

基于手持相机的文档图像拼接算法

苗立刚

(东北大学秦皇岛分校, 秦皇岛 066004)

摘要 为了把手持相机拍摄的多幅文档图像拼接成一幅大的图像,提出了一种基于全局对准模型的文档图像拼接算法。该算法首先通过估计文档图像的消息点坐标来校正透视失真,使相邻图像的几何关系可以用仿射变换表示;然后采用随机采样方法调整特征点之间的距离,使其尽可能均匀地分布在整个重叠区域内;接着利用所有重叠图像对的局部对准约束通过建立文档图像拼接的全局对准模型来有效地消除误差积累;最后利用二值函数对图像进行剪切,以减小重叠区内的对准误差。实验结果表明,该方法无需事先标定摄像机的内外参数和限制相机的位置,不仅具有较高的对准精度,且可有效地拼接手持相机拍摄的各种文档图像。

关键词 文档图像拼接 图像对准 误差积累 透视失真

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2009)08-1656-07

Hand-held Camera Based Document Image Mosaicing Algorithm

MIAO Li-gang

(Department of Automation, Northeastern University at Qinhuangdao, Qinhuangdao 066004)

Abstract This paper presents a global alignment model based image mosaicing method for camera-captured document images, and it can be used to combine multiple overlapping document images into one large image. It corrects the perspective distortion with the estimated vanishing points, and there exists only an affine transform between two adjacent images. Then, it adjusts the distance of featurepoints to distribute them as evenly as possible in the overlapping regions. Thirdly, it uses local alignment constraints of all the overlapping image pairs to construct global alignment model, thus, to eliminate the error accumulation. In order to reduce alignment error of overlapping area, a binary weighted function is used to blend the overlapping region of image pairs. This method is unique because it does not require the calibration of the internal/external camera parameters in advance and does not restricting the camera position, thus allowing greater flexibility than scanner-based or fixed-camera-based approaches. It can produce a high resolution and accurate full page mosaic from small image patches of a document.

Keywords document image mosaicing, image alignment, error accumulation, perspective distortion

1 引言

扫描仪是获取数字文档图像的一种传统方式,其特点是光照稳定,并具有较高的成像分辨率。近年来,手持成像设备的普及程度越来越高,如数码照相机、数码摄像机、摄像头、个人数字助理(PDA),以及可拍照手机等,它们具有便于携带、成像简单、

快速,以及非接触等特点。目前,基于手持相机的文档图像分析已经成为光学字符识别(OCR)技术的一个重要的研究方向^[1],并且在许多手持设备中得到了越来越多的应用。

普通扫描仪和手持相机的视场范围相对有限,而对于幅面较大的文档,如工程图纸、新闻报纸、地图等,则无法一次获取大幅面的高分辨率文档图像。但在许多应用中,可以通过扫描多幅图像来覆盖整

收稿日期:2007-12-26;改回日期:2008-05-07

第一作者简介:苗立刚(1974~),男,讲师。2007年获中国科学院自动化研究所工学博士学位。主要研究方向为图像处理、模式识别、计算机视觉等。E-mail:ligangmiao@yahoo.com.cn; miaolg@hotmail.com

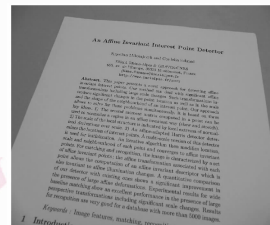
个文档页面,并把它们拼接成一幅高分辨率的文档图像。相对于自然场景图像的拼接来说,文档图像拼接对精度的要求比较高。文献[2]、[3]的拼接方法主要针对平台式扫描仪获取的文档图像,由于光照比较均匀,并且可以用2维欧氏变换来表示相邻图像的几何关系,因此容易取得较高的拼接精度。Mirmehdi等人通过限定相机的位置来获取文档图像^[4],并假定相邻图像的几何关系为平移变换。Nakao等人把摄像头安装在鼠标上,先使成像平面正面朝向文档,然后通过页面上拖动摄像头来获取文档图像序列^[5]。Zappala等人则把摄像机固定在文档的正上方,通过在桌面上移动文档来采集图像^[6]。这些方法由于通过硬件的支持,避免了文档图像的尺度变化和透视失真,从而降低了文档图像拼接的难度。

利用手持相机可以方便、快速地获取文档图像,但是其拼接难度较大。相对于扫描仪而言,手持相机的图像拼接主要存在以下3方面的问题:(1)相机位置的变化会不可避免地产生透视失真;(2)图像受光照变化的影响较大,图像的重叠范围不易控制;(3)当拼接图像的数量较大时,产生的误差积累比较严重。在现有的手持相机拍摄的文档图像拼接算法中,Liang等人提出了一种基于主成分分析和尺度不变特征变换(PCA-SIFT)的文档图像拼接算法^[7],其缺点是没有考虑误差积累问题。Sato等人提出了一种视频序列的文档图像拼接算法^[8],它要求参考帧平行于文档平面,并且需要标定相机的内参数。文献[9]、[10]通过估计相机外部参数来消除透视失真的影响,但由于它仍需要事先标定相机的内参数,并且要确保内参数在图像采集过程中保持不变,因而增加了操作的不便性和算法的复杂度。

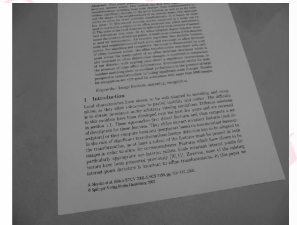
针对手持相机拍摄的文档图像,提出了一种新的文档图像拼接算法,它既不需要对摄像机的姿态和运动做太多的限制,也不需要事先标定相机的内外参数。该方法首先利用文档图像的格式信息来分步校正透视失真,使相邻图像的几何关系可以用仿射变换表示;然后利用尺度不变特征变换(SIFT)方法进行文档图像对准,并且自动找出文档图像的重叠关系;接着利用所有重叠图像对的局部对准约束来建立全局对准模型,并联合求解所有图像的全局对准参数,用来消除拼接过程的误差积累;最后利用二值函数来对图像重叠区域进行剪切,以减小重叠区内小的对准误差。

2 透视失真校正

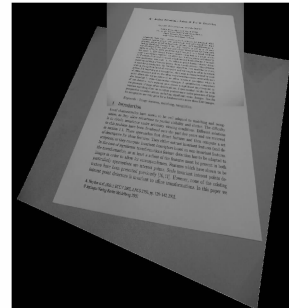
在用手持相机拍摄文档图像的过程中,由于相机光轴经常偏离文档平面的法线方向,从而导致文档图像不可避免地产生透视失真。图像拼接时需要先选取一幅图像作为参考帧,然后把其他图像对准到参考帧上。如图1所示,如果参考帧的图像平面与相机平面不能保持平行,那么将会导致最终的拼接结果出现严重的透视失真(图1(c))。因此,在图像拼接之前需要进行透视校正,以恢复文档图像的一个正面平行视图。



(a) 第1幅透视失真的文档图像



(b) 第2幅透视失真的文档图像



(c) 两幅失真文档图像的拼接结果

图1 文档图像的透视失真

Fig. 1 Perspective distortion of document images

根据射影几何理论,利用水平和竖直消隐点可以恢复文档图像的仿射结构。目前的透视校正算法主要利用文档的格式信息来估计消隐点的坐标。一般情况下,由于文档图像的竖直消隐点的定位精度低于水平消隐点,直接利用消隐线进行校正会产生较大的误差,为此,本文利用文献[11]的分层校正方法来消除透视失真。文档图像透视失真的示意图如图2(a)所示,该方法先利用文本行的信息来估计水平消隐点的坐标,它可以消除旋转和水平方向的透视失真,使各个文本行沿水平方向保持平行(见图2(b));然后利用竖直笔划边界、段落或页面边界等信息来估计部分校正图像的竖直消隐点的坐标,它可以消除垂直方向的透视失真(图2(c));最后通

过估计剪切角度,来校正剪切失真,即可得到文档图像的一个正面平行视图(图 2(d)).分层校正方法不仅明显改进了消隐点的定位精度,并且具有较高的鲁棒性,有关该方法的详细步骤请参考文献[11].

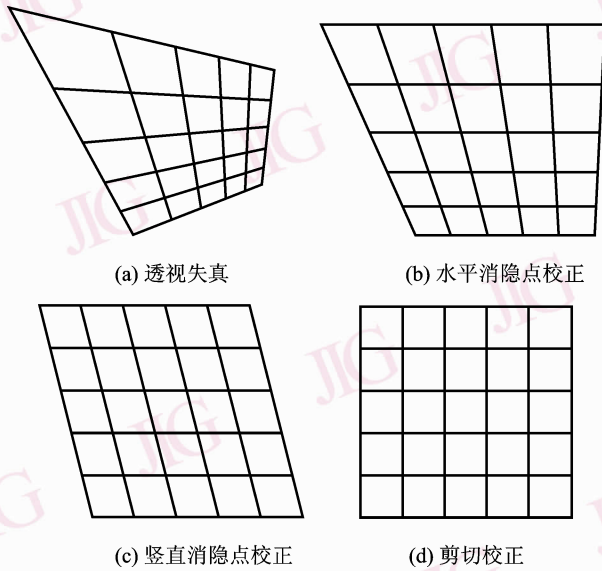


图 2 透视失真的分层校正

Fig. 2 Hierarchical correction of perspective distortion

3 文档图像对准

由于经过透视校正后,就可以认为相邻两幅文档图像之间只存在平移和两个方向的尺度变化,因而图像对准模型就可以用下面的仿射变换表示:

$$\bar{\mathbf{x}} = \mathbf{A}\mathbf{x} \quad (1)$$

其中, $\mathbf{x} = (x, y, 1)^T$ 和 $\bar{\mathbf{x}} = (\bar{x}, \bar{y}, 1)^T$ 表示两幅图像对应点的齐次坐标, \mathbf{A} 表示仿射变换矩阵:

$$\mathbf{A} = \begin{bmatrix} s_x & 0 & t_x \\ 0 & s_y & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

其中, s_x 和 s_y 分别表示水平和垂直方向的缩放因子, t_x 和 t_y 分别表示水平和垂直方向的平移参数。

图像对准方法主要分为基于特征的方法和基于区域的方法两类。由于文档图像的光照变化和尺度伸缩较大,这两种方法一般无法取得比较理想的对准结果。近年来,基于不变特征的匹配方法取得了较大的进展,其中 Lowe 提出的 SIFT 方法^[12] 由于可以处理图像间的平移、旋转、尺度伸缩,以及光照变化等情况下的特征点匹配问题,因而在许多领域取得了广泛的应用。在文档图像对准过程中,首先进

行 SIFT 特征点的检测和匹配,并利用仿射变换约束来消除其中的外点,即可求解相邻图像的对准参数。

图像对准的精度还依赖于特征点分布的均匀程度。特征点的位置与图像的内容有较大关系,通常在信息丰富的区域,其特征点的密度比较大;而在结构相对单一的区域,则几乎没有特征点的分布。由于文档图像主要由相对单一的字符组成,因此造成特征点对分布不均匀的程度比较高。由于仿射变换的参数求解会过于依赖分布比较稠密的特征点对,从而容易产生对准误差。为了精确地计算对准参数,本文对特征点之间的距离进行了调整,使其尽可能均匀地分布在重叠区域内。

本文提出了一种随机采样的方法,用来调整特征点对分布的均匀程度。用 T_{dis} 表示相邻特征点的最小距离的阈值, \mathbf{S} 表示原始的特征点对 $\mathbf{x}_k \leftrightarrow \bar{\mathbf{x}}_k$ ($k = 1, \dots, N$) 的集合, \mathbf{T} 表示间距调整后的特征点对的集合。该采样方法,首先从 \mathbf{S} 中随机选取一个特征点对移入 \mathbf{T} 中,将其作为 \mathbf{T} 的初始值;然后,从 \mathbf{S} 的其他特征点对中随机选取一个特征点对 $\mathbf{x}_i \leftrightarrow \bar{\mathbf{x}}_i$, 并用 d_i 和 \tilde{d}_i 分别表示 \mathbf{x}_i 和 $\bar{\mathbf{x}}_i$ 与 \mathbf{T} 中所有特征点对的最小距离,即

$$d_i = \min_j \|\mathbf{x}_i - \mathbf{y}_j\| \quad (2)$$

$$\tilde{d}_i = \min_j \|\bar{\mathbf{x}}_i - \bar{\mathbf{y}}_j\| \quad (3)$$

其中, $(\mathbf{x}_i, \bar{\mathbf{x}}_i) \in \mathbf{S}$, $(\mathbf{y}_j, \bar{\mathbf{y}}_j) \in \mathbf{T}$ 。

由于受透视投影和镜头失真的影响, d_i 和 \tilde{d}_i 可能相差较大,本文用两者的最大值 $\max(d_i, \tilde{d}_i)$ 表示特征点对之间的距离。如果该距离大于阈值 T_{dis} , 则将其从 \mathbf{S} 移入 \mathbf{T} 中;否则,重新从集合 \mathbf{S} 中选取其他特征点对进行比较。如果 T_{dis} 的初始值太大,或经过若干次迭代之后得到的特征点的数量太少,则可以通过逐渐降低阈值来重新进行搜索,直到集合 \mathbf{T} 中有适当数量的特征点对出现。

4 全局对准模型

图像拼接需要确定图像之间的匹配关系,本文大致按照一种从上到下、从左到右的扫描顺序采集文档图像,这样每幅图像只需要与相邻几幅图像进行匹配即可。本文采用 RANSAC 方法先将所有的特征点对分为内点和外点^[13], 然后根据内点和外点的分布判断该图像对是否匹配。设重叠区内总的匹配特征点对的数量为 n_{total} , 其中内点的数量为 n_{in} 。

当内点数量大于一定比例时,就认为当前图像对匹配,否则就认为该图像对不匹配。本文直接利用 n_{in}/n_{total} 的比值进行判断,其阈值设为 0.22,该方法可以取得比较好的匹配验证效果。

由于受文档图像表面的起伏和成像过程中的视差、镜头失真,以及噪声等因素的影响,导致图像对准过程总会存在微小的对准误差。当图像数量较大时,依次连接两幅图像的拼接方法还会产生误差积累。为此,本文从总体上考虑所有重叠图像对的对准约束,并提出了一种文档图像拼接的全局对准模型。该方法首先选定一幅图像作为参考图像,然后把其他所有图像投影到参考图像所在的坐标系,并利用所有特征点对的投影误差来建立目标函数。若用 $u_i^{(k)} \leftrightarrow u_j^{(k)}$ 表示第 i 幅图像和第 j 幅图像的第 k 个特征点对,则所有重叠图像对的特征点的投影误差可以表示为

$$E = \sum_{i=1}^n \sum_{j \in M(i)} \sum_{k \in N(i,j)} \|A_i u_i^{(k)} - A_j u_j^{(k)}\|^2 \quad (4)$$

其中, n 表示图像的数量, $M(i)$ 表示与第 i 幅图像匹配的图像集合, $N(i,j)$ 表示第 i 幅图像和第 j 幅图像中的匹配特征点对的集合, A_i 表示第 i 幅图像到参考帧的仿射变换。

如果选取第 1 幅图像作为参考帧,则 A_1 的取值为单位矩阵。由于该目标函数(式(4))是一个凸函数,可利用共轭梯度法同时求解所有图像的全局对准参数 $A_i (i=2, \dots, n)$,从而可以把这些图像合成一幅大幅面的文档图像。

5 文档图像融合

在文档图像的采集过程中,一方面由于受光照变化、曝光时间差异等因素的影响,会导致相邻图像产生明显的颜色(或亮度)不连续性;另一方面,由于文档图像表面的起伏、镜头失真和透视失真校正等产生的误差,致使图像重叠区的边界处总会存在较小的对准误差。尽管图像融合可以消除重叠区颜色的不一致性,但是它会造成重叠区边界附近产生错位,虽然这对于自然场景图像而言,并不会对视觉造成较大的影响,但是,由于文档图像主要由黑白分明的字符组成,因此重叠区域边界的模糊和错位会严重影响文档图像的视觉效果和字符识别率。

一般情况下,图像重叠区中心位置的对准误差可以忽略不计,而从重叠区中心到重叠区周围,对准

误差则逐渐增大。为了减小错位的影响,可利用一个二值加权函数来对文档图像的周边区域进行适当的剪切,并把每幅图像依次叠加到前面的部分拼接图像中,这样就可以每次只考虑相邻两幅图像的融合。在某一水平(x 方向)扫描线上,如果图像 I 和部分拼接图像 \tilde{I} 在区间 (x_1, x_2) 上重叠,则 \tilde{I} 在扫描线上点 x 的取值为

$$\tilde{I}(x) = w(x) \tilde{I}(x) + (1 - w(x)) I(x) \quad (5)$$

其中, $I(x)$ 表示当前图像扫描线上点 x 的像素值, $\tilde{I}(x)$ 表示部分拼接图像的像素值, $w(x)$ 表示二值加权函数,其取值范围为 0 和 1,即

$$w(x) = \begin{cases} 1 & x \leq \frac{x_1 + x_2}{2} \\ 0 & x > \frac{x_1 + x_2}{2} \end{cases} \quad (6)$$

上式表明,拼接图像在重叠区域的中心位置的左端保持不变,而在中心位置的右端则为新图像。因此,虽然它对重叠区的对准误差具有较强的抑制能力,但是它无法消除亮度的不连续性。一般情况下,文档图像的灰度直方图符合双模式分布,但光照的变化却使两种模式产生偏移。文献[7]采用了局部直方图规整化的方法,使得直方图的分布变为黑白两种模式。尽管这种方法可以避免亮度的不连续性对 OCR 的不利影响,但并不适合人的视觉观察。因此,在尽量减小错位的情况下,如何消除亮度的不连续性仍然是一个需要进一步研究的课题。

6 实验结果

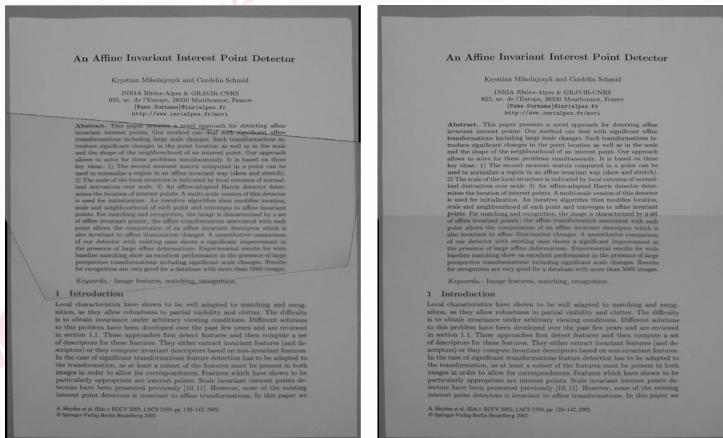
本文对用数码相机拍摄的文档图像进行了实验,采集的文档图像大小为 $1\,728 \times 2\,304$ pixels。实验时,首先对图 1 中的两幅文档图像进行透视校正,校正结果如图 3 所示,其中的文档边界基本上保持水平和竖直方向;然后利用本文方法进行拼接,得到的拼接图像大小为 $2\,680 \times 2\,180$ pixels,其中图 4(a)是这两幅图像直接加权平均的结果,由该图可以看出,重叠区内的亮度可以平滑过渡,但是存在明显的对准误差。图 4(b)是本文方法的融合结果,虽然图像存在亮度的差异,但是它减小了重叠区的文字错位的影响。图 4(c)中第 1 行的两幅图像分别是图 3 中的两幅文档图像的子图像,图 4(c)中第 2 行的两幅图像则分别对应线性加权平均的文档图像和本文方法融合的文档图像。



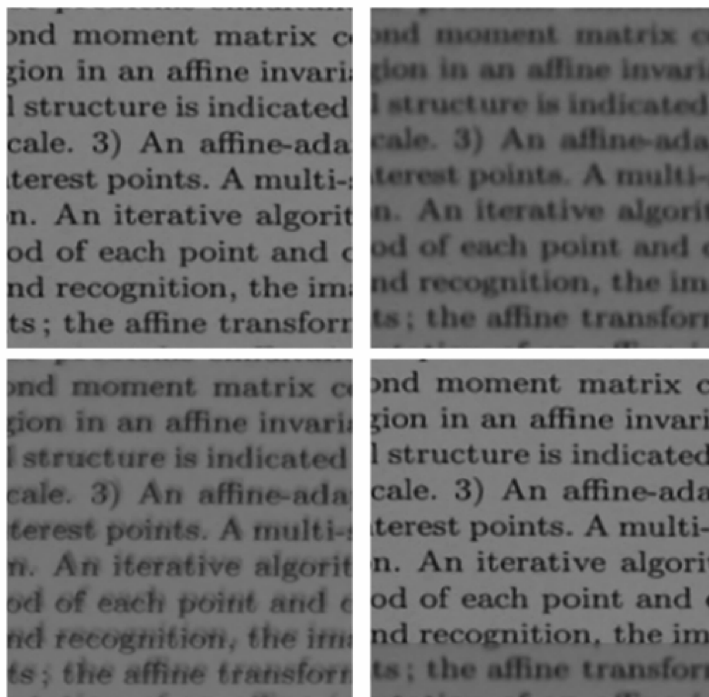
(a)图1 (a) 校正结果 (b)图1 (b) 校正结果

图 3 透视校正后的文档图像

Fig. 3 Document images after perspective correction



(a)线性加权平均后的文档图像 (b)二值函数剪切后的文档图像



(c)局部子图像

图 4 两幅文档图像的拼接结果

Fig. 4 Mosaicing result of two document images

7 结 论

本文研究了手持相机拍摄的文档图像拼接算法中的若干关键问题,并提出了一种行之有效的解决方案。该方法根据文档图像的各种格式信息来估计消隐点的坐标,并利用分层校正的方法消除透视失真。为了提高文档图像对准的精度,还采用随机采样的方法对 SIFT 特征点之间的距离进行调整,使得这些特征点对尽可能均匀地分布在两幅图像的重叠区域内。在全局对准阶段,则根据所有重叠图像对的局部对准约束,提出了一种文档图像拼接的全局对准模型,用来有效地消除误差积累。最后采用二值函数对图像重叠区域进行剪切。实验表明,该图像拼接方法不但具有较高的拼接精度,并且无需事先标定摄像机的内外参数和限制相机的位置,从而大大减小了用户的干预。

需要指出的是,本文直接利用针孔模型表示手持相机的成像过程。实际上,几乎所有的镜头都存在一定程度的镜头失真,因此进一步的研究还需要考虑镜头失真的校正问题。该方法的另一个不足之处是计算速度相对较慢,将来还需要对现有的算法进行优化,以便嵌入到各种手持成像设备之中,如 PDA、手机等等。

参考文献 (References)

- 1 Liang J, Doermann D S, Li H. Camera-based analysis of text and documents: a survey [J]. *International Journal on Document Analysis and Recognition*, 2005, 7(2-3):84-104.
- 2 Isgro F, Pilo M. A fast and robust image registration method based on an early consensus paradigm[J]. *Pattern Recognition Letters*, 2004, 25(8):943-954.
- 3 Nagabhushan P. Sliding window based approach for document image mosaicing[J]. *Image Vision Computing*, 2006, 24(1):94-100.
- 4 Mirmehdi M, Clark P, Lam J. Extracting low resolution text with an active camera for OCR [A]. In: *Proceedings of the IX Spanish Symposium on Pattern Recognition and Image Processing [C]*, Valencia, Spanish, 2001:43-48.
- 5 Nakao T, Kashitani A, Kaneyoshi A. Scanning a document with a small camera attached to a mouse [A]. In: *Proceedings of the 4th Workshop on Applications of Computer Vision [C]*, Princeton, New Jersey, USA, 1998:63-68.
- 6 Zappala A, Gee A, Taylor M J. Document mosaicing[J]. *Image and Vision Computing*, 1999, 17(8):585-595.
- 7 Liang J, DeMenthon D, Doermann D. Camera-based document image mosaicing[A]. In: *Proceedings of 18th International Conference on Pattern Recognition [C]*, Hong Kong, China, 2006, 2:476-479.
- 8 Sato T, Ikeda S, Kanbara M, et al. High-resolution video mosaicing for documents and photos by estimating camera motion [A]. In: *Proceedings of SPIE Conference on Computational Imaging II [C]*, San Jose, California, USA, 2004, 5299:246-253.
- 9 Iketani A, Sato T, Ikeda S, et al. Super-resolved video mosaicing for documents based on extrinsic camera parameter estimation [A]. In: *Proceedings of the 7th Asian Conference on Computer Vision [C]*, Hyderabad, India, 2006, 2:101-110.
- 10 Nakajima N, Iketani A, Sato T, et al. Video mosaicing for document imaging [A]. In: *Proceedings of the 2nd International Workshop on Camera-based Document Analysis and Recognition [C]*, Curitiba, Brazil, 2007:171-178.
- 11 Miao L G, Peng S L. Perspective rectification of document images based on morphology [A]. In: *Proceedings of the 2006 International Conference on Computational Intelligence and Security [C]*, Guangzhou, China, 2007:1805-1808.
- 12 Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2):91-110.
- 13 Brown M, Lowe D G. Recognising panoramas [A]. In: *Proceedings of the 9th International Conference on Computer Vision [C]*, Nice, France, 2003:1218-1225.