基于隐马尔可夫模型的色情镜头建模及其应用*

刘安安[1,2,3]，宋砚[2,3]，李锦涛[2,3]，张勇东[2,3]，张冬明[2,3]，杨兆选[1]

（1 天津大学电子信息工程学院，天津，300072）

（2 中国科学院计算技术研究所 虚拟现实技术实验室，北京 100910）

（3 中国科学院计算技术研究所 智能信息处理重点实验室，北京 100910）

摘　要: 本文提出了一种新颖的基于隐马尔可夫模型的色情镜头建模方法及其应用。通过对图像中肤色区域的检测，特征的提取，我们实现了对图像的表征，然后通过隐马尔可夫模型充分利用视频的时序特性，建立了色情镜头模型。在此基础上，我们介绍了该模型在视频色情内容过滤上的应用。大量实验证明了该模型的准确性和鲁棒性。

关键词: 隐马尔可夫模型 (HMM)，混合高斯模型（GMM），色情，过滤

HMM Based Sex Shot Modeling and its Application

Anan Liu[1,2,3], Yan Song[2,3], Jintao Li[2,3], Yongdong Zhang[2,3], Dongming Zhang[2,3], Zhaoxuan Yang[1]

(1 Department of Elevtronic Engineering, Tianjin University, Tianjin, 200072)

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

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

Abstract: This paper presents an innovative Hidden Markov Model (HMM) based sex shot modeling method and its application. Firstly, we implement the detection of skin regions and visual feature extraction on video frames. Then we found the sex shot model based on HMM considering the sequence characteristics of video. Furthermore, we introduce the application of the model in sex content filter for videos. The experiment results demonstrate that the accuracy and robustness of the model.

Keywords: Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Sex, Filter

随着互连网的广泛应用，用户在获得大量有用信息的同时也可能遇到许多不良信息，其中尤其以色情视频最为严重。搜索引擎因其能够将大量成人网站地址提供给用户而恶化了这一现象。因此，进行色情信息的检测和过滤有重要的意义。

传统的色情视频检测主要存在两种过滤方式。第一种是通过检测色情网站的网址进行封堵。但是，由于色情网站的网址通常是变化的，该系统需要大量的人力来搜集色情网站的网址，所以网址封堵的方式并不是一种可行的方法。另一种是通过视频相应的文本内容进行检测。尽管视频对应的文本内容分析方法更为准确，但是该方法对文本信息有很强的依赖性，而色情视频通常并没有相应文本介绍，因此该方法并不是可靠的方法。

基于以上分析，根据图片和视频本身视觉特征检测其内容是更可靠和可行的解决方式。WIPE 系统[1]通过预处理和特征匹配两步实现色情图片检测。Forsyth 等[2]研究人员通过肤色检测和人体形状匹配的方法进行检测。Bosson 等[3]建立的色情检测系统则在肤色检测的基础上分析了 K 近临分类器，多层感知分类器（MLP）和支持向量机分类器（SVM）等的性能，并提出 MLP 在检测中具有最好的分类效果。尽管目前已有大量的研究人员从事该方面研究,但是，已有工作主要集中在如何建立理想的肤色模型和准确检测色情图片，再通过检测视频是否包含一定数目的色情图片来判定视频是否包含色情内容，而没有人专门根据视

* 基金项目：973 项目(2007CB311100), 863 项目(2007AA01Z416), 国家自然科学基金项目 (60773056), 北京科技新星项目(2007B071), 中科院计算所知识创新项目(20076031).

频的特性直接检测色情视频。因此，目前主要存在如下三个关键的因素制约色情视频的检测：1. 人种的多样性，光线的变化等因素导致很难建立完善的肤色模型实现肤色点的准确检测；2. 人体姿态多样性，遮挡等外界因素使得难以实现色情图片的准确检测；3. 对于不同视频难于建立通用的规则通过色情图片的检测来实现色情视频的准确检测。针对现有问题，本文提出了一种新颖的基于隐马尔可夫模型的色情视频检测方法，用于对色情视频及视频中的色情镜头进行检测。该方法创新点在于，该方法不严重依赖于准确的肤色模型，准确的色情图像检测并且不需要复杂的肤色区域分割算法，充分利用了视频的时序特性，增强了算法的鲁棒性，提高了检测的准确率，为不良视频封堵提供技术基础。

本文主要包括如下内容：第 1 部分介绍了基于隐马尔可夫模型的色情镜头建模方法；第 2 部分详细介绍了实验结果；第 3 部分介绍了该模型在视频色情内容过滤中的应用；最后一部分总结全文。

1．基于隐马尔可夫模型的色情镜头建模

基于隐马尔可夫模型的色情镜头建模分为肤色区域检测，图像特征提取和模型建立三个部分，分别详细介绍如下。

1.1 肤色区域提取：

肤色区域即由代表肤色的象素点所组成的若干连通区域。因此，肤色区域的提取主要由肤色点提取实现。我们通过混合高斯模型建立了肤色模型和非肤色模型，并通过贝叶斯分类器进行检测。

1.1.1 模型建立

每个象素点的颜色可以被认为是一个观测 X，并且 X 可以被认为是由特定的单高斯概率密度函数集合 G={ $p_i(X;\theta)$，i=1,2,……}中有限个元素的加权平均组合而成，即象素点可以由混合高斯模型（GMM）表示。因此观测 X 的概率密度函数 $p(X;\phi)$ 可以表示为如下形式：

$$\begin{cases} p(X;\phi)=\sum_{i=1}^{q}\pi_i \bullet p_i(X;\theta)=\sum_{i=1}^{q}\pi_i \bullet p_i(X\mid i;\theta)=\sum_{i=1}^{q}\pi_i \bullet \frac{1}{(2\pi)^{d/2}\mid\sum_i\mid^{1/2}}\exp^{-\frac{1}{2}(X-\mu_i)^T(\sum_i)^{-1}(X-\mu_i)} \\ \sum_{i=1}^{q}\pi_i=1 \end{cases} \quad (1)$$

其中，π_i 为权重，θ 为由均值和协方差矩阵组成的参数向量。未知参数 π_i 和 θ 可以通过期望最大化算法（Expectation Maximization, 简记为：EM）[4]，利用训练集中标注的肤色点和非肤色点样本来计算得到。这样，我们通过混合 GMM 可以分别建立肤色模型 M_1 和非肤色模型 M_2。

1.1.2 肤色点检测

模型建立后，可以以最小损失原则通过贝叶斯分类器[5，6]实现肤色点检测，即象素点（X）分类（W_1：肤色类；W_2：非肤色类）。用 C_{ij} 表示实际 $X\in W_i$，判决 $X\in W_j$ 时的损失。当 $i=j$ 时，正确分类；否则，错误分类。用 $R_i(x)$ 表示判决 $X\in W_i$ 时的条件损失，则对于本文中象素点分类问题（$i=1,2$），$R_i(x)$ 计算如下：

$$\begin{cases} R_1(x)=C_{11}*p(w_1\mid X)+C_{12}*p(w_2\mid X) \\ R_2(x)=C_{21}*p(w_1\mid X)+C_{22}*p(w_2\mid X) \end{cases} \quad (2)$$

其中，$p(w_i\mid X)$ 表示 X 为类 i 的后验概率。因此，分类规则如下：

$$\begin{cases} R_1(x)<R_2(x)\Rightarrow X\in w_1 \\ R_1(x)>R_2(x)\Rightarrow X\in w_2 \end{cases} \quad (3)$$

结合（2）和（3），我们得到如下结果：

$$\begin{cases}(C_{12}-C_{22})*p(w_2\mid X)<(C_{21}-C_{11})*p(w_1\mid X)\Rightarrow X\in w_1\\(C_{12}-C_{22})*p(w_2\mid X)>(C_{21}-C_{11})*p(w_1\mid X)\Rightarrow X\in w_2\end{cases} \quad (4)$$

通过结合贝叶斯规则，

$$p(w_i\mid X)=p(X\mid w_i)*p(w_i)/p(X) \quad (5)$$

基于最小损失的贝叶斯决策表示如下：

$$\begin{cases}p(X\mid w_1)/p(X\mid w_2)>\tau\Rightarrow X\in w_1\\p(X\mid w_1)/p(X\mid w_2)<\tau\Rightarrow X\in w_2\end{cases}\ （其中，\ \tau=\frac{(C_{12}-C_{22})}{(C_{21}-C_{11})}*\frac{p(w_2)}{p(w_1)}）\quad (6)$$

其中，$p(X\mid w_i)$ 为肤色与非肤色的条件概率密度函数，即分别由 M_1 和 M_2 表示；$p(w_i)$ 为类 w_i 的先验概率；τ 表示可调阈值。由于正确分类的损失（C_{11} 和 C_{22}）通常被设定为 0，因此，阈值主要由 C_{12} 和 C_{21} 控制，在我们的实验中，根据大量实验，我们令 $C_{12}=C_{21}$，此时阈值为 1。

1.2 图像特征提取

将肤色点用“1”标记，将非肤色点用“0”标记，实现肤色区域提取，并得到原图像对应的二值图像。将此图像做连通区域分割[7]，得到原图像对应的二值模版图象 $Mask_1$。计算各个象素点对应的似然值 P：

$$P=\frac{P_{w1}}{P_{w1}+P_{w2}} \quad (7)$$

并将各个象素点对应 P 归一化作为该象素点的灰度值，得到原图像对应的灰度模版图像 $Mask_2$。我们提取如下图像纹理和形状特征来表征图像：

（a）整幅图像的平均肤色概率 $P_{\mathrm{Im}age}$：

$$P_{\mathrm{Im}age}=\frac{1}{M\times N}\sum_{j=1}^{N}\sum_{i=1}^{M}Gray(i,j) \quad (8)$$

其中，$Gray(i,j)$ 表示图像 $Mask_2$ 中位置 (i,j) 的灰度值，M 和 N 分别表示图像的宽和高。

（b）所有肤色区域内平均肤色概率：计算图像 $Mask_1$ 中灰度为“1”的点对应图像 $Mask_2$ 的象素点的灰度的平均值。

（c）所有肤色区域外平均肤色概率：计算图像 $Mask_1$ 中灰度为“0”的点对应图像 $Mask_2$ 的象素点的灰度的平均值。

（d）计算最大连通区域面积：计算图像 $Mask_1$ 中灰度为“1”的各个连通区域包含的象素点最多的区域所对应的象素点数。

（e）计算最大连通区域的平均肤色概率：计算（d）中得到的最大连通区域在图像 $Mask_2$ 对应象素点的平均灰度值。

（f）计算最大连通区域的中心的坐标 (r,c)：

$$\begin{cases}r=\frac{1}{A}\sum_{s\in R}r_s\bullet y_s\\c=\frac{1}{A}\sum_{s\in R}c_s\bullet y_s\\A=\sum_{s\in R}y_s\end{cases} \quad (9)$$

其中，y_s 表示各象素点灰度值，r_s 和 c_s 分别表示象素点 S 的横坐标和纵坐标，R 表示最大连通区域。

（g）计算最大连通区域的二阶矩：

二阶行矩：$\mu_{rr}=\frac{1}{A}\sum_{s\in R}(r_s-r)^2 y_s$ (10)

二阶列矩：$\mu_{cc}=\frac{1}{A}\sum_{s\in R}(c_s-c)^2 y_s$ (11)

二阶混合矩：$\mu_{rc}=\frac{1}{A}\sum_{s\in R}(r_s-r)(c_s-c)y_s$ (12)

用椭圆来拟合最大连通区域，计算该椭圆区域的长轴 l_{major}，短轴 $l_{\min or}$ 和长轴绕纵轴逆时针旋转的角度 θ，存在如下四种情况：

- 当 μ_{rc}=0，且 $\mu_{rr}>\mu_{cc}$ 时：

$$\begin{cases} l_{major}=4\sqrt{\mu_{rr}} \\ l_{\min or}=4\sqrt{\mu_{cc}} \\ \theta=-90^\circ \end{cases} \tag{13}$$

- 当 μ_{rc}=0，且 $\mu_{rr}\le\mu_{cc}$ 时：

$$\begin{cases} l_{major}=4\sqrt{\mu_{cc}} \\ l_{\min or}=4\sqrt{\mu_{rr}} \\ \theta=0^\circ \end{cases} \tag{14}$$

- 当 $\mu_{rc}\neq 0$，且 $\mu_{rr}\le\mu_{cc}$ 时：

$$\begin{cases} l_{major}=\sqrt{8((\mu_{rr}+\mu_{cc})+\sqrt{(\mu_{rr}-\mu_{cc})^2+4\mu_{rc}^2})} \\ l_{\min or}=\sqrt{8((\mu_{rr}+\mu_{cc})-\sqrt{(\mu_{rr}-\mu_{cc})^2+4\mu_{rc}^2})} \\ \theta=\arctan\dfrac{-2\mu_{rc}}{(\mu_{rr}-\mu_{cc})+\sqrt{(\mu_{cc}-\mu_{rr})^2+4\mu_{rc}^2}} \end{cases} \tag{15}$$

- 当 $\mu_{rc}\neq 0$，且 $\mu_{rr}>\mu_{cc}$ 时：

$$\begin{cases} l_{major}=\sqrt{8((\mu_{rr}+\mu_{cc})+\sqrt{(\mu_{rr}-\mu_{cc})^2+4\mu_{rc}^2})} \\ l_{\min or}=\sqrt{8((\mu_{rr}+\mu_{cc})-\sqrt{(\mu_{rr}-\mu_{cc})^2+4\mu_{rc}^2})} \\ \theta=\arctan\dfrac{(\mu_{rr}-\mu_{cc})+\sqrt{(\mu_{cc}-\mu_{rr})^2+4\mu_{rc}^2}_{rc}}{-2\mu} \end{cases} \tag{16}$$

（h）计算拟合椭圆的面积 A_e：

$$A_e=4\pi\sqrt{\mu_{rr}\mu_{cc}-\mu_{rc}^2} \tag{17}$$

1.3 色情镜头建模

色情镜头通常包含连续的带有大量肤色区域的图像，因此我们需要将图像特征和视频的时序特性结合起来，利用 HMM 建立色情镜头模型。

HMM[4]是一个双重随机过程，由两个组成部分：a.马尔可夫链：描述状态的转移，用转移概率描述；b.一般的随机过程：描述状态与观察序列间的关系，用观察值的概率描述。HMM 可以用一个五元组来表示：

$$\lambda = (N, M, \pi, A, B) \tag{18}$$

其中：N 为状态数目；π 为起始状态概率，$\pi = \{\pi_1, \pi_2, ..., \pi_N\}$；A 为状态转移矩阵 $A = \{a_{ij}\} = p(Q_t = j \mid Q_{t-1} = i)$；B 为状态输出概率 $B = \{b_j(o_t)\} = p(O_t = o_t \mid Q_t = j)$，是观测事件对应状态的概率分布，也就是对于一个状态输出每个观测事件的概率；M 为每个状态对应的观察事件数目。

HMM 可以用来对连续状态转化进行描述，我们通过训练样本，即给定的观测序列 $O = \{O_1, O_2, ..., O_T\}$ 通过 EM 算法得到该模型参数 λ，使得：

$$\lambda^* = \arg\max_{\lambda} P(O \mid \lambda) \tag{19}$$

这样，该 HMM 模型可以被用于描述给定类别的观测序列的状态转移关系。

2．实验结果

在模型建立过程中，我们采用了M. JONES[10]提供的数据集，人工标注了肤色点和非肤色点，并将每个象素点的RGB分量构成特征向量用于模型训练，对于每个GMM模型选用16个单高斯模型构成。

在色情镜头模型建立过程中，我们选取了10个色情视频片段（每个长2分钟），对于每个视频，提取各帧图像的上述视觉特征构成特征向量，并将一个视频中所有帧对应的特征向量构成码本用于色情镜头模型的建立。

在测试过程中，我们选取了20个色情视频片段和20个非色情视频片段，每个长约3分钟。对每个视频片段的连续帧提取上述特征组成观测序列作为HMM的输入，得到该镜头为色情镜头的概率值P，若P大于阈值Th，则该镜头被判定为色情镜头，否则为非色情镜头。通过大量实验，我们将Th选取为0.6，此时查全率和查准率均达到100%。作为对比实验，我们采用[3]中所提取的色情图片检测方法，并规定当被检测视频包含超过一定数目N的色情图片时将该视频判定为色情视频。实验中发现，当N=5时此方法结果最理想，但是此时查准率为82%，而查全率仅为45%。可见该方法具有很高准确性。此外，由于随机选取的测试视频完全独立于模型训练视频，所以该方法具有很高的鲁棒性。

3．应用

在视频结构化和色情镜头建模基础上，我们建立了视频色情内容过滤系统，如图 1 所示。

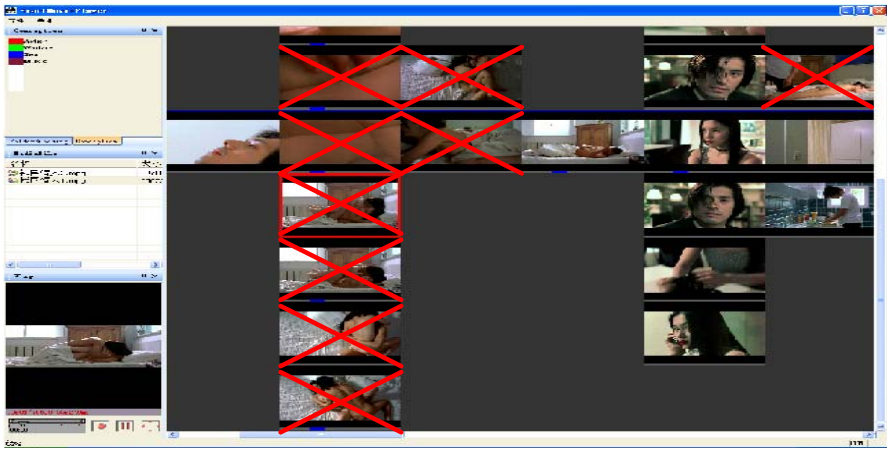

图 1. 视频色情内容过滤系统

对于输入视频，我们先进行镜头分割和关键帧提取[8]以及场景分类[9]，然后通过色情镜头模型对于不同镜头进行检测。如图 1 所示，每个视频用二维笛卡儿坐标系层次化展示，其中，水平方向表示视频场景的时序顺序，垂直方向为每个场景内镜头顺序。对于色情镜头，

我们用“叉”进行标注。这样，我们就实现了色情镜头的自动检测。在此基础上，我们支持各种视频封堵技术，从而实现视频色情内容过滤。

4．总结

本文提出了一种新颖的基于隐马尔可夫模型的色情镜头建模方法及其应用。通过对图像中肤色区域的检测，特征的提取，色情镜头模型的建立，我们实现了视频色情内容的检测。大量实验证明了该模型的准确性和鲁棒性。在次基础上，我们建立了视频色情内容过滤系统，为不良视频封堵提供技术基础。

参考文献：

1．James Ze Wang, Jia Li, Gio Wiederhold, and Oscar Firschein. System for screening bjectionable images. Images, Computer Communications Journal, 21(15):1355_1360, 1998.

2．M.M. Fleck, D.A. Forsyth, and C. Bregler. Finding naked people. In Proc. European Conf. on Computer Vision, pages 593_602. B. Buxton, R. Cipolla, Springer-Verlag, Berlin, Germany, 1996.

3．A. Bosson, G.C. Cawley, Y. Chan, and R. Harvey. Non-retrieval: blocking pornographic images. proceedings of the intl. conf. on image and video retrieval,london, uk, 2002. Lecture Notes in Computer Science, 2383:50_60, 2002.

4．模式分类。Richard O.Duda, Peter E.Hart, David G.Stork，机械工业出版社，2005。

5．Zarit B D。Super B J。Quek F K H，Comparison of five color models in skin pixel classification[A]．In：Proceedings of International W orkshop on Recognition， Amilysis and Tracking of Faces and Gestures in Real-Time Systems，ICCV’99[C]，Corfu，Greece，1999：58 ～63．

6．Vezhnevets V，Sazonov V，Andreeva A．A survey on pixel—based skin color detection techniques[A ．In；Proceedings of 13th International Conference of Computer Graphics and Visualization Graphicon 2003 [C]，Moscow，Russia，September，2003：85—92．

7．计算机视觉，贾云得，北京：科学出版社,2000。

8． Yueting Zhuang, Yong Rui, Thomas S. Huang et al. Adaptive key frame extraction using unsupervised clustering. Image Processing, ICIP 1998.

9．Zeeshan Rasheed, Mubarak Shah. Detection and Representation of Scenes in Videos. IEEE Transaction on Multimedia, Vol7, NO.6, December, 2005.

10．M. Jones, James M. Reng, Statistical Color Models with Application to Skin Detection, International Journal of Computer Vision 46(1), 81–96, 2002.