

基于 HOG - LBP 特征和 SVM 分类器的视频摘要方法

程海鹰, 王凤随, 朱树明

(安徽工程大学电气工程学院, 安徽 芜湖 241000)

摘要:为了提高视频摘要质量、高效地获取视频信息,提出了一种基于 HOG - LBP 特征和 SVM 分类器的视频摘要方法。首先,从分解到的视频序列中提取形状和纹理特征。其次,利用训练好的支持向量机对待检测图像进行分类,并将分类正确的帧暂定为关键帧。最后,运用图像余弦相似度方法剔除其中的冗余帧,并将最终保留下的关键帧按照原始视频中出现的时间顺序依次组合起来,形成视频摘要。实验结果表明,提出的算法在三个不同场景视频中提取关键帧的平均正确率达到 94.08%,平均错误率仅为 23.18%。算法提取到的关键帧准确率较高且内容具有代表性,视频摘要质量得到有效改善。

关键词:视频摘要;HOG - LBP 特征提取;SVM 分类器;余弦相似度

中图分类号:TP391

文献标志码:A

引言

随着互联网和多媒体技术日新月异的进步,数字视频广泛应用于生活的各个方面,人们很难从海量视频中获取到有用信息,而视频摘要技术就可以帮助用户解决这个问题。视频摘要是一项缓解视频存储浏览等方面的关键技术,它是对完整视频的提炼和概括总结,将原始视频的主要信息简明扼要地呈现给目标用户。视频摘要一般包括静态视频摘要和动态视频摘要,静态视频摘要主要通过提取具有代表性的关键帧组合形成视频摘要,以标题、海报和故事板等形式展示;动态视频主要通过视频段的分割,选取和整合形成视频摘要,多数以精彩集锦和全局缩略的形式展现^[1-3]。视频从本质上来

说是由一帧帧连续的图像序列按照时间发生的先后顺序拼接而成,而提取关键帧就是在连续时间点上的采样且去除冗余信息的过程。当前在安防监控视频中,受用户关注度较高的是人和车辆,而在制作视频摘要的过程中,重点都是在于关键帧的提取。Zhao 等^[4]利用平均灰度累积直方图和边缘直方图的方法来提取关键帧,该方法虽然能较准确地提取关键帧,但内容覆盖率不高。Guan 等^[5]提出通过计算 SIFT 特征点来建立表示全局信息的特征点池,从而选取可以覆盖特征点池的视频帧作为关键帧,该方法虽然能真实全面地反映原始视频信息,但所提取关键帧的冗余度高。司若妍等^[6]提出基于 HSV - LBP 和 K - means 聚类的方法提取关键帧,该方法能自适应得到视频的阈值,但准确率不高且实时性不

收稿日期:2018-05-07

基金项目:安徽省自然科学基金项目(1708085MF154);安徽高校省级自然科学基金资助重点项目(KJ2015A071)

作者简介:程海鹰(1993-),女,安徽桐城人,硕士生,主要从事视频摘要方法方面的研究,(E-mail)2927015227@qq.com;

王凤随(1981-),男,副教授,博士,硕士生导师,主要从事视频通信、计算机视觉方面的研究,(E-mail)fswang@ahpu.edu.cn

强。刘长征等^[7]提出利用改进向量机和滑动窗口的方法提取关键帧,该方法提高了关键帧提取速度,但所提取的关键帧冗余度高,影响视频摘要的真实性,效果不佳。

本文提出的基于 HOG-LBP 特征和 SVM 分类器的视频摘要方法。首先,从分解到的视频序列中提取形状特征和纹理特征。其次,利用训练好的支持向量机(Support Vector Machines, SVM)对用户感兴趣的图像进行分类,将分类正确且不冗余的序列定位为关键帧。最后将保留下的关键帧按照原始视频中出现的时间顺序依次组合起来,形成视频摘要。

1 理论基础

1.1 HOG-LBP

方向梯度直方图(Histograms of Oriented Gradients, HOG)可以表示图像的形状特征,不同的物体形状不一样,这样可以作为区分目标物体的重要参考。HOG 特征是模式识别和计算机视觉领域很常用的一种特征描述子^[8],能将局部区域内目标的梯度结构和边缘信息很好地表示出来。HOG 特征最开始是由 Navneet Dalal 和 Bill Triggs 等人于 2005 年提出来的^[9-11],他们的方法是将一幅图像分割成很多个细胞单元(cell),再从 cell 中提取特征,而不是直接从图像整体中提取,以此反映图像中不同像素之间的梯度。

局部二值模式(Local Binary Pattern, LBP)可以用来表征局部纹理特征。不同的物体具有不同的物理属性,这种属性差异会导致物体的灰度信息不同,而 LBP 特征能够很好地表示这种灰度信息。LBP 算子最开始是由 Ojala T 等人于 1994 年提出在 3×3 的邻域内定义的,固定的半径范围无法满足图像尺寸和纹理的需要,为了改善其不足,Ojala T 等人用任意的圆形邻域代替 3×3 的正方形邻域,改进后的圆形邻域内像素点的个数 P 可以是任意的,半径 R 也可以是任意的,记为 LBP_P^R ^[12]。

1.2 SVM 分类器

SVM 是一个有监督的学习模型,它的特点是在分类模型和模型参数的选择上始终优先结构最小化,一般用

于二类别分类问题。在视频检测特征分类中,针对低维空间的线性不可分问题,通过核函数映射到高维空间达到线性可分,再进行线性分割实现特征分类^[13-14],而核函数是一个对称函数 $K: R_n \times R_n \rightarrow R$,它是将两个 R_n 空间中的 n 维向量映射成一个实数。本文选取径向基核函数(RBF)作为核函数,通过 SVM 分类器对待提取的图像特征进行分类。核参数的选择受训练数据大小的影响,需要进行相应的优化。往往一个理想的 SVM 分类器需要大量的具有代表性的训练样本,比如在视频监控当中就要选取大量的车辆和行人样本,而选取的样本要确保有大小、光照、遮挡等情况不一的图片若干,以此保证样本的多样性,使得算法更具有鲁棒性。

2 本文算法

为了能有效地进行分类,选取合适的分类特征参数也是极其重要的。由于 HOG 特征是在图像的局部单元上进行提取,它对图像几何和光学的形变都能保持很好的不变性,LBP 算子具有灰度不变性,光照对其基本没有影响,改进后的圆形 LBP 算子对图像旋转具有不变性。本文算法兼顾了两者在特征提取上的优势,将 HOG 形状特征向量和 LBP 纹理特征因子融合在一起,可有效地减小形变和光照对实验结果产生的误差。

2.1 特征提取

由于图像特征分明,比视频更容易操作,从图像中能捕获到更多的细节信息。故从分解到的视频序列中提取 HOG-LBP 特征。对于 HOG 特征提取,本文灰度化处理后得到归一化图像的大小为 128×128 ,细胞单元大小设置为 8×8 ,一个细胞单元的宽度为 8 个像素,块(block)大小设置为 16×16 ,参数设置完毕后进行 HOG 特征计算。

首先采用 Gamma 校正法对图像颜色空间进行归一化:

$$I(x, y) = I(x, y)^{\text{gamma}} \quad (1)$$

再计算每个像素的梯度:

$$M_x(x, y) = N(x + 1, y) - N(x - 1, y) \quad (2)$$

$$M_y(x, y) = N(x, y + 1) - N(x, y - 1) \quad (3)$$

其中: $M_x(x,y)$ 表示水平方向梯度, $M_y(x,y)$ 表示垂直方向梯度, $N(x,y)$ 表示像素值。

由式(2)和式(3)计算出像素点 (x,y) 处的幅值和方向值为:

$$M(x,y) = \sqrt{M_x(x,y)^2 + M_y(x,y)^2} \quad (4)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{M_y(x,y)}{M_x(x,y)}\right) \quad (5)$$

然后将图像划分为若干细胞单元,计算细胞单元的梯度直方图。将每 4 个细胞单元构成一个 block,将 4 个 9 维的特征向量组合成一块,那么每个 block 的特征向量为 36 维,最后把所有的 block 特征组合在一块形成整幅图像的 HOG 特征。HOG 特征提取的流程图如图 1 所示。综上所述,得到整幅图像的 HOG 特征向量的总维数大小为 $(16-1) \times (16-1) \times 9 \times 2 \times 2 = 8100$ 维。

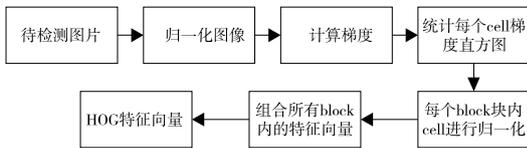


图 1 HOG 特征提取流程图

由于 HOG 特征只能表征形状信息,为了更好地表征图像信息,本文在原本的 HOG 特征基础上,增加了 LBP 纹理算子。

首先选取所要计算区域的中心像素,如图 2 所示,设定这个中心像素的灰度值为阈值,然后将周围圆形邻域的像素灰度值与该阈值进行一一对比,如果大于阈值,该像素点位置标记为 1,如果小于阈值则标记为 0。这样便可以得到一串二进制序列,再对不同位置的像素值进行加权求和,就可以得到该区域的 LBP 值。

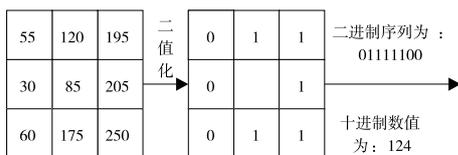


图 2 基本 LBP 算子计算示意图

一个 LBP 算子存在多个二进制模式,对于 LBP_p^r ,其二进制模式有 2^p 种,也就是说二进制模式的种类会随着圆形邻域内采样点数目的增加而呈现指数式的增长。故本文在基本 LBP 算子的基础上,运用等价模式

的 LBP 算子对算法进行降维,以减少计算量。LBP 算子的等价模式是:对于某个二进制模式,如果它的二进制数从 1 到 0 或从 0 到 1 的情况最多出现两次,那么这个二进制模式所对应的二进制就称为一个等价模式类。例如 00000000,11111111,00111111,10000111 分别有 0 次,0 次,1 次,2 次跳变,这些都归为等价模式类。对于像 10011011 有四次跳变的二进制数来说,被归为混合模式类。通过等价模式的改进,模式数量由原先的 2^p 降到了 $P(P-1)+2$,故原先 LBP 算子中的 256 种降到了 58 种,减少了 198 种,达到降维的目的,减小了计算量。

2.2 SVM 训练及分类

在监控视频当中,选择分类器的训练样本时要保证样本的多样性,例如样本可以大小不一、光照强度和遮挡面积不同等。SVM 分类器训练模型如图 3 所示。

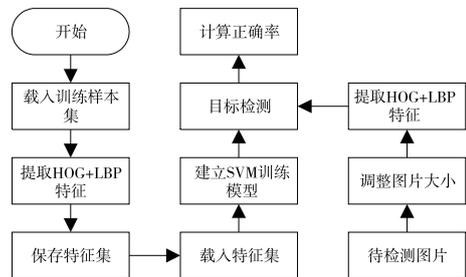


图 3 SVM 分类器训练模型

2.3 冗余帧的剔除

在通过 SVM 分类器正确分类后的图片中,发现有不少车辆和人物相似或重复,比如有的车辆一直处于静止状态中,这会导致不少冗余帧的存在。为了精确得到关键帧,将执行剔除冗余帧的步骤,利用余弦相似度方法将所有待定的关键帧之间进行对比,筛查过滤冗余帧。

在余弦相似度方法中,将图像的灰度直方图划分成 64 个区,连续 4 个灰度等级划分为一个区,对每个区的 4 个灰度等级进行求和运算,这样得到的 64 个数据表征一幅图像的一个向量。通过对比两幅图像向量的余弦值和余弦夹角的大小来判断两幅图像的相似度。在分解的视频帧中,余弦值越接近 1,表明夹角越小,相似度越高。在本文中,若余弦值接近 1,则这两帧会被看作冗

余,第二帧会被剔除。

综上所述,提出算法提取关键帧的步骤如下:

Step1:从分解到的视频序列中分别提取 HOG 和 LBP 特征,训练 SVM 分类器。

Step2:利用训练后的 SVM 分类器对待检测图片进行分类,将分类正确图片暂定为关键帧。

Step3:利用图像的余弦相似度方法对暂定的关键帧再进行过滤筛选步骤,剔除其中的冗余帧。

3 实验结果及分析

本文算法实验硬件环境是普通台式电脑,主要配置是 Inter Core i3-7100(3.90 GHz)的四核处理器和4 GB 内存,软件环境是 Windows10 64 位,仿真运行工具为 Matlab2017a。其中 SVM 训练库采用的是 VOC2007 图片数据集,从 4952 张图库中筛选出适合的训练样本集,在筛选的过程保证同一样本的多变性,从不同的方向视角状态出发,建立了一个简易的人物库和车辆库。如图 4 与图 5 所示。



图4 车辆库的部分内容



图5 人物库的部分内容

基于 SVM 分类器对视频帧分别进行单一特征提取和 HOG-LBP 特征提取,实验结果见表 1。从表 1 中可以对分类的正确率、错误率以及所耗费的时间。

表1 单一特征提取与 HOG-LBP 特征提取实验结果对比

方法	训练样本/帧	测试样本/帧	准确率/%	错误率/%	耗时/s
HOG + SVM	323	65	90.76	6.15	28.9
LBP + SVM	323	65	81.53	13.85	31.2
HOG + LBP + SVM	323	65	93.84	3.07	32.5

由表 1 数据可知,在 SVM 分类器、训练和测试样本完全相同的情况下,三种特征提取方式所耗费的时间相当,但 HOG-LBP 特征提取方式比 HOG、LBP 单一特征提取正确率分别提高了 3.08%、12.31%,而错误率分别降低了 3.08%、10.78%。故 HOG-LBP 特征提取效果更好,表明 HOG-LBP 结合的特征提取优于单一特征提取。

本文基于 HOG-LBP 特征,利用 SVM 分类器将分类正确且不冗余的帧作为视频的关键帧。采用准确率(Accuracy Rate, AR),错误率(Error Rate, ER)这两个指标作为算法生成的视频摘要的评价标准,计算公式为:

$$AR = \frac{L_r}{L_m} \quad (6)$$

$$ER = \frac{L_r}{L_m} \quad (7)$$

其中: L_m 为 10 个用户都选择的关键帧集合, L_r 为算法生成且与 L_m 相匹配的关键帧集合, L_r 为算法生成但不匹配的关键帧集合。

为了验证本文算法的有效性,另外选择了街道、路口、校园三个不同场景的视频作为测试数据。测试视频属性见表 2。同时为了使实验结果更加具有说服力,在同一场景下还将本文算法与文献[7]算法作了对比,并采取了人为主观的评价模式,实验结果见表 3。

表 2 测试视频属性

场景	视频时间/s	分辨率	总帧数/帧	帧率/fps
街道	113.1	1280 × 720	3314	30
路口	121.9	1280 × 720	3656	30
校园	64.3	1280 × 720	1928	30

表 3 两种算法在不同场景时对应的准确率和错误率

算法	场景	L_m /帧	L_r /帧	L_r^- /帧	准确率/%	错误率/%
本文算法	街道	597	563	144	94.30	21.42
	路口	630	590	150	93.65	23.80
	校园	333	314	72	94.29	21.62
文献[7]	街道	597	558	182	93.47	30.49
	路口	630	584	193	92.70	30.63
	校园	333	313	99	93.99	29.73

从表 3 可知,对于三个不同场景的视频,10 个用户选择的关键帧作为真实值,在真实值相同的情况下,本文所使用的算法提取的关键帧数目比文献[7]要少,但是与真实值重合的数目却要多,反映了本文算法更具有针对性,更贴近真实值。本文算法与文献[7]算法相比,三个场景关键帧提取的准确率平均提高了 2.08%,错误率下降了 21.31%,从而提高了关键帧的质量。由于监控视频的所有视频帧的背景都是静止不变的,本文算法比文献[7]算法增加了剔除冗余帧的步骤,减少了冗余帧的干扰,使提取到的关键帧更具有代表性,将关键帧按照原始视频出现的时间顺序依次组合起来,更能反映原始视频的真实内容,视频摘要效果更佳。

4 结束语

本文提出了一种基于 HOG-LBP 特征和 SVM 分类器的视频摘要方法,该算法结合 HOG 和 LBP 特征提取的优势,利用 SVM 分类器对监控视频图像进行分类,将分类准确且不冗余的视频帧作为关键帧,最后,将提取到的所有关键帧按照原始视频中出现的时间顺序依次组合起来,形成视频摘要。实验结果表明,本文算法与文献[7]算法相比,三个场景的准确率平均提高了 2.08%,错误率平均下降了 21.31%,提高了关键帧质量,为用户提取监控视频摘要提供了有效依据。

参考文献:

- [1] 王娟,蒋兴浩,孙铨锋.视频摘要技术综述[J].中国图象图形学报,2014,19(12):1685-1695.
- [2] 欧阳建权,李锦涛,张勇东.视频摘要技术综述[J].计算机工程,2005,31(10):7-9.
- [3] 贺宏遵.视频摘要技术的研究[D].长沙:中南大学,2009.
- [4] ZHAO X,LIU JW,HU G D,et al.Adaptive key-frame selection based on image features in Distributed Video Coding[C]//Proceeding of 2013 International Conference on Computational Problem-Solving(ICCP),Jiuzhai, China,October 26-28,2013:245-248.
- [5] GUAN G,WANG Z Y,LU S Y,et al.Keypoint-based keyframe selection[J].IEEE Transactions on Circuits and Systems for Video Technology,2013,23(4):729-734.
- [6] 司若妍,张明.基于 K-means 聚类算法的视频关键帧提取的研究[J].现代计算机,2016(20):59-63.
- [7] 刘长征,周荣华.基于改进支持向量机的视频摘要技术[J].实验室研究与探索,2016,35(1):85-88,101.
- [8] 陶建峰.基于多特征融合的行人检测方法研究[D].南京:南京理工大学,2013.

- [9] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C] // Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR '05), San Diego, CA, USA, June 20-25, 2005: 886-893.
- [10] DALAL N. Finding people in image and videos [D]. Grenoble, FRA: Institution National Polytechnique de Grenoble, 2006.
- [11] DALAL N, TRIGGS B, SCHMID C. Human detection using oriented histograms of flow and appearance [C] // Proceedings of the 9th European Conference on Computer Vision (ECCV'06), Graz, Austria, May 7-13, 2006: 428-441.
- [12] 张亚洲. 视频摘要算法研究 [D]. 杭州: 杭州电子科技大学, 2017.
- [13] 孙娅斌. 基于支持向量机的纹理图像分类算法 [J]. 计算机仿真, 2012, 29(5): 286-290.
- [14] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016.

Video Summarization Based on HOG-LBP Feature and SVM Classifier

CHENG Haiying, WANG Fengsui, ZHU Shuming

(College of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China)

Abstract: In order to improve the quality of video summarization, and get video information efficiently, a video summarization based on HOG-LBP feature and SVM classifier is proposed. Firstly, shape and texture features are extracted from the decomposed video sequence. Secondly, the trained support vector machine is used to classify the detected images and tentatively classify the correct frames as tentative frames. Finally, the image cosine similarity method is used to eliminate the redundant frames, and the key frames that are last retained are combined in the order of time in the original video to form a video summarization. Experimental results show that the proposed algorithm can extract key frames in three different scenes with an average accuracy of 94.08%, and the average error rate is only 23.18%. The accuracy of key frames extracted by the algorithm is high and the contents are representative. The quality of the video summarization is effectively improved.

Key words: video summarization; HOG-LBP feature extraction; SVM classifier; cosine similarity