

基于改进 SIFT 特征和图转换匹配的图像匹配算法

张官亮^{1,2}, 邹焕新¹, 秦先祥¹, 林小平¹

(1. 国防科学技术大学 电子科学与工程学院, 长沙 410073; 2. 武警乌鲁木齐指挥学院 教研部, 乌鲁木齐 830049)

摘要: 针对 SIFT 特征在纹理丰富的图像中提取较多的伪点和不稳定的点而影响图像匹配的问题, 提出了一种基于 Harris 阈值准则的局部不变特征图像匹配算法。该算法在提取 SIFT 不变特征的基础上, 利用 Harris 阈值准则对所提取到的不变特征进行选择, 剔除了图像区域中大量可区分性较差的特征点, 从而得到了相对稳定和可区分性较好的特征点。其次, 结合不变特征矢量与图转换匹配(GTM)的方法对提取到的稳定特征点进行了精确匹配。实验对比结果表明, 用取得稳定的特征点, 进而结合一种好的匹配策略, 能够更加增强图像匹配的高效性和鲁棒性。

关键词: 图像匹配; 特征点提取; SIFT 特征; 高斯差分尺度空间; Harris 阈值准则; 自相关矩阵; 图转换匹配

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2013)09-2861-04

doi:10.3969/j.issn.1001-3695.2013.09.077

Algorithm of image matching based on improved SIFT feature and graph transformation matching

ZHANG Guan-liang^{1,2}, ZOU Huan-xin¹, QIN Xian-xiang¹, LIN Xiao-ping¹

(1. College of Electronic Science & Engineering, National University of Defense Technology, Changsha 410073, China; 2. Dept. of Urumqi Command College of Armed Police, Urumqi 830049, China)

Abstract: As the SIFT operator might extract more false keypoints in the image with various texture, which would affect the result of image matching, this paper proposed a new algorithm of image matching based on SIFT local invariant feature of Harris threshold criterion. On the basis of extracting SIFT invariant features, the extracted invariant feature was selected based on Harris threshold criterion. Therefore, there leaved some more robust and well separable features because the worse separable features were rejected in some region of close-grained image. Finally, it used the vector of invariant feature and graph transformation matching method to match accurately. The experimental results demonstrate that the image matching is high efficiency and robust if it combines stable features with a better matching strategy.

Key words: image matching; feature point extraction; scale invariant feature transform (SIFT) feature; DOG scale-space; Harris threshold criterion; autocorrelation matrix; graph transformation matching (GTM)

0 引言

图像匹配是对从同一区域不同时间、不同的观测角度或者不同的传感器, 得到的图像进行对准的过程。它在遥感领域的应用非常广泛^[1-3], 现有的图像匹配技术一般归为两类^[4]: 基于区域和基于特征两种匹配方法。基于区域的匹配方法不需要对图像显著的特征进行检测, 也不需要算法进行优化。但是其对剧烈的变换比较敏感, 易受噪声和光照的影响。而基于特征点的方法能够通过两幅图像的共同特征, 比如控制点、线和轮廓对图像进行匹配, 对噪声和扰动具有稳定性。

局部不变特征目前广泛应用于计算机视觉中的目标检测、目标识别和图像检索。其本质优点就是在几何变换下的不变性。文献[5]对当前的局部特征点进行了全面的回顾。在各种局部特征点检测和描述算子中, SIFT^[6,7] 算法匹配能力较强, 可以处理两幅图像之间发生平移、旋转、仿射变换情况下的匹配问题, 甚至在某种程度上对任意角度拍摄的图像也具备较

为稳定的特征匹配能力。但其存在计算量大、时间复杂度高和算法耗时长以及在噪声和模糊等情况下的鲁棒性比较有限的缺点。针对这些缺点, 出现了一些改进算法, 如 PCA-SIFT、N-SIFT、SIFT-GM 和 SIFT-GMEP^[8], 这些算法在一定程度上对其有所改进, 但各自都有优劣。本文首先在提取 SIFT 特征的基础上, 利用 Harris 阈值准则对其进行选择, 去掉部分伪点和不稳定的点, 使较为稳定的特征点保留下来。在提取稳定特征点的基础上, 通过对特征点之间的关系, 用图变换匹配(GTM)的方法建立 K-NN 图, 经过多次迭代, 得出图像匹配结果。最后将本文算法同经典 BBF 匹配算法及 RANSAC 算法进行了比较, 实验结果验证了本文算法的可行性和有效性。

1 鲁棒的局部特征点

SIFT 算法于 1999 年首次被 Lowe 提出, 2004 年总结完善^[7]。虽然 SIFT 中的 DOG 算子在几何变换中具有许多优点, 但是提取的特征点在附加噪声后和图像模糊时稳定性不是很

收稿日期: 2012-12-05; 修回日期: 2013-01-15

作者简介: 张官亮(1985-), 男, 陕西渭南人, 硕士, 主要研究方向为图形与图像处理技术(zgl8504@163.com); 邹焕新(1973-), 男, 副教授, 博士, 主要研究方向为多源卫星信息融合处理、SAR 图像解译与目标识别; 秦先祥(1986-), 男, 博士研究生, 主要研究方向为计算机视觉与智能信息处理、遥感图像处理; 林小平(1984-), 男, 硕士, 主要研究方向为计算机视觉与智能信息处理。

好。基于此,需要提取更加稳定的特征点进行匹配。

1.1 SIFT 特征

为了不重复前人工作,主要对 SIFT 特征进行简要概述。

1) 尺度空间极值点检测及定位

局部特征点检测要利用 DOG 算子提取具有尺度不变性的特征点。建立高斯差分(DOG)尺度空间过程如下:一幅二维图像的尺度空间定义为

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

其中 $G(x, y, \sigma)$ 是尺度可变高斯函数, (x, y) 是空间坐标, σ 是尺度坐标。为了在尺度空间检测到稳定的关键(特征)点, Lowe 提出了高斯差分尺度空间(DOG scale-space),即利用不同尺度的高斯差分核与图像卷积生成,如式(2)所示。

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (2)$$

在高斯差分尺度空间中每一个采样点要在其 26 邻域内比较,以确保在尺度空间和二维图像空间都检测到极值点。如图 1 所示,中间的检测点与它同尺度的 8 个邻域点和上下相邻尺度对应的 18 个点共 26 个点比较。根据关键点周围像素点的梯度分布情况,然后确定关键点的位置和尺度。

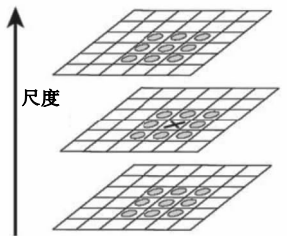


图1 尺度空间极值点检测

2) 关键点方向分配和描述子的产生

每个关键点指定方向参数由关键点邻域像素的梯度方向分布特性决定,使其具备旋转不变性。对于每一个采样图像 $L(x, y)$, 梯度大小 $m(x, y)$ 、方向 $\theta(x, y)$ 可以通过像素差分得到:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (3)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (4)$$

其中 L 所用的尺度为每个关键点各自所在的尺度。计算过程中,以关键点为中心的邻域窗口内进行采样,并用直方图统计邻域像素的梯度方向。在描述子产生中,以关键点为中心取 16×16 个像素的窗口,进而分成 16 个 4×4 的子块,在每个子块上每 45° 一个方向,共 8 个方向,然后计算这 8 个方向的梯度大小和梯度方向直方图。因此,共可以得到 128 个方向描述符,这个 1×128 的向量定义为特征描述符向量。最后将此特征描述符向量归一化并对梯度幅值进行限制,最终完成特征点提取的过程。

1.2 用 Harris 阈值准则提取稳定特征点

在各种特征点算子中^[5,9], Harris 算子能够在各种变形和几何变换中有稳定的表现: Harris 算子对图像旋转、灰度变化、噪声影响和视点变换不敏感,在文献[9]的角点特征的比较实验中, Harris 角点特征的重复性和区分性被证明是最好的。因此本文提出了用 Harris 准则来选择最稳定的 SIFT 特征点来进行图像匹配。

Harris 算子^[10]是通过微分运算和自相关矩阵对 Moravec 的角点检测算法进行的改进。对于一幅图像 $I(x, y)$, 给出点 (x, y) 的自相关矩阵 M , M 的特征值是自相关函数的一阶曲率。Harris 算子的表达式如下:

$$M = \sum_{x,y} \omega(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (5)$$

$$H = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 = \det(M) - k(t_r)^2 \quad (6)$$

其中: I_x 为 x 方向的梯度; I_y 为 y 方向的梯度; $\omega(x, y)$ 为高斯核窗口函数; $\det(M)$ 为矩阵的行列式; t_r 为矩阵的迹; k 为参数, 本文选取 $k = 0.06$; H 表示图中相应某个像素点的兴趣值, 由于 $\det(M)$ 在边缘处较小在角点处较大, 而 t_r 在边缘与角点处一致, 因此, 当像素点 (x, y) 的 H 值为局部最大时, 即为角点。

给定 SIFT 点集 $P = \{p_i(x, y, \sigma, \theta)\}_{i=1}^N$, 此处 x, y 为坐标, σ, θ 分别为尺度和方向参数。然后, 计算 Harris 相应的函数值 $H_i^r(x, y)$, σ 是用来计算自相关矩阵 M 的高斯核窗口的标准偏差, 设置阈值选取稳定的 SIFT 特征点, 如下:

$$\text{Thre} = \frac{\alpha}{N} \sum_{i=1}^N H_i^r(x, y) \quad (7)$$

其中: α 是选择稳定特征点可调整的常数, 一般其取值范围为 $0.1 \sim 0.5$, 本文选取 $\alpha = 0.5$ 。由于不同的图像纹理性和可区分性都不同, 可以根据不同的图像, 计算不同的 Thre 值, 然后通过此 Thre 值, 就可得到图像在有噪声、模糊、尺度变换和旋转等情况下稳定的关键点。如图 2(a1) ~ (c1) 为经典图库 boat 图及其变换图像; 图 2(d1) 为 boat6 图, 为原图的尺度和旋转变换; 图 2(a2) ~ (d2) 为与图 2(a1) ~ (d1) 对应图像提取 SIFT 特征后的图, 可以看出图像提取的特征比较杂乱, 同时也产生了许多不稳定极值点; 图 2(a3) ~ (d3) 是用 Harris 阈值准则对 SIFT 特征进行选择后的图像, 比较发现, 用 Harris 阈值选取的点明显比 SIFT 特征点少了许多, 而且特征点都集中在纹理较为丰富的地方。这是因为此阈值用更高的梯度分布保持了最稳定的局部特征, 又剔除了较低梯度分布的易于被噪声或模糊等引入产生的特征点。因此, 基于 Harris 阈值的方法是自适应并且能够产生相对比较稳定的关键点。

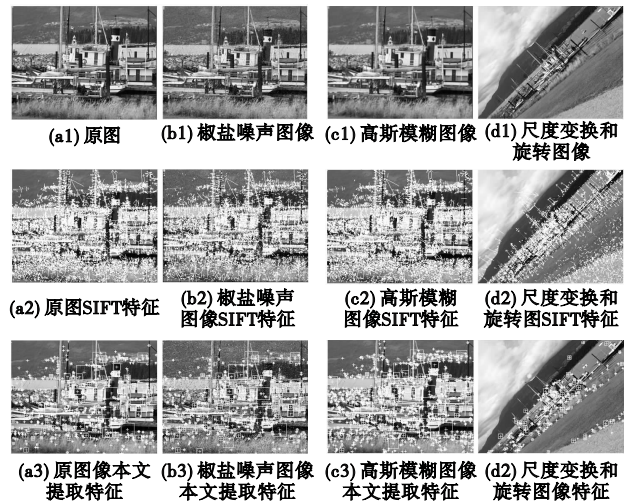


图2 传统SIFT特征和本文方法提取的特征

为了进一步体现 Harris 阈值准则能够选取稳定的关键点, 定义函数 f 对 SIFT 和本文算法的表现进行评估。定义 Q_0 为原图像选取关键点的点集, Q_d 为原图像在各种变换后选取的关键点的点集, 所以 f 为

$$f = \frac{|Q_0 \cap Q_d|}{|Q_0 \cup Q_d|} \quad (8)$$

从式(8)可以看出,当 f 趋向于1的时候,表明从原图像及其加噪或模糊图像得到了几乎相同的关键点集,这样 f 就成为图像在各种情况下检测稳定关键点的评判标准。

然后将 SIFT 与本文方法(SIFT-Harris)进行对比,如图3,分别对原图像进行了加噪(高斯噪声($\sigma=0.1$)、椒盐噪声($d=0.05$)、斑点噪声($\text{var}=0.05$))、模糊(高斯模糊($\sigma=2.5$)、运动模糊($\text{len}=2$, $\theta=45^\circ$))、尺度变换($\text{factor}=0.5$)和旋转($\theta=45^\circ$)。从图中运用 SIFT、SIFT-Harris 和 SCC^[11] 方法的 f 值可以明显地看出,在大多数情况下,本文方法比 SIFT 表现良好, SIFT 特征点只在椒盐噪声和旋转中表现较好。除此之外,可以注意到 SCC 算法的 f 值比原 SIFT 方法中 f 值较低,表现稍差,而且整体远比提出的 SIFT-Harris 方法差。所以可以得出,提出的 SIFT-Harris 方法可以得到更加稳定的特征点。

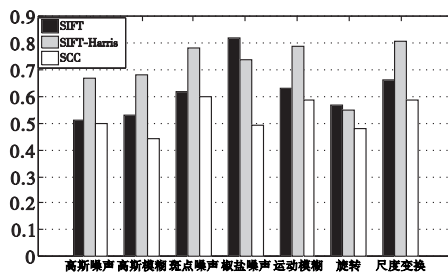


图3 SIFT、本文方法和SCC在各种图像变换下的 f 值比较

2 图像匹配

文献[7]中,Lowe 提出了用 Best-Bin-First(BBF) 算法对图像进行匹配。BBF 算法是基于改进的 K-d 树的有序搜索算法,匹配效率较高,但在纹理较丰富的图像进行匹配时,由于特征点较多,发生的误匹配也较多。RANSAC 算法^[12]充分利用了所有的测量数据和特征的位置信息,并根据阈值把它们分成内点和外点,用内点数据比较准确的特点来进行参数估计而剔除了不准确的数据,但是此算法在出格点超过 50% 时匹配效果较差^[13]。Aguilar 等人^[14]提出了图转换匹配(GTM)的点匹配算法,此算法对两幅图像中的点分别从候选匹配点中寻找近邻图,计算其他点到某一顶点的所有距离,通过中值距离约束建立此点的 K-NN 图,然后用 K-NN 图来描述局部特征,达到删除误匹配的目的。GTM 方法主要流程如下:

a) 对于两幅图像中有相互对应的点集(包括一定数目的出格点): $P=\{p_i\}$ 和 $Q=\{q_i\}$, p_i 点与 q_i 点相匹配。用 K-NN 对图 $G_p=(V_p, E_p)$ 作了如下定义:

因为每个匹配点 $P=\{p_i\}$ 中有顶点 $V_p=\{v_1, \dots, v_N\}$,当 p_i 为 p_i 最近邻的一个点并且 $\|p_i - p_j\| \leq \eta$ 时,存在无向边 $e(i, j)$, η 定义为

$$\eta = \text{median}_{(l,m) \in V_p \times V_p} \|p_l - p_m\| \quad (9)$$

η 为所有点到某一顶点 v_i 距离的中值,若其他点到 v_i 的距离大于 η ,则去除此点,然后通过 η 在两幅图像中的点集中构造 K-NN 图 G_p 和 G_q , G_p 存在 $N \times N$ 邻接矩阵 A_p ,并且当 $(i, j) \in E_p$ 时, $A_p(i, j)=1$,否则 $A_p(i, j)=0$ 。同样的方法,对于 G_q 也相应地存在 A_q 。如果此两图是同构的,即所有相应的匹配是正确的。

b) 通过以下步骤进行迭代:

(a) 从匹配的点集中选择出格点(w^{out}),出格点的选择要

根据计算邻接矩阵的残差 $T=|A_p - A_q|$,然后选择两图中邻接矩阵残差的最大值,即为 w^{out} :

$$w^{\text{out}} = \underset{j=1, \dots, N}{\text{argmax}} \sum_{i=1}^N T(i, j) \quad (10)$$

(b) 出格点确定后,将两个相对应的顶点同时去除, N 就相应地减小为 $N-1$ 。

(c) 然后重新构造 K-NN 图,继续进行下一次迭代,直到达到 $T(i, j)=0$ 时,算法迭代停止。

3 实验结果及分析

为了验证本文改进方法的有效性,用两幅图像进行实验,如图4(a)为2010年甘肃舟曲发生泥石流时无人机的真实航拍图像,来源于文献[15],尺寸均为 925×617 ,两图分别为不同高度和不同视角拍摄的,存在尺度变换。实验是在 CPU Intel Core i5 2.80 GHz,内存为 2.96 GB,操作系统为 Windows 7,仿真平台为 MATLAB 7.1 的 PC 机上完成的。图4(b)为传统 SIFT 方法提取两图的特征,(c)为本文方法提取的特征,(d)为 Lowe 的 BBF 算法^[7]的匹配结果,(e)为 RANSAC 算法^[12]的匹配结果,(f)为 GTM 算法的匹配结果。

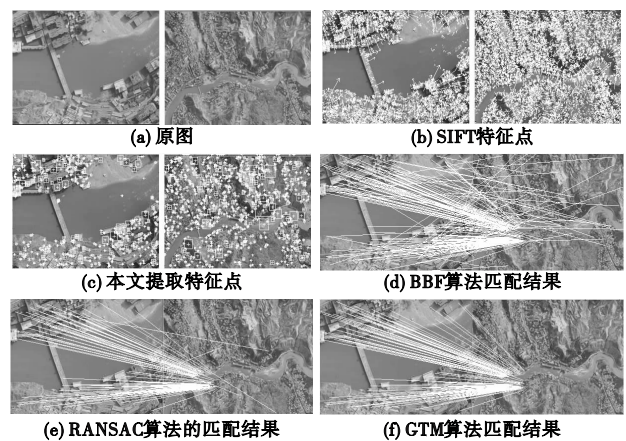


图4 原图像、提取特征图像及三种算法匹配结果

通过实验得出数据如表1所示。

表1 真实图像匹配实验对比结果

参数	两图的特征点个数	匹配点对数	正确匹配点对数	误匹配点对数	匹配率/%	匹配耗时/s
BBF 算法 (SIFT 特征点)	4212/10116	176	127	49	72.2	19.95
RANSAC 算法 (本文提取特征点)	627/1089	104	83	21	79.8	10.53
GTM 算法 (本文提取特征点)	627/1089	131	118	13	90.1	17.26

从表1中可以看出 SIFT 方法两幅图分别检测到的点数为 4 212 和 10 116 个,本文方法两幅图分别检测到的特征点数为 627 和 1 089 个,特征点数量大大减少,并且基本上为可分性较好且较为稳定的点;用 BBF 算法进行匹配,匹配点的对数为 176 对,正确匹配点对为 127 对,误匹配点对为 49 对,匹配率为 72.2%,其运算时间为 19.95 s;RANSAC 算法匹配点的对数为 104 对,正确匹配点对为 83 对,误匹配点对为 21 对,匹配率为 79.8%,其运算时间为 10.53 s;GTM 算法匹配点的对数为 131 对,正确匹配点对为 118 对,误匹配点对为 13 对,虽然其运算时间为 17.53 s,但其正确匹配率远远高于前两者,为 90.1%,所以,用改进的 SIFT 特征结合 GTM 算法的图像匹配方法精度更高,效果更好。实验结果表明,在图像纹理较为丰富的图中

取得稳定的特征点,进而用一种好的匹配策略是图像特征点匹配的关键。

4 结束语

由于 Harris 算子在图像中的稳定表现,本文在提取 SIFT 不变特征的基础上,利用 Harris 阈值准则对所提取到的不变特征进行选择,剔除了纹理丰富的图像中大量可区分性较差的特征点,从而得到了相对稳定和可区分性较好的特征点,为后面的图像匹配奠定了基础。相比传统 BBF 和 RANSAC 算法,用 GTM 算法结合本文改进的 SIFT 特征,具有更高的匹配精度,匹配效果较好。

参考文献:

- [1] TIAN Hao, LI Guo-hui, JIA Li, *et al.* Local descriptor for object recognition in high resolution remote sensing images[C]//Proc of International Conference on Energy Systems and Electrical Power. Singapore: Procedia, 2011: 683-691.
- [2] MISRA I, MOORTHY S M, DHAR D, *et al.* An automatic satellite image registration technique based on Harris corner detection and random sample consensus outlier rejection model[C]//Proc of the 1st International Conference on Recent Advances in Information Technology. 2012: 68-73.
- [3] ZHANG Ji-xian, LI Guo-sheng, ZENG Yu. High precision method for multi-source remote sensing image automatic registration[J]. *Journal of Remote Sensing*, 2005, 9(1): 73-75.
- [4] 丁文益, 王平, 黄庆, 等. 一种适应仿射变换的全自动景象匹配算法[J]. *重庆理工大学学报: 自然科学版*, 2010, 24(3): 71-75.
- [5] 孙浩, 王程, 王润生. 局部不变特征综述[J]. *中国图象图形学报*, 2011, 16(2): 141-151.
- [6] LOWE D G. Object recognition from local scale - invariant features [C]//Proc of the 7th International Conference on Computer Vision. New York: ACM Press, 1999: 1150-1157.
- [7] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [8] 焦斌亮, 樊曼曼. 基于改进 SIFT 算法的多源遥感影像配准研究[J]. *激光与红外*, 2011, 12(41): 1383-1386.
- [9] SCHMID C, MOHR R, BAUCKHAGE C. Evaluation of interest point detectors[J]. *International Journal of Computer Vision*, 2000, 37(2): 151-172.
- [10] HARRIS C, STEPHENS M. A combined corner and edge detector [C]//Proc of the 4th Alvey Vision Conference. New York: ACM Press, 1988: 147-151.
- [11] RATTARANGSI A, CHIN R T. Scale-based detection of corners of planar curves[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1992, 14(4): 430-449.
- [12] 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.
- [13] LIU Zhao-xia, AN Ju-bai, JING Yu. A simple and robust feature point matching algorithm based on restricted spatial order constraints for aerial image registration[J]. *IEEE Trans on Geoscience and Remote Sensing*, 2012, 50(2): 514-527.
- [14] AGUILAR W, FRAUEL Y, ESCOLANO F, *et al.* A robust graph transformation matching for non-rigid registration[J]. *Image and Vision Computing*, 2009, 27(7): 897-910.
- [15] 国家测绘局. <http://www.sbsm.gov.cn/article/mtbd/201008/20100800071205.shtml> [EB/OL].
- (上接第 2860 页)
- [3] UNGER J. Incident light field[D]. Norrköping: Linköping University, 2009.
- [4] GROSSBERG M, NAYAR S. Determining the camera response from images: what is knowable[J]. *IEEE Trans of Pattern Analysis and Machine Intelligence*, 2003, 21(11): 1455-1467.
- [5] LI Xiao-guang, LAM K, SHEN Lan-sun. An adaptive algorithm for the display of high-dynamic range images[J]. *Journal of Visual Communication and Image Representation*, 2007, 18(5): 397-405.
- [6] KANG S, UYTENDAELE M, WINDER S. High dynamic range video[J]. *ACM Trans on Graphics*, 2003, 22(3): 319-325.
- [7] KRYMSKI A. High dynamic range imager with a rolling shutter: United States, 7397509[P]. 2008-07-08.
- [8] NAYAR S, MITSUNAGA T. Apparatus and method for high dynamic range imaging using spatially varying exposures: United States, 6864916[P]. 2005-03-08.
- [9] NAYAR S, BRANZOI V. Adaptive dynamic range imaging: optical control of pixel exposures over space and time[C]//Proc of the 9th IEEE International Conference on Computer Vision. [S. l.]: IEEE Press, 2003: 1168-1175.
- [10] AGGARWAL M, AHUJA N. Split aperture imaging for high dynamic range[J]. *International Journal of Computer Vision*, 2004, 58(1): 7-17.
- [11] TOCCI M D, KISER C, TOCCI N, *et al.* A versatile HDR video production system[J]. *ACM Trans on Graphics*, 2011, 30(4): 41.
- [12] ZHANG Zheng-you. A flexible new technique for camera calibration [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334.
- [13] LEVOY M. Light fields and computational imaging[J]. *IEEE Computer*, 2006, 39(8): 46-55.
- [14] ZELNIK-MANOR L, IRANI M. Multi-view subspace constraints on homographies[C]//Proc of IEEE International Conference on Computer Vision. [S. l.]: IEEE Press, 1999: 710-715.
- [15] 华顺刚, 王丽丹, 段宗璞. 同一场景不同曝光图像的配准及 HDR 图像合成[J]. *计算机辅助设计与图像学报*, 2007, 19(4): 528-534.
- [16] 李将云. 基于形心的二值图像多尺度表示方法[J]. *浙江大学学报: 理学版*, 2003, 30(3): 631-633.
- [17] DRAGO F, MYZKOWSKI K, ANNEN T, *et al.* Adaptive logarithmic mapping for displaying high contrast scenes [J]. *Computer Graphics Forum*, 2003, 22(3): 419-426.
- [18] DUAN J, BRESSAN M, DANCE C, *et al.* Tone-mapping high dynamic range images by novel histogram adjustment[J]. *Pattern Recognition*, 2010, 43(5): 1847-1862.
- [19] REINHARD E, STARK M, SHIRLEY P, *et al.* Photographic tone reproduction for digital images [J]. *ACM Trans on Graphics (TOG)*, 2002, 21(3): 267-276.