

文章编号: 1005—0523(2007)04—0059—05

基于均值聚类的车牌定位技术研究

蒋先刚, 郝 励, 刘玄第, 黄书明, 王亚伟

(华东交通大学 基础科学学院, 江西 南昌 330013)

摘要: 采用强制综合聚类中心的均值聚类思想对车牌图像进行区域分类, 在 HLS 颜色空间下, 利用数理统计的方法得到聚类颜色中心, 并对目标图像进行强制聚类, 以最终获得只包括车牌四种颜色的待处理的图像, 然后利用形态学、人工神经网络等技术对车牌进行车牌识别. 该方法在实践中进行了验证, 比基于二值化和灰度化处理思路的定位效果更好, 能对多种不同环境下的拍摄的车牌进行有效的识别.

关键词: 均值聚类; HLS 颜色空间; 数学形态学

中图分类号: TU45

文献标识码: A

1 概述

车牌定位主要利用车牌边缘、形状、颜色等特性, 结合形态学、小波变换、人工神经网络等技术对车牌进行车牌定位, 定位之后再通过神经网络等分类方法识别出每个字符, 从而通过计算机自动识别出图片上的车牌号码.

本文中采用的方法是在 HLS 颜色空间下, 用自定义聚类中心的方法将车牌图像中的像素点分为白、黑、黄和蓝这四种颜色群, 结合数字图像学中的腐蚀和膨胀等技术找出具有车牌特征的区域, 从而达到定位车牌的目的. 所谓强制综合聚类中心的均值聚类, 就是均值聚类的一种改进, 使聚类的初始中心和聚类结果向四种特定的车牌颜色聚合, 这种改进使得算法的迭代次数降低, 算法的针对性更强. 而自定义聚类和以往的灰度化或二值化处理车牌的定位方法相比其优点在于处理的始终是彩色图像, 而不是灰度图像, 这就避免了由于环境的复杂而带来的定位的不确定性, 图 1 是本文车牌定位的分析流程图.

2 基于强制综合聚类中心的均值聚类的方法

2.1 颜色空间的变换

根据不同的应用环境, 图像颜色的表示方法也不同, 人们最熟悉的是 RGB 颜色空间, 由于它的各个颜色间具有非线性相关性, 且这三种颜色不易表现出周围环境的影响, 比如光照、反射等, 在复杂背景的环境下, 不易进行颜色提取, 在这里我们用到了另一种易于区分的颜色空间, 即 HLS 颜色空间.

HLS 颜色空间中 $H(0 \leq H \leq 360)$, $L(0 \leq L \leq 1)$, $S(0 \leq S \leq 1)$ 分别代表颜色的色调 (Hue), 流明 (Luminance) 和饱和度 (Saturation), 和它类似的有 HSV 颜色空间, 它们都是用 Munsell 三维空间坐标系统表示, 坐标之间的心理感知具独立性, 因此, 可以独立感知各颜色分量的变化; 且这种颜色模型具有线性伸缩性, 可感知的颜色差是与颜色分量的相应样值上的欧几里德距离成比例的.

2.2 图像聚类中心的数理统计

图像数理统计是通过对 800 张不同颜色车牌图片进行数学上的统计, 从而能够得到中国车牌常用底色的四群的初始中心值, 其结果见表 1.

表 1

颜色统计值	Hue	Lumens	Saturation
白 色	219.70936	0.535206288	0.033237215
黄 色	49.598122	0.432360512	0.806774168
蓝 色	204.96797	0.306038829	0.760536935
黑 色	245.93227	0.199116123	0.029748341

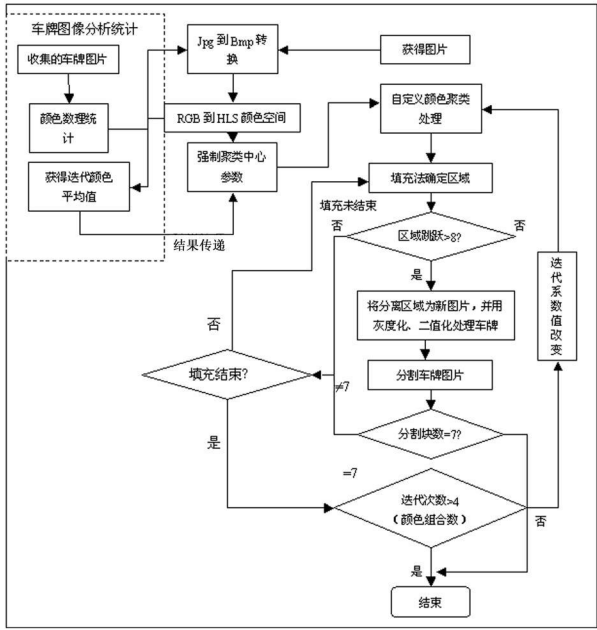


图 1 车牌定位分析流程图

2.3 分割后彩色区域的进一步处理

数学形态学(Mathematical morphology)是一门建立在图论和拓扑学基础之上的图像分析学科.其基本的运算包括:二值腐蚀和膨胀、二值开、闭运算、骨架抽取、灰值腐蚀和膨胀、灰值开闭运算等.

均值聚类处理后的车身和背景图像只包含四种中国车牌的颜色,为了消除断点、裂缝和孤立的小像素,可对其继续进行腐蚀运算和膨胀运算,图 2—图 5 是这些处理技术的效果比较.腐蚀运算和膨胀运算的方法如下:

如 X 和 B 是中的集合,X 被 B 膨胀定义为:

$$E=X\oplus B=\{z|\ [(B)_z\cap X]\subseteq X\}$$

其中 B 叫做膨胀的结构元素

另外定义为集合 B 的反射,即

$$B=\{w|w=-b,b\in B\}$$

$(B)_z$ 表示集合 B 平移到点 $z=(x,y)$,定义为:

$$(B)_z=\{c|c=b+z,b\in B\}$$

即用 B 膨胀 X 得到的集合 E 可以将其看作是 B 映射的位移与 X 至少有一个元素相交时,结构元素 B 的原点位置的集合.

膨胀运算相当于一个低通滤波器,经常应用于

连接图像断点,裂缝等场合.



图 2 原始图像

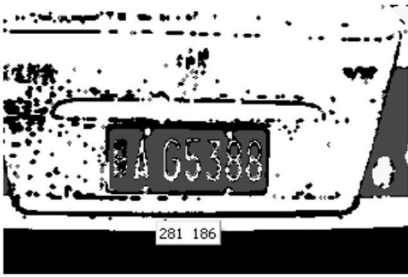


图 3 进行膨胀处理后的结果



图 4 原始图像



图 5 进行腐蚀处理后的结果

对 Z 中的集合 X 和 B,使用 B 对 X 进行腐蚀,用 $X\ominus B$ 表示,并定义为:

$$E=X\ominus B=\{z|\ (B)_z\subseteq X\}$$

这个公式说明,使用 B 对 X 进行腐蚀是 B 平移 z 后得到 $(B)_z$ 包含于 X,记下这个点 z,所以满足上述条件的 z 点组成的集合 E 称作 X 被 B 腐蚀(Erosion)的结果.

腐蚀运算经常用来从二值图像中消除不相关的细节,即可以滤除一孤立的小的像素.

3 新的车牌定位技术的理论技术及效率比较

K 均值聚类算法能够使聚类区域中所有样品到聚类中心的距离平方和最小,其原理为先取 K 个初始距离中心,计算每个样品到这 K 个中心的距离,找出最小距离并把样品归入最近的聚类中心,修改中心点的值为本类所有样品的均值,再计算各个样品到 K 个中心的距离,重新归类,修改中心点,直到新的距离中心等于上一次的中心点结束.

K—均值聚类算法如下:

1) 选取聚类块数 K;

2) 将句子中的每个词按照基于中心词扩展的方法分配到不同的聚类块中(有指导方法) K 个向量 $C_1, C_2, \dots, C_k$ 作为聚类中心;

3) 将每个样本向量 $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$ (n 为输入向量的维数),按下列欧式距离归入中心为的类中: $\|x_i - C_i\| = \min \|x_i - c_j\|$;

4) 重新调整聚类中心.令 $C = [C_1, C_2, \dots, C_k]^T$,

其中 C_{im} 由下式计算得出: $C_{im} = \frac{\sum_{i \in cluster_i} x_{ilm}}{N_i}$ 其中是第 i 个聚类块中的向量数;

5) 如果步骤(4)中的聚类中心 ($C_i, i=1, 2, 3, \dots, k$) 不再变化,就终止,否则转步骤 4).

下面介绍改进的自定义值聚类法及实现技术,由上面原理可知道 K 均值聚类首先选 K 个聚类中心,一般为颜色空间的 K 均分,然后迭代计算,本文的改进就是所选的颜色中心是自定义的,即为上面数理统计的颜色值,然后进行迭代,由于在 HLS 空间,而 H,L 和 S 值相距很大,因此在选择聚类参数时,采用下列方法:

$$dcolor = \sqrt{[\alpha(h_i - hcolor)]^2 + [\beta(l_i - lcolor)]^2 + [\gamma(s_i - scol)]^2}$$

$i=1, 2, 3, 4.$

dcolor:颜色距离差

α, β, γ :HLS 各自在颜色空间距离中的权值

h_i, l_i, s_i :四种颜色的 HLS 空间值 hcolor, lcolor, scol-

or:图片颜色的 HLS 空间值

在此处 α, β, γ 的值依经验分别取为 0.8、260 和 300.图 6 是车牌颜色 HLS 聚类中心统计和分析的界面.

其具体计算步骤为:

1) 取出图片上的像素点并将其从 RGB 空间之值转换成 HLS 空间之值.

2) 将得到的值与数理统计得到的四类数值进行比较计算(利用上面公式)并得到相应的 dcolor.

3) 将四个 dcolor 进行比较,取出其最小值.

4) 通过最小值所属类得到所求像素点所属类(蓝、黄、白、黑).

5) 将所求像素点的颜色值更改为所属类的颜色值,即:黑(0,0,0),蓝(0,0,255),黄(255,255,0),白(255,255,255).

6) 统计各个类别的中心,即对属于该类别所有点去均值.

7) 将新得到的均值与数理统计得到的均值进行比较,其比较公式是:

$$\det a[i] = abs(olddcolor[i] - newcolor[i])$$
$$if(\det a[i] < \det) then \det a[i] = 1$$
$$else \det a[i] = 0;$$
$$flag = \det a[2] * \det a[2] * \dots * \det a[12];$$

($i=1, 2, \dots, 12$)

其中:

olddcolor[i]:聚类前的 HLS 权值.

newcolor[i]:聚类计算后得到新的 HLS 权值

如

det=1:聚类中心判定系数

新

如果 flag=0,则将得到的聚类中心替换前一迭代中心,并转步骤 1).

如果 flag=1,聚类分割完成,结束聚类计算.

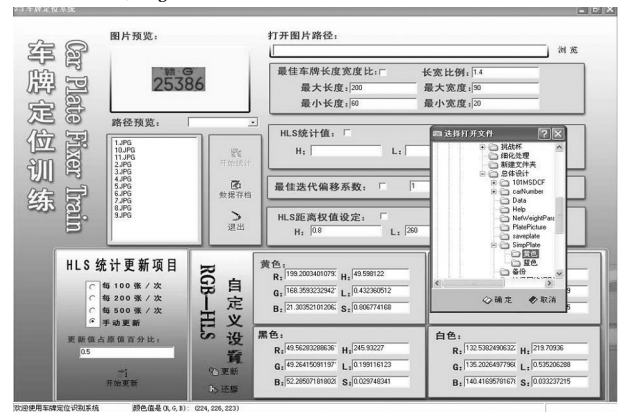


图 6 车牌颜色分析、统计和数据更新界面

从上面步骤可以看出聚类所得到四类的值不仅 RGB 颜色值不同,而且 R+G+B 的值也不同,而且由于目标值是数理统计得到的值,这点不仅可以使得聚类中心的判断可以不需要那么严格,而且其迭代次数也可大大地降低,基本上经过 1~2 次迭代就可以得到很好的聚类效果.下面就是聚类处理后

的效果比较图.



图 7 蓝色车牌处理前的图片



图 8 蓝色车牌处理后的图片



图 9 蓝色车牌处理前的图片



图 10 蓝色车牌处理后的图片

如果车身和车牌是同样的颜色的话,或者其颜色分群不是很明显的话(如图 12、15)那么在定位处理时候就很困难了,但是如果进行第二轮迭代,效果具有非常显著的差别(如图 13 和 16),原因是,因为 H 代表的是色度值当车的颜色和车牌的色度相同,或者即使颜色不同而由于是距离的计算缺陷例如颜色取值于(0, 0, 255)和(255, 0, 0)时,如用简单的距离计算时,结果是相同的,对于 RGB 颜色是蓝色和

红色的情况,在聚类计算时,通过减少一个变量会得到更好的效果.

上面是经第一轮迭代的效果图,如果第一轮迭代效果不是很好的话,那么还可以进行第二轮迭代,而在迭代次数上增加可达到更好的效果.

由于 HLS 一共三个变量,其排列组合共有 7 种迭代选择余地,这里列出作者经过多次实验得到的 HLS 迭代轮数较佳的顺序排列表(见表 2).



图 11 黄色车牌处理前图片



图 12 黄色车牌 HLS 处理后图片



图 13 黄色车牌 LS 处理后图片



图 14 蓝色车牌处理前图片



图 15 蓝色车牌 HLS 处理后图片



图 16 蓝色车牌 LS 处理后图片

表 2

迭代组合	Hue	Lmens	Saturation
1	✓	✓	✓
2	×	✓	✓
3	×	✓	×
4	×	×	✓

考虑到最复杂的情况, 经过上面四轮迭代以后, 一般的车牌的位置可以确定了.

5 结论

通过强制综合聚类中心的彩色均值聚类, 使聚类的初始中心和聚类结果向四种特定的车颜色聚

合, 这种改进使得算法的迭代次数降低, 算法的针对性更强. 避免了环境的复杂而产生的车牌图片定位的不确定性, 使车牌定位的精确性大大提高, 该方法也可应用于车流量的分析中, 通过完善数据的更新系统, 可以避免因为地域和季节的的不同而带来的车牌定位识别率的下降.

参考文献:

[1] 秧淑莹. 图像模式识别 VC++ 技术实现[M]. 北京: 清华大学出版社, 2005.
[2] 蒋先刚, 吴小林, 等. 基于灰色形态学的车牌定位技术研究[J]. 华东交通大学学报, 2006, 4.

The Technology Research of Orientation of Vehicle Plate Number by Forced Integrated K—means Clustering

JIANG Xian-gang, HAO Li, LIU Xun-di, HUANG Shu-ming, WANG Ya-wei

(School of Basic Science, East China Jiaotong University, Nanchang 330013, China)

Abstract: The paper researches a new way to recognize the car plate number by forced integrated K—means clustering center. We can get K—means clustering center by statistics and analysis in HLS color space, then recognize the car plate number by morphology, BP neural net technologies. This methodology has been verified and it has turned out to be better than the way based on the image of two—value and gray processing. It can recognize the car number effectively under the complex circumstances

Key words: K—means clustering, HLS color space, mathematic morphology.

简讯三

华东交通大学入选教育部大学英语教学改革示范点项目学校

教育部高等教育司近日正式批准我校为全国大学英语教学改革第二批示范点项目学校. 本批示范点是根据专家组深入各备选学校实地考察评估, 经过综合评估和研究批准的, 共有包括北京科技大学、首都师范大学、天津大学等在内的 34 所高校获得批准成为第二批示范点项目学校. 江西省入围本批示范点项目的仅有我校和江西财经大学两所高校.

外国语学院 唐 涛
2007 年 7 月 4 日