视频字幕的自动检测与去除

季丽琴 王加俊
(苏州大学电子信息学院, 苏州 215021)

摘 要 由于视频中固化的字幕影响了不同语种间视频的交流和处理,为此提出了一种基于 CEMA 算法和纹理修复技术的自动检测与去除视频内字幕的方法。首先,运用 CEMA 算法检测出视频中的字幕,然后,结合纹理修复技术,将检测出来的字幕从原图中去除,同时,恢复原图中被字幕所遮挡的背景区域。实验结果表明,该方法能较好地检测和去除视频图像内的字幕。

关键词 彩色边缘检测 数学形态学 文字检测 文字去除 图像修复
中图法分类号: TP391.4 **文献标识码**: A **文章编号**: 1006-8961(2008)03-0461-06

Automatic Text Detection and Removal in Video Images

JI Li-qin, WANG Jia-jun
(*School of Electronics and Information Engineering, Soochow University, Suzhou* 215021)

Abstract The texts embedded in video images have obstructed video intercommunication and processing among different languages. This paper proposes an approach for automatic text detection and removal in video images based on CEMA and texture restoration. First, text regions in the video images are detected by a CEMA-based algorithm. Second, we remove the detected texts from the original video images by texture-based restoration technology; meanwhile, restore the background occluded by texts in video images. Experiment results show that the proposed method performs fairly well.

Keywords color edge detection, mathematical morphology, text detection, text removal, image inpainting

1 引 言

随着计算机技术、多媒体技术以及通信技术的飞速发展,视频已经成为人们日常生活中不可缺少的娱乐活动之一。老式的视频由于技术限制,字幕被固化于视频中,成为视频的一部分。但在实际应用中人们发现,固化的字幕严重阻碍了不同语种间的视频交流。因此,如果将字幕从视频中去除,并修复被字幕所遮挡的背景,然后添加上多语种的字幕,则将对不同语种间的视频交流、对视频的再次使用是非常有价值的。

视频字幕的自动检测与去除和图像中斑点、线形刮痕的检测与去除及其相似。但视频中的字幕往往在尺寸上大于斑点和刮痕,所以不适合采用与处理斑点或刮痕的相同的方法来去除视频中的字幕。Bertalmio 等人在 2000 年提出了一种基于投影迭代传播的 BSCB 算法[1]来修复用户指定的区域(如文字区域);Chan 等人在 2001 年提出了基于 Total Variational 模型[2]和基于曲率的扩散模型(curvature-driven diffusion)[3]的修复算法。这些算法主要利用边界信息向待修复区域内进行各向异性的迭代扩散,计算量大,在修复区域较小时有较好的效果。Chang Woo Lee 等人提出利用相邻帧的信息来填充待修复帧[4],这对于持续多帧不变的字幕显然不合适。Criminisi 等人提出基于纹理生成的修复方法[5],在待修复区域的边界通过块匹配的方式选择合适的纹理填充。这种方法计算量小,对大区域也

基金项目:国家自然科学基金资助项目(30300088)
收稿日期:2006-03-24;**改回日期**:2006-10-26
第一作者简介:季丽琴(1980~),女。2006 年获苏州大学通信与信息系统专业硕士学位。主要研究方向为数字图像处理。E-mail: jiliqin2003@163.com

有较好的修复效果。

基于上述思想,本文利用视频中字幕的边缘信息,提出先用基于 CEMA(color-edge detection, morphology, logic operator "AND", CEMA)的文字检测算法来对视频中的字幕进行精确地定位,然后用纹理修复算法去除字幕,同时,填补原视频中被字幕所遮挡的背景区域,使视频达到或接近未加字幕前的视觉效果。实验结果表明,本文方法能较好地去除视频图像内的字幕。

2 字幕检测

从视频图像中检测文字的相关工作在各方面已有很多进展:Keechul Jung 等人提出一种综合运用纹理和连通组元分析的方法[6]来定位文字,组建基于多层感知器(MLP)的纹理分类器和基于连通分量(CC)的滤波器,整个算法复杂,且需要足够的训练样本;Kim 等人提出用支持向量机(SVM)的纹理分类器来检测视频中的文字[7],该方法的检测结果虽然较好,但是计算量大,且同样需要大量的训练样本。Li 等人通过小波变换提取特征并采用神经网络分类器来检测文字[8]等等。

视频中的字幕一般位于视频的下部,呈水平排列,相同字幕的帧间位置基本不变,基于以上特点,本文提出基于 CEMA 的字幕检测算法,处理过程如图 1 所示。

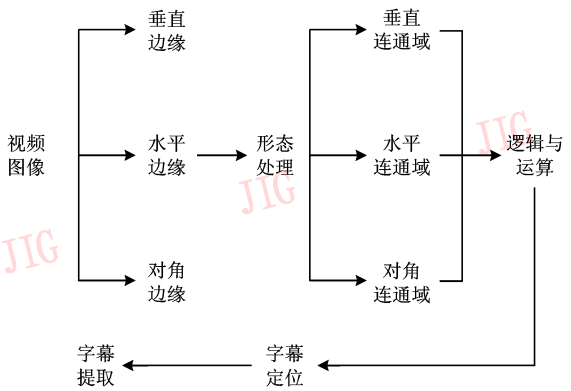


图 1 字幕检测算法 CEMA 的流程图

Fig. 1 Flowchart of the algorithm CEMA on text detection

2.1 垂直、水平对角边缘的提取与二值化

由于视频中的字幕与背景有较强的对比性,表现为在字幕与背景的交界处,存在十分明显的高频区域,因此可以用提取边缘的方法来大致估计出字幕可能存在的区域。文献[8]将小波分析推广到 2 维情形,即通过多分辨率分析和 Mallat 塔式分解方法,得到文字在垂直、水平以及对角方向的分量图。在分析了上述方法的基础上,本文提出用 3 个简单的不同方向的彩色边缘检测模板来代替小波变换,如图 2 所示。

0	0	0
-3	3	0
0	0	0

(a) 垂直方向

0	-3	0
0	3	0
0	0	0

(b) 水平方向

-3	0	0
0	3	0
0	0	0

(c) 对角方向

图 2 3 个不同方向的边缘检测模板

Fig. 2 Three edge detection masks in different directions

这 3 个检测模板分别作用于彩色图像的红、绿、蓝 3 个分量上提取边缘,以像素点(i, j)为例,定义垂直方向模板的检测(其他方向的检测算子雷同)如下:

$$\boldsymbol{R}_{\mathrm{V}}(i, j) = \sum_{k=1}^{9} e_{\mathrm{V}}(k) * \boldsymbol{R}(k) \tag{1}$$

$$\boldsymbol{G}_{\mathrm{V}}(i, j) = \sum_{k=1}^{9} e_{\mathrm{V}}(k) * \boldsymbol{G}(k) \tag{2}$$

$$\boldsymbol{B}_{\mathrm{V}}(i, j) = \sum_{k=1}^{9} e_{\mathrm{V}}(k) * \boldsymbol{B}(k) \tag{3}$$

其中,$\boldsymbol{R}_{\mathrm{V}}(i, j)$、$\boldsymbol{G}_{\mathrm{V}}(i, j)$、$\boldsymbol{B}_{\mathrm{V}}(i, j)$分别是利用垂直检测模板在像素点$(i, j)$处获得的红、绿、蓝分量;$\boldsymbol{R}(k)$、$\boldsymbol{G}(k)$和$\boldsymbol{B}(k)$分别是在像素点$(i, j)$处及它的 8 邻域内读取到的红、绿、蓝分量;$e_{\mathrm{V}}(k)$是垂直边缘检测模板中的数值。图 3 给出了检测模板扫描像素点(i, j)周围的像素的顺序。

(a) 模板扫描像素的顺序 (b) 当前像素点(i, j)

图 3 检测模板扫描像素的顺序

Fig. 3 The scanning order of pixels in detection masks

通过以上模板的检测,得到垂直、水平和对角方向的彩色边缘图像$\boldsymbol{E}_{\mathrm{V}}(x, y)$、$\boldsymbol{E}_{\mathrm{H}}(x, y)$、$\boldsymbol{E}_{\mathrm{D}}(x, y)$,从而避免了运用复杂的小波变换来提取文字边缘特征,体现了本文方法的可行性和简单性。

对于字幕提取来说,边缘图像的二值化也是很重要的问题。如果阈值过大,可能会漏掉一些文字边缘,而阈值过小,则可能会使较多的非文字边缘被当作文字边缘处理,导致误检较多。本文针对不同图像的具体情况动态选取阈值,采用最小误差法求阈值的算法[9],分别对 $E_V(x,y)$、$E_H(x,y)$、$E_D(x,y)$进行二值化,得到二值边缘图像 $\hat{E}_V(x,y)$,$\hat{E}_H(x,y)$,$\hat{E}_D(x,y)$,实验结果表明,基于最小误差法求阈值的方法取得了很好的效果。

2.2　形态处理

形态学可将图像信号与其几何形状联系起来,利用一定形态的结构元素度量和提取图像中的对应形状和结构,本文采用形态学处理来提取字幕在视频图像中对应的形状。

形态学最基本的概念是腐蚀和膨胀,以及由它们组合而成的各种形态操作算子。设 $\boldsymbol{\Omega}$ 为 2 维欧几里德空间,图像 $\boldsymbol{A}$ 是 $\boldsymbol{\Omega}$ 的一个子集,结构元素 $\boldsymbol{S}$ 也是 $\boldsymbol{\Omega}$ 的一个子集,$\boldsymbol{b}\in\boldsymbol{\Omega}$ 是欧式空间的一个点,定义 4 个基本运算:

(1) 膨胀　$\boldsymbol{A}\oplus\boldsymbol{S}=\{\boldsymbol{a}+\boldsymbol{b}\mid\boldsymbol{a}\in\boldsymbol{A},\boldsymbol{b}\in\boldsymbol{S}\}$

(2) 腐蚀　$\boldsymbol{A}\ominus\boldsymbol{S}=\{\boldsymbol{z}\in\boldsymbol{\Omega}\mid\boldsymbol{S}^{\boldsymbol{z}}\subseteq\boldsymbol{A}\}$,其中,$\boldsymbol{S}^{\boldsymbol{z}}$ 表示为 $\boldsymbol{S}$ 被 $\boldsymbol{z}$ 平移后的结果。

(3) 开运算　$A\circ\boldsymbol{S}=(\boldsymbol{A}\ominus\boldsymbol{S})\oplus\boldsymbol{S}$

(4) 闭运算　$A\bullet\boldsymbol{S}=(\boldsymbol{A}\oplus\boldsymbol{S})\ominus\boldsymbol{S}$

其中,膨胀具有扩大目标区域的作用,腐蚀具有收缩目标区域的作用,开运算可删除目标区域中的小分支,闭运算可填补目标区域中的空洞。基于以上 4 个运算,本文的形态处理流程为

(1) 1 次闭　采用 8 连通结构元素填补边缘图像中文字区域的空洞;

(2) 1 次开　采用 8 连通结构元素删除文字区域中的小分支;

(3) 6 次水平膨胀　因视频中文字一般是水平方向分布,所以为了有效地形成文字连通域,采用水平结构元素,即 $\boldsymbol{S}=\{1,1,1,1,1,1\}$。

(4) 3 次水平腐蚀　因膨胀后文字区域明显大于实际的文字区域的大小,所以需要进行水平方向的腐蚀,采用水平结构元素,即 $\boldsymbol{S}=\{1,1,1,1,1\}$。

实验发现,若水平结构元素的尺寸太大,会导致无效的膨胀重叠现象,增大计算量,而水平结构元素尺寸过小,将不能有效地形成连通候选字幕区域,所以结构元素 $\boldsymbol{S}$ 的选择对于候选连通字幕区域的形成与字幕区域的提取至关重要。实验结果表明,本文所采用的结构元素能很好地形成各个方向的候选连通字幕区域 $\boldsymbol{R}_V(x,y)$、$\boldsymbol{R}_H(x,y)$、$\boldsymbol{R}_D(x,y)$。

2.3　基于逻辑与运算的字幕区域定位

为了定位最终的字幕区域,本文提出了将 3 幅连通域图进行逻辑与运算,因为此运算综合了垂直、水平和对角方向的字幕信息,能有力地保证字幕区域存在的准确性和精确性。而实验也证明,经过此运算后,去掉了很大部分噪声区域,得到较精确的连通域图 $\boldsymbol{R}_I(x,y)$,用公式表示为

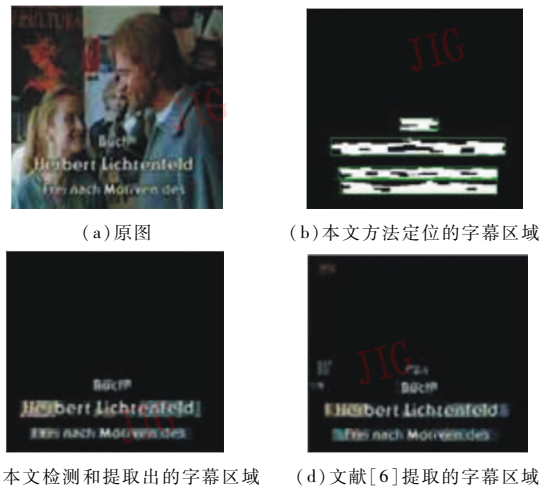
$$\boldsymbol{R}_I(x,y)=\boldsymbol{R}_V(x,y)\cap\boldsymbol{R}_H(x,y)\cap\boldsymbol{R}_D(x,y)\quad(4)$$

但 $\boldsymbol{R}_I(x,y)$ 中仍可能存在一些虚假的不含字幕的连通区域,需对所有连通区域做进一步分析。和文献[10]中采用的 7 个判定规则来判断所有连通区域相比,本文在进行逻辑与运算之后,仅需运用一个判定规则,即可得到确定的只有文字区域的连通域图 $\boldsymbol{R}_I'(x,y)$。即采取递归算法[11]来统计各连通区域的白色像素总数(*pixelNum*),若 *pixelNum* 小于 *areapixel*(*areapixel* = 图像高度 × 图像宽度/150),则将虚假的不含字幕的区域从图像中删除。可见,本文的字幕区域定位方法简单且非常有效。

2.4　字幕区域的提取

本文采用递归算法求出字幕区域的最小外接矩形,此算法的大致思想为:首先定义一个全局 Crect 类型数组变量 textLocation[*n*](*n* 为图像中的字幕区域数),*top*、*left*、*bottom*、*right* 分别为 textLocation[*n*] 的 4 个成员变量,代表外接矩形的左上角和右下角的坐标。若像素点 (x,y) 为白色,则扫描点 (x,y) 的 4 邻域像素点,若 4 邻域中仍存在白色像素点,则调整全局变量 textLocation[*n*] 结构中的 *top*、*left*、*bottom*、*right* 的大小,然后再进行递归调用。算法结束返回的 textLocation[*n*] 即为所求的最小外接矩形。

图 4 给出了视频字幕检测的中间结果。根据图 4(c)可以看出,本文提出的基于 CEMA 的字幕检测算法不但简单,而且对于字幕区域的定位很精确,而从图 4(d)中可看出,提取出的字幕区域中有许多是非字幕区域,而原图中本是文字的,却没有提取出来。可见,本文提出的方法要优于文献[6]所采用的方法。



(a)原图 (b)本文方法定位的字幕区域

(c)本文检测和提取出的字幕区域 (d)文献[6]提取的字幕区域

图 4 字幕区域提取的比较

Fig. 4 Comparison of the extracted text regions

2.5 字幕区域的后处理

纹理修复的目标区域并非是整个字幕区域,而只是字幕区域中的字幕像素,不包括字幕区域中的背景。一般情况下,要区分字幕像素还是背景区域的前提,是字幕像素的周围存在一些高亮度的像素。如图 5(a)所示,在字幕像素“n”的边缘,存在一些黑色像素,这些像素突出了“n”的存在,起到了区分像素“n”和具有相似颜色的背景区域的作用。因此,在去除字幕像素的时候,除了要精确地去除字幕像素外,还要把这些边缘黑色像素考虑进去,即也要将它们去除。因此,本文在采用八叉树颜色量化算法[12]对字幕区域二值化后(如图 5(b)所示),运用形态处理的方法,即采用 3×3 的结构元素,膨胀已提取出来的字幕像素,直到完全包含字幕的所有边缘像素(如图 5(c)所示),这样才能保证可以精确地去除视频图像中的所有字幕,完全地修复被字幕遮挡的背景区域。

(a)边缘存在高亮度像素的例子

(b)字幕区域的二值化结果 (d)形成较大的待修复区域

(e)纯绿色的待修复区域

图 5 字幕区域的后处理

Fig. 5 Post processing of the text region

3 视频修复

在修复视频前,本文将形成的待修复区域映射到原彩色视频中(如图 5(d)所示),并将所要修复的区域用一种颜色表示(文中采用纯绿色),以示区别。然后,使用 Criminisi 等人提出的纹理修复算法[5]进行视频修复。

如图 6 所示,设整幅图像为 $\boldsymbol{I}$,待修复的目标区域表示为 $\boldsymbol{\Omega}$,区域的边界为 $\boldsymbol{\delta\Omega}$,$\boldsymbol{\Phi}$ 为视频中的非修复区域,$\boldsymbol{\Psi}_p$ 是以像素点 $\boldsymbol{p}$ 为中心点的待修复正方形模板。首先,修复算法需要计算出 $\boldsymbol{\Psi}_p$ 的优先级 $\boldsymbol{P}(\boldsymbol{p})$,其定义如下:

$$P(\boldsymbol{p}) = C(\boldsymbol{p}) \times D(\boldsymbol{p}) \tag{5}$$

式中,$C(\boldsymbol{p})$称为待修复模板 $\boldsymbol{\Psi}_p$ 的置信度,$D(\boldsymbol{p})$为 $\boldsymbol{\Psi}_p$ 的数据信息项,代表 $\boldsymbol{p}$ 处的等幅透线与待修复区域边界 $\boldsymbol{\delta\Omega}$ 碰撞的强度。它们各自的定义如下所示:

$$C(\boldsymbol{p}) = \frac{\sum_{\boldsymbol{q}\in\boldsymbol{\Psi}_p\cap\bar{\boldsymbol{\Omega}}} C(\boldsymbol{q})}{|\boldsymbol{\Psi}_p|} \tag{6}$$

式中,规定函数 $C(\boldsymbol{p})$的初始值为

$$\begin{cases} C(\boldsymbol{p}) = 0 & \forall \boldsymbol{p} \in \boldsymbol{\Omega} \\ C(\boldsymbol{p}) = 1 & \forall \boldsymbol{p} \in \boldsymbol{I} - \boldsymbol{\Omega} \end{cases} \tag{7}$$

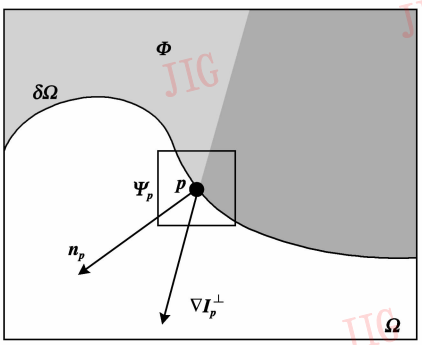


图 6 纹理修复示意图

Fig. 6 Schematic illustration of the texture-based inpainting

式中,q 为模板 Ψ_p 中无需修复的像素,$|\Psi_p|$ 是模板 Ψ_p 的面积。

$$D(p) = \frac{|\nabla I_p^{\perp} \cdot n_p|}{\alpha} \tag{8}$$

式中,$\nabla I_p^{\perp}$ 是 p 处的等幅透线(包括方向和强度),n_p 是点 p 在待修复区域边界 $\delta\Omega$ 的法向量。α 代表归一化因子(如对于一个典型的灰度图像来说 $\alpha = 255$)。此后,依次求出以待修复区域边界上的各像素点为中心的模板的优先级,从而得到优先级最大、最先修复的模板 $\Psi_{\hat{p}}$。用公式表示为

$$\Psi_{\hat{p}} \mid \hat{p} = \arg\max_{p \in \delta\Omega} P(p) \tag{9}$$

然后,从非修复区域 Φ 中找出与待修复模板 $\Psi_{\hat{p}}$ 最相似的匹配模板 $\Psi_{\hat{q}}$,用公式表示为

$$\Psi_{\hat{q}} \mid \hat{q} = \arg\max_{\Psi_q \in \Phi} d(\Psi_{\hat{p}}, \Psi_q) \tag{10}$$

式中,距离 $d(\Psi_{\hat{p}}, \Psi_q)$ 定义为两个模板中无需修复的像素之间的差值和;最后,将匹配模板 $\Psi_{\hat{q}}$ 覆盖待修复模板 $\Psi_{\hat{p}}$ 的区域,从而实现了视频的一次修复。为了完全地修复视频,需要重复以上操作,直至视频图像中不存在待修复区域。

4　实验结果和比较

基于上述技术,本文在 Windows 2000 环境下使用 Visual C++6.0 实现了视频字幕的去除,以图 4(a)为例,得到的修复过程中的结果及最后的修复结果如图 7 所示。

(a) 修复过程 1

(b) 修复过程 2

(c) 修复过程 3

(d) 最终的修复结果

图 7　修复过程中的结果

Fig. 7　The results in the course of inpainting

通过目视图 7(d),可以清楚地看出原来视频中被字幕遮挡的背景都被很好地修复了,所以,本文方法很好地完成了视频修复的目标。图 8 给出了文献[4]方法和本文方法的一个比较。

(a) 原图

(b) 文献[4]方法修复的结果

(c) 本文方法修复的结果

图 8　两种方法修复的结果

Fig. 8　The results of inpainting with two different methods

从图 8 可看出,本文方法与文献[4]方法的修复结果相差无几,但是,文献[4]方法是利用相邻帧的信息来填充待修复帧,这对于持续多帧不变的字幕显然不合适,而且计算量大,而本文提出先用基于 CEMA 的算法检测出视频字幕,然后结合了算法简单,计算量小的纹理修复方法,很好地去除了字幕,恢复了背景。所以本文方法要优于文献[4]方法提出的方法。

为了体现本文方法的鲁棒性,对视频中的倾斜字幕也做了一个实验。如图 9 所示。

通过目视,发现修复后的视频(图 9(c)所示)与未加入字幕的原有视频(图 9(a)所示)基本一致。因此,本文方法较好地完成了视频修复的目标。

(a) 原视频(未加入字幕) (b) 人工加入倾斜字幕后的视频 (c) 修复后的视频

图 9 倾斜字幕的去除

Fig. 9 Removal of the inclined texts

5 结 论

本文先用基于 CEMA 的字幕检测算法,提取出视频中的字幕,该方法简单且有效。然后,结合纹理修复技术,将检测出来的字幕从原图中去除,同时恢复原图中被字幕所遮挡的背景区域。实验结果表明,该方法能较好地检测和去除视频图像内的字幕。

由于本文处理的视频对象都是静止的,即视频字幕是静止的,视频背景也是静止的,所以今后进一步的工作可以考虑如何将本文方法推广到静止的字幕覆在运动的背景上或者运动的字幕覆在静止和运动的背景上;同时进一步改进本文的纹理修复方法,将该算法和其他图像修复算法综合起来,以便更快、更好地修复视频。

参考文献(References)

1 Bertalmio M. Image inpainting[A]. In: Preceedings of ACM SIGGAPH' 2000[C], New Orleans, Louisiana, USA, 2000: 417 ~ 424.

2 Chan T, Shen F. Mathematical models for local nontexture inpaintings[J]. SIAM Journal on Applied Mathematics, 2002, **62**(3): 1019 ~ 1043.

3 Chan T, Shen J. Non-Texture Inpainting by Curvature Driven Diffusions[R]. TR00-35, Department of Mathematics, University of California-Los Angeles, Los Angeles, California, USA, 2000.

4 Chang W L. Automatic text detection and removal in Video Sequences [J]. Pattern Recognition Letters, 2003, **24**(15): 2607 ~ 2623.

5 Criminisi A, Perez P, Toyama K. Object removal by exemplar based inpainting[A]. In: Preceedings of Computer Society Conference on Computer Vision and Pattern Recognition[C], Madison, Wisconsin, USA, 2003: 721 ~ 728.

6 Keechul J, Jung H H. Hybrid approach to efficient text extraction in complex color images [J]. Pattern Recognition Letters, 2004, **25**(6): 679 ~ 699.

7 Kim K I, Jung K, Park S H, *et al*. Support vector machines for texture classification[J]. In: IEEE Transactions on Image Processing, 2002, **24**(11): 1542 ~ 1550.

8 Li H, Doermann D, Kia O. Automatic text detection and tracking in digital video[J]. IEEE Transactions on Image Processing, 2000, **9**(1): 147 ~ 156.

9 Liu Q. The research of text extraction based on color images[D]. Chengdu: Communication University of SouthWest China, 2003. [刘倩. 基于彩色图像的文本区域提取研究[D]. 成都: 西南交通大学, 2003.]

10 Zhang Yin, Pan Yun-he. The noval approach of text extraction in color images and videos[J]. Jounal of Computer-aided design & Computer Graphics, 2002, **14**(1): 36 ~ 40. [张引,潘云鹤. 面向彩色图像和视频的文本提取新方法[J]. 计算机辅助设计与图形学学报, 2002, **14**(1): 36 ~ 40.]

11 Xu Hui. The Choicenesses of practical project cases in Digital image with Visual C + + (First Edition)[M]. Beijing: Posts&Telecom Press, 2004: 345 ~ 347. [徐慧著. Visual C + + 数字图像实用工程案例精选(第一版)[M]. 北京: 人民邮电出版社, 2004: 345 ~ 347.]

12 Zhou Chang-fa. The Mastery of image processing programme with Visual C + + (First Edition)[M]. Beijing: Publishing House of Electronics Industry, 2004: 201 ~ 211. [周长发著. 精通 Visual C + + 图像处理编程(第一版)[M]. 北京: 电子工业出版社, 2004: 201 ~ 211.]