

像已经成了人们迫切需要解决的问题。

为了保护青少年的身心健康，同时也是为了维护网络的健康发展，不同的国家采取了不同的措施，并且颁布了相应的法律法规来规范网络运营商和网民的个人行为。2000年，美国通过了“儿童互联网保护法”的法案 CIPA (children Internet protection act)^[2]，将其加入宪法中，并于2001年4月生效，该法案要求美国的16000所公众图书馆和84000所学校图书馆必须在其计算机中安装有过滤软件来防止对色情等非法网站的访问，对于不履行该法案的图书馆将不对其进行财政拨款。2004年，为了从源头上制止不良内容的传播，建立我国互联网行业自律机制，规范从业者行为，我国制定了《中国互联网行业自律公约》，目前签约企业已超过1500家。2004年6月，“违法和不良信息举报中心”网站开通，从此揭开了我国围剿色情网站行动的序幕，并且取得了相当大的成绩。但是这一切并不意味着我们能够把网络色情轻易扫出互联网，我们应该清楚地认识到，由于互联网的全球性，国内的法律并不能约束其它国家和地区，因此必须辅以必要的技术手段进行互联网内容识别并过滤，尽可能给未成年人营造一个清洁的网络空间。不良图像识别技术是过滤网络中色情图像信息的关键技术，如何判断一幅图像是不是色情图像是一个新的课题，因而引起了人们的广泛关注。由于图像比文本具有更丰富的信息，因而，色情图像比色情文本信息的危害更大，以图像分析与图像理解技术为支撑的基于内容的不良图像识别技术正在成为色情信息过滤技术研究的一个重要方面。基于内容的不良图像的识别和检测技术近年来已引起人们的极大关注和兴趣。

不良图像信息识别就是判定一幅图像是否是具有不良信息的图像，不良图像信息识别技术主要是通过图像分析技术，将图像中的肤色区域分割出来。该技术的关键是如何准确的检测出肤色区域，以及判断该肤色区域是不是色情区域。因此，不良图像信息识别就是一个特殊的基于内容的图像识别技术。它是建立网络色情图像过滤系统的关键技术。本文中所指的不良图像是指现今人们常说的色情图片或成人图片等，不包括成人色情漫画和黑白图片。

1.2 基于内容的不良图像识别技术研究现状

为了保护青少年的身心健康，同时也是为了维护网络的健康发展，不少软件厂商针对网络色情泛滥这一问题开发了不同的软件。目前，已经开发出来的软件包有

CyberPatrol、NetNanny、CyberSitter, 美籍华人施颖裕推出的网络色情防范系统软件“网络色情锁”(X-sTOP)和美国 sun 微系统公司开发的“电子巡逻者”, 清华大学开发的“五行卫士”等。由于这些软件包中的算法不是基于 IP 地址的过滤就是基于网页中文本内容的判断, 并不能有效地识别检测出以图像方式来传播的色情信息。由于图像比文本具有更丰富的信息, 不良图像比不良文本的危害更大, 因此, 以图像分析与图像理解技术为支撑的基于内容的图像识别引起了广泛的研究与应用。下面就目前国内基于内容的图像识别技术的研究现状、应用与方向进行简单的介绍。

1. 2. 1 国外研究现状

由于国外网络起步较早, 因而基于内容的图像识别技术的研究也取得了一定的成绩, 下面介绍了一些国外开发的过滤系统以及采用的技术。

国外一些大学和研究机构在这方面进行了探索性的研究, 比较典型的有 1996 年美国 Berkely 大学的 Forsyth 等^[3]利用计算机视觉和图像理解技术对不良图像识别进行了研究, 通过对肤色分割和人体姿态的几何特征检测判别图像是否含有色情内容。其基本思想是将人体看作是按照一定规则的若干柱状区域的组合。Forsyth 算法主要由皮肤过滤器和人形分组器组成。其中, 皮肤过滤器主要是根据皮肤区域无纹理且属于一定颜色集合的特点, 采用基于颜色和纹理的特征抽取技术来确定图像中的皮肤区域。人形分组器主要是在皮肤区域内寻找柱状投影区, 并将柱状按人体肢体分组、连接并识别。实验表明, 该算法在具有 565 幅不良图像和 4289 幅正常图像的数据库中对于不良人体图像的识别率达 57%, 而对于正常图像的错分率为 3.4%。此外, 通常情况下一幅经皮肤过滤器确定为可疑图像的处理, 人形分组器需要执行 6 分钟, 由此可见, 该系统在以下方面还存在着较大的差距, 主要表现在仅能够处理单一类型图像, 缺乏适应性和通用性, 处理速度较慢, 图像的识别率较低, 准确性和智能性不高。

eVe (evision visual engine) 是 eVision 公司开发的图像、视频搜索引擎, 可用于不良图像信息的识别, 其实现过程分为三步: 1. 肤色过滤, 基于颜色和纹理特征判断图像中是否存在大的皮肤区域, 如果存在, 转第二步, 否则判定为正常; 2. 分割图像, 计算皮肤区域的视觉特征; 3. 特征比较, 计算待判定图像的视觉特征与预先求出的基准图像集的视觉特征的相似度, 如果有任一相似度超过 70, 则判定为不良; 如果有某一相似度在 40% -70%之间, 则暂时搁置, 由人工判定; 如果相似度均在 40%以