

# 疑似色情视频中卡通视频的过滤方法

徐杰<sup>1,2,3</sup>, 张勇东<sup>1,2</sup>, 张冬明<sup>1,2</sup>, 汪文英<sup>1,2,3</sup>

(1 中国科学院计算技术研究所 虚拟现实技术实验室, 北京 100190)

(2 中国科学院计算技术研究所, 北京 100190)

(3 中国科学院研究生院, 北京 100049)

**摘要:** 在基于肤色的互联网色情视频检测研究中发现, 误检的色情视频中卡通视频占很大比例。针对这种情况, 本文提出了综合利用卡通视频及色情视频的特征, 过滤疑似色情视频中卡通视频的方法。通过分析卡通视频和非卡通视频在视觉方面的差异以及卡通视频和色情视频中主要对象在姿态上的差异, 利用图像颜色直方图、图像梯度、Daubechies 小波和矩构造特征向量, 通过支持向量机来过滤疑似色情视频中的卡通视频。实验证明, 该方法可以有效的检出疑似色情视频中的卡通视频, 显著降低色情视频检测系统的误检率。

**关键词:** 视频检测; 卡通; Daubechies 小波; 矩; SVM

## Techniques for Filtering Cartoon Video from Suspected Pornographic Video

Jie Xu<sup>1,2,3</sup>, Yongdong Zhang<sup>1,2</sup>, Dongming Zhang<sup>1,2</sup>, Wenying Wang<sup>1,2,3</sup>

(1 Virtual Reality Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

(2 Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

(3 Graduate School of the Chinese Academy of Sciences, Beijing 100049)

**Abstract:** In the study of Internet pornographic video detection based on skin, we found that cartoon video account for a large proportion in the misjudged pornographic video. In view of this situation, an approach which utilizes the comprehensive features of cartoon video and pornographic video to filter cartoon video from suspected pornographic video was proposed. By analyzing the difference between cartoon video and non-cartoon video in the vision and gesture of main targets, we utilize color histogram, image gradient, Daubechies wavelet and moments etc. to construct feature vector. Then we use SVM (Support Vector Machine) to filter cartoon video from suspected pornographic video. Experiments show that our method can significantly reduce the rate of false positives of pornographic video detection system.

**key words:** Video Detection; Cartoon; Daubechies Wavelet; Moments; SVM

## 1 引言

### 1.1 研究背景及研究现状

在基于肤色的互联网色情视频检测的研究中发现, 误检的色情视频中卡通视频占很大的比例 (图 1)。这是由于卡通视频中卡通人物皮肤区域以及一些背景的饱和度, 色调以及纹理都跟人体的肤色非常相似, 基于肤色的色情视频检测技术很难将卡通人物跟人体区分开来。所以卡通视频的检测, 对提高色情视频检测系统的查准率起着非常重要的作用。

---

基金资助: 973 项目资助, 项目编号 2007CB311100

863 项目资助, 项目编号 2007AA01Z416

作者简介: 徐杰 (1982-), 男, 山西, 博士研究生, Email: xujie@ict.ac.cn

目前,国内外对卡通视频检测的研究并不是很多。Roach[1]等提出了利用运动特征进行卡通视频检测,但是卡通视频帧的运动特征与动画制作的方法有关,卡通制作的方法越来越先进,现代卡通视频的运动特征跟传统卡通视频有很大的差别,这种方法有着一定的局限性。Ianeva[2]等给出了卡通视频的定义,认为卡通视频颜色简单强烈,种类少,局部区域往往纹理简单,颜色一致,并被黑边包围,而且卡通视频中有着很多的文字信息,在此定义的基础上,提取了饱和度等一系列特征进行检测。但是现在的卡通种类越来越多,越来越趋近于真实世界,这样的一些特征在很多卡通视频中体现不出来,所以此方法不具有普遍适用性。Glasberg[3]等将纹理、颜色以及运动特征结合起来构成特征向量,各分量之间的权重用非线性的sigmoid函数构建并用多层感知器来对MPEG-2视频进行分类,对卡通视频能达到80%的查准率。高新波[4]等利用MPEG-7描述子等8组视觉特征来构造405维的卡通视频的特征空间,用带有主动相关反馈技术的支持向量机(SVM)进行分类,其存在的缺点是特征维数太高,特征之间有冗余,一些特征在很多卡通视频中得不到体现,起不到应有的作用。[5][6]中的方法都是结合了音视频进行卡通视频检测,并不适合于基于肤色的色情视频检测系统。

卡通视频具有多样性而且制作卡通的技术不断发展,越来越逼近于真实的世界。例如传统的特别是国产的卡通视频中人物、物体有着明显的黑色轮廓,但是现代电脑制作的卡通一般都没有黑色边缘了;传统的卡通往往颜色鲜艳,饱和度高,但是像日本漫画家宫崎骏的电影卡通片的饱和度却低于一般的自然视频;迪斯尼系列的卡通电影的饱和度,人物形态介于传统卡通和自然视频之间[4],甚至有的很难跟自然视频区分开来,这样更加造成了卡通视频的检测难度。通过以上的分析,可以看出很难用统一有效的特征来描述多种多样的卡通视频。



图1 误检的色情视频

Fig.1 False positive pornographic videos

## 1.2 本文所做的工作

本文的工作重点在于将包含有类似人体肤色的卡通视频与色情视频实时区分开来,从而降低基于肤色的色情视频检测系统的误检率。

基于1.1的分析,没有统一有效的特征来对卡通视频帧进行描述,所以本文提出了利用卡通视频的视觉特征和视频中主要对象的姿态特征来构造特征向量,利用SVM分类器来过滤出疑似色情视频中的卡通视频。

一般认为,传统的卡通视频颜色强烈,饱和度高,而且卡通人物制作简单,肤色区域往往采用大面积均匀着色,在肤色局部区域,图像梯度无变化。将颜色直方图、彩色程度和肤色梯度结合起来,可以检出传统的、饱和度比较高的、色彩强烈的并且含有明显肤色区域的卡通视频。

迪斯尼风格的卡通跟真实世界非常相似。这种类型的疑似色情卡通视频,在视觉上跟色情视频的区分度比较小,难以利用[2]中定义的卡通的特征进行检测。所以,本文提出了利用Daubechies小波和矩对视频帧中的主要对象提取形状特征,利用主要对象的姿态来将卡通视频跟色情视频区分出来。

经过试验,本文提出的方法可以有效的提高基于肤色的色情检测系统的性能。

## 2 HSV颜色直方图

HSV 模型是 1978 年被 Alvy Ray Smith 创造定义的一个色彩空间，由以下三部分组成（图 2）：

Hue 色相：例如红、黄、蓝、橙、绿、紫等颜色，范围变化由 0-360 度。Saturation 饱和度：范围从 0-100% (0.1-1.0)，S=1 色彩最纯净（不含任何白光），色彩也最鲜艳。Value 值：范围从 0-100% (0.0-1.0)，V=0 为最黑（R、G、B=0），V=1 为最亮（R、G、B=255）。

HSV 颜色空间比 RGB 颜色空间更符合人类的眼睛视觉系统，所以应用在许多图像处理算法中。

本文利用 [8] 中的方法对视频帧进行非均匀量化，得到 36 维的特征向量。

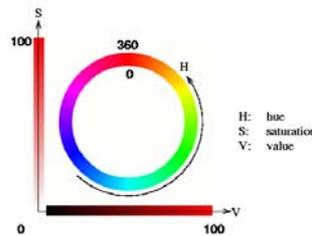


图 2 HSV 颜色空间

Fig.2 HSV Color Space

## 3 图像的彩色程度

一般认为，传统卡通视频的色彩强烈。本文利用 [4] 中提出的方法来表征图像的彩色程度。

在 RGB 空间中，对于图像中的任意像素点  $p$ ，定义： $p_{rg} = R - G$ ， $p_{yb} = 0.5(R + G) - b$ 。对于一幅图像  $I$ ，定义如下参量：

$$\begin{aligned} u_{rg} &= \frac{1}{|I|} \sum_{p \in I} p_{rg}, u_{yb} = \frac{1}{|I|} \sum_{p \in I} p_{yb} \\ \sigma_{rg}^2 &= \frac{1}{|I|} \sum_{p \in I} (p_{rg}^2 - u_{rg}^2), \sigma_{yb}^2 = \frac{1}{|I|} \sum_{p \in I} (p_{yb}^2 - u_{yb}^2), \\ \sigma_{rgyb} &= \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2}, u_{rgyb} = \sqrt{u_{rg}^2 + u_{yb}^2} \end{aligned} \quad (1)$$

图像的彩色程度指标用下式表示：

$$M = \sigma_{rgyb} + 0.3u_{rgyb} \quad (2)$$

## 4 图像中肤色区域的梯度

对于一个向量场  $T(x, y, z)$ ，在三维直角坐标系中，在点  $(x, y, z)$  处的梯度定义为：

$$\text{Grad}T(x, y, z) = \nabla T = \left( \frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}, \frac{\partial T}{\partial z} \right) \quad (3)$$

本文利用图像中某点周围 4 个方向上像素点像素值的差分来近似不同方向的偏导数，从而计算出梯度向量  $\text{Grad}T$ 。

设图像  $I$ ,  $p$  点的坐标为  $(x, y)$ ,

$$\text{Grad}T = (I(x-1, y+1) - I(x+1, y-1), I(x-1, y-1) - I(x+1, y+1), I(x-1, y) - I(x+1, y), I(x, y+1) - I(x, y-1)) \quad (4)$$

用模  $|\text{Grad}T|$  来反映点  $p$  处的亮度变化幅度。

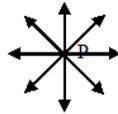


图 3  $p$  点处的梯度

Fig.3 The gradient of Point  $p$

卡通人物的皮肤往往是大面积均匀着色, 在比较小的区域内, 梯度模值几乎一样。而现实世界中人体皮肤区域的梯度随光照而变化。

在本文中, 在肤色掩码图像上, 构建  $16 \times 16$  的网格, 求网格交点处肤色点的梯度的模, 然后取均值(图 4)。

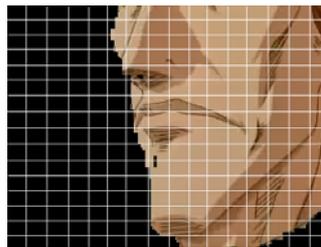


图 4 求肤色掩码图像梯度模的均值

Fig.4 The average gradient of skin mask

## 5 Daubechies小波

Daubechies 小波已被广泛应用于信号处理、图像压缩、语音识别、微分方程求解等众多领域。[7]中给出了 Daubechies 小波的详细定义和性质。本文主要利用 Daubechies 小波在边缘提取中的优良特性, 来降低背景噪音的影响, 从而来提取出图片中对象轮廓的大致走势(图 5)。



图 5 利用小波提取卡通视频的边缘

Fig.5 Edge extracted by wavelet

## 6 矩

矩是基于区域的物体形状表示方法。 $f(x, y)$  是  $xy$  平面上的连续函数,  $(p+q)$  阶矩定义为:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy, \quad p, q \in N \cup \{0\} \quad (5)$$

中心矩定义为:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(x - \frac{m_{10}}{m_{00}}\right)^p \left(y - \frac{m_{01}}{m_{00}}\right)^q f(x, y) dx dy \quad (6)$$

对于离散的图像, 中心矩定义为:

$$\mu_{pq} = \sum_x \sum_y \left(x - \frac{m_{10}}{m_{00}}\right)^p \left(y - \frac{m_{01}}{m_{00}}\right)^q f(x, y) \quad (7)$$

对该中心矩进行标准化操作:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \gamma = \frac{p+q+2}{2} \quad (8)$$

对于这些矩, [9]提出了一系列分别具有变换、旋转和缩放无关性的 7 个矩:

$$\begin{aligned} \phi_1 &= \mu_{2,0} + \mu_{0,2} \\ \phi_2 &= (\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2 \\ \phi_3 &= (\mu_{3,0} - 3\mu_{1,2})^2 + (\mu_{0,3} - 3\mu_{2,1})^2 \\ \phi_4 &= (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{0,3} + \mu_{2,1})^2 \\ \phi_5 &= (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2] \\ &\quad + (\mu_{0,3} - 3\mu_{2,1})(\mu_{0,3} + \mu_{2,1})[(\mu_{0,3} + \mu_{2,1})^2 - 3(\mu_{3,0} + \mu_{1,2})^2] \\ \phi_6 &= (\mu_{2,0} - \mu_{0,2})[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{0,3} + \mu_{2,1})^2] \\ &\quad + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{0,3} + \mu_{2,1}) \\ \phi_7 &= (3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2] \\ &\quad + (\mu_{3,0} - 3\mu_{2,1})(\mu_{0,3} + \mu_{2,1})[(\mu_{0,3} + \mu_{2,1})^2 - 3(\mu_{3,0} + \mu_{1,2})^2] \end{aligned} \quad (9)$$

## 7 支持向量机

支持向量机 SVM (Support Vector Machines) 是根据统计学理论的 VC 维 (Vapnik Chervonenks Deminsion) 理论, 在结构风险最小化基础上, 较好的解决小样本、分线性和高维数等实际问题。SVM 通过把数据映射到高维空间, 使本来在低维空间上线性不可分问题在高维空间上线性可分, 从而找到最优分类面。

SVM 算法的输入是一个集合:

$$T = \{(x_i, y_i) \mid x_i \in R^d, y_i = \pm 1, \text{其中 } i = 1, 2, \dots, n\} \quad (10)$$

其中  $y_i$  是样本  $x_i$  的所属的标志,  $d$  是样本的维数,  $n$  是训练样本的个数。

SVM 首先将输入空间  $R^d$  中的元素  $x$  依映射  $\Phi$  投影到一个特征空间  $\Gamma = R^m (m \geq d)$ , 对应元素  $z = \Phi(x)$ , 使训练样本在特征空间  $\Gamma$  中线性可分。SVM 的目标是在特征空间  $\Gamma$  中找到一个最优分类面  $o(x) = w^T z + b$ , 满足:

$$w^T z_i + b \begin{cases} \geq 1, y_i = 1 \\ \leq -1, y_i = -1 \end{cases}$$

$$\text{其中 } z_i = \Phi(x_i), \varpi = \sum_{i=1}^N \partial_i y_i z_i. \quad (11)$$

## 8 实验及实验结果分析

在实验中, 根据 2 对原始视频帧提取 36 维的特征向量; 根据 3 对原始视频帧提取 1 维特征向量; 经过肤色检测后, 根据 4 对肤色掩码图像提取 1 维的梯度特征; 通过对视频帧进行 Daubechies 小波边缘检测后, 根据 6 提取 5 阶以下中心矩和具有变换、旋转和缩放变换无关性的 7 个矩共 28 维特征向量。所有这些特征向量组合成 66 维特征向量, 作为 SVM 的输入。

训练测试的视频帧库包括 1687 幅视频帧, 其中包括 536 张卡通视频帧、1151 幅色情视频帧。为了减少由于数据分布不均匀可能造成的测试结果不准确, 采用了 8 次交叉实验。

首先将 536 幅卡通视频帧中的帧序号被 8 除余 1 的选出, 作为训练集合中的正例, 其余的作为测试视频帧集合的正例部分; 同理将 1151 幅色情视频帧中帧序号被 8 除余 1 的选出, 作为训练集合中的反例, 其余的作为测试视频帧集合的反例部分。这样就选出了第一组训练集合, 以及与其对应的测试集合。以此类推, 可以构造出由序号被 8 除余 2、余 3 等视频帧组成的训练测试集合对, 一共 8 组。

实验表明, 本文提出的方法能有效区分卡通视频和色情视频, 对卡通视频帧查全率达到 84.3%, 误检率平均为 4.5%。通过本文的方法, 可以有效的过滤疑似色情视频帧中的卡通视频, 显著降低基于肤色的色情视频检测系统的误检率。

### 参考文献:

- [1] Roach M, Mason J S, and Pawlewski M. Motion-based classification of cartoons. International Conference on Intelligent Multimedia, Video and Speech Processing, Hong Kong, May 2-4, 2001: 146-149
- [2] Ianeva T I, de Vries A P, and rohrig H. Detecting cartoons: A case study in automatic video-genre classification. Proc. IEEE International Conference on Multimedia and Expo. Baltimore, Maryland, Jul. 6-9, 2003, 1:449-452
- [3] R. Glasberg, K. Elazouzi and T. Sikora. Cartoon-Recognition using Visual-Descriptors and a Multilayer-Perception, WIAMIS, Montreux, April 13-15, 2005
- [4] 高新波, 田春娜, 张娜. 一种基于 SVM 主动学习的卡通视频检测方法. 电子与信息学报, 2007, Vol. 29 No. 6
- [5] L.Q. Xu, Y.Li, Video classification using spatial-temporal features and PCA, IEEE ICME V3, pp. 485-8, 2003
- [6] Antonio Rama, Rancesc Tarres, Laura Sanchez. Cartoon detection using fuzzy integral, IEEE WIAMIS, 2007
- [7] Ingrid Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, 1992
- [8] Zhang L, Lin F Z, and Zhang B. A CBIR method based on color-spatial feature. Proceedings of the IEEE Region 10 Conference on TENCON. Korea. Sept. 15-17, 1999
- [9] M. K. Hu, Visual Pattern Recognition by Moment Invariants, IRE(IEEE) Trans. Info. Theory, Vol. IT-8, pp. 179-187