

中图法分类号: TP391.41 文献标志码: A 文章编号: 1006-8961(2011)07-1226-08

论文索引信息: 刘毅志, 杨颖, 唐胜, 林守勋. 基于视觉注意模型 VAMAI 的敏感图像检测方法 [J]. 中国图象图形学报, 2011, 16(7): 1226-1233

基于视觉注意模型 VAMAI 的敏感图像检测方法

刘毅志^{1),2)}, 杨颖³⁾, 唐胜¹⁾, 林守勋¹⁾

¹⁾(中国科学院计算技术研究所前瞻研究实验室, 北京 100190)

²⁾(中国科学院研究生院, 北京 100049)

³⁾(中国农业大学信息与电气工程学院, 北京 100083)

摘要: 基于内容的敏感图像检测方法是过滤互联网上敏感信息的有效手段。然而, 基于全局特征的检测方法误检率偏高, 现有的基于 BoW (bag-of-visual-words) 的检测方法速度较慢。为了快速准确地检测敏感图像, 本文提出基于视觉注意模型 VAMAI (visual attention model for adult images) 的敏感图像检测方法, 包括构造面向敏感图像的视觉注意模型 VAMAI、基于感兴趣区域和 SURF (speeded up robust features) 的视觉词表算法、全局特征选择及其与 BoW 的融合三部分。首先, 结合显著图模型、肤色分类模型和人脸检测模型, 构造 VAMAI, 用于较准确地提取感兴趣区域。然后, 基于感兴趣区域和 SURF 构建视觉词表, 用于提高基于 BoW 的检测方法的检测速度与检测精度。最后, 比较多种全局特征的性能, 从中选择颜色矩, 将它与 BoW 的支持向量机分类结果进行后融合。实验结果表明: VAMAI 能够较准确地检测感兴趣区域; 从检测速度和检测精度两方面显著地提高了敏感图像的检测性能。

关键词: 敏感图像检测; 视觉注意模型; 感兴趣区域; 视觉单词; SURF 特征

Visual attention model VAMAI for adult image detection

Liu Yizhi^{1),2)}, Yang Ying³⁾, Tang Sheng¹⁾, Lin Shouxun¹⁾

¹⁾ (Advanced Computing Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190 China)

²⁾ (Graduate University of Chinese Academy of Sciences, Beijing 100049 China)

³⁾ (Institute of Information and Electrical Engineering, Chinese Agricultural University, Beijing 100083 China)

Abstract: Content-based adult image detection is an effective approach of filtering objectionable information proliferating on the Internet. However, the false positive is high for adult image detection based on global features; and the detection speed is low for adult image detection based on bag-of-visual-words (BoW) models. To detect adult images rapidly and accurately, we propose a novel method which includes three steps. First, we present VAMAI, the abbreviation of visual attention model for adult images, specialized in capturing the region of interest (ROI). VAMAI integrates saliency map model, skin-color model and face detection model to obtain ROI more precisely than ever before. Second, a codebook algorithm based on ROI and SURF (speeded up robust features) is proposed to improve the speed and the accuracy of BoW based adult image detection. Furthermore, we select color moments as the global feature after evaluating some commonly-used global features. Then we use "late fusion" to combine the results of color moments and BoW which are classified by supported vector machines. Experiments show that VAMAI is able to detect ROI precisely and our method can remarkably improve the performance of adult image detection in terms of speed and accuracy.

Keywords: adult image detection; visual attention model; region of interest; visual words; speeded up robust features

收稿日期: 2010-05-27; 修回日期: 2010-10-13

基金项目: 国家重点基础研究发展计划(973)项目(2007CB311100); 国家自然科学基金项目(60873165, 60802028); 北京市科技新星计划项目(2007B071); 北京市教育委员会共建项目专项资助。

第一作者简介: 刘毅志(1973—), 男。中国科学院计算技术研究所计算机应用技术专业博士研究生, 主要研究方向为基于内容的图像与视频过滤。E-mail: liuyizhi@ict.ac.cn。

0 引言

随着互联网络技术的迅速发展、数字图像获取设备的日益便捷和计算机存储介质成本的不断降低,互联网正在转变成一个海量的图像数据库。由于图像数据良莠并存,其中的敏感内容对青少年的毒害极大,所以,人们希望敏感内容过滤技术能够精益求精,使其性能尽善尽美。

基于内容的敏感图像检测方法是最有效的过滤技术之一,它利用计算机视觉技术、图像处理和图像理解的方法,判断图片内容是否是敏感的^[1]。其中,特征选择、特征提取和分类器是决定性能的关键因素。鉴于支持向量机(SVM)的优越性能,许多敏感图像检测系统^[2-5]都采用 SVM 分类器。特征提取的速度与选择的特征密切相关。一般而言,全局颜色特征与全局纹理特征的提取速度较快,而形状特征与局部特征的提取速度较慢。根据特征的种类与提取的区域,将基于内容的敏感图像检测方法分为3类:1)基于全局特征(形状、颜色和纹理等)的检测方法,从整幅图像中提取特征;2)基于 BoW (bag-of-visual-words)的检测方法,从图像中的兴趣点中提取局部特征;3)基于兴趣区域(ROI)的检测方法,从 ROI 中提取更能表征“敏感部位或敏感姿态”的特征。下面简要予以综述与分析。

最早的工作是 Forsyth 等人^[6]根据人体形状设定组合规则来构建人体轮廓模型,再依据肤色区域能否组合成为人体轮廓来检测敏感图像。郑清芳等人^[7]使用基于边缘的 Zernike 矩以检测有害标志物。这些方法由于使用形状特征,因此速度很慢。基于全局颜色特征和全局纹理特征的检测方法速度较快,但误检率偏高。例如:Jones 和 Rehg^[8]提出的统计模型,根据皮肤表面抽取的特征,采用神经网络分类器识别敏感图像;Daubechies 小波变换、归一化中心矩和颜色直方图被用于构成语义匹配向量以分类图像;Zheng 等人^[9]运用肤色检测的最大熵模型来过滤敏感图像。AIRS^[10]用 MPEG-7 描述子来识别和分级敏感图像。WIPE 系统^[11]将直方图分析,纹理分析和形状匹配方法,用于截获敏感图像。曾炜等人^[3]采用多种全局特征开发了图像卫士系统。Rowley 等人^[4]在 Google 中利用 27 个视觉特征来过滤互联网上的敏感图像。唐胜等人^[5]采用隐含狄利克雷分布(Latent Dirichlet Allocation)模型依

据全局特征将图像分成若干子类后,再用 SVM 分类。

基于 BoW 的敏感图像检测方法^[2, 12-13]可以降低误检率,因为局部特征能够描述语义丰富的图像区块(patch),对于姿态遮挡、尺度伸缩、视角转换和光照变化等具有较好的鲁棒性。敏感图像检测的关键在于图像中是否存在敏感部位或敏感姿态,因此,基于 BoW 的检测方法具有广阔的应用前景。但是,现有方法难以适应大规模敏感图像检测的需求,主要原因是:1)采用 SIFT (scale-invariant feature transform)作为局部特征,特征提取非常耗时;2)用于聚类视觉词表的区块中,描述“敏感部位或敏感姿态”的区块(简称敏感区块)所占的比例太小,使得视觉词表难以准确地表达敏感图像的语义;3)现有方法中特征类型单一,很少讨论与全局特征的融合。

ROI 是指图像中包含“敏感部位或敏感姿态”的较小区域。现有的 ROI 提取方法大多基于肤色检测^[14-16],尽管计算快捷,但却难以区分与肤色相近的非敏感事物。譬如: Xu 等人^[14]运用 JSEG 算法将图像分割成任意形状的区域后,基于全局特征和高斯混合模型获得肤色区域。Shih 等人^[15]通过检测肤色像素获取 ROI。Yang 等人^[16]把图像划分为 4×4 的网格,根据周围点与网格的投票得到 ROI。近年出现了基于 ROI 和 BoW 的检测方法,即从 ROI 中提取局部特征,再用基于 BoW 的方法进行检测。王宇石等人^[12]提出兴趣点检测子 FCDoG (fast color differences of Gaussian),并用于提取 ROI。文献^[13]通过图像分割将相关信息密集分布的区域作为 ROI。

如何准确地检测 ROI 是一个重要的挑战性问题。由于视觉注意模型非常符合人们的视觉感知,所以,依靠视觉注意模型(VAA)获取注意区域的方法值得借鉴。通用的视觉注意模型中,最具代表性的是 Itti 等人^[17]和 Ma 等人^[18]提出的显著图模型(SMM)。Itti 等人^[17]提出的显著图是一幅表明图像各点显著性的 2 维图像,显著图中的点和输入图像的像素有拓扑上的对应关系。首先采用不同尺度的 DoG (differences of Gaussian) 函数对图像的不同特征滤波,把得到的各个特征的响应求和,作为该图像的显著图;然后使用动态神经网络采用 WTA (winner take all) 算法选取显著值最大的点作为注意点。Ma 等人^[18]认为,对比是决定人的视觉感知的最重要因

素,进而提出基于对比的视觉注意模型。先构造一个基于对比的显著图,再用模糊增长算法寻找到图像中的三级注意区域。然而,将通用视觉注意模型直接用于敏感图像检测时,我们发现:1) Ma 等人基于对比的显著图提取算法^[18]在提取视觉注意区域时的准确率不是很高,尤其是对于含有人物的图像来说,往往是把人物的整体当成了背景,而不是将人物突显出来。同时,由于该算法的预处理过程需要对图像进行块量化,抹杀了图像的一些细节,从而影响了注意区域轮廓的清晰性;2) Itti 等人的算法^[17]虽然在提取视觉注意区域时的精度和准确率很高,但是需要计算图像的颜色、纹理和方向等特征,计算速度较慢。

为了快速准确地检测敏感图像,提出基于视觉注意模型 VAMAI (visual attention model for adult images) 的敏感图像检测方法,从 3 个方面着手提高敏感图像的检测性能(包括检测速度和检测精度): 1) 构造面向敏感图像的视觉注意模型 VAMAI。它结合显著图模型、肤色分类模型和人脸检测模型,用于更准确地提取 ROI; 2) 构建基于 ROI 和 SURF (speeded up robust features) 的视觉词表算法(如图 1 所示)。利用 ROI 来选择用于聚类视觉词表的区块,提高敏感区块所占的比例,使视觉词表能够更准确地表达敏感图像的语义,达到提高检测精度的目的。此外,最新的研究^[19-20]表明,与 SIFT 相比, SURF 不仅特征提取速度较快,而且对于光照变化、尺度伸缩、仿射变化和旋转等也具有相当的鲁棒性。为了在不降低检测精度的前提下提高检测速度,故采用 SURF 取代 SIFT; 3) 比较多种全局特征的性能,从中选择颜色矩,将它与 BoW 的 SVM 分类结果进行后融合,进一步提高检测精度。实验结果表明,本文方法从检测速度和检测精度两方面显著地提高了敏感图像的检测性能。

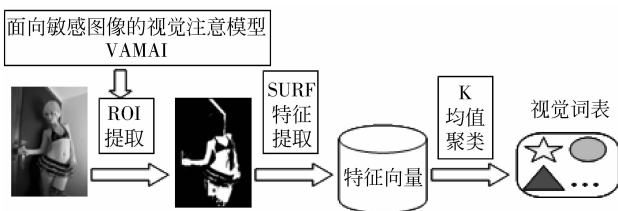


图 1 视觉注意模型 VAMAI 与视觉词表算法

Fig. 1 VAMAI and the codebook algorithm

1 VAMAI

如图 2 所示,构造一个面向敏感图像的视觉注意模型 VAMAI,用于更准确地提取 ROI。其构造思想是:首先结合 Itti 和 Ma 等人的显著图提取算法,建立适合人体图像的显著图模型(SMM),获得粗粒度(图像块)上的显著图;其次,应用 Garcia 等人^[21]像素级的肤色分类模型(SCM)获得细粒度(像素级)上的肤色区域;然后,借助 Viola 等人^[22]的人脸检测模型(FDM)获取人脸区域;最后,将这些区域二值化,分别得到显著区域、肤色区域和人脸区域,VAMAI 获取的 ROI 就是显著区域、肤色区域和非人脸区域的交集。

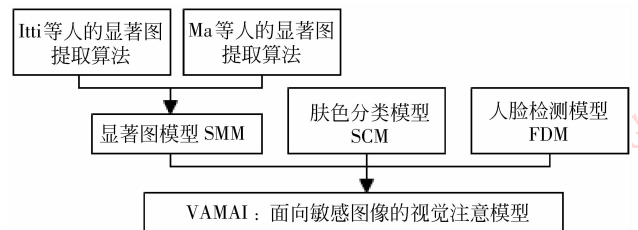


图 2 VAMAI 的构造过程

Fig. 2 The method of constructing VAMAI

定义 假设 R_{VAMAI} 表示用面向敏感图像的视觉注意模型获得的区域, R_{SMM} 表示用显著图模型获得的显著区域, R_{SCM} 表示用肤色分类模型获得的肤色区域, R_{FDM} 表示用人脸检测模型获得的人脸区域,则 VAMAI 模型可以表示为

$$R_{VAMAI} = R_{SMM} \cap R_{SCM} \cap \bar{R}_{FDM} \quad (1)$$

图 3 给出了一些 VAMAI 获取 ROI 效果示例。其中,由 Itti 等人方法^[17]和 Ma 等人方法^[18]获取的显著区域较大且模糊;由(a)一(d)可以看出,肤色检测难以区分近似肤色的非敏感事物,借助显著图可以有效地减少这一现象;而(e)一(j)中显著图的轮廓较模糊,而肤色图使得 ROI 的轮廓相对清晰;人脸检测也有一定效果,但(a)中的人脸没有检测出来。由图 3 可以看出:1) 对于敏感图像,与文献^[17-18]的显著图模型相比,显著图模型获得的显著区域能够更好地捕获敏感部位所在区域;2) VAMAI 能够较精确地勾勒出 ROI 的轮廓。根据 4.2 节的评估结果,VAMAI 检测 ROI 的准确率可以达到 91.33%。



图 3 VAMAI 获取 ROI 的效果示例图

Fig. 3 Some examples of capturing ROI based on VAMAI

1.1 显著图模型

结合 Itti 和 Ma 等人的显著图提取算法,构建适合人体图像的显著图模型的过程如下:

1) 原始图像预处理,包括统一图像尺寸和图像块量化。先把原始图像转换为统一尺寸的图像;再将图像量化成 $n \times n$ 像素的小块。由于图像的大小固定,且以量化块为单位而不是以像素为单位加以处理,提高了计算效率。

2) 构造图像的九级尺度的 2 维高斯金字塔 $I(\sigma), \sigma \in \{0, 1, \dots, 8\}$, 分别对各级金字塔计算各个量化块的对比值,得到多尺度的基于对比的特征图。

(1) 将图像的每个量化块看成一个感知单元 (i, j) , 然后计算它邻域内的对比值 $C_{i,j}$ 。在式(2)中, $p_{i,j}(i, j \in \{0, \dots, n\})$ 和 q 表示量化块的感知信号(本文只计算亮度和颜色特征)。 d 是 $p_{i,j}$ 和 q 之间的距离(本文用欧氏距离度量)。 θ 是指感知单元 (i, j) 的邻域。

$$C_{i,j} = \sum_{q \in \theta} d(p_{i,j}, q) \quad (2)$$

(2) 图像的亮度特征是由 $I = (r + g + b)/3$ 得到,并为其产生 9 个尺度的高斯金字塔。在中心尺度 c 和外周尺度 s 之间用中心-外周差操作(记作

Θ)来计算亮度特征图。对于 σ 个不同尺度的亮度图为

$$I(c, s) = |I(c) \Theta I(s)| \quad (3)$$

式中, $c \in \{2, 3, 4\}, s = c + \delta, \delta \in \{3, 4\}$ 。

(3) 图像的颜色特征包含红色、绿色、蓝色和黄色 4 个颜色通道图,形成 4 个色度金字塔 $R(\sigma), G(\sigma), B(\sigma)$ 和 $Y(\sigma)$ 。

$$\text{红色通道 } R = r - (g + b)/2 \quad (4)$$

$$\text{绿色通道 } G = g - (r + b)/2 \quad (5)$$

$$\text{蓝色通道 } B = b - (r + g)/2 \quad (6)$$

$$\text{黄色通道 } Y = (r + g)/2 - |r - g|/2 - b \quad (7)$$

由于人的视觉皮层中存在红绿、绿红、黄蓝和蓝黄颜色对比对,所以计算 $RG(c, s)$ 和 $BY(c, s)$ 来表示人们对颜色的感知。

$$RG(c, s) = |(R(c) - G(c)) \Theta (G(s) - R(s))| \quad (8)$$

$$BY(c, s) = |(B(c) - Y(c)) \Theta (Y(s) - B(s))| \quad (9)$$

3) 将各级特征图归一化后进行叠加,形成最终的显著图。

将特征图内每个量化块邻域内的对比值 $C_{i,j}$ 归一化到 $[0, 255]$, 归一化操作表示为 $N(\cdot)$ 。这样得到亮度显著图 \bar{I} 和颜色显著图 \bar{C} , 叠加结果为最终的显著图 S 。

$$\bar{I} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(I(c, s)) \quad (10)$$

$$\bar{C} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [N(RG(c, s)) + N(BY(c, s))] \quad (11)$$

$$S = \frac{1}{2}(N(\bar{I}) + N(\bar{C})) \quad (12)$$

式中, \oplus 表示叠加运算。

1.2 肤色分类模型

应用 Garcia 等人^[21]提出的静态阈值肤色分类模型提取肤色区域。在 YC_bC_r 颜色空间上, 颜色接近于肤色的像素与 C_b 、 C_r 分量的关联程度大于 Y 分量。故必须将输入图像从 RGB 空间转换到 YC_bC_r 空间。当像素呈肤色时, 其 C_b 和 C_r 分量满足以下条件:

$$\begin{cases} C_r \geq \max\{-2(C_b + 24), -(C_b + 17), \\ -4(C_b + 32), 2.5(C_b + \theta_1), \theta_3, 0.5(\theta_4 - C_b)\} \\ C_r \leq \min\{(220 - C_b)/6, 4(\theta_2 - C_b)/3\} \end{cases} \quad (13)$$

式中的 θ_1 、 θ_2 、 θ_3 和 θ_4 则满足如下约束:

$$\begin{cases} \theta_1 = \begin{cases} -2 + (256 - Y)/16 & Y > 128 \\ 6 & \text{其他} \end{cases} \\ \theta_2 = \begin{cases} 20 - (256 - Y)/16 & Y > 128 \\ 12 & \text{其他} \end{cases} \\ \theta_3 = \begin{cases} 6 & Y > 128 \\ 2 + Y/32 & \text{其他} \end{cases} \\ \theta_4 = \begin{cases} -8 & Y > 128 \\ -16 + Y/16 & \text{其他} \end{cases} \end{cases} \quad (14)$$

1.3 人脸检测模型

在所有的人脸检测模型中, Viola 等人^[22]提出的 Haar-Like 算法最为经典。该算法采用 Haar 特

征, 使用 Adaboost 方法训练人脸检测分类器, 可用于实时检测。

Haar 特征块是一系列由黑白两色区域组成的矩形。Haar 特征是将 Haar 特征块中的白色区域灰度值减去黑色区域灰度值后得到的结果。在人脸的特定区域, 不同部位对应的某些 Haar 特征值往往是很固定的。也就是说, 某些人脸部位可以用相应的 Haar 特征块来覆盖。然后, 用 Adaboost 方法将多个 Haar 特征融合, 可获得一个较强的人脸分类器, 在候选区域中筛选出准确的人脸区域。

2 基于 ROI 和 SURF 的视觉词表算法

视觉词表是 BoW 的核心, 所以, 视觉词表算法对基于 BoW 方法的性能至关重要。Jurie 等人^[23]认为, 视觉词表算法包括区块选择、区块描述和量化算法三部分。除了上述三部分之外, 视觉词表算法还应该包括区域选择方法。因为 Nowak 等人^[24]认为, 区块采样策略应该关注图像中信息丰富的区域。区域选择是指选择两类区域: 视觉词表的采样区域和用于训练或测试的视觉单词直方图的采样区域。表 1 详细比较了本文的视觉词表算法与传统算法, 区别之处在旁边用星号(*)加以标注。

基于 ROI 和 SURF 构建视觉词表, 用于提高基于 BoW 的敏感图像检测方法的检测速度和检测精度。使用 VAMAI 获取 ROI, 从中选择聚类视觉词表的区块, 提高敏感区块在总区块中所占的比例, 使视觉词表能够更准确地表达敏感图像的语义, 达到提高检测精度的目的。

表 1 本文的视觉词表算法与传统算法的区别

Tab. 1 The differences of our codebook algorithm and the traditional codebook algorithm

| | 视觉词表算法 | 传统算法 | 本文算法 |
|------|--|--------------|---------------|
| 局部特征 | 区块检测子 区块描述子 | DoG SIFT | * SURF |
| 区域选择 | 视觉词表的采样区域(用于聚类) 视觉单词直方图的采样区域(用于训练或测试) | 整幅图像 整幅图像 | * ROI 整幅图像 |

基于 BoW 的敏感图像检测系统的耗费时间主要包括局部特征的提取时间与视觉单词直方图检测时间两部分。当视觉词表中视觉单词数量接近时, 视觉单词直方图检测的时间也相差不大。SURF 与 SIFT 对于光照变化、尺度伸缩、仿射变化和旋转等都具有较强的鲁棒性。在检测关键点时, SIFT 一般

使用单一高斯函数对不同尺度的图像进行卷积, 而 SURF 使用多种滤波器对原始图像进行卷积, 同时采用模板模拟高斯滤波器函数, 大大减少了时间的开销。在计算主方向时, SIFT 以关键点附近像素的梯度主方向为准, SURF 由于采用积分图像而使计算更简洁。因此, 用 SURF 代替 SIFT, 能够在不降低

检测精度的前提下,极大地减少了特征提取时间,从而提高了检测速度。在 Linux PC (Pentium IV, 3 GHz) 条件下,对于 Graffiti 图像集中 300×240 像素的第 1 幅图像,提取 64 维 SURF 特征需要 354 ms,提取 128 维 SURF 特征花费 391 ms,提取 SIFT 特征 (128 维) 则需 1 036 ms^[19]。考虑到检测精度与检测速度的折中,这里使用 128 维的 SURF 特征。

3 全局特征选择及其与 BoW 的融合

融合全局特征与 BoW,采用高斯 RBF 核 (Gaussian Radial Basis Function Kernel) SVM,进一步提高检测精度。特征融合方法有前融合和后融合两种。前融合是指将多个特征依次连接得到新特征,然后用分类器进行判决。后融合是指将多种特征的分类结果,运用直接平均、最大值、最小值和乘积等方法得到最终的判决结果。

全局特征包括 3 种全局颜色特征和 3 种全局纹理特征:166 维的颜色直方图 (CH)、166 维的颜色相关图 (CC)、225 维的颜色矩 (CM)、320 维的边缘纹理 (EH)、96 维的共生纹理 (TC) 和 108 维的小波纹理 (HW)。首先,逐一比较常见的 3 种全局颜色特征的性能;然后,选择性能最好的全局颜色特征与其他 3 种全局纹理特征分别进行前融合;最后,从上述 6 种情况中,选择性能最佳的全局特征或全局特征组合,与 BoW 进行融合。

4 实验结果与分析

4.1 基于内容的敏感图像检测方法

敏感图像检测的性能不仅取决于检测方法,而且与测试图像集密切相关。为了保证实用性,从互联网和 Corel Gallery 库收集了 9 万幅图像。如表 2 所示,训练集有 5 万幅图像,测试集有 4 万幅图像,各自均包含 1 万幅敏感图像和 1 万幅人脸、手足和泳装等易误检的非敏感图像;非敏感图像中还包含大量颜色接近肤色的易误检的其他图像。

表 2 本文图像库的统计数据
Tab.2 Statistics of our dataset

| | 训练集/幅 | 测试集/幅 |
|-------|--------|--------|
| 敏感图像 | 10 000 | 10 000 |
| 非敏感图像 | 人体图像 | 10 000 |
| | 其他图像 | 20 000 |

由于敏感图像检测领域没有公共图像集,所以,采用 Zheng 等人的最大熵肤色分类模型^[9]和基于全局特征的 SVM 分类结果两个基准。CH 和 EH 分别是最常用且有效的全局颜色特征和全局纹理特征,所以选择 CH 与 EH 前融合再用 SVM 分类的检测方法 (即 CH + EH) 作为另一个基准。本文实验使用的机器配置是 1.86 GHz 的双核 CPU 和 2 G 内存,在 visual studio 2003 环境下实现。性能评价指标选用 ROC (receiver operating characteristic) 曲线。

4.2 评估视觉注意模型 VAMAI

由于人类视觉感知的主观性,视觉注意模型没有标准的、客观的评估方法^[18]。借鉴文献 [18] 的评估方法,将图像集分成 3 组,每组 350 幅,请 20 位非计算机专家对其进行评价。对于 VAMAI 获取 ROI 的准确程度,给予“GOOD”、“ACCEPT”与“FAILED”3 种评价。其中,“ACCEPT”是指获得的 ROI 包含了原图像中的一部分“敏感部位或敏感姿态”,但是仍有一部分没有包含;“GOOD”和“ACCEPT”都属于“检测准确”的情况。

由表 3 可以得出如下结论:VAMAI 能够较准确地发现“敏感部位或敏感姿态”所在区域,准确率为 91.33%,但不能完全囊括图像中的“敏感部位或敏感姿态”。结合图 3 还可以发现,VAMAI 能够较精确地勾勒出 ROI 的轮廓。

表 3 VAMAI 获取 ROI 的效果评估

Tab.3 Evaluation of the effects of capturing ROI based on VAMAI

| 图像集 | GOOD | ACCEPT | FAILED | /% |
|---------------|-------|--------|--------|----|
| 第 1 组 (350 幅) | 41.71 | 49.43 | 8.86 | |
| 第 2 组 (350 幅) | 41.43 | 51.14 | 7.43 | |
| 第 3 组 (350 幅) | 46.29 | 44.00 | 9.71 | |
| 平均 | 91.33 | | 8.67 | |

4.3 基于 BoW 的敏感图像检测方法

比较 3 种视觉词表算法的敏感图像检测结果:1) 基于整幅图像内 SIFT 特征的视觉词表算法 (即现有的视觉词表算法),用整幅图像内所有区块的 SIFT 特征聚类生成视觉词表;2) 基于整幅图像内 SURF 特征的视觉词表算法,即在现有的视觉词表算法中替换局部特征;3) 基于肤色区域内 SURF 特征的视觉词表算法,应用 Garcia 等人的肤色分类模型^[21]获取肤色区域,将肤色区域内区块的 SURF 特征聚类生成视觉词表;4) 基于 VAMAI 和 SURF 的视觉词表算法,用 VAMAI 获取 ROI 后,将 ROI 内区块

的 SURF 特征聚类生成视觉词表。为了便于比较,用于聚类视觉词表的区块数目均保持在 58 万左右,采用 K 均值聚类算法获得 600 个视觉单词。如图 4 所示。

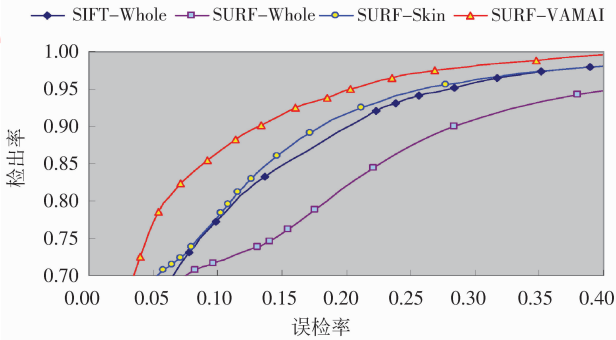


图 4 基于不同视觉词表算法的敏感图像检测的 ROC 曲线
Fig. 4 ROC curves of adult image detection based on different codebook algorithms

综合分析图 4 和表 4,可以得到以下结论:1) 基于 ROI 和 SURF 的视觉词表算法(SURF-VAMAI 和 SURF-Skin)优于现有的视觉词表算法,而且本文的视觉词表算法 SURF-VAMAI 效果显著。从图 4 可以看出,如果直接用 SURF 代替 SIFT,效果较差;SURF-Skin 采用肤色区域作为 ROI 也能获得较优的性能,如图 4 中标记圆圈的曲线;运用 VAMAI 获取的 ROI 能够显著提高检测性能,如图 4 中标记三角形的曲线。2) 本文的视觉词表算法比现有的视觉词表算法速度有明显提高。表 4 中给出了采用这 4 种视觉词表算法进行敏感图像检测所需耗费的时间。容易看出,特征检测时间相差无几,关键是 SURF 提取时间只有 SIFT 提取时间的 24%。

表 4 基于不同视觉词表算法的敏感图像检测时间开销
Tab. 4 The consuming time of adult image detection based on different codebook algorithms

| | 特征提取 | 特征检测 | 系统执行 / (ms/幅) |
|------------|-------|------|---------------|
| SIFT-Whole | 1 626 | 682 | 2 308 |
| SURF-Whole | 390 | 779 | 1 169 |
| SURF-Skin | 390 | 670 | 1 060 |
| SURF-VAMAI | 390 | 698 | 1 088 |

4.4 全局特征选择

如图 5 所示,比较了多种全局特征及其组合的性能,选择其中性能最优者(即颜色矩)。比较常见的 3 种全局颜色特征的性能,选择性能最好的颜色矩与其他 3 种全局纹理特征进行前融合。

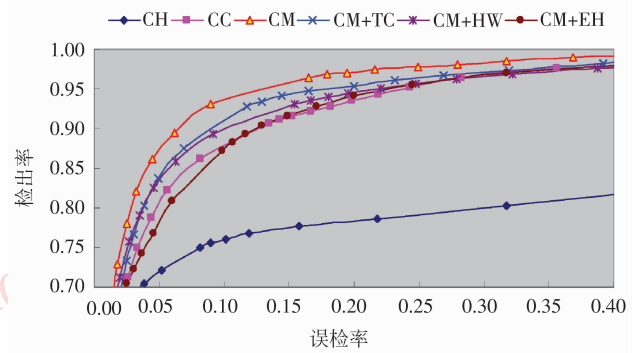


图 5 多种全局特征的 ROC 曲线
Fig. 4 ROC curves based on different global features

4.5 融合全局特征与 BoW 的性能

比较颜色矩与图 4 中 SURF-VAMAI 的多种融合方法,“平均法”后融合的 ROC 曲线最佳。图 6 中,标记三角形的曲线就是 CM 与 SURF-VAMAI “平均法”后融合的结果,标记方框的曲线是 Zheng 等人方法^[9]的结果,标记圆圈的曲线表示 CH + EH 的 SVM 分类结果。其他点表示了部分相关工作的性能。由图 6 可知,本文方法性能最优越。

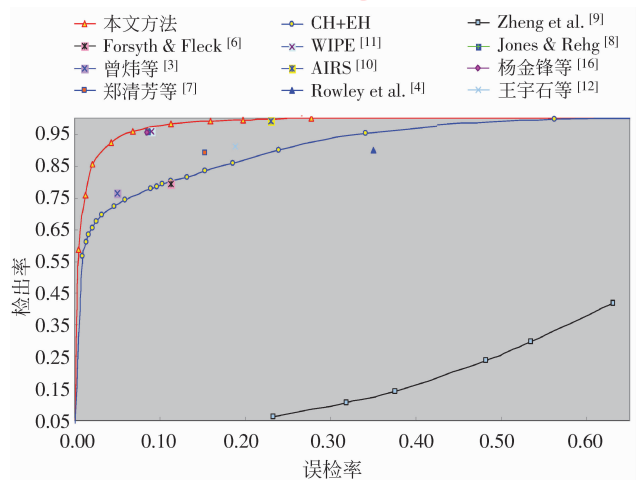


图 6 本文方法的敏感图像检测性能
Fig. 6 The performance of our method

5 结论

如何准确地检测敏感部位的所在区域,如何从检测速度与检测精度两方面提高敏感图像检测性能,是本文的研究重点。为了快速准确地检测敏感图像,提出基于视觉注意模型 VAMAI 的敏感图像检测方法。它包括 3 个方面:

1) 构造面向敏感图像的视觉注意模型 VAMAI,

用于较准确地提取 ROI;2) 基于 ROI 和 SURF 的视觉词表算法能够显著提高敏感图像检测速度与检测精度;3) 融合颜色矩与 BoW 的 SVM 分类结果, 进一步提高检测精度。实验结果表明, 本文方法从检测速度和检测精度两方面显著地提高了敏感图像检测性能。当然, 本文方法仍存在不足。对于“敏感部位或敏感姿态”所在区域, VAMAI 的查准率较高, 但查全率偏低。下一步工作将力图提高 VAMAI 的查全率。

参考文献 (References)

- [1] Zheng Qingfang. Research on Key Techniques for Image Retrieval and Filtering in Image Management [D]. Beijing: Institute of Computing Technology of Chinese Academy of Sciences, 2007. [郑清芳. 图像管理中检索与过滤的关键技术研究[D]. 北京: 中国科学院计算技术研究所, 2007.]
- [2] Deselaers T, Pimenidis L, Ney H. Bag-of-visual-words models for adult image classification and filtering [C] // Proceedings of the 19th International Conference on Pattern Recognition. Tampa, USA: University of South Florida, 2008: 1-4.
- [3] Zeng W, Gao W, Zhang T, et al. Image guarder: an intelligent detector for adult images [C] // Proceedings of the 6th Asian Conference of Computer Vision. Jeju Island, Korea: Asian Federation of Computer Vision Society, 2004: 198-203.
- [4] Rowley H A, Yushi J, Baluja S. Large scale image-based adult-content filtering [C] // Proceedings of the 1st International Conference on Computer Vision Theory and Applications. Berlin: Springer, 2006: 290-296.
- [5] Tang S, Li J, Zhang Y, et al. PornProbe: an LDA-SVM based pornography detection system [C] // Proceedings of the 17th ACM International Conference on Multimedia. New York: ACM Press, 2009.
- [6] Forsyth D A, Fleck M M. Body plans [C] // Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society Press, 1997: 678-683.
- [7] Zheng Q F, Zeng W, Gao W, et al. Shape-based adult images detection [C] // Proceedings of the 3rd International Conference on Image and Graphics, Washington, DC: IEEE Computer Society Press, 2004: 150-153.
- [8] Jones M J, Rehg J M. Statistical color models with application to skin detection [J]. International Journal of Computer Vision, 2002, 46(1): 81-96.
- [9] Zheng H. Maximum Entropy Modeling for Skin Detection: With an Application to Internet Filtering [D]. France: Univeristé des Sciences et Technologies de Lille, 2004.
- [10] Yoo S J. Intelligent multimedia information retrieval for identifying and rating adult images [C] // Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems. Berlin: Springer-Verlag, 2004: 164-170.
- [11] Wang J Z, Li J, Wiederhold G, et al. System for screening objectionable images [J]. Computation Communication, 1998, 21: 1355-1360.
- [12] Wang Yushi, Li Yuanning, Gao Wen. Detecting pornographic images with visual words [J]. Transactions of Beijing Institute of Technology, 2008, 28(5): 410-413. [王宇石, 李远宁, 高文. 基于局部视觉单词分布的敏感图像检测[J]. 北京理工大学学报, 2008, 28(5): 410-413.]
- [13] Wang Yushi, Gao Wen. Pornographic image detection based on visual words [J]. Chinese High Technology Letters, 2009, 19(10): 1041-1047. [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.]
- [14] Xu Y, Li B, Xue X Y, et al. Region-based pornographic image detection [C] // Proceedings of the 7th IEEE Workshop on Multimedia Signal Processing. [s. l.]: IEEEExplore Digital Library, 2005: 1-4.
- [15] Shih J, Lee C, Yang C. An adult image identification system employing image retrieval technique [J]. Pattern Recognition Letters, 2007, 28: 2367-2374.
- [16] Yang Jinfeng, Fu Zhouyu, Tan Tieniu, et al. A novel algorithm for content-based image recognition and filtering [J]. Journal of China Institute of Communication, 2004, 25(7): 93-106. [杨金锋, 傅周宇, 谭铁牛, 等. 一种新型的基于内容的图像识别与过滤方法[J]. 通信学报, 2004, 25(7): 93-106.]
- [17] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
- [18] Ma Y F, Zhang H J. Contrast-based image attention analysis by using fuzzy growing [C] // Proceedings of the 11st ACM International Conference on Multimedia. New York: ACM Press, 2003: 374-381.
- [19] Bay H, Tuytelaars T, Gool L V. SURF: speeded up robust features [C] // Proceedings of the 9th European Conference on Computer Vision. Berlin: Springer, 2006: 404-417.
- [20] Juan L, Gwon O. A comparison of SIFT, PCA-SIFT and SURF [J]. International Journal of Image Processing, 2009, 3(4): 143-152.
- [21] Garcia C, Tziritas G. Face detection using quantized skin color regions merging and wavelet packet analysis [J]. IEEE Transactions on Multimedia, 1999, 1: 264-277.
- [22] Viola P, Jones M J. Robust real-time face detection [J]. International Journal of Computer, 2004, 57(2): 137-154.
- [23] Jurie F, Triggs B. Creating efficient codebooks for visual recognition [C] // Proceedings of the 10th IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society Press, 2005.
- [24] Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification [C] // Proceedings of the 9th European Conference on Computer Vision. Berlin: Springer, 2006: 490-503.