.
- [4] GUO Feng, YANG Jie, CHEN Yi-lei, et al. Research on Image Detection and Matching Based on SIFT Features[C]// in Proceedings of the IEEE International Conference on Control and Robotics Engineering, 2018: 130—134.
- [5] XU Jian-peng, LIN Sheng, TENG Ao-xue. Improved SIFT Algorithm Based on Adaptive Contrast Threshold[C]// IEEE International Conference on Computer and Technology Applications, 2018: 151—155.
- [6] 李秋实, 雷建军, 周志远, 等. 基于 SIFT 匹配的多

- 视点立体图像零视差调整[J]. 红外与激光工程, 2015, 44(2): 177—280.
- LI Qiu-shi, LEI Jian-jun, ZHOU Zhi-yuan, et al. Zero-disparity Adjustment of Multiview Stereoscopic Images Based on SIFT Matching[J]. Infrared and Laser Engineering, 2015, 44(2): 177—280.
- [7] MUNICH M E, PIRJANIAN P, BERNARDO E D, et al. SIFT-ing Through Features with ViPR: Application of Visual Pattern Recognition to Robotics and Automation[J]. IEEE Robotics and Automation Magazine, 2006, 26(3): 72—77.
- [8] BAY H, ESS A, TUYTELAARS T, et al. SURF: Speeded Up Robust Features[J]. Computer Vision and Image Understanding (CVIU), 2008, 110(3): 346—359.
- [9] 索春宝, 杨东清, 刘云鹏. 多种角度比较 SIFT、SURF、BRISK、ORB、FREAK 算法[J]. 北京测绘, 2014(4): 23—26.
- SUO Chun-bao, YANG Dong-qing, LIU Yun-peng. Compare SIFT, SURF, BRISK, ORB and FREAK Algorithms from Different Ways[J]. Beijing Surveying and Mapping, 2014(4): 23—26.
- [10] IGHAYENE M, GAO Qiang, BENLEFKI T. Making Bayesian Tracking and Matching by the BRISK Interest Points Detector/Descriptor Cooperate for Robust Object Tracking[C]// in Proceedings of the IEEE International Conference on Signal and Processing, 2016: 731—735.
- [11] ROSTEN E, DRUMMOND T. Machine Learning for High Speed Corner Detection[C]// In 9th European Conference on Computer Vision, 2006, 1: 430—443.
- [12] BIADGIE Y, SOHN K A. Feature Detector Using Adaptive Accelerated Segment Test[C]// International Conference on Information Science & Applications (ICISA), 2014: 1—4.
- [13] CHOI S G, HAN S W. New Binary Descriptors Based on BRISK Sampling Pattern for Image Retrieval[C]// In Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC), 2014: 575—576.
- [14] KASHIF M, DESERNO T M, HAAK D, et al. Feature Description with SIFT, SURF, BRIEF, BRISK, or FREAK? A General Question Answered for Bone Age Assessment[J]. Computers in Biology and Medicine, 2015, 68: 67—75.
- [15] AGHAJANI K, YOUSEFPOUR R, SHIRPOUR M, et al. Intensity Based Image Registration by Minimizing the Complexity of Weighted Subtraction under Illumination Changes[J]. Biomedical Signal Processing & Control, 2016, 25: 35—45.