

基于高层语义视觉词袋的色情图像过滤模型

吕林涛¹, 赵呈轩¹, 尚进¹, 杨宇祥²

(1. 西安理工大学 计算机科学与工程学院, 西安 710048; 2. 西安理工大学 机械与精密仪器工程学院, 西安 710048)

(lvtintao@xaut.edu.cn)

摘要:针对目前色情图像过滤算法对比基尼图像和类肤色图像误检率过高,且不能有效过滤带有淫秽动作的多人色情图像的缺点,提出一种基于高层语义视觉词袋的色情图像过滤模型。该模型首先通过改进的 SURF 算法提取色情场景局部特征点,然后融合视觉单词的上下文和空间相关高层语义特征,从而构建色情图像的高层语义词典。实验结果表明,该模型检测带有淫秽动作的多人色情图像准确率可达 87.6%,明显高于现有的视觉词袋色情图像过滤算法。

关键词:色情图像;过滤;图像高层语义;语义树;视觉词袋;鲁棒特征加速

中图分类号:TP393.08;TN911.73 **文献标志码:**A

Pornographic images filtering model based on high-level semantic bag-of-visual-words

LÜ Lin-tao¹, ZHAO Cheng-xuan¹, SHANG Jin¹, YANG Yu-xiang²

(1. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an Shaanxi 710048, China;

2. School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an Shaanxi 710048, China)

Abstract: Current pornographic images filtering algorithms have some shortcomings, such as high false positive rate toward the bikinis images and insufficiency when filtering pornographic images with pornographic actions. The paper proposed a new pornographic image filtering model based on High-level Semantic Bag-of-Visual-Words (BoVW). Firstly, local feature points in sex scene were detected using the Speeded-Up Robust Features (SURF) algorithm and then high-level semantic dictionary was constructed by fusing the context of the visual vocabularies and spatial-related high-level semantic features of pornographic images. The experimental results show that the model has an accuracy up to 87.6% when testing the multi-person pornographic images, which is significantly higher than the existing pornographic images filtering algorithm based on BoVW.

Key words: Pornographic image; filtering; image high-level semantics; semantic tree; Bag-of-Visual-Words (BoVW); Speeded-Up Robust Features (SURF)

0 引言

快速有效地监测和过滤网络色情信息已经成为网络信息安全中的一项重要研究课题。现存的色情图像过滤技术虽然对网络上明显的裸体图像可以实现较好的过滤效果,但不能有效过滤带有淫秽动作的多人色情图像,并且对含有比基尼的正常图像误检率较高。

随着视觉词袋 (Bag-of-Visual-Words, BoVW) 算法在图像检索和场景分类领域的成功应用^[1-2], 2008 年 Thoms Deselaers 等人^[3]首次提出采用视觉词袋算法提取图像高层特征过滤色情图像,该算法大大提高了色情图像的识别效率,但忽略了肤色特征。2009 年 Ana P. B. Lopes 等人^[4-5]在此基础上加入色调信息提出了 HUE-SIFT 特征描述算子,采用固定词汇大小分配,最后使用基于支持向量机 (Support Vector Machine, SVM) 分类算法进行色情图像的分类。但该算法对词袋算法本身缺乏优化,算法需要付出昂贵的计算时间。

2009 年 YIZHI LIU 等人^[6]提出了构建 SURF (Speeded-Up Robust Features) 视觉词汇的方法来提取肤色区域的局部特征算子以减少 SIFT 算法的计算时间^[7],虽然 SURF 的计算效率较 SIFT 算法有较大提高,但是视觉词汇聚类过程中采用简单的 K-Mean 聚类算法,造成语义缺失。2009 年王宇石等人^[8-9]采用视觉单词和语义映射算法检测色情图像,该算法虽然采用了语义映射模式,但是未能对高层语义视觉词组进行提取,因此算法仍存在一定的局限性。

针对目前色情图像过滤算法误检率过高等问题,本文提出一种基于高层语义视觉词袋的色情图像过滤模型。该模型在传统低层视觉特征基础上,利用语义分析和理解相关技术,提取图像中包含的高层语义特征,从语义层面来识别不良图像,不仅符合机器视觉认知过程,而且可有效弥补现有算法比基尼图像误检率过高、不能有效过滤带有淫秽动作的不良图像过滤技术的不足。实验证明,算法可以有效过滤带有淫秽动作的多人不良图像,较同类算法识别率高,误检率低。

收稿日期:2010-12-17;修回日期:2011-03-14。

基金项目:陕西省自然科学研究计划项目(2010JK740;2010JM8019);西安理工大学学科联合研究项目(102-210914)。

作者简介:吕林涛(1955-),男,陕西西安人,教授,主要研究方向:网络与信息安全、数据挖掘;赵呈轩(1982-),男,河南新乡人,硕士研究生,主要研究方向:网络与信息安全;尚进(1984-),男,河南郑州人,硕士研究生,主要研究方向:网络与信息安全;杨宇祥(1974-),男,湖南湘潭人,副教授,博士,主要研究方向:生物医学信号检测与处理。

1 高层语义视觉词袋的色情图像过滤模型

本文提出的基于高层语义视觉词袋的色情图像过滤模型如图1所示。该模型主要包括特征提取、高层语义词典构建和分类器构建三部分。

1) 特征提取采用文献[6]中的SURF(4×4)算法提取局部特征;

2) 高层语义词典构建通过随机森林算法^[1]将提取的视觉基元的进行分类产生底层视觉词汇,并将中级词汇再次分类、融合上下文关系和空间相关特征^[10-12],最终构建具有高区分度的高层语义视觉词典;

3) 分类器采用文献[1]中的 χ^2 -kernel SVM分类算法。

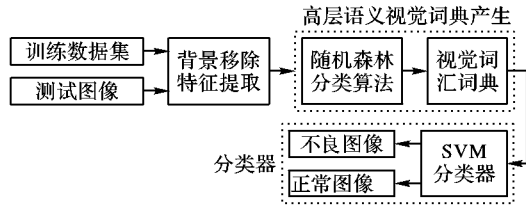


图1 高层语义视觉词袋的色情图像过滤模型

2 高层语义词典构建算法

高层语义词典构建算法主要通过视觉基元分类和视觉单词分类两阶段完成。视觉基元分类算法主要由随机森林算法和消除干扰词算法完成;视觉单词分类主要由上下文相关视觉词汇分配算法和空间相关视觉词汇分配算法完成。最终完成高层语义视觉词典的构建。高层语义词典构建算法流程如图2所示。

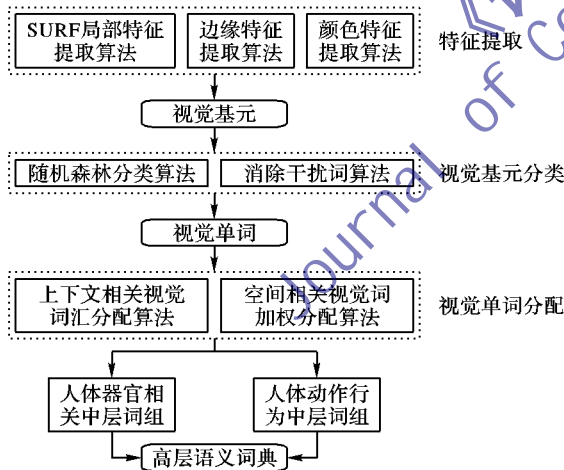


图2 高层语义词典构建算法流程

2.1 空间相关视觉词汇的权重分配算法

色情图像最明显的特点是有大面积的肤色区域和敏感器官的暴露,视觉词汇在分配的过程中如果对色情图像空间相关的视觉词汇加权分配,可以有效识别敏感区域和提高视觉词汇的区分度。本文采用对色情图像空间相关的视觉词汇加权分配^[10]。算法描述如下:

1) 从兴趣点集合提取出的视觉词汇集合 $w_l(1 < l < k, k$ 是视觉词典集合中视觉词汇的数量);

2) 计算视觉词汇 w_l 在高斯 β_i 条件下加权频率 $T_{fw\beta_i}$,

$$T_{fw\beta_i} = \sum_{m=1}^{n_i} P(\beta_i/Z_m);$$

3) 计算视觉词汇 w_l 的平均权重 $T_{fw\beta_i}, T_{fw\beta_i} =$

$$\sum_{i=1}^{n_{w_l}} (T_{fw\beta_i})/n_{w_l};$$

4) 计算 w_l 的权重 $I_{fw\beta_i}, I_{fw\beta_i} = \ln \frac{n}{n_{w_l}};$

5) 计算视觉词汇 w_l 最终空间权重 $S_{w_{w_l}}, S_{w_{w_l}} = T_{fw\beta_i} \times I_{fw\beta_i}。$

2.2 人体器官特征的空间拓扑结构算法

传统的视觉词袋算法通过编码局部突出片段的视觉基元获取视觉单词,但是却忽略了图像对象间的空间拓扑结构,本文将利用人体器官特征的空间拓扑结构表达的三角关系提取色情图像高区分度特征。经研究发现,人体器官特征本身具有一定的空间拓扑结构,利用人体器官特征的这一特点可以准确判定图像是否含有人体躯干部位。对于图像中包含单人的色情图像身体倾斜度不大于90°,人体暴露的胸部特征和人脸特征基本成三角形形状,如图3所示。

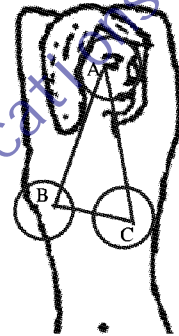


图3 色情图像人体器官特征空间拓扑结构

本文采用简化2D字符串^[11]构建高层语义视觉词汇间的空间拓扑结构,假设 V 是人体器官特征对象集合,集合内每一个对象对应一个高层视觉词组,本文用“<”表示左右或上下空间关系。如图3,人体器官特征对象形成字符串为“B < A < C”。对于满足此类关系的图像将被识别为色情图像。

2.3 多人色情图像高层语义树构建算法

对于包含多人的色情图像,除具有大面积的肤色区域信息,更重要的是包含一些特定的淫秽动作,主要有亲吻、拥抱以及躯体部位的粘合。亲吻主要涉及嘴部和其他敏感器官,拥抱主要涉及肢体,躯体粘合主要涉及人体躯干和敏感器官。上述三种行为语义基本构成多人色情图像中的行为语义特征,色情图像高层语义树构建过程如图4所示。

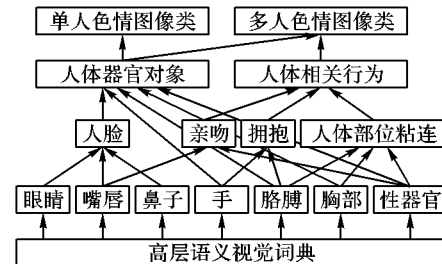


图4 色情图像高层语义树构建过程

在图4中,语义模型包括:人体器官对象模型和人体相关行为模型。由于色情图像相关的语义词汇中有些词汇出现的概率较高,因此,本文在文献[12]提出的加权分配语义词汇算法基础上,去除场景语义模型,只对人体器官对象视觉词组和人体行为相关视觉词组加权分配。对于数据集 V 中任意一张图像 C_i ,如果人体器官对象视觉词组概率为 P_o ,人体行

为相关视觉词组概率为 P_A , 可以通过公式 $P_{C_i} = \alpha P_O + \beta P_A$, $\alpha + \beta = 1$, 计算 C_i 的敏感系数。当敏感系数大于阈值 t 时判定此图色情图像。

3 实验结果分析对比

为了验证高层语义视觉特征在色情图像过滤中的效果, 本文通过采用网络收集的色情图像和 Pascal VOC2010 图像库正常图像作为训练和测试数据集合, 并与本领域著名文献 [4,6,9] 中方法完成实验结果对比。

3.1 实验采样数据集

实验图像来自网络和 Pascal VOC2010 图像库, 其中在网络上收集的色情图像 5 800 张, 可以分为 A 类: 包含明显的裸体图像; B 类包含比基尼图像; C 类: 包含带有淫秽动作的多人色情图像; 正常图像 5 000 张均取自 Pascal VOC2010 图像库。

训练数据集和测试数据集分别为 3 000 和 5 560 张图像, 从三类色情图像中各随机取出 1 500 张进行训练, 另外手工标注了 2 000 张色情图像的人体躯干和敏感区域(性器官), 用来产生各类低层视觉语义单词。

3.2 实验结果对比

本文采用随机森林算法从 1 932 900 个关键点中分类产生了 7 883 个 SURF 视觉单词, 并根据词汇的上下文关系和空间相关特征进行再次分类, 产生 5 120 个中级视觉词组, 同时将中级词组分为 2 150 个人体器官相关和 2 980 个人体动作相关的高层语义词汇。最后对高层语义词汇编码形成色情图像高层语义词典。

由于色情图像中包含大面积的肤色区域和人体敏感器官, 此类局部特征点具有明显的上下文相关和空间相关特性。表 1 结果表明: 加入上下文相关和空间相关特性的视觉词汇分配算法的特征区分度要明显高于传统视觉词汇分配算法; 三类图像的过滤正确率本文算法均高于其他算法。视觉词袋算法的时间复杂度主要取决于特征提取和视觉词汇分配算法的时间复杂度, 本文采用文献 [1] 中的随机森林算法对视觉词汇进行分配, 缩短了视觉词汇分配的计算时间。实验结果表明, 在特征提取和词汇分配两个阶段的总计算时间, 本文算法的平均计算时间均低于其他三类算法。四类算法特征提取和词汇分配平均计算时间对比如图 5 所示。

表 1 四类色情图像过滤算法正确率对比 %

方法序号	方法	裸体图像类	比基尼类	带有淫秽动作多人色情图像类
1	文献[6]方法	85.4	70.2	67.3
2	文献[4]方法	81.6	85.1	65.2
3	文献[9]方法	89.9	80.2	68.4
4	本文方法	92.7	90.1	87.6

4 结语

基于视觉词袋的色情图像检测是一种很有前景的方法, 因为通过理解图像表达的高层语义对色情图像进行过滤不仅可以过滤明显的裸体图像和多人色情图像, 而且也符合机器视觉和认知理论。本文用具有高分度度的高层语义特征进行分类检测速度和准确度都较现有算法有较大提高, 但在过滤 90°侧身的色情图像检测方面有待进一步研究。

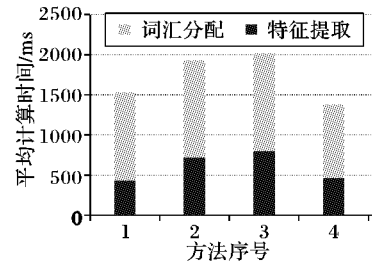


图 5 四类过滤算法特征提取和词汇分配平均计算时间对比

参考文献:

- [1] UIJLINGS J R R, SMEULDERS A W H, SCHA R J H. Real-time visual concept classification [J]. IEEE Transactions on Multimedia, 2010, 12(7): 665 - 681.
- [2] WU LEI, HOI S C H, YU NENGHAI. Semantics-preserving bag-of-words models and applications [J]. IEEE Transactions on Image Processing, 2010, 19(7): 1908 - 1920.
- [3] DESELAERS T, PIMENIDIS L, NRY H. Bag-of-visual-words models for adult image classification and filtering [C]// ICPR 2008: 19th International Conference on Pattern Recognition. Washington, DC: IEEE Computer Society, 2008: 1 - 4.
- [4] LOPES A P B, de AVILA S E F, PEIXOTO A N A, et al. A bag-of-features approach based on HUE-SIFT descriptor for nude detection [C]// EUSIPCO 2009: Proceedings of the 17th European Signal Processing Conference. Glasgow, Scotland: Mendeley, 2009: 1552 - 1556.
- [5] LOPES A P B, de AVILA S E F, PEIXOTO A N A, et al. Nude detection in video using bag-of-visual-features [C]// SIBGRAPI'09: Proceedings of the 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing. Washington, DC: IEEE Computer Society, 2009: 224 - 231.
- [6] LIU YIZHI, XIE HONGTAO. Constructing SURF visual-words for pornographic images detection [C]// ICCIT'09: 12th International Conference on Computers and Information Technology. Washington, DC: IEEE, 2009: 404 - 407.
- [7] LUO JUAN, GWON O. A comparison of SIFT, PCA-SIFT and SURF [J]. International Journal of Image Processing, 2009, 3(4): 143 - 152.
- [8] 王宇石, 李远宁, 高文. 基于局部视觉单词分布的成人图像检测 [J]. 北京理工大学学报, 2008, 28(5): 410 - 413.
- [9] 王宇石, 高文. 基于视觉单词和语义映射的色情图像检测算法 [J]. 高技术通讯, 2009, 19(10): 1041 - 1047.
- [10] LI TENG, MEI TAO, KWEON I, et al. Contextual bag-of-words for visual categorization [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 21(4): 381 - 392.
- [11] WANG MEI, WU YANLING, LI GUANGDA, et al. Object recognition via adaptive multi-level feature integration [C]// APWEB 2010: 12th International Asia-Pacific Web Conference. Washington, DC: IEEE, 2010: 253 - 259.
- [12] CHEN MIANSU, FU PING, LI YONG, et al. Condensed semantic tree model for image category representation [C]// 2010 ICCAE: Proceedings of the 2nd International Conference on Computer and Automation Engineering. Chongqing: [s. n.], 2010, 4: 358 - 362.