

基于几何矩特征与纹理度量的图像匹配算法

金欣¹ 钟洪²

¹(赣南师范大学科技学院 江西 赣州 341000)

²(赣南师范大学数学与计算机科学学院 江西 赣州 341000)

摘要 为提高多种几何内容变化下的特征点匹配准确度,引入双边滤波器,设计了几何矩特征耦合纹理度量的图像匹配算法。引入双边滤波器,对图像进行预处理,以去除图像中噪声,接着利用 Hessian 算子,准确检测图像特征。利用图像灰度信息,计算图像特征圆域内的几何矩,以形成鲁棒性较强的特征向量。通过平均梯度模型,对特征点邻域的纹理特征进行度量,并将其联合欧氏测量结果,对图像特征实施匹配。借助互相关模型,对匹配特征的相关程度进行测量,以检测特征点的匹配正确性,优化特征匹配结果。实验数据表明:较当前特征匹配方法而言,在多种几何内容变化下,所提技术具有更高的匹配准确度与鲁棒性。

关键词 图像匹配 Hessian 算子 双边滤波 几何矩特征 纹理度量 平均梯度模型

中图分类号 TP391

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.03.033

IMAGE MATCHING METHOD BASED ON GEOMETRIC MOMENT FEATURE COUPLED WITH TEXTURE MEASUREMENT

Jin Xin¹ Zhong Hong²

¹(College of Science and Technology, Gannan Normal University, Ganzhou 341000, Jiangxi, China)

²(School of Mathematics and Computer Science, Gannan Normal University, Ganzhou 341000, Jiangxi, China)

Abstract In order to improve the matching accuracy of feature points under the variety of geometric contents, based on the introduction of bilateral filtering, this paper designs an image matching algorithm using image geometric moment feature and texture measurement. A bilateral filter was introduced to preprocess the image and filter out the noise contained in the image. Hessian operator was used to detect the image features. By using the gray value of the pixels, the geometric moments in the image feature circle were calculated to form a robust feature vector. The average gradient model was used to measure the texture features in the neighborhood of feature points, and the texture measurement results were combined with the Euclidean measurement results to accurately match the image features. With the help of cross-correlation model, the correlation degree of matching features was measured to detect the matching correctness of feature points and optimize the matching results. Experimental data show that the proposed method has higher matching accuracy and robustness than the current feature matching method under the change of various geometric contents.

Keywords Image matching Hessian operator Bilateral filtering Geometric moment feature Texture measurement Average gradient model

0 引言

数字图像的出现为人们的生活带来了信息传递的便利性^[1]。人们利用数字图像匹配技术,使得多项工作都变得更为简单和可靠。例如:人们通过图像匹配技术可实现目标的追踪、医疗的诊断以及门禁的安防等^[2-3]。由此可见,对图像匹配技术的研究是一项和人们实际生活紧密相连的重要工作。

在人们长期对图像匹配技术的研究下,出现了多种图像匹配算法。如 Yan 等^[4]为了提高匹配率,在将彩色图像转换为灰度图像时,增加了颜色偏移量和曝光偏移量,并在此基础上通过 SIFT 方法来获取图像特征,并计算图像的梯度信息对图像特征进行描述,进而实现图像的匹配。该算法中利用 SIFT 方法提取图像特征时,存在较多的伪特征,限制了匹配精度。Bharkad 等^[5]通过 radon 变换提取图像特征,通过离散余弦方法匹配图像特征,进而完成图像匹配。由于离散余弦方法匹配图像特征时,没有考虑其纹理特征,限制了其匹配准确度。如 Jiang 等^[6]借助 delaunay 三角剖分及其对应的图形,形成随机分布特征点的邻域结构,在此构造的邻域结构的基础上,实现了光度约束和几何约束两种约束,结合层次消去和左右检查策略,有效地传递了异常值对内联判断的影响,保证了最终匹配的高精度。虽然该方法从图像的几何约束出发,设计图像匹配方法,以提高算法的鲁棒性,但该方法没有考虑图像中所含噪声的影响,降低了匹配结果的准确度。

为了降低噪声对匹配过程的干扰,本文设计了几何矩特征耦合纹理度量的图像匹配算法。在提取图像特征前,利用双边滤波方法将初始图像中的噪声进行滤波,消除噪声的干扰。随后,采用了 Hessian 算子提取了滤波图像的特征。在以图像 Haar 小波信息为依据形成的主方向上,求取了图像的几何矩信息,以形成特征向量。利用欧氏测量模型和平均梯度模型,分别测量了图像的距离特征和纹理特征,从而对图像特征进行匹配。利用互相关模型测量了图像间的相关程度,甄别真伪匹配点对,优化匹配结果。最后,在多种几何内容变换下测试了所提算法的匹配性能。

1 算法设计

本文图像匹配算法的过程如图 1 所示。本文算法大致可分为图像预处理和检测图像特征、求取特征向量以及图像特征匹配、匹配特征的优化四个部分。

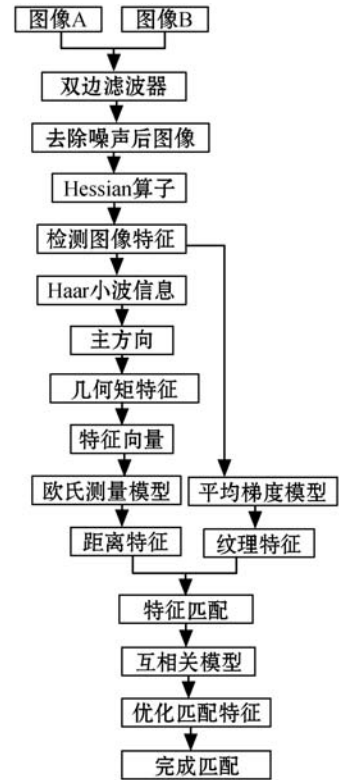


图 1 本文图像匹配算法的过程

1) 在图像预处理和检测图像特征时,利用双边滤波去除噪声,再采用 Hessian 算子检测图像特征。去噪后的图像有助于提高 Hessian 算子对图像特征的检测准确性,以减少算法的误匹配。

2) 在求取特征向量时,利用 Haar 小波信息计算主方向,并通过计算图像的几何矩特征,形成了特征向量。

3) 在图像特征匹配时,利用欧氏测量模型,通过特征向量,计算了图像的距离特征。引入平均梯度模型,计算了图像的纹理特征。联合距离特征和纹理特征完成了图像特征的匹配。

4) 在匹配特征的优化时,采用互相关模型,从考虑匹配特征的相关程度出发,找出伪匹配点对,净化匹配结果。

1.1 图像预处理和特征检测

为了适应图像中的噪声干扰,在此引入双边滤波方法,去除图像噪声,以获取纯净的源图像。双边滤波利用像素点的邻域灰度特征来去除噪声,其不仅去噪能力强,而且还能较好地保持图像的边缘特征^[7]。

双边滤波借助空间域高斯滤波核 σ_d 和灰度域高斯滤波核 σ_r , 来对图像实行非线性滤波。令 $p_0(x_0, y_0)$ 是位于尺寸为 ω 的图像 $f(x, y)$ 中,灰度值为 h_0 的像素点。则获取滤波后图像 $\hat{f}(x, y)$ 的过程为^[8]:

$$\hat{f}(x, y) = \sum_{\omega_b \in \omega} \frac{1}{w_{\omega_b}} \left[\sum_{p \in \omega_b} G_{\sigma_d}(\|p(x, y) - p_0(x_0, y_0)\|) \times G_{\sigma_r}(\|h - h_0\|) h_0 \right] \quad (1)$$

式中: ω_b 和 w_{ω_b} 分别为滤波窗口大小和滤波函数; $p(x, y)$ 是 $f(x, y)$ 中, 灰度值为 h 的像素点; G_{σ_d} 和 G_{σ_r} 分别为空间域函数及灰度域函数。 G_{σ_d} 和 G_{σ_r} 及 w_{ω_b} 的方程为:

$$G_{\sigma_d} = \exp\{ \lVert -(x, y) - (x_0, y_0) \rVert^2 / 2 \cdot \sigma_d^2 \} \quad (2)$$

$$G_{\sigma_r} = \exp\{ -[(h - h_0)^2 / 2] \cdot \sigma_r^2 \} \quad (3)$$

$$w_{\omega_b} = \sum_{(x, y) \in \omega_b} G_{\sigma_d}[\lVert (x, y) - (x_0, y_0) \rVert] \times G_{\sigma_r}(h - h_0) \quad (4)$$

式中: (x, y) 和 (x_0, y_0) 分别为 $p(x, y)$ 和 $p_0(x_0, y_0)$ 的坐标。

利用式(1)对输入图像进行去噪预处理后, 将引入 Hessian 算子, 从去噪后图像中准确地检测图像特征。

对于图像 $\hat{f}(x, y)$ 上尺度为 β 的像素点 $p(x, y)$, 其 Hessian 矩阵为^[9]:

$$H(x, y, \beta) = \begin{bmatrix} L_{xx}(x, y, \beta) & L_{xy}(x, y, \beta) \\ L_{xy}(x, y, \beta) & L_{yy}(x, y, \beta) \end{bmatrix} \quad (5)$$

式中: $L_{xx}(x, y, \beta)$ 为高斯函数 GS 与 $p(x, y)$ 的卷积。GS 的方程为:

$$g(x, y, \beta) = \frac{1}{2\pi\beta^2} e^{-(x^2+y^2)/2\beta^2} \quad (6)$$

$L_{xy}(x, y, \beta)$ 和 $L_{yy}(x, y, \beta)$ 具有类似意义。

在 Hessian 算子检测图像特征的过程中, 利用了 9×9 大小的箱式滤波器来逼近高斯函数 GS。为了便于描述, 将箱式滤波器与 $p(x, y)$ 的卷积结果记为 $w \times w$ 、 D_{xy} 、 D_{yy} 。利用 $\beta = 1.2$ 时的 $L_{xy}(x, y, \beta)$ 和 $L_{xx}(x, y, \beta)$ 值来计算箱式滤波器近似 GS 时产生的误差^[10]:

$$\frac{\lVert L_{xy}(1.2) \rVert \lVert D_{xx}(9) \rVert}{\lVert L_{xx}(1.2) \rVert \lVert D_{xy}(9) \rVert} \approx 0.9 \quad (7)$$

最后, 利用式(7)中的误差可求得 $p(x, y)$ 的 Hessian 行列式 $\det(H)$:

$$\det(H) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (8)$$

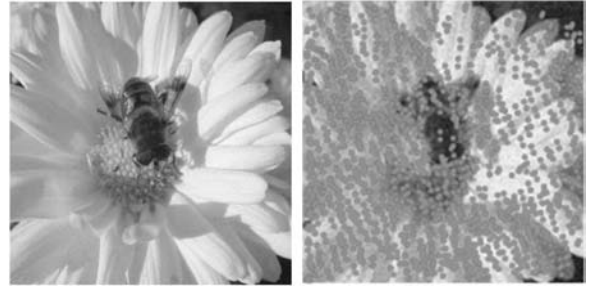
若 $\det(H) > 0$, 则 $p(x, y)$ 就为图像特征。

图 2 为经过双边滤波去噪和 Hessian 算子特征检测的结果。从图 2(c) 可见, 双边滤波方法能够较好地对比图 2(b) 中的噪声进行滤除。从图 2(d) 可见, Hessian 算子能够准确地检测去噪后图像的特征。



(a) 图像 A

(b) 图像 B



(c) 双边滤波去噪结果 (d) Hessian 算子特征检测结果

图 2 图像去噪预处理和特征检测结果

1.2 求取特征向量

图像的特征向量有助于对图像特征进行描述。为此, 将借助像素点的灰度信息来构造几何矩特征模型, 以计算图像的特征向量。而几何矩特征具有较强的旋转以及平移稳健性, 有助于提高算法的鲁棒性。

首先, 引入 SURF 方法^[11], 通过计算图像特征扇形区域中的 Haar 小波累加值, 选取最大累加值的方向为主方向。

然后, 利用像素点的灰度值, 构造几何矩特征模型。对于灰度值为 $h(x, y)$ 的像素点 $p(x, y)$, 其 $i + j$ 阶几何矩特征模型 $G_{i,j}$ 可表述为^[12-13]:

$$G_{i,j} = \iint x^i y^j h(x, y) dx dy \quad (9)$$

式中: $i = 0, 1, \dots, n$ 和 $j = 0, 1, \dots, n$ 为不同的阶数。

由式(9)可求得 0 阶几何矩 $G_{0,0}$ 、1 阶几何矩 $G_{0,1}$ 和 $G_{1,0}$:

$$G_{0,0} = \iint h(x, y) dx dy \quad (10)$$

$$G_{0,1} = \iint y h(x, y) dx dy \quad (11)$$

$$G_{1,0} = \iint x h(x, y) dx dy \quad (12)$$

最后, 按照 SURF 方法建立特征点的邻域圆^[11]。按图 3 所示方法, 沿着主方向以 60° 为间隔把该邻域圆分为六等分。利用式(10) - 式(12)求取每个等分中像素点的 $G_{0,0}$ 、 $G_{0,1}$ 和 $G_{1,0}$ 值。并分别计算 $G_{0,0}$ 、 $G_{0,1}$ 和 $G_{1,0}$ 值的总和, 由此单个等分中就含有三个元素 $\sum G_{0,0}$ 、 $\sum G_{0,1}$ 、 $\sum G_{1,0}$ 。将六个等分中的元素结合起来, 并进行归一化运算, 由此便可得到图像特征含有 18 个元素的特征向量 V 。特征向量的示意图如图 4 所示。

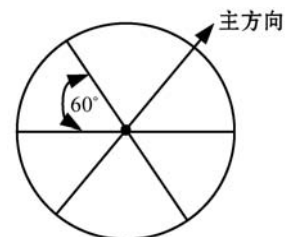


图 3 邻域圆均分 6 等分的示意图

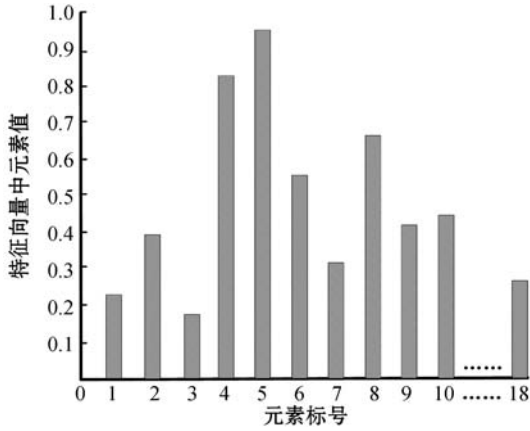


图4 特征向量示意图

1.3 图像特征匹配

利用图像特征间的距离信息对特征点进行匹配的方法较为常见。两个特征点 A 和 B 间的欧氏测量模型为^[14]:

$$Eu(A, B) = \sqrt{\sum_{i=1}^{i=18} |VA_i - VB_i|^2} \quad (13)$$

但是,仅通过特征点间距离信息来匹配图像特征,匹配条件单一且没有考虑特征点的邻域特征,易出现匹配错误。对于图像而言,其纹理特征是刻画图像细节的重要因素,而图像的平均梯度值便可反映出图像的纹理特征。对此,本文将利用平均梯度模型来度量图像的纹理特征,并将其引入到图像的匹配过程,以提高匹配准确性。

在特征点 $w \times w$ 大小的邻域窗口 $R(x, y)$ 中,平均梯度模型可表述为^[15]:

$$K = \frac{1}{w \times w} \sum_{x=1}^w \sum_{y=1}^w 0.5 \left(\frac{\partial R(x, y)}{\partial x}^2 + \frac{\partial R(x, y)}{\partial y}^2 \right)^{\frac{1}{2}} \quad (14)$$

利用式(14)度量出 A 和 B 的邻域纹理信息 K_A 和 K_B ,并将其加入到图像特征的匹配过程,形成匹配函数:

$$B = \arg[\min(Eu(A, B)) \oplus \min(|K_A - K_B|)] \quad (15)$$

式中: \oplus 为同或运算符。

从式(15)可见,在所有特征点中,只有当特征点 B 与特征点 A 的距离最小,而且纹理差异最小时,其方可被视为是 A 的匹配点。

通过图 2(a) 和图 2(b) 的距离特征和纹理特征,得到的匹配结果如图 5 所示。从图 5 可见通过图像距离特征和纹理特征匹配的结果准确度较高。



图5 图像特征匹配结果

1.4 匹配特征的优化

匹配后的图像特征具有较强的相关性,若某对匹配点的相关性不是最强的,则表明其匹配有误。对此,本文借助互相关模型,来甄别特征匹配的真伪,优化匹配准确度。

令 $hA(x, y)$ 和 $hB(x, y)$ 分别为匹配点 A 和 B 矩形窗口的灰度值。此时 A 和 B 的互相关模型为^[16]:

$$HX(A, B) = \frac{\sum_{x=1}^{2M+1} \sum_{y=1}^{2M+1} hA(x, y) hB(x, y)}{\sqrt{\sum_{x=1}^{2M+1} \sum_{y=1}^{2M+1} hA^2(x, y)} \sqrt{\sum_{x=1}^{2M+1} \sum_{y=1}^{2M+1} hB^2(x, y)}} \quad (16)$$

式中: $2M+1$ 为 A 和 B 矩形窗口的大小。

利用式(16)求取点 A 与所有特征点的互相关值,如有 A 与 B 的 $HX(A, B)$ 值最大,说明 A 和 B 匹配正确,应给予保留。反之 A 和 B 为伪匹配点对,应给予删除。

利用匹配点间相关程度优化匹配特征的结果如图 6 所示。可以看出,经过优化后的匹配特征较图 5 更为准确。



图6 匹配特征优化结果

2 实验

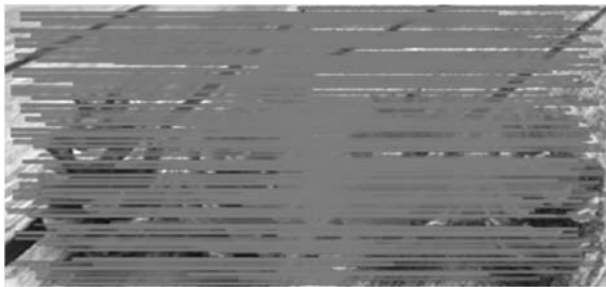
利用本文算法在 INTEL5 双核 CPU、500 GB 硬盘的 DELL PC 上,借助 MATLAB 2017a 软件搭建的仿真环境进行实验。同时选用了文献[17]和文献[18]作为实验对比,以更为直观地突出本文算法的优势。

实验中,采用本文算法及文献[17]和文献[18]分别对含噪、旋转以及缩放图像进行了匹配测试。其中三种算法对含噪图像的匹配结果及数据如图 7 和表 1 所示。从图 7 可以发现,图 7(c)中具有的错误匹配和漏匹配现象最为严重,图 7(d)中的错误匹配和漏匹配比图 7(e)中更多。通过对比表 1 中的数据也可以发现,图 7(e)中的匹配结果较图 7(c)和图 7(d)中的匹配结果最为理想。图 8 展示了各算法对旋转图像的匹配结果。同时表 2 也反映了图 8 中各算法匹配结果的数据。通过观察图 8 中各算法的匹配结果,以及表 2 中各算法匹配结果对应的数据可以发现,文献[17]的

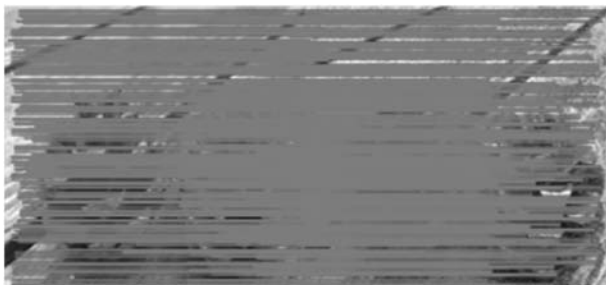
匹配结果中匹配点数最少,文献[18]的匹配结果中含有较多的误匹配,本文算法的匹配点最多及准确率最高。不同算法对缩放图像的匹配结果如图 9 所示,图 9 中不同算法匹配结果的数据如表 3 所示。通过比较图 9 中各算法的匹配结果可以发现,较图 9(c)和图 9(d)中所示匹配结果,图 9(e)所示的匹配结果中正确匹配点最多。观察表 3 中的数据也发现本文算法匹配结果中正确匹配点最多。说明本文算法的匹配特性较好,能对多种变换图像进行较好地匹配。因为本文算法采用了双边滤波去除了输入图像中的噪声,为后续匹配处理提供了较为纯净的源图像,同时本文算法还采用对旋转等变换抵御性较强的几何矩特征获取了图像特征的特征向量,从而提高了算法的匹配性能。



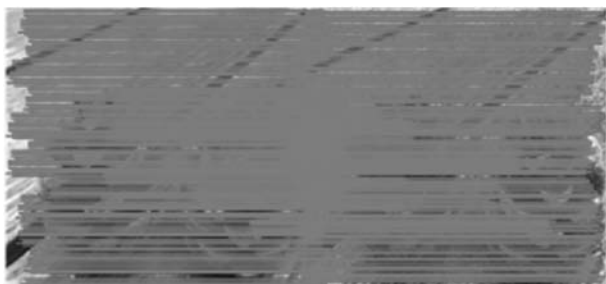
(a) 图像 A (b) 图像 B



(c) 文献[17]算法



(d) 文献[18]算法

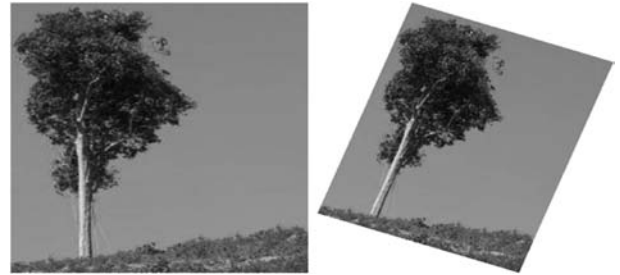


(e) 本文算法

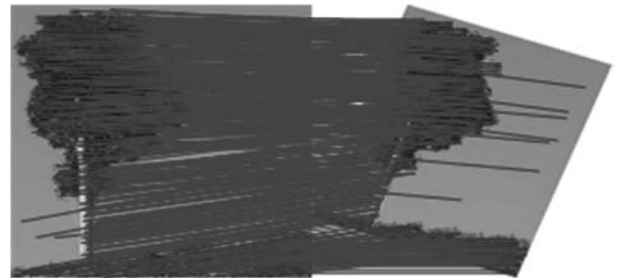
图 7 三种算法对加噪图像的匹配结果

表 1 图 7 中各算法匹配结果对应的数据

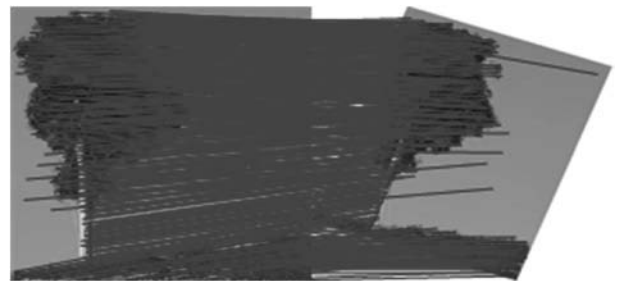
参数	文献[17]算法	文献[18]算法	本文算法
总特征点数	400	400	400
正确匹配特征点数	367	377	389
错误匹配特征点数	12	10	6
匹配准确率/%	91.75	94.25	97.25



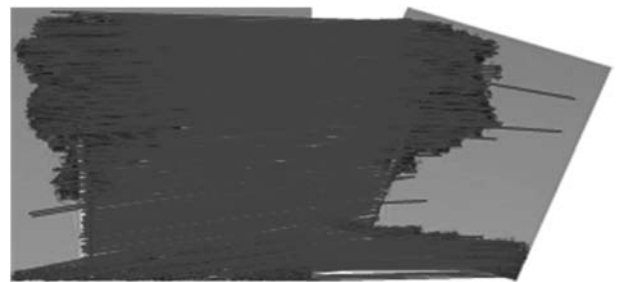
(a) 图像 A (b) 图像 B



(c) 文献[17]算法



(d) 文献[18]算法

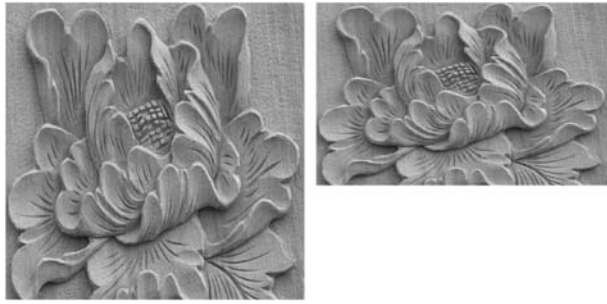


(e) 本文算法

图 8 三种算法对旋转图像的匹配结果

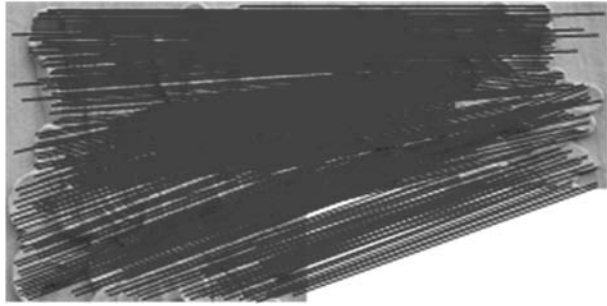
表 2 图 8 中各算法匹配结果对应的数据

参数	文献[17]算法	文献[18]算法	本文算法
总特征点数	383	383	383
正确匹配特征点数	329	348	366
错误匹配特征点数	21	16	11
匹配准确率/%	85.90	90.86	95.56

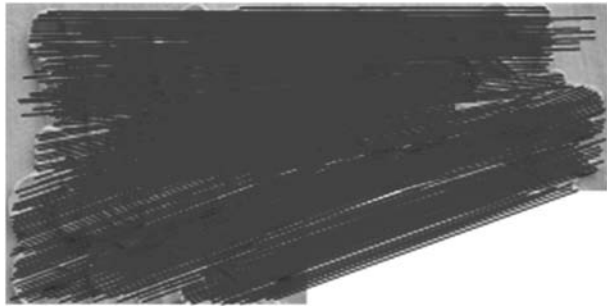


(a) 图像 A

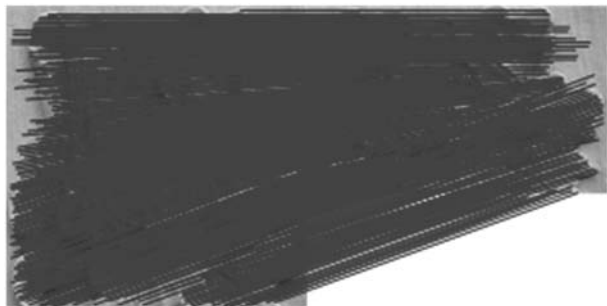
(b) 图像 B



(c) 文献[17]算法



(d) 文献[18]算法



(e) 本文算法

图9 三种算法对缩放图像的匹配结果

表3 图9中各算法匹配结果对应的数据

参数	文献[17]算法	文献[18]算法	本文算法
总特征点数	394	394	394
正确匹配特征点数	332	351	372
错误匹配特征点数	28	22	14
匹配准确率/%	84.26	89.09	94.42

将图 10 所示的目标图像做不同角度的旋转变换, 以获得多角度的匹配图像, 接着利用本文算法及文献 [17] 和文献 [18] 算法对其进行匹配实验。观察不同

算法匹配结果对应的准确率, 以分析其对应的匹配特性。



图10 目标图像

图 11 显示了各算法匹配结果的准确率。可见, 本文算法对旋转度数为 100° 图像匹配结果的准确率为 90.09% , 文献 [17] 和文献 [18] 算法的匹配准确率分别为 80.23% 、 84.64% 。通过对比其他旋转度数下各算法匹配结果的准确率可以发现, 本文算法的准确率始终比其他两种算法的准确率高, 说明本文算法的匹配特性较好, 可以对图像进行准确匹配。因为本文算法采用了鲁棒性较好的几何矩特征来描绘了特征点的特征向量, 并将平均梯度模型引入到图像特征的匹配过程, 考虑了图像的纹理信息。同时本文算法还对匹配点对的相关程度进行了度量, 以优化了匹配结果, 从而使得本文算法的匹配特性较好。文献 [17] 算法中采用 DOG 方法来检测图像的特征, 在图像梯度信息的基础上获取特征向量, 并利用特征距离信息来完成特征匹配, 最后通过像素点间均值信息改进的 ransac 方法优化匹配结果。由于该方法中检测图像特征前没有进行去噪处理, 而且仅依靠距离信息匹配图像特征易出现误匹配, 从而使得文献 [17] 算法的匹配特性不佳。文献 [18] 算法利用直方图线性变换的方法对图像去噪, 并构建了图像特征的二值化特征向量, 并采用阈值调节下的距离比较方法完成了图像匹配。由于二值化特征向量对旋转等变换的抵御性较差, 而且阈值调节下的距离比较方法的自适应性较低, 易出现匹配出错, 由此限制了文献 [18] 算法的匹配特性。

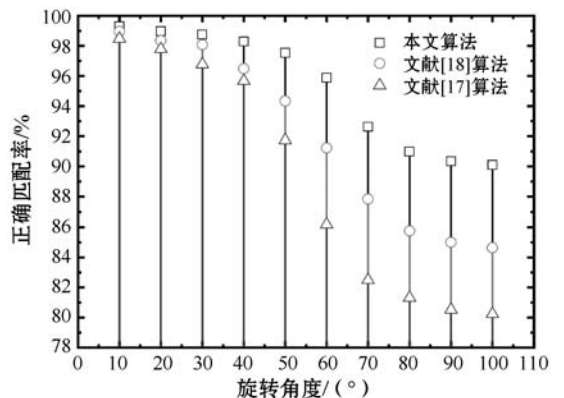


图11 不同算法匹配结果对应的准确率

3 结 语

本文引入了双边滤波算法去除图像噪声,利用 Hessian 算子准确检测滤波图像的特征;随后,利用鲁棒性能较好的几何矩特征求取了特征点的特征向量;利用平均梯度模型度量了图像的纹理信息,并将其与距离信息相结合,完成了图像特征的匹配;再利用匹配点对的相关程度,甄别了匹配点对的真伪情况,优化了匹配结果。实验结果显示,本文算法对多种变换图像的匹配效果较好,能够准确完成图像匹配。

参 考 文 献

- [1] Chen Y, Shang L. Improved SIFT image registration algorithm on characteristic statistical distributions and consistency constraint[J]. *Optik*,2016,127(2):900-911.
- [2] 陈虹,肖越,肖成龙,等. 基于 SIFT 算子融合最大相异系数的自适应图像匹配算法[J]. *计算机应用*,2018,38(5):1410-1414.
- [3] 陈剑虹,韩小珍. 结合 FAST-SURF 和改进 k-d 树最近邻查找的图像配准[J]. *西安理工大学学报*,2016,32(2):213-217.
- [4] Zhao Y, Zhai Y W, Dubois E, et al. Image matching algorithm based on SIFT using color and exposure information [J]. *Journal of Systems Engineering and Electronics*,2016,27(3):691-699.
- [5] Bharkad S, Kokare M. Rotation-invariant fingerprint matching using radon and DCT[J]. *Sadhana*,2017,42(3):2025-2039.
- [6] Jiang S, Jiang W H. Reliable image matching via photometric and geometric constraints structured by Delaunay triangulation [J]. *ISPRS Journal of Photogrammetry & Remote Sensing*,2019,153(1):1-20.
- [7] Ch M, Riaz M, Iltaf N, et al. Weighted image fusion using cross bilateral filter and non-subsampled contourlet transform [J]. *Multidimensional Systems and Signal Processing*,2019,30(4):2199-2210.
- [8] Biradar N, Dewal M L, Rohit M, et al. Echocardiographic image denoising using extreme total variation bilateral filter [J]. *Optik*,2016,127(1):30-38.
- [9] 刘盛清,孙季丰,余家林,等. 基于多尺度特征融合 Hessian 稀疏编码的图像分类算法[J]. *计算机应用*,2017,37(12):3517-3522.
- [10] Liu P F, Xiao L, Li T. Normal curvature-induced variational model for image restoration[J]. *IET Image Processing*,2018,12(5):679-689.
- [11] 杨顺波,龙永红,向昭宇,等. 基于 SURF 算法的双目视觉立体匹配研究[J]. *湖南工业大学学报*,2019,33(3):75-80.
- [12] 张伟,何金国. HU 不变矩的构造和推广[J]. *计算机应用*,2010,30(9):2449-2452.
- [13] Kumar A. Deblurring of motion blurred images using histogram of oriented gradients and geometric moments[J]. *Signal Processing: Image Communication*,2017,55(1):55-65.
- [14] 邹国良,张石明,陈长吉,等. 基于改进最小距离法的图像匹配并行算法设计[J]. *计算机工程*,2018,44(9):218-223.
- [15] Gopal G, Smriti S, Vishal K, et al. Palmprint recognition using novel oriented Gabor gradients [J]. *Journal of Intelligent & Fuzzy Systems*,2018,35(5):5345-5351.
- [16] 高东阳,赵西安,潘昕. 基于稀疏快速傅里叶的互相关图像配准[J]. *北京建筑大学学报*,2017,32(2):25-28,64.
- [17] Hossein-Nejad Z, Mehdi N. An adaptive image registration method based on SIFT features and RANSAC transform[J]. *Computers and Electrical Engineering*,2017,62(1):524-537.
- [18] Zhang C F, Wu Y Z, Liu N, et al. Enhanced SURF-based image matching using pre- and post-processing [C]//International Forum of Digital TV and Wireless Multimedia Communication,2016:83-92.
-
- (上接第 193 页)
- [14] Sriram A, Jun H, Gaur Y, et al. Robust speech recognition using generative adversarial networks [C]//IEEE International Conference on Acoustics, Speech and Signal Processing,2018:5639-5643.
- [15] Johnson J, Gupta A, Li F. Image generation from scene graphs [C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:1219-1228.
- [16] Gao L, Guo Z, Zhang H W, et al. Video captioning with attention-based LSTM and semantic consistency [J]. *IEEE Transactions on Multimedia*,2017,19(9):2045-2055.
- [17] Xu T, Zhang P C, Huang Q Y, et al. AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks [C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:1316-1324.
- [18] Xu H J, Kate S. Ask, attend and answer; Exploring question-guided spatial attention for visual question answering [C]//European Conference on Computer Vision,2016:451-466.
- [19] Ma S, Fu J, Chen C W, et al. DA-GAN: Instance-level image translation by deep attention generative adversarial networks [C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:5657-5666.
- [20] The Caltech-UCSD birds-200-2011 dataset [EB/OL]. [2020-12-21]. <https://gwern.net/doc/ai/dataset/2011-wah.pdf>.
- [21] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs [C]//30th International Conference on Neural Information Processing Systems, 2016: 2234-2242.