

一种高精度的体育商标分割算法

岳英杰 刘盼 徐荣青*

(南京邮电大学电子与光学工程学院 江苏 南京 210046)

摘要 在体育比赛中,俱乐部和赞助商希望观众实时获取宣传信息,可通过对运动员衣服上的商标识别来实现。图像分割是商标识别的一个关键部分,目前的分割算法对于背景复杂的图片分割精度不高,针对这一问题,提出一种高精度的商标分割算法。该算法融合了 Graph Cut 和 Grab Cut 两种算法的思想,利用 Graph Cut 将图片分为不同区域,通过设定双阈值判定不同区域属于背景或前景,在此基础上初始化图片的掩膜,再进行 Grab Cut 迭代,完成图像分割。实验结果表明,该算法与 Grab Cut 算法相比,有效解决了欠分割以及过分割的问题,并且保留了图像细节,克服了背景噪声,算法分割精度高达 94%。

关键词 商标 Grab Cut 图像分割 掩膜 双阈值

中图分类号 TP3 文献标志码 A DOI:10.3969/j.issn.1000-386x.2022.12.037

A HIGH PRECISION SEGMENTATION ALGORITHM OF SPORTS TRADEMARK

Yue Yingjie Liu Pan Xu Rongqing*

(College of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210046, Jiangsu, China)

Abstract In sports competition, the spectators must rely on the trademark recognition technology to obtain the club and sponsor information in real time through the trademark on the athletes' clothes. Image segmentation is a key part of trademark recognition, but the accuracy of traditional image segmentation algorithm is not high. To solve this problem, an efficient trademark segmentation algorithm is proposed. This algorithm combined the ideas of graph cut and grab cut. The image was divided into different regions by graph cut, and the probability that the region belongs to the background and foreground was calculated by using double thresholds, so as to advance ahead. The pedestrian intervened to initialize the mask of grab cut, and performed the iteration of grab cut to complete the image segmentation. The experimental results show that compared with the grab cut algorithm, the algorithm effectively solves the problem of under segmentation and over segmentation, and effectively retains the details of the image, overcomes the background noise, and the segmentation accuracy of the algorithm is up to 97%.

Keywords Trademark Grab cut Image segmentation Mask Double threshold

0 引言

随着信息服务的蓬勃发展,商标识别的应用越来越广泛,例如在电子商务平台监测用户参与度^[1-2],在日常生活中特定领域的标志物识别(例如车辆标志^[3]和交通标志识别^[4])。在体育领域,赞助商每年在体育营销上花费数百万欧元,俱乐部赞助球队,每年要在球队上投入一大笔金额。因此赞助商和俱乐部希望以更有效的途径让更多的观众直接、方便地获取赞助商和

俱乐部的宣传信息,通过运动员商标识别来连接到主页 URL 是一种新的推介方式,所以从运动员穿着的衣物上进行商标识别是非常有必要的。早些年间,Chat-topadhyay 等^[5]在体育比赛中,实时将场地旁边的广告牌识别,以完成宣传赞助商的目标。Kim 等^[6]提出了一种基于文字的实时监测与识别运动场地广告牌的算法。

进行商标识别的前提是对商标进行分割。将商标分割的目的是为了分割出无关背景^[7],只将商标作为分类器的训练内容,有效地减少了背景等无关项的干扰,有利于训练准确性的提升。但是由于运动员在球

场上剧烈运动,这就导致衣服上的商标具有旋转、平移、缩放的特性,并且受环境的影响,例如光照不均匀、图像背景复杂、图片模糊情况的影响,所以传统图像分割算法易造成欠分割或者过分割的情况。本文针对上述复杂情况下的图片,提出一种 Graph Cut 与 Grab Cut 相结合的改进算法,为后续商标识别提供了技术支持。

1 相关算法

1.1 Graph Cut

Graph Cut 现在已经成为解决一类能量最小化问题(例如计算机视觉中的图像分割)的主要选择方法之一。Graph Cut 最初是由 Boykov 等^[8]在解决马尔可夫随机场的能量最小化问题时提出来的,并且他们还得出了如何使用 Graph Cut 得到近似最优解的方法。

Graph Cut 是基于图论的分割,如图 1 所示,黑色的点代表像素点,整幅图像被重构为图 1 所示的结构,第一类边为相邻像素之间的边,称为“n-links”,第二类边为每个像素与 s 和 t 连成的边,称为“t-links”。图像分割的结果就是将 s 与 t 分开,其中与 s 相连的为前景像素,与 t 相连的为背景像素。Graph Cut 使能量函数 $E(A)$ 最小完成分割。

$$E(A) = \lambda R(A) + R(B) \quad (1)$$

式中: $A = \{A_1, A_2, \dots, A_p\}$ 为每个像素的标签 label, A_i 为 0(背景)或 1(前景); λ 是区域项和边界项之间的重要因子,决定它们对能量的影响大小; $R(A)$ 为区域项; $R(B)$ 为边界项。

$$R(A) = \sum_{p \in P} R_p(A_p) \quad (2)$$

$$B(A) = \sum_{p \in P} \sum_{|p,q| \in N} B_{|p,q|} \delta(A_p, A_q) \quad (3)$$

式中: $\delta(A_p, A_q) = \begin{cases} 1 & A_p \neq A_q \\ 0 & A_p = A_q \end{cases}$, p 和 q 为相邻的像素。

衡量两个像素的相似性有很多方法,此时我们利用两个像素 RGB 三通道的欧氏距离来作为计算方法。

$$Ed_{RGB} = \sqrt{(r_1 - r_2)^2 + (g_1 - g_2)^2 + (b_1 - b_2)^2} \quad (4)$$

$r, g, b \in [0, 255]$

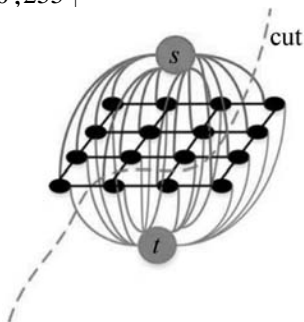


图 1 Graph Cut 连接图

图片重构的结果如图 2 所示,可以看出,图片依照轮廓被分为了很多不同的区域,其中分割区域的多少受设为相似区域的像素距离阈值影响。



图 2 Graph Cut 图片区域重构

1.2 Grab Cut

Rother 等^[9]在 SIGGRAP 2004 中通过将 Graph Cut 算法迭代,提出了一个基于 Graph Cut 迭代的 Grab Cut 算法。通过构造前景中和背景中的两个标记的能量函数,将分割问题转化为两个标记问题。该算法最终可以确定每个像素是属于背景还是前景,以便它可以使能量最小化。

Grab Cut 算法的步骤如图 3 所示。首先,用户选择一个感兴趣矩形(ROI)。然后与原始图像具有相同像素的掩膜图像被初始化,用来记录图像分割的结果。掩膜图像的每个像素点只有四种取值,分别为 0(表示明确属于背景的像素)、1(表示明确属于前景的像素)、2(表示可能属于背景的像素)、3(表示可能属于前景的像素)。然后,利用式(5) - 式(7)计算被称为边界能量项像素的 n-links,并根据用户选择的感兴趣区域矩形,建立前景和背景的 GMM 模型。

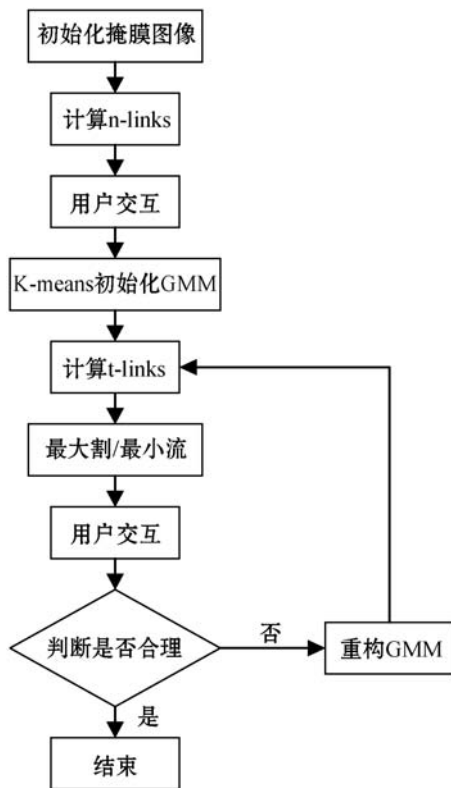


图 3 Grab Cut 算法流程

$$D(x) = \sum_{i=1}^K \pi_i g_i(x; \mu_i, \Sigma_i) \quad (5)$$

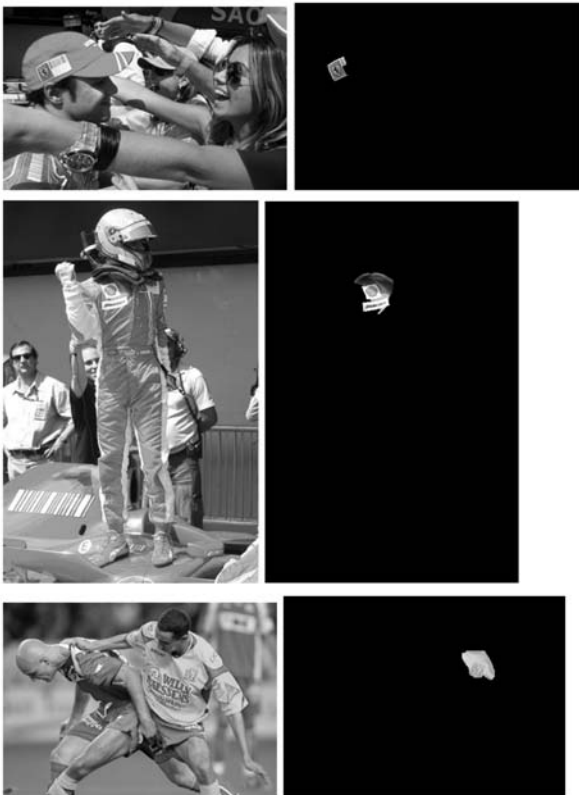
$$\sum_{i=1}^K \pi_i = 1 \text{ 且 } 0 \leq \pi_i \leq 1 \quad (6)$$

$$g(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp \left[-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right] \quad (7)$$

式中: μ 为三通道的高斯分量的均值向量; Σ 为 3×3 的协方差矩阵。

在算法继续往下进行之前,紧接着利用 K-means 聚类算法^[10]将组成 GMM 模型的所有像素点分为 k 类。基于此聚类的结果,GMM 模型的参数 θ 和 t-links 也可以被计算出来。值得注意的是,在使区域能量也就是 t-links 最小化过程中,应该多次迭代 k 和 θ 以满足最小化。因此,s-t 网络被建立,并且通过最小割/最大流算法将图像分割。

在图 4 中,(a)是原始图像,(b)是 Grab Cut 分割的结果(GMM 模型中的参数 $K = 500$)。可以看出,Grab Cut 的一个缺点是当商标与背景颜色相似时,分割后会出现一些图像碎片噪声,也就是我们所说的欠分割情况。这种情况会导致前景提取的不精确,为接下来的商标识别增加很多干扰和影响。因此,改进此分割算法显得尤为重要。



(a) 原始图像 (b) Grab Cut 分割结果

图 4 Grab Cut 分割结果

2 改进的商标分割算法

根据前文的分析,Graph Cut 算法在分割图像时,算法利用了物体位置和边缘的信息,当一幅图像中存在很多噪声时这种方法是十分有效的。从图 4(b)中我们可以看出,Grab Cut 算法存在很多不足,当图片的背景与前景颜色相近时,这幅图片很可能被欠分割。并且受到环境因素、光照等情况的影响,尽管某些部分与目标前景明显不同,但是 Grab Cut 算法不能很好地分开与前景无关的背景干扰像素。

所以在进行图像分割之前,有效地进行预处理是接下来工作的根本。因为 Graph Cut 利用了图片的边缘信息,所以本文考虑使用锐化处理。锐化处理的目的是为了增强图片的边缘,使得目标物体的边缘鲜明,以便于提取目标的边缘,对图像进行分割。本文采用拉普拉斯算子进行图像锐化,它是根据相邻像素之间的变化程度来计算的。一阶函数微分描述的是函数变化的方向,二阶函数微分描述的是函数变化的速度,用图像来解释就是一阶函数微分能确定图片的边缘是否存在,而二阶函数微分能确定图片边缘的位置。

考虑到锐化是在 Graph Cut 算法之前进行的,而前文描述的 Graph Cut 算法在计算相邻像素的连通值时采用的是 8 连通法。如图 5 所示,所谓的八连通法就是在计算周围像素到中心像素的影响时,考虑上、下、左、右、左上、右上、左下、右下—共 8 个像素。

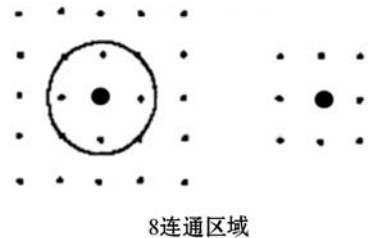


图 5 图片中像素的 8 连通示意图

故拉普拉斯算子我们采用预定的 $\nabla = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$

来计算,得到的锐化图像 $g(x)$ 为:

$$g(x) = \begin{cases} f(x,y) - \nabla^2 f(x,y) & \nabla^2 f(x,y) < 0 \\ f(x,y) + \nabla^2 f(x,y) & \nabla^2 f(x,y) \geq 0 \end{cases} \quad (8)$$

本文考虑了锐化步骤在增强图像边缘的同时,也会将噪声增强,所以在进行锐化处理之前将图片先进行高斯滤波。

另外针对分割不精确,Grab Cut 还提供人工修改掩膜的权利,又多了一次交互过程,本文将修改掩膜的

步骤提前进行,因为图像分割的实质是迭代使得图像每个部分的能量函数最小化,此时输出的最优结果为最佳分割。而在 Graph Cut 算法中,仅涉及两类能量,第一类是像素点与图像区域不同的能量,称为点与类的能量;第二类是像素点与像素点之间不同的能量,称为点与点的能量。两个能量之间所占的比例关系使用 Potts 模型参数 K 来控制比例,当点与点的能量所占比例高时,像素局部平滑约束就越高,图像就越平滑,图像分割保留的细节就越少。所以利用双阈值首先将图片进行初步过滤,直接将图片的某些部分视为前景或背景的部分,再送入 Grab Cut 进行迭代。

根据上述情况,本文提出的改进算法步骤如下:

(1) 对图像先进行高斯滤波处理,去除噪声。

(2) 对高斯滤波之后的图像进行拉普拉斯算子锐化处理,增强图像边缘。

(3) 对整幅图像使用 Graph Cut 算法,统计出每个像素的标签值。

(4) 用户进行交互,将目标前景用矩形框圈出,并统计矩形框中有多少种标签和属于该标签有多少数量的像素。

(5) 计算矩形框中每种标签的像素占整幅图像属于该种标签的像素比例。

(6) 设置阈值 H_1 和 H_2 ,两个阈值的设定根据特定的图片应用场景,根据图片分割的平滑约束度,利用 K 值来设定阈值 H_1 和 H_2 :

$$H_1 = \frac{K}{K + \frac{1}{P} \sum_{m=0}^P \text{Connect}_{\text{size}}} \quad (9)$$

$$H_2 = H_1 + \frac{\frac{1}{P} \sum_{m=0}^P \text{Connect}_{\text{size}}}{\max(\text{Connect}_{\text{size}})} \quad (10)$$

式中: K 代表 Graph Cut 算法设定的 Potts 模型参数; P 代表用户交互后得出的矩形框中像素点的总数; $\text{Connect}_{\text{size}}$ 代表像素点与其相邻 8 邻域像素点的连通平均值。

即 H_1 为低阈值, H_2 为高阈值,比例小于 H_1 ,则将矩形框中的属于该标签的像素视为背景的一部分,若比例大于 H_1 小于 H_2 视为可能的背景,若比例大于 H_2 视为可能的前景。

(7) 根据判决比例和矩形框中的标签值,生成掩膜图像。

(8) 利用生成的掩膜图像去初始化 Grab Cut 算法,完成图像分割。

假设矩形图像中 N 个标签,每个标签 $k=1,2,\dots,$

N ,每个标签中像素的数量是 S_k 。用户交互完成后,在矩形框中有 i 个标签,属于这 i 个标签的像素为 $R_i(i=1,2,\dots,N,i \in K)$ 。通常,用户交互后的矩形中像素所占所有像素标签的比例定义为:

$$L_k = \frac{R_k}{S_k} \quad k=1,2,\dots,N,k \in i \quad (11)$$

L_k 的含义是矩形框中的一个像素属于背景的可能性。若 $L_k=1$ 的含义就是属于该标签的所有像素都处于矩形框中。 L_k 的值越接近于 0,越能说明该像素属于背景。所以,当计算出的 L_k 低于设定的阈值 H_1 和 H_2 ,在掩膜中该点的取值由可能的前景变为可能的背景或背景,从而减少背景像素的干扰。

3 实验结果分析

通过实验我们发现,阈值的大小会严重影响分割的结果,不同应用场景下阈值不同。针对此应用场景,通过式(9)和式(10)计算得出 $H_1=0.3$ 、 $H_2=0.8$ 。图 6 是使用改进后的算法分割的结果。可以看出,改进之后的算法相比较于传统 Grab Cut 算法在精度上具有明显的提升,有效地满足了欠分割的问题。另外我们从实验得出,当阈值设置过低时,改进的算法不是很有效。若将阈值 H_1 和 H_2 设为 0,那么改进后算法的结果将与传统 Grab Cut 一致,算法改进将变得没有意义。

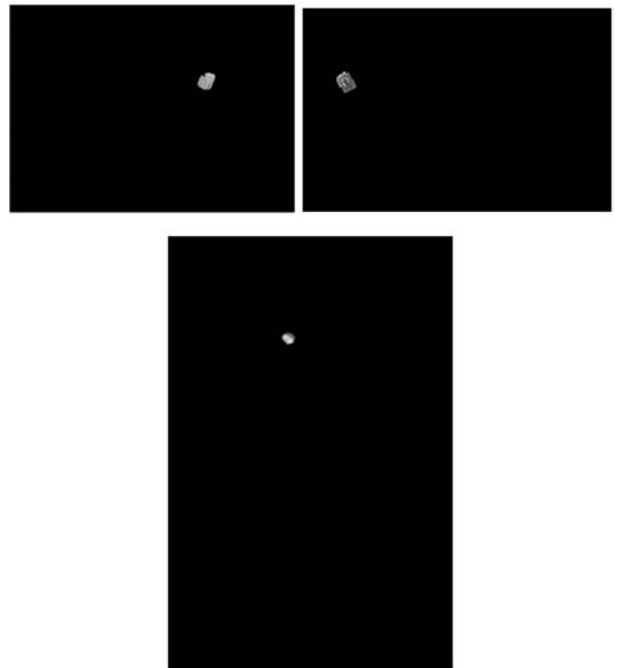


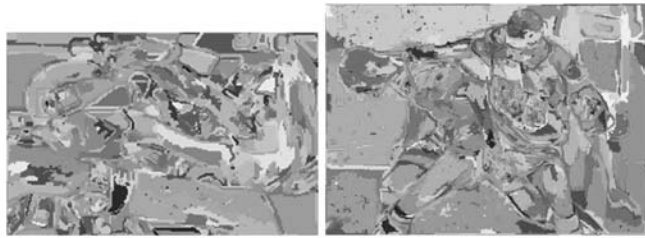
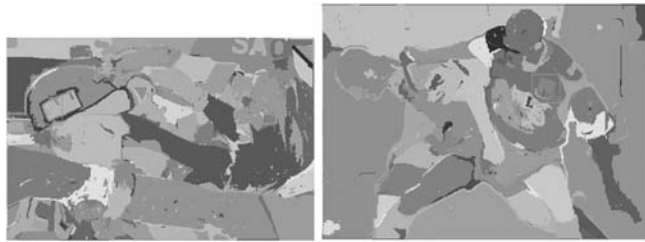
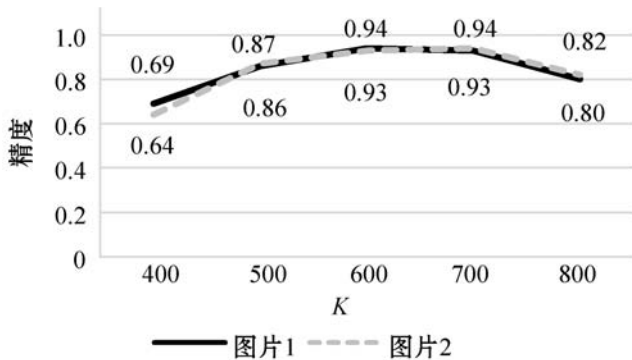
图 6 改进算法分割结果

另外,Graph Cut 初始化设置分割区域数量的多少也会对分割结果产生重要的影响。而分割区域的数量与 K 值的大小有关。 K 值越大,则两个像素相隔的距离越近,一幅图像被分成的区域越多;反之 K 值越小,

两个像素相隔的距离就远,一幅图像被分成的区域越少。图 7(b)展示了设置过多的分割区域数量,算法的精度将会下降。图 8 展示了不同的 K 值对分割精度的影响,可以看出, K 值在一定范围内不会对分割精度产生很大的影响,但是将图片分割成更细小的区域意味着需要更多的时间,所以在保证大概分割区域数量对精度影响很小的情况下,尽量减小 K 值,使分割区域尽可能少,这样分割的速度会有一定程度上的提升。



(a) 原始图像

(b) $K = 800$ (c) $K = 600$ 图 7 不同 K 值对图像的影响图 8 K 值对算法精度的影响

4 算法评价

本文从 BelgaLogos 公开数据集中选择了 100 幅图片,并用 Photoshop 图像抓取工具分割它们的背景作为完美分割的参考图像。评价分割算法最重要的指标是

精确度,本文使用标准的 Jaccard 相似度指数对改进算法的精确度进行验证。Jaccard 相似性指数是衡量两个集合的相似性的指标,利用式(12)计算出每幅图片的 Jaccard 相似指数,再取平均,得到算法的精确度结果。

$$J(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \quad (12)$$

式中: S_1 和 S_2 指的是两个集合,在本文中, S_1 为改进算法的分割结果图像, S_2 为参考图像。

$$Accuracy = \frac{\sum_{i=0}^{99} J_i(S_1, S_2)}{100} \quad (13)$$

式中: $i=0, 1, \dots, 99$ 代表每幅图片; $J_i(S_1, S_2)$ 代表每幅图片的 Jaccard 指数。

改进算法的分割结果展示在表 1 中。可以看出改进后的算法对分割精度有了很大的提升,但是精度的提升受到了特定图片的限制。另外,本文将算法改进结果与 Wicaksono 等^[11]提出的 K-means 改进算法相比较,在精度方面有相对程度上的提高。

表 1 算法结果

算法	Accuracy
Graph Cut	0.78
Grab Cut	0.86
K-means 分割算法	0.89
改进算法	0.94

同时,从不同的性能指标验证分割方法是非常必要的,本文使用峰值信噪比(PSNR)指标来评判分割的性能,一定程度上验证了改进算法的说服力。利用式(14)计算得出的均方误差 MSE,是反映估计量与被估计量之间差异程度的一种度量。

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (14)$$

式中: m 代表图像的宽度像素数量; n 代表图像的长度像素数量; I 代表算法分割之后的图像; K 代表完美分割的参考图像。PSNR 用来检测图像的质量,利用式(15)计算得出 PSNR 值。

$$PSNR = 10 \times \log_{10} \left(\frac{MAXI^2}{MSE} \right) \quad (15)$$

式中: $MAXI$ 是表示图像点颜色的最大数值。PSNR 数值越大,代表图像的质量越好。如图 9 所示,本文分别求得 Graph Cut、Grab Cut、k-means 分割算法和改进的 Grab Cut 算法的 PSNR 值,并以柱状图的形式展示。

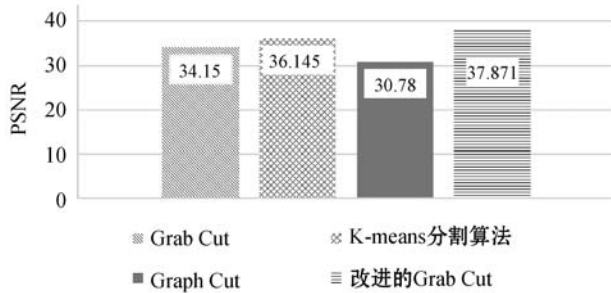


图9 不同算法的 PSNR 比较

由图9可得,本文的改进 Grab Cut 算法具有最大的 PSNR 值,一定程度上验证了本文算法有一定程度上的性能提升,但是当目标物体比背景更大的时候,分割精度更高,因此,不是所有的图片都能体现出改进算法的优点。

5 结 语

本文提出一种基于体育商标的改进图像分割算法。实验结果展示了该算法的精确度,使用改进后的算法能使得商标分割准确率为 94%,基本上满足准确分割体育商标的需要。但是,算法的缺点是增加了运行时间。本文的研究意义在于,发现了可以利用像素与像素之间的关系来改进 Grab Cut 的不足之处。

毫无疑问,当商标图片背景特别复杂时,对分割的精度会造成重大的影响,但是本文通过改进 Grab Cut 算法,使得在复杂背景下使 Grab Cut 分割精度有了一定的提升,有效解决欠分割和过分割的问题。

参 考 文 献

- [1] Kaur H, Paruthi M, Islam J, et al. The role of brand community identification and reward on consumer brand engagement and brand loyalty in virtual brand communities [J]. *Telematics and Informatics*, 2020, 46: 101321.
- [2] Critchlow N, MacKintosh A M, Hooper L, et al. Participation with alcohol marketing and user-created promotion on social media, and the association with higher-risk alcohol consumption and brand identification among adolescents in the UK [J]. *Addiction Research Theory*, 2019, 27(6): 515–526.
- [3] 沃焱,郭正凯,江丽萍,等.一种车牌识别方法:CN102880859A [P]. 2013-01-06.
- [4] 邓向林.车载通讯的交通信号自动识别研究[J]. *企业技术开发*, 2016, 35(19): 28–30.
- [5] Chattopadhyay T, Sinha A. Recognition of trademarks from sports videos for channel hyperlinking in consumer end [C]// 2009 IEEE 13th International Symposium on Consumer Electronics, 2009: 943–947.
- [6] Kim Y S, Kim W Y. Content-based trademark retrieval system using a visually salient feature [J]. *Image and Vision*

Computing, 1998, 16(12–13): 931–939.

- [7] 孙浩荣.基于 SVM 的食物图像分类算法的研究[D].西安:西安电子科技大学,2016.
- [8] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(11): 1222–1239.
- [9] Rother C, Kolmogorov V, Blake A. Interactive foreground extraction using iterated graph cuts [J]. *ACM Transactions on Graphics*, 2004, 23(3): 309–314.
- [10] Vij S, Sharma S, Marwaha C. Performance evaluation of color image segmentation using K means clustering and watershed technique [C]// 2013 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013: 1–4.
- [11] Wicaksono Y A, Rizaldy A, Fahriah S, et al. Improve image segmentation based on closed form matting using K-means clustering [C]// 2017 International Seminar on Application for Technology of Information and Communication (iSemantic), 2017: 26–30.

(上接第 200 页)

- [4] Huang R, Padoem J, Chen C. YOLO-LITE: A real-time object detection algorithm optimized for Non-GPU computers [C]// 2018 IEEE International Conference on Big Data (Big Data), 2018.
- [5] Chollet F. Xception: Deep learning with depthwise separable convolutions [EB]. arXiv:1610.02357, 2016.
- [6] 王晓东.基于注意力机制的三维超声影像的多尺度目标识别的研究[D].西安:西安电子科技大学,2018.
- [7] 王海龙.基于深度学习的快速目标检测算法研究[D].哈尔滨:哈尔滨工业大学,2018.
- [8] 崔家华,张云洲,王争,等.面向嵌入式平台的轻量级目标检测网络[J]. *光学学报*, 2019, 39(4): 307–313.
- [9] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517–6525.
- [10] Huang G, Liu Z, Maaten L, et al. Densely connected convolutional networks [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2261–2269.
- [11] 余良凯.基于深度学习的机场场面目标检测[D].成都:电子科技大学,2018.
- [12] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [EB]. arXiv:1704.04861, 2017.
- [13] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [EB]. arXiv:1612.03144, 2016.
- [14] 祝思君.基于深度学习的无人机遥感图像目标识别方法研究[D].北京:北京建筑大学,2018.
- [15] Redmon J, Farhadi A. YOLOv3: An incremental improvement [EB]. arXiv:1804.02767, 2018.