

# 彩色文档图像的倾斜自动校正算法

何希平<sup>1),2)</sup> 李云峰<sup>1)</sup> 朱庆生<sup>1)</sup>

<sup>1)</sup>(重庆大学计算机学院,重庆 400044) <sup>2)</sup>(重庆工商大学计算机学院,重庆 400067)

**摘要** 对彩色扫描文档进行倾斜校正是对其进行 OCR 等处理所必须首先经历的步骤,为了对彩色倾斜文档图像进行准确、高效校正,提出了一种新的彩色文档图像的倾斜自动校正算法,该算法包括倾斜检测算法和倾斜校正算法。其中,倾斜检测算法解决了准确获得图像的倾斜角的问题;而倾斜校正算法则除了完成图像的旋转变换外,还通过色彩补偿解决了由于整数运算所造成的“锯齿”现象。倾斜检测是通过纹理复杂性分析实现的,而色彩补偿则是基于颜色线性相关的双线性插值算法。实验表明,该算法较好地解决了彩色文档图像的倾斜自动校正问题,它对于具有单一背景的彩色文档图像是准确、高效、实用的。

**关键词** 彩色文档图像 倾斜 检测 自动校正 纹理复杂性 颜色补偿

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2006)03-0367-05

## An Efficient Algorithm for Automatic Skew-Correction of Color Document Image

HE Xi-ping<sup>1),2)</sup>, LI Yun-feng<sup>1)</sup>, ZHU Qing-sheng<sup>1)</sup>

<sup>1)</sup>(College of Computer, Chongqing University, Chongqing 400044)

<sup>2)</sup>(College of Computer, Chongqing Technology and Business University, Chongqing 400067)

**Abstract** The automatic skew-correction of scanned color document image is a necessary step undergone before some processing such as OCR (optical character recognition). For the purpose of accurate and efficient automatic correction, two algorithms, namely skew-detection and skew-adjustment, are proposed. The skew-detection algorithm, which is based on the analysis of texture complexity of document image with homogeneous background, solves the problem of how the accurate skew angle can be found. Besides the rotation transform, the skew-adjustment algorithm includes a dual linear interpolation algorithm for color compensation based on linear dependence among the colors in the neighborhood to accomplish the skew-correction of a color image and the elimination of the “Saw tooth” phenomenon resulted from the integer-operation during the coordinates transformation. In the end, the authors illustrate their experimental results, which show that the presented algorithm is exact, efficient, and practical for color documents whose background are not complex.

**Keywords** color document image, skew, detection, automatic correction, texture complexity, color compensation

## 1 引言

对彩色扫描文档进行倾斜校正是对其进行处理所必须首先经历的步骤,而且如何在时间代价不高的条件下,准确、高效地实现自动校正,一直是扫描仪生产商在文档处理方面所不断努力实现的目标。

目前,倾斜检测的方法都是针对二值图像或灰度图像<sup>[1-3]</sup>的,其主要可分为基于 Hough 变换的方法、基于交叉相关性的方法、基于 Fourier 变换的方法、基于投影的方法和  $k$ -最近邻簇方法 5 类。

其中,Hough 变换的方法是最常用的倾斜检测方法,其是利用式(1)所示的 Hough 变换来提取直线

收稿日期:2003-09-02;改回日期:2005-06-23

第一作者简介:何希平(1968 ~ ),男,副教授。2002 年获重庆大学计算机软件与理论硕士学位,现为重庆大学博士研究生。主要研究方向为图像分析处理、数据压缩,发表论文 20 余篇。E-mail: jsjxhp@ctbu.edu.cn

$$\rho = x \cos \theta + y \sin \theta \quad (1)$$

它是一种变换域提取直线的方法,它是先把坐标平面  $O-xy$  上的直线变换到极坐标平面  $O-\rho\theta$  上的点,且在  $O-\rho\theta$  上,与  $O-xy$  中的共线点对应的所有曲线相交于一点,而在  $O-\rho\theta$  平面上描出图像中所有目标像素对应于式(1)的曲线;然后统计  $O-\rho\theta$  平面上各网格点上曲线经过的次数,其中数值大者,其对应的平行直线条数越多。

由于基本 Hough 变换算法的计算量非常大,于是人们研究出了一些专门用于倾斜检测的 Hough 变换改进算法<sup>[4]</sup>,这些算法的核心思想是减少变换的数据量。

交叉相关性算法是一种基于等距离的竖直(水平)平行线上像素点相关性的方法。由于倾斜的原因,致使文本行在两条平行线上的像素点会发生平移。该方法先以平行线上像素点来建立相关矩阵,然后对相关矩阵在竖直方向上进行投影,并使投影图的全局最大值对应于倾角。该方法虽然准确性高,但由于相关矩阵的计算量非常大,而且为了确定全局最大值,在某些情况下还需要回溯计算相关矩阵和投影,因此,计算量会更大。

基于 Fourier 变换的方法是利用页面倾角对应于使 Fourier 空间密度最大的方向角特性,将文档图像中的所有像素点进行 Fourier 变换,由于计算量非常大,故目前很少使用。

基于投影图的方法是利用投影图的某些特性来进行判断。图像向  $y$  轴方向的投影图的偏差值是倾斜角的一个评价函数,其全局最大值对应于倾角。

$k$ -最近邻簇方法是先找出所有连通区域的中心点的  $k$  个最近邻,然后计算每对近邻点的矢量方向,并统计生成直方图,直方图的峰值就对应于整个页面的倾角,其计算复杂度为  $O(n^2)$ ,  $n$  为连通区个数。

彩色文档图像的倾斜自动校正算法包括倾斜检测算法和倾斜校正算法,其中倾斜检测算法要解决准确获得图像的倾斜角问题;而倾斜校正算法则要完成图像的旋转变换,且要解决由于整数运算所带来的“锯齿”现象的补偿问题,但至今尚无就此提出好的解决办法的参考文献。在文档图像中,文本块的像素对比度比图片块和背景块的像素对比度变化幅度更大,其中背景像素的对比度变化相对而言比较平滑、连续。图像中的每个文本像素对文本方向的确定都起到至关重要的作用。由此可见,针对彩

色文档图像,如何有效地分析文档图像的纹理特征,以及如何正确地对文本、文本背景与其他块进行分类、标记,是解决倾斜检测问题的关键。本文提出的基于纹理复杂性分析的倾斜检测方法有效地利用了文本图像的上述纹理特征,通过对 334 幅样图进行测试的结果表明,不但其性能稳定、准确可靠,而且时间效率也比各种 Hough 变换算法高。

## 2 基于纹理复杂性分析的倾斜检测算法

本算法的指导思想是:以最少的计算量、最快的处理速度、最强的健壮性和最高的准确性来保证实现倾斜检测。为保证实现最高的准确性,需要先通过平滑滤波技术来消除图像噪声,然后将彩色图像转换为灰度图像,并在此基础上对图像像素进行分类,以便抽取出文本与文本背景像素;最后分析图像的纹理复杂性,进而得出文本方向,以用于确定图像的倾斜角。

### 2.1 图像预处理与图像转换

平滑噪声可以在空间域中进行,其基本方法是求像素的灰度的平均值或中值。为了既平滑噪声,又保护图像信号,已有了一些改进技术,比如,在频率域中运用低通滤波技术。而邻域平均技术则是空间域中的平滑噪声技术,其非常适合于去除通过扫描得到的图像中的颗粒噪声。

对于给定图像  $I$  中的每个像素点  $(x_0, y_0)$ ,其邻域  $N_{x_0, y_0}$  含  $M$  个像素,邻域像素灰度平均值为

$$\bar{p}(x_0, y_0) = \frac{1}{M} \sum_{(x, y) \in N_{x_0, y_0}} p(x, y) \quad (2)$$

用其替换  $p(x_0, y_0)$ ,即用像素邻域内各像素灰度的平均值来代替该像素原来的灰度。如  $N$  取  $3 \times 3$  大小的邻域,像素点  $(x_0, y_0)$  位于  $N$  的中心,则式(2)表示为

$$\bar{p}(x_0, y_0) = \frac{1}{9} \sum_{i=-1}^1 \sum_{j=-1}^1 p(x_0 + i, y_0 + j) \quad (3)$$

而按下述灰度值计算则可将彩色图像转换为灰度图像,即

$$p = [0.299 \times R + 0.587 \times G + 0.114 \times B]$$

其中的  $R, G, B$  为彩色图像像素的红、绿、蓝 3 色分量;  $[\cdot]$  表示对  $\cdot$  取整,即不超过  $\cdot$  的最大整数,下同。

### 2.2 文本背景抽取

由于灰度图像灰度变化的相对连续性,以及文

档图像中相邻同类纹理像素灰度的相近性,对于同类相邻纹理像素灰度值  $p_i$  与  $p_j$ , 有

$$|p_i - p_j| < \delta$$

由于文档图像的文本背景通常具有局部一致连续变化的特点,且文本背景在整个图像中占了相当大的比例,因此,对各像素的分类,尤其是文本背景抽取,宜在充分参考其邻域特征的前提下采用近似划分的策略。本文的算法如下:

首先将像素灰度分为  $l$  (可取为 16) 个区段,并按式(4)计算各像素的灰度级  $g(x, y)$ ,再用该灰度级  $g(x, y)$  来构造原图像的灰度级别图  $G$ 。

$$g(x, y) = p(x, y) \times l \div 255 \quad (4)$$

然后,用数学形态学的逻辑与运算来检测图像背景,结构元素  $S$  为  $4 \times 4$  的窗口。检测方法为:先将灰度级别图  $G$  分成  $4 \times 4$  大小的块  $N_{i,j}$ ,并对  $N_{i,j}$  进行标记,标记方法是:将结构元素窗口  $S$  在图像表面上按从左至右、从上至下的顺序逐块滑动,同时按式(6)确定  $N_{i,j}$  的类别标记  $c(i, j)$ ,并按式(7)计算类别标记  $c$  的重复频数  $f(c)$

$$\text{sign}(x, y) = \begin{cases} 0 & (x \neq y) \\ 1 & (x = y) \end{cases} \quad (5)$$

$$c(i, j) = g(x, y) \times \prod_{\substack{x \neq m \text{ 或 } y \neq n, \\ (x, y) \in N_{i,j}, (m, n) \in N_{i,j}}} \text{sign}(g(x, y), g(m, n)) \quad (6)$$

$$f(c) = \sum_{N_{i,j}} \prod_{\substack{x \neq m \text{ 或 } y \neq n, c(i, j) = c, \\ (x, y) \in N_{i,j}, (m, n) \in N_{i,j}}} \text{sign}(g(x, y), g(m, n)) \quad (7)$$

由于式(7)巧妙地屏蔽了不连续尖峰噪声的干扰,因而可求得块类别标记频数直方图的最大峰值

$$f_{\max} = \max_{c \in \{c(i, j)\}} \{f(c)\} \quad (8)$$

若  $f_{\max} = f(G_{\text{back}})$ , 则  $G_{\text{back}}$  即为背景灰度类别标记,进而可求得表示背景像素与非背景像素标识的矩阵  $B$ , 其元素为

$$b(x, y) = \begin{cases} 1 & c\left(\left[\frac{x}{4}\right], \left[\frac{y}{4}\right]\right) = G_{\text{back}} \\ 0 & \text{其他} \end{cases} \quad (9)$$

其中,  $[\cdot]$  表示取整。

### 2.3 图像倾斜角检测

本文把沿某直线  $L$  的方向  $\theta$  上像素由背景变为非背景的次数称为该直线的复杂性  $V_L$  (如图 1 所示),而将该方向上所有直线的复杂性的方差称为该文档图像沿该方向的复杂性  $V_\theta$ 。设  $N_\theta$  为  $\theta$  方向

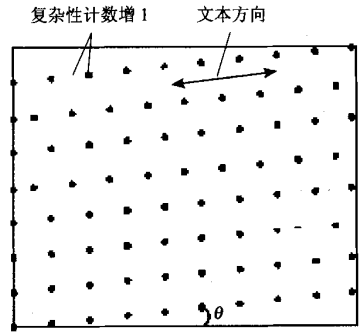


图 1 文本方向复杂性示意图

Fig. 1 Sketch map of complexity in text orientation

上的直线条数,则由式(9)知,  $V_L, V_\theta$  可按式(10)、(11)计算

$$V_L = \sum_{(x-1, k) \in L, (x, y) \in L} (1 - b(x - 1, k)) b(x, y) \quad (10)$$

$$V_\theta = \frac{1}{N_\theta} \sum_L (V_L - \bar{V}_L)^2 \quad (11)$$

显然,由于文本行方向的复杂性变化幅度最大、频率最高,所以复杂性最大的方向即为文本行方向  $\theta$ , 此方向角  $\theta$  就是文档图像的倾斜角  $\gamma$ , 即

$$\gamma = \theta, V_\theta = \max_{\beta} \{V_\beta\} \quad (12)$$

而且,正是由于用于确定文本方向的是方差最大值,而线图、连续图像沿文本方向的复杂性变化却较均匀,且振幅相对于文本串也较小,所以上述的文本方向复杂性分析并不受插入的表格等线图的干扰,也不受插入的连续图像的影响。由此可见,文档图像的每个文本方向的文本像素对图像倾斜角的检测都起着至关重要的作用,而且对各方向的复杂性进行分析是检测文档图像倾斜角的有效方法。

据此,可对灰度图以  $0.1^\circ$  为步长,沿  $-15^\circ$  到  $15^\circ$  的方向,递增搜索所有方向的最大复杂性,而在每一方向上,则先求出各条直线  $L$  的复杂性  $V_L$ ,再求这些  $V_L$  的方差(方向复杂性)  $V_\theta$ 。其算法描述如下:

```

for each angle  $\theta \in [-15^\circ, 15^\circ]$ 
  for each line  $L \in \{L_0, L_1, \dots, L_{N_\theta-1}\}$ 
    for each  $(x, y) \in L, (x - 1, k) \in L,$ 
       $V_L = V_L + (1 - b(x - 1, k)) b(x, y)$ 
    endfor
  endfor

```

求该方向的复杂性  $V_\theta$ :  $V_\theta = \frac{1}{N_\theta} \sum_L (V_L - \bar{V}_L)^2$

若该  $V_\theta$  为目前所得到的  $V_\theta$  值中最大者,则  $\gamma = \theta$

endfor

为提高检测效率,本文采用基于直线拟合的方法来生成像素点的水平、竖直偏移量,以快速实现三角函数的求值计算,进而快速定位直线上的各点。

### 3 图像倾斜校正

当检测到扫描图像发生倾斜时,为得到正常的原图像,应将源图像绕图像中心旋转一定角度,并把旋转结果进行适当平移,以使整个旋转结果图像落入屏幕的可显示区域内。

#### 3.1 源图像绕中心旋转的变换公式

设屏幕坐标系(如图 2 所示)下的源图像(宽为  $W$ ,高为  $H$ )的像素点为  $p(x_0, y_0)$ ,经绕图像中心  $\hat{O}(a, b)$  旋转  $\theta$  角(不会很大;逆时针为正)后变换为  $p(x_1, y_1)$ 。它可看成是经如下 3 步完成:(1)将坐标系沿水平和垂直方向作平移,使  $O(0, 0)$  变成  $\hat{O}(a, b)$ ; (2)将源图像绕  $\hat{O}$  旋转  $\theta$  角; (3)对旋转结果图像进行平移变换,则绕图像中心旋转的算法如下:

平移变换可表示为

$$x_0 = \hat{x}_0 + a, y_0 = \hat{y}_0 + b \quad (13)$$

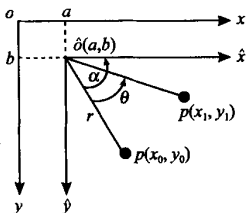


图 2 屏幕坐标系

Fig. 2 Reference frame on screen

$\hat{x}\hat{y}$  坐标系下的旋转变换可表示为  
旋转前

$$\hat{x}_0 = r \cos \alpha, \hat{y}_0 = r \sin \alpha \quad (14)$$

旋转后

$$\begin{cases} \hat{x}_1 = r \cos(\alpha - \theta) = r \cos \alpha \cos \theta + r \sin \alpha \sin \theta \\ \hat{y}_1 = r \sin(\alpha - \theta) = r \sin \alpha \cos \theta - r \cos \alpha \sin \theta \end{cases}$$

结合式(14)得

$$\hat{x}_1 = \hat{x}_0 \cos \theta + \hat{y}_0 \sin \theta, \hat{y}_1 = \hat{y}_0 \cos \theta - \hat{x}_0 \sin \theta \quad (15)$$

所以

$$\hat{x}_0 = \hat{x}_1 \cos \theta - \hat{y}_1 \sin \theta, \hat{y}_0 = \hat{x}_1 \sin \theta + \hat{y}_1 \cos \theta \quad (16)$$

由式(13)和式(16)得

$$x_0 = \hat{x}_1 \cos \theta - \hat{y}_1 \sin \theta + a, y_0 = \hat{y}_1 \cos \theta + \hat{x}_1 \sin \theta + b \quad (17)$$

由于设结果图像的中心为  $(c, d)$ , 则旋转后的

平移变换可表示为  $\hat{x}_1 = x_1 - c, \hat{y}_1 = y_1 - d$  所以

$$\begin{cases} x_0 = (x_1 - c) \cos \theta - (y_1 - d) \sin \theta + a \\ y_0 = (y_1 - d) \cos \theta + (x_1 - c) \sin \theta + b \end{cases} \quad (18)$$

其中,  $a = W/2, b = H/2, c = \hat{W}/2, d = \hat{H}/2$ 。  $\hat{W}, \hat{H}$  分别为结果图像的宽度、高度,而  $\hat{W}, \hat{H}$  则可根据原图像的 4 个顶点在  $\hat{x}\hat{y}$  坐标系下的旋转结果位置坐标  $(\hat{x}_1, \hat{y}_1), (\hat{x}_2, \hat{y}_2), (\hat{x}_3, \hat{y}_3), (\hat{x}_4, \hat{y}_4)$  确定:

$$\hat{W} = \max_{1 \leq i \neq j \leq 4} \{ |\hat{x}_i - \hat{x}_j| \}, \hat{H} = \max_{1 \leq i \neq j \leq 4} \{ |\hat{y}_i - \hat{y}_j| \} \quad (19)$$

而位置坐标  $(\hat{x}_i, \hat{y}_i) (i = 1, 2, 3, 4)$  则可按式(20)计算

$$\begin{cases} x_1 = (x_0 - a) \cos \theta + (y_0 - b) \sin \theta + c \\ y_1 = (y_0 - b) \cos \theta - (x_0 - a) \sin \theta + d \end{cases} \quad (20)$$

因此,对于彩色结果图像中的每一像素值  $P(x_1, y_1)$ ,  $P$  包括红绿蓝 3 个分量,则按式(18)和式(19)确定的映射关系来取其在源图像中的像素值  $P(x_0, y_0)$  即可。编程实现时,应事先算出  $\cos \theta$  和  $\sin \theta$  的值,以提高时间效率。本文采用基于直线拟合的方法来生成像素点的水平、竖直偏移量,以便快速地对倾斜页面进行校正。

#### 3.2 色彩线性相关补偿算法

由于按式(18)计算得到的屏幕坐标  $(x_0, y_0)$  并不一定正好是源图像中网格点的坐标,因此,应对其取整

$$x = [x_0 + 0.5], y = [y_0 + 0.5] \quad (21)$$

如果每一像素  $(x_1, y_1)$  按式(21)在源图像中直接取颜色数据  $P(x, y)$ , 则势必产生明显的误差。为此,记  $P(x, y)$  为  $P_{NW}$ , 且  $P(x_0, y_0)$  的另 3 个相邻像素点(见图 3)为  $P_{NE}(x+1, y), P_{SW}(x, y+1), P_{SE}(x+1, y+1)$ , 则  $P(x_1, y_1)$  的源颜色数据可基于距离的线性相关性,通过  $P_{NW}, P_{NE}, P_{SW}, P_{SE}$  点的颜色分量  $C$  进行双线性插值平滑来得到( $C$  表示颜色分量  $R, G, B$  之一)。其快速算法为如下的分 3 步完成的一阶线性插值算法:

设

$$\varepsilon_{PEW} = x_0 - x, \varepsilon_{PNS} = y_0 - y \quad (22)$$

则

$$T_C = C_{P_{NW}} + \varepsilon_{PEW} (C_{P_{NE}} - C_{P_{NW}}) \quad (23)$$

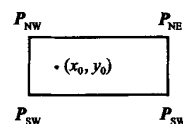


图 3 4 邻点示意图

Fig. 3 Four neighboring pixels

$$B_C = C_{P_{SW}} + \varepsilon_{P_{EW}}(C_{P_{SE}} - C_{P_{SW}}) \quad (24)$$

$$C_P = T_C + \varepsilon_{P_{NS}}(B_C - T_C) \quad (25)$$

如果按式(25)计算像素  $P = (x_1, y_1)$  的各颜色分量  $C_P$ , 则图像过渡更平滑、更自然。计算时, 应直接从式(22)递推到式(25), 共进行 3 次乘法、8 次加减法, 而不能全部代入式(25)后再计算, 否则, 将至少进行 4 次乘法、12 次加减法, 效率会明显下降。

### 4 实验结果与分析

由于本文提出的倾斜检测方法是基于最大纹理复杂性分析的统计方法, 而图片、背景的复杂性都不及文本高, 因此, 即使文档图像中插入了图片, 只要文本处于检测区域内, 而且没有太多的线图干扰, 则不论页面是否颠倒, 其检测结果都是准确可靠的。尤其是当文档图像的绝大多数成分都为单一背景的文本时, 检测结果非常准确(见图版 I 图 1 ~ 图 3)。笔者把文中提到的快速 Hough 变换方法、交叉相关性算法以及从网站 <http://web.tiscalinet.it/delphipage> 上可以得到的由共享组件所提供的算法与本文提出的算法进行了比较, 结果表明, 本文的算法在稳定性和准确性方面都具有明显的优势。表 1 给出了 4 种算法作用于图版 I 图 1 的结果对比。

表 1 几种算法的校正效果对比

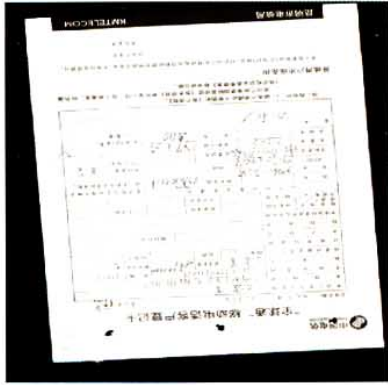
Tab.1 Effect comparison of several correction algorithm

方法	检测角度(°)	绝对误差(°)	耗时(s)
快速 Hough 变换	3.000	0.2	10.072
交叉相关	3.150	0.050	12.53
共享组件	3.125	0.075	5.678
本文方法	3.208	0.008	8.620

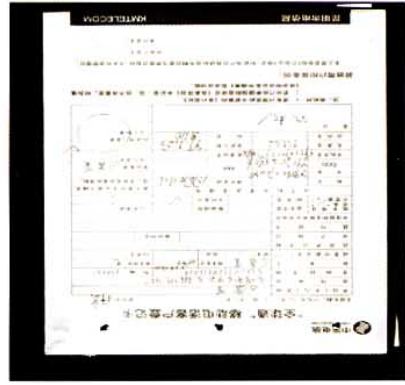
为提高运算速度, 结合实际发生倾斜的最大可能, 可将检测的范围缩小到  $[-5^\circ, 5^\circ]$ ; 对于 300dpi 的图像分辨率, 可以将其降到 75dpi, 这样, 由于可以大大降低计算处理的数据量, 从而使计算处理的速度提高约一个数量级, 而检测结果仍是准确可靠的。通过应用到与企业合作的研究项目中来进行检验, 该算法的准确、高效和实用性也得到了证实, 但由于本文的算法是针对文档图像的, 因此, 对于具有复杂背景的、含文字较少的图像(如杂志封面、画册等), 其检测算法的准确性会明显下降; 另外该算法也对文字排列方向不只包含纵向或横向的文档失效。

### 参考文献 (References)

- Steinherz T, Intrator N, Rivlin E. Skew detection via principal components analysis [A]. In: Proceeding of ICDAR '99 [C], Bangalore, India, 1999:153 ~ 156.
- CHEN Ming, DING Xiao-qing. A Robust skew detection algorithm for grayscale document image[A]. In: Proceedings of ICDAR '99[C], Bangalore, India, 1999:617 ~ 620.
- Okun O. Severe document skew detection[A]. In: SPIE Conference on Mathematical Modeling and Estimation Techniques in Computer Vision[C], San Diego, CA, USA, 1998,3457:263 ~ 274.
- Perantonis S J, Gatos B, Papamarkos N. Block decomposition and segmentation for fast Hough transform evaluation [J]. Pattern Recognition, 1999, 32(5):811 ~ 824.

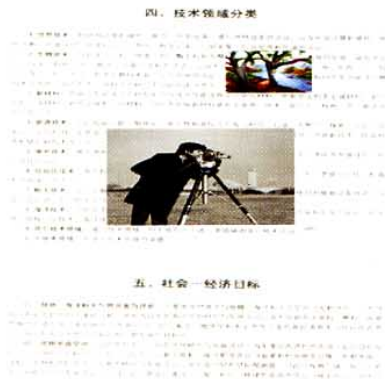


(a) 原图像

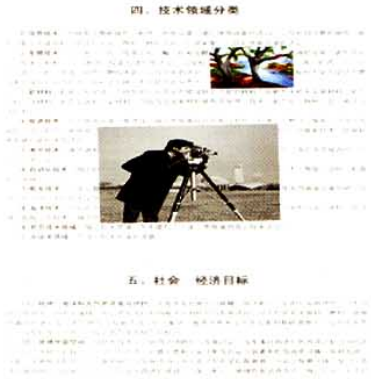


(b) 校正结果

图1 表格文本图像  
Fig.1 Table document image



(a) 原图像



(b) 校正结果

图2 图文混排彩色文本图像  
Fig.2 Color document image mixed with picture



(a) 原图像



(b) 校正结果

图3 彩色图书封面文本图像  
Fig.3 Color document image of the back cover of a book