

文章编号:1007-5321(2017)04-0068-06

DOI:10.13190/j.jbupt.2017.04.011

动态规划算法在手机拍照文档图像中的应用

王茂森¹, 牛少彰²

(北京邮电大学 智能通信软件与多媒体北京市重点实验室, 北京 100876)

摘要: 提出了一种动态规划算法解决手机拍照文档图像的边界检测问题. 根据手机拍照文档图像的特点, 设计新的动态规划模型, 提出了利用逻辑回归二分类算解决文档区域 4 个边的组合问题. 虽然手机拍照的文档图像一般比较模糊, 光照不均匀, 畸变, 但是这些算法很好地解决了手机拍照文档图像的边界检测和组合问题.

关键词: 霍夫变换; 动态规划; 非极大值抑制; 逻辑回归

中图分类号: TN911.22

文献标志码: A

Using Dynamic Programming for Detecting Document Area of Mobile Images

WANG Mao-sen¹, NIU Shao-zhang²

(Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: The detection algorithm of the edge of the document area in the mobile image was presented. According to the characteristics of mobile image, a new dynamic programming model was designed. A combination of four edges of document area was proposed by using the two classification of logistic regression. Although the mobile images are generally blurred, uneven illuminated and distorted, these algorithms can solve the problem of boundary detection and combination of document area of mobile image well.

Key words: Hough transforms; dynamic programming; non-maximum suppression; logical regression

随着智能手机的普及和手机摄像功能越来越强大, 一些传统数码相机和扫描仪功能逐渐被手机拍照功能代替. 目前有很多应用程序通过手机拍照纸质文档, 把文档区域从背景图像中分离出来, 进行图像增强, 可以达到扫描仪的效果. 虽然手机拍照文档图像简单、方便和经济, 但是由于手机拍照的文档图像中, 携带与文档无关的背景, 需要把文档区域从背景图像中分离出来. 从背景图像中分割文档区域, 属于图像分割技术. 手机拍照图像的主要特点是: 文档区域作为背景图像的目标在图像中所占比例较大, 通常大于 70%; 纸质文档的物理边界是矩

形, 边界一般是直线. 所以对文档区域的分割方法一般采用基于边缘的分割方法, 比如通过霍夫变换检测直线^[1-2]的方法确定文档区域的边缘, 实现目标的分割. Zhang 等^[3]在白板扫描和图像增强的论文中用霍夫变换从背景图像中寻找白板的边界, 抠出白板区域, 实验结果也表明霍夫变换解决了白板图像的搜边问题, 但是论文中使用的拍照设备是高分辨率的数码相机, 并且相机是固定的, 图像质量远高于手机拍照图像, 白板边界复杂度却低于手机图像. 相对而言, 手机拍照具有以下特点:

1) 由于手机是移动手持设备, 拍照过程中容易

收稿日期: 2017-04-25

基金项目: 国家自然科学基金项目(61370195, U1536121)

作者简介: 王茂森(1971—), 男, 博士生, E-mail: wmsbeijing@126.com; 牛少彰(1963—), 男, 教授, 博士生导师.

发生抖动,造成拍照图像不清晰,经常会模糊,如图 1(a)所示;

2) 通常拍照是在自然场景进行,光线不均匀,例如在晚上,依靠室内灯光照射拍照,造成图像光线不均匀,有阴影产生,如果打开手机自带闪光灯,文档图像会包含亮斑,如图 1(b)(c)所示;

3) 拍照的纸质文档,经常是书本的某一页,文档所在平面并非水平,拍照文本图像会发生几何畸变,如图 1(d)所示。

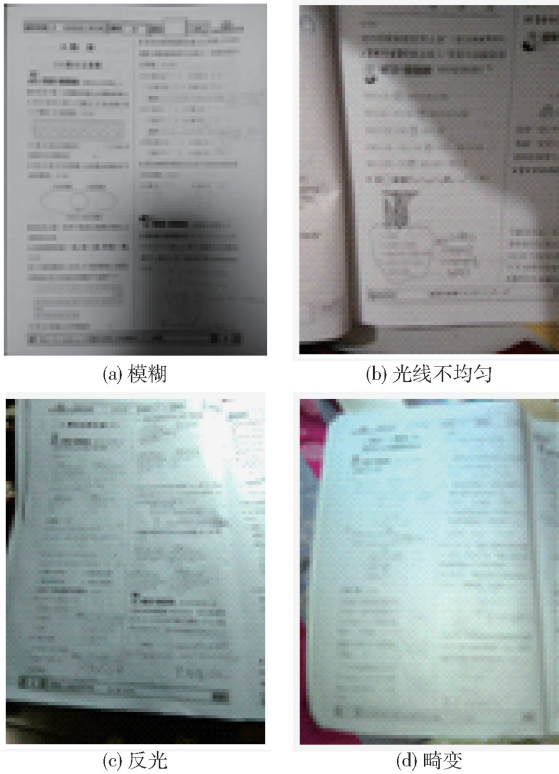


图 1 手机拍照产生的低质量图像

手机拍照产生的图像一般比较模糊,噪点多,加上文本内容本身有很多文字、图表、图形等。实验表明,图像经过霍夫变换产生的线段比较多,而且会产生大量的短线段,如图 2(a)所示;对于弯曲的边界,则会产生多条线段,如图 2(b)所示。另外,标准霍夫变换算法时间复杂度 $O(N^3)$,即使经过优化,计算量仅可以降低到 $O(N^2 \lg_2 N)$ ^[4],当图像比较大,算法时间会很长,无法满足用户的实时需求。实验证明,霍夫变换并不适应于手机拍照的图像的边界检测。

1 边界检测算法

为了解决手机拍照图像中文档区域边界检测问

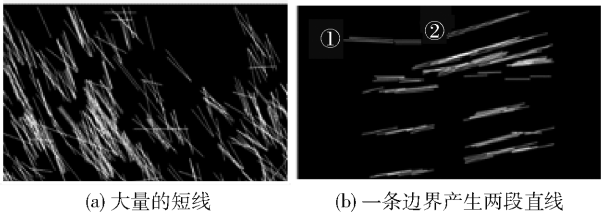


图 2 霍夫变换检测

题,采用动态规划算法^[5]。动态规划算法是由 Bellman 等在研究多阶段决策过程 (multistep decision process) 的优化问题时提出的。

在计算机视觉领域,动态规划有很多应用场景, Amini^[6]利用动态规划算法,通过建立主动轮廓模型实现目标物体的分割; Sonka 等^[7]利用动态规划检测直线; Merlet 等^[8]利用动态规划在遥感卫星图像中,检测道路方向; Ungru^[9]用动态规划在医学图像实现目标分割; Rosado-Tora 等^[10]提出基于极坐标下的成本函数的动态规划算法,实现目标物体的自动分割。

对于 N 行 M 列的图像,图像数字化后,文档的边界由离散的像素点组成,图像经过二值化处理后,像素点只有黑白 2 种状态,如图 3 所示。假设文档的边界由黑色像素点组成,在手机拍照图像中,文档区域占整个图像的 70% 以上,如果一条路径是文档区域的上边界,这条路径一定是从左到右,接近水平,经过前景点最多的路径。如图 3 所示,图中标出的路径 1,虽然不在一条直线上,但是相比于图中路径 2,更有可能是文档的边界。如果把图像的从左到右的每一列作为一个阶段,求文档区域上边界的问题,转化求一个最长路径问题,符合动态规划分阶段、最优化原理。

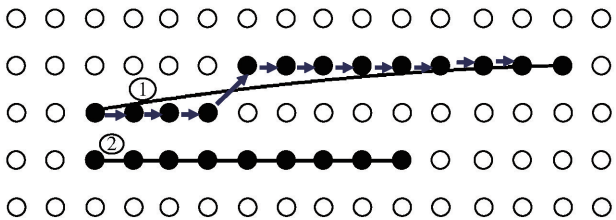


图 3 图像的离散化表示

对于文档区域边界,它的起点位置不能被确定,以搜索文档区域的上边界为例,需要从左到右遍历所有的像素点,计算每个像素点的成本函数,这个成本函数值代表文档区域边界经过此像素点的可能性,用 $E(i, j)$ 来表示。

考虑文档区域边界的特点,每个像素点成本函

数由两部分组成,到达此点的路径最优函数 $C(i,j)$ 和本身图像特征 $d(i,j)$.

根据动态规划思想,设计最优优化函数为

$$E(i,j) = C(i,j) + d(i,j) \quad (1)$$

其中 $C(i,j) = \max \{ C(i-1,l) + g(l,j) \}$

1.1 $C(i,j)$ 的计算

以文档区域的上边界为例,它有以下特征:从左到右、接近水平方向. 计算路径权重的时候,当前点的路径只能来自前一行相邻的元素. 如图4所示,点 $P_{i,j}$,路径选择只能从它周边的8邻域中选择3邻域,左上角点 $P_{i-1,j-1}$,左边点 $P_{i,j-1}$,左下角点 $P_{i+1,j-1}$.

对于像素点 $P_{i,j}$,左上角点 $P_{i-1,j-1}$ 和左下角的 $P_{i+1,j-1}$ 点到 $P_{i,j}$ 路径的权重相同,设定 Δ_1 与 Δ_3 相等. 为了体现上边界的近水平的特点,左边点 $P_{i,j-1}$ 到 $P_{i,j}$ 路径的权重应该大于左上角点 $P_{i-1,j-1}$ 和左下角的 $P_{i+1,j-1}$ 点到 $P_{i,j}$ 路径的权重,设定 $\Delta_2 > (\Delta_1 = \Delta_3)$.

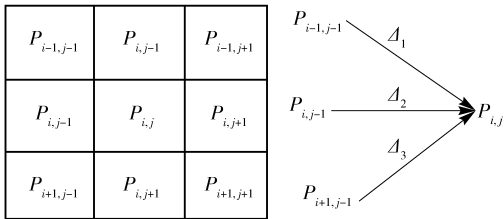


图4 图像像素点的邻域

$$\left. \begin{aligned} g(P_{i,j}, P_{i-1,j-1}) &= \Delta_1 \\ g(P_{i,j}, P_{i,j-1}) &= \Delta_2 \\ g(P_{i,j}, P_{i+1,j-1}) &= \Delta_3 \end{aligned} \right\} \quad (2)$$

经过大量实验分析, $\Delta_1, \Delta_2, \Delta_3$ 参数的取值为 $\Delta_1 = \Delta_3 = 0, \Delta_2 = 1$, 可以获得比较好的结果.

1.2 $d(i,j)$ 的计算

对于二值图像,图像像素值只有2个值,0和1,0代表背景点,1代表前景点,文档的边界会经过最多的前景点.

$$d(i,j) = \begin{cases} 0, & \text{若 } P_{i,j} = 0 \\ \theta, & \text{若 } P_{i,j} = 1 (\theta > 0) \end{cases} \quad (3)$$

通过实验证明 θ 与 Δ_2 取值相同,文档边界近水平特征与边界是前景点权重相同,可以获得比较好的结果.

在实验中,取图像上半部分的 $1/3$ 作为上边界的搜索区域,从上到下,从左到右计算图像每个像素点的 $E(i,j)$. 如果把 $E(i,j)$ 看作像素点能量,图像

能量光谱如图5(a)所示,图5(b)展示了最后一列每个像素的 $E(i,j)$ 形成的曲线图. 对最后一列的能量值,用非极大值抑制 (non-maximum-suppression) 求出局部的极大值,选取这些局部极大值点向左回溯得到可能的路径,如图6(b)所示,通过局部极大值回溯的路径,有可能是多条路径,取能量值最大的3条路径.

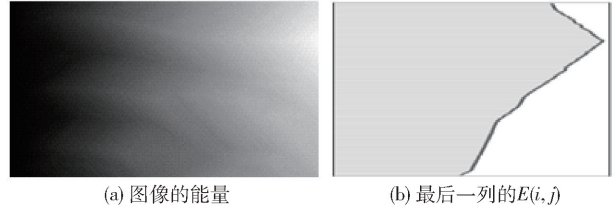


图5 图像能量光谱

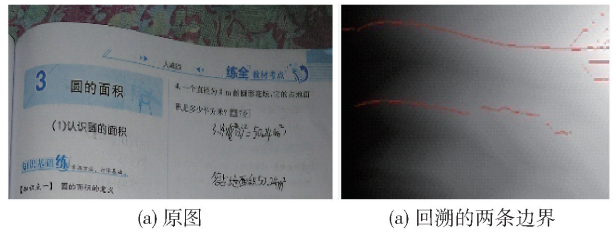


图6 边界回溯路径

1.3 最长路径计算及回溯方法

以多阶段决策图来具体说明如何通过动态规划求最长路径问题,如图7所示,将图划分为 $V_1 \sim V_5$ 5个阶段,求文档区域边界问题,可以类似为求多段图的最长路径问题. 求文档区域的上边界类似在图7中求从 $V_1 \sim V_5$ 的一条最大成本的路径问题.

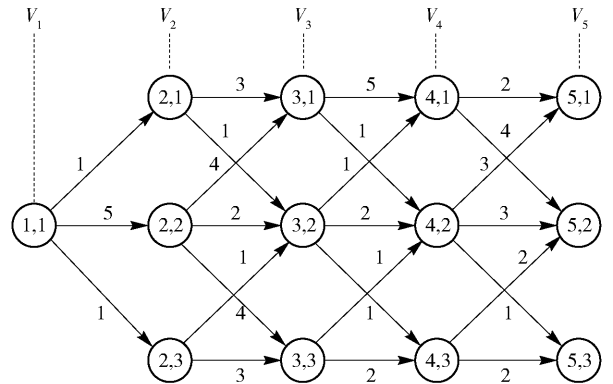


图7 多阶段决策图

用 $C(i,j)$ 表示路径最优函数,在 $C(i,j)$ 中表示 V_i ,代表决策的阶段, j 表示 V_i 阶段中的结点号, $C(i,j)$ 表示在 i 阶段,经过结点 j 的这条路径的成本.

$C(i,j)$ 的定义为

$$C(i,j) = \max\{C(i-1,l) + g(l,j)\} \quad (4)$$

其中: l 为 $i-1$ 阶段与结点 j 相连所有结点, $g(l,j)$ 为 $i-1$ 阶段所有结点与结点 j 连接的权重,如图8所示.

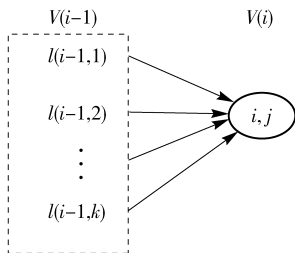


图8 $g(l,j)$ 结点的关系

分阶段递推结果

1) $K=2$ 阶段

$$C(2,1) = g(1,1) = 1 \quad (5)$$

$$C(2,2) = g(1,2) = 5 \quad (6)$$

$$C(2,3) = g(1,3) = 1 \quad (7)$$

2) $K=3$ 阶段

$$C(3,1) =$$

$$\max(C(2,1) + g(1,1), C(2,2) + g(2,1)) = \max(C(2,1) + 3, C(2,2) + 4) = 9 \quad (8)$$

$$C(3,2) = \max(C(2,1) + g(1,2),$$

$$C(2,2) + g(2,2), C(3,2) + g(3,2)) =$$

$$\max(C(2,1) + 1, C(2,2) + 2, C(2,3) + 1) = 7 \quad (9)$$

$$C(3,3) = \max(C(2,2) + g(2,3),$$

$$C(2,3) + g(3,3)) =$$

$$\max(C(2,2) + 4, C(2,3) + 3) = 9 \quad (10)$$

3) $K=4$ 阶段

$$C(4,1) = \max(C(3,1) + g(1,1),$$

$$C(3,2) + g(2,1)) =$$

$$\max(C(3,1) + 5, C(3,2) + 1) = 14 \quad (14)$$

$$C(4,2) = \max(C(3,1) + g(1,2),$$

$$C(3,2) + g(2,2), C(3,3) + g(3,2)) =$$

$$\max(C(3,1) + 4, C(3,2) + 3, C(3,3) + 4) = 10 \quad (12)$$

$$C(4,3) = \max(C(3,2) + g(2,3),$$

$$C(3,3) + g(3,3)) =$$

$$\max(C(3,2) + 1, C(3,3) + 2) = 8 \quad (13)$$

4) $K=5$ 阶段

$$C(5,1) = \max(C(4,1) + g(1,1),$$

$$C(4,2) + g(2,1)) =$$

$$\max(C(4,1) + 2, C(4,2) + 4) = 16 \quad (14)$$

$$C(5,2) = \max(C(4,1) + g(1,2),$$

$$C(4,2) + g(2,2), C(4,3) + g(3,2)) =$$

$$\max(C(4,1) + 4, C(4,2) + 3, C(4,3) + 1) = 18 \quad (15)$$

$$C(5,3) = \max(C(3,2) + g(2,3),$$

$$C(3,3) + g(3,3)) =$$

$$\max(C(3,2) + 1, C(3,3) + 2) = 11 \quad (16)$$

在 $K=5$ 阶段,最大值是 $C(5,2)$,路径长度最大成本为18. 通过向后回溯,求最长路径经过的结点.

在式(15)中,使 $C(5,2)$ 取得最大值是 $C(4,1)$,在 $K=4$ 阶段中,能使 $C(4,1)$ 取得最大值是 $C(3,1)$,在 $K=3$ 阶段中,能使 $C(3,1)$ 取得最大值是 $C(2,2)$,回溯求得的路径是: $(5,2), (4,1), (3,1), (2,2), (1,1)$.

1.4 文档边界检测算法流程

1) 图像宽度为 $M(m=1,2,\dots,M)$,高度为 $N(n=1,2,\dots,N)$,第1列所有结点的初始能量 $C(P_{i,1}) = D(P_{i,1}) (i=1,2,\dots,n)$.

2) 对于所有的 $m=1,2,\dots,M$ 重复步骤3).

3) 在图像中 m 列的 (n,n) 点,对于前 $m-1$ 列,重复步骤4).

4) 设

$$E(i,j) = \max_i [E(i-1,j) + g(l,j)] + d(i,j) \quad (17)$$

5) 求最后一列像素的能量函数 $E(P_{i,M}, i=1, 2, \dots, n)$,通过极大抑制算法^[11],求出能量值为局部极大值的点,从这些局部极大值点回溯得到的路径,有可能是文档区域的边界,但是回溯的路径有可能多条.

2 边界组合算法

对于文档区域的上、下、左、右4个边界,分别通过动态规划算法来检测,由于手机拍照图像的质量不高,通过动态规划回溯的路径可能不仅仅是一条,一般的原则保留1~3条路径作为候选. 如果要确定一条路径是否是文档的区域,要通过其他边界关联来确定,因为只有4条边界合理组合,才能形成文档区域. 假定每条边界有3条候选边,则可能形成 $(3 \times 3 \times 3 \times 3 = 81)$ 81个组合,为了快速判定一个组合是文档区域,通过逻辑回归的二分类算法来处理这个问题.

通过动态规划算法检测的路径,由离散的点组成,这些点未必在一条直线上,通过直线拟合法,把检测到的文档区域的边界转化为直线,求得 4 条边界的交点: $P_1(x_1, y_1), P_2(x_2, y_2), P_3(x_3, y_3), P_4(x_4, y_4)$,如图 9 所示。

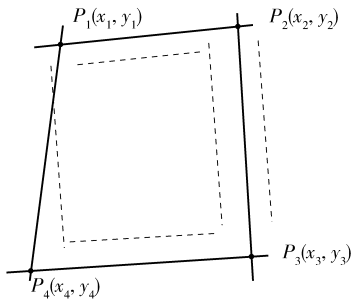


图 9 文档区域的边界交点

如果 4 条边界形成了文档区域,则这些边界有下列特点:

- 相对的 2 个边接近平行, $180^\circ \pm 30^\circ$,平行特征: x_1 ;
- 相对的 2 个边距离相对较远,距离特征: x_2 ;
- 相临边的角度接近 $90^\circ \pm 30^\circ$,角度特征: x_3 ;
- 组成四边形面积大于整个图像面积 60%,面积特征: x_4 。

1) 根据上述特征设计逻辑回归的假设函数

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

因变量用 $h_{\theta}(x)$:

$h_{\theta}(x) = 1$, 是文档区域边界。

$h_{\theta}(x) = 0$, 非文档区域边界。

自变量 x_1, x_2, x_3, x_4 计算如下:

$$x_1 = \left(\frac{|y_2 - y_1|}{|x_2 - x_1|} - \frac{|y_3 - y_4|}{|x_3 - x_4|} \right) \left(\frac{|y_4 - y_1|}{|x_4 - x_1|} - \frac{|y_3 - y_2|}{|x_3 - x_2|} \right) \quad (18)$$

$$x_2 = \frac{|x_2 - x_1| + |x_3 - x_4|}{2w} \quad (19)$$

其中 w 为图像宽度

$$x_3 = \tan^{-1} \frac{y_1}{x_1} - \tan^{-1} \frac{y_2}{x_2} \quad (20)$$

$$x_4 = \frac{\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 2 \end{vmatrix} + \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_4 & y_4 & 1 \end{vmatrix}}{wh} \quad (21)$$

其中 w, h 分别为图像宽度和高度。

2) 建立价值函数

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

3) 用梯度下降法不断修正 θ 值,直至收敛。

$$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$$

其中 α 为学习速率。

4) 求得最终 $(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$

通过 1 000 个正样本和 1 000 个负样本,利用 Matlab 工具箱逻辑回归函数,计算出

$$h_{\theta}(x) = 0.0347 + 0.1867x_1 - 1.6754x_2 + 0.07896x_3 + 0.02367x_4$$

3 实验

3.1 PC 机上测试

PC 机的配置:酷睿双核 i5 处理器,主频 2.7 GHz,内存 32 GB。

1) 边界检测测试

对不同场景下图片,包括清晰、模糊、非水平畸变、光照不均匀的情况进行测试。统计边界检测的准确率和运行时间。

测试图片的总数量:4 000 张

清晰的图片:1 000 张

比较模糊图像:1 000 张

光照不均匀:1 000 张

非水平畸变:1 000 张

测试图像分辨率大小:1 600 × 1 200,从以下 2 个方面测试算法效果。

找边成功率:能从背景图像中找到候选的边,每个边界取 TOP 3,即 0 ~ 3 个边。

运行时间:检测到文档区域边界的时间。

测试结果如表 1 所示。

表 1 PC 机上测试结果

类型	图像数量/张	找边成功率/%	平均运行时间/ms
清晰	1 000	98.56	60
模糊	1 000	94.70	65
不均匀光照	1 000	96.56	75
畸变	1 000	94.56	78
平均	4 000	95.85	70

2) 边界组合测试

输入自变量 x_1, x_2, x_3, x_4 ,通过二分类的逻辑回归算法计算 $h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$,判定选定的 4 个边是否为文档区域。测试结果如表 2 所示。

表 2 逻辑回归的判定结果

类型	图像数量/张	准确率/%	平均运行时间/ms
清晰	200	95.56	20
模糊	300	92.70	43
不均匀光照	300	91.56	35
畸变	200	92.56	38
平均	1 000	93.00	34

3.2 在手机上测试

测试方法:模拟用户的拍照场景,测试图像数据来源手机拍照的视频流。

手机型号:华为 DIG-AL00
CPU 型号:骁龙 435 (MSM8940)
CPU 参数:八核,1.4 GHz
存储 ROM:32 GB, RAM:3 GB

边界检测算法在手机上运行,速度非常重要,如果边界检测的速度较慢,用户感觉反应迟钝,体验效果差。为了提高边界检测的速度,图像大小缩放到 800 × 600。与网易的有道云笔记^[12]相比,确定文档区域边界的准确率和运行速度超过同类产品,测试结果如表 3 所示。

表 3 手机上测试结果

算法	确定文档区域的准确率/%	帧率/(帧·s ⁻¹)
本文	780/1 000 = 78	10.05
有道	720/1 000 = 72	9

在手机上测试算法的准确率低于 PC 上准确率,主要有以下原因:

- 1)在测试过程中,手机处于运动状态,手机的焦距不准,模糊图像所占的比例比较大;
- 2)手机拍照过程中,不断地调整相机的角度和远近,由于图像是实时抓取,有些视频流图像帧中包含部分文档的区域,文档区域的边界不完整,边界检测的失败率比较高;
- 3)测试数据中准确率,不仅包含边界检测算法,而且包含边界组合的算法。

在实际应用场景下,手机上找边过程与用户是实时交互的,用户可以判断找边是否准确,当找边运算速度达到 10 帧/s,这个准确率是完全可以满足用户拍照需求的。

4 结束语

通过实验证明,动态规划算法可以很好地解决

手机拍照图像的文档区域边界检测问题,用逻辑回归的二分类算法解决边界组合问题。但是对于动态规划中的路径权重分配以及参数选择问题如何做到最优是下一步要重点研究的问题。

参考文献:

[1] Duda R Oand Hart P E. Using hough transform to detect lines and curves in pictures [J]. Communications of the ACM, 1972, 15(1): 11-15.

[2] Xu L, Oja E. Randomized hough transform (RHT): basic mechanisms, algorithms, and computational complexities [J]. CVGIP: Image Understanding, 1993, 57(2): 131-154.

[3] Zhang Zhenyou, Li Weihe. White board scanning and image enhancement[J]. Digital Signal Processing, 2007, 17(2): 414-432

[4] 孙丰容,刘积仁. 快速霍夫变换算法 [J]. 计算机学报,2001, 24(10): 1102-1109.
Sun Fengrong, Liu Jiren. Fast Hough transform algorithm [J]. Journal of Computer Science, 2001, 24(10): 1102-1109.

[5] Bellmann R E. Dynamic programming [M]. Princeton, NJ: Princeton University Press, 1957.

[6] Amini A A. Using dynamic programming for solving variational problems in vision [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(9): 855-867.

[7] Sonka M, Hlavac V. Image processing, analysis and machine vision [J]. Journal of Electronic Imaging, 2008(82): 685-686.

[8] Merlet N, Zerubia J. New prospects in line detection by dynamic programming [J]. IEEE Transactions on Pattern Analysis and machine intelligence, 1996, 18(4): 426-431.

[9] Ungru K, Jiang X Y. Dynamic programming based segmentation in biomedical imaging [J]. Computational and Structural Biotechnology Journal, 2017(15): 255-264.

[10] Rosado-Toro J A. Dynamic programming using polar variance for image segmentation [J]. IEEE Transactions on Image Processing, 2016, 25(12): 5857-5866.

[11] Zaytseva E, Vitriù J. A search based approach to non maximum suppression in face detection [C]// 2012 19th IEEE International Conference on Image Processing. 2012: 1469-1472.

[12] 北京网易有道计算机系统有限公司. 有道云笔记 [EB/OL]. (2017-05-10) [2017-06-10].
http://note.youdao.com/download.html#androi.