인터넷에서의 유해 이미지 컨텐츠 등급 분류 기법
(Classification Method of Harmful Image Content Rates in Internet)

남 택 용 † 정 치 윤 † 한 치 문 ††
(Taek Yong Nam) (Chi Yoon Jeong) (Chimoon Han)

요 약 본 논문은 인터넷 등을 통해 유입되는 유해 이미지를 그 특징을 이용하여 무해, 선정, 유해(누드), 심한 유해(성인물)과 같은 이미지 컨텐츠의 등급으로 선별하기 위한 이미지 특징 추출 방법과 이미지 분류 기술을 제시한 것이다. 이를 위해 본 논문에서는 입력 이미지에서 유해 정보임을 인식하기 위한 피부 영역 검출 기법을 제시한다. 또한, 노이즈를 줄이고 효과적으로 유해성 정도를 추출하기 위해 관심 영역을 설정하고 그 관심 영역 안에서만 특징을 정의하는 관심 영역 검출 알고리즘을 제안한다. 그리고 이미지를 4 종류의 등급으로 선별하기 위해 유해 이미지 분류 모델을 생성하는 다중 SVM 학습 기법과 생성된 분류 모델을 이용하여 입력 데이타의 유해 등급을 분류하는 다중 SVM 분류 기법을 제시한다. 특히 피부색 영역 이미지의 형태 정보와 피부색 비율 이미지의 색깔정보를 합하여 만든 피부색 가능성 분포 이미지를 제시하고, 이 피부색 가능성 분포 이미지를 축소하여 학습 과정에서 특징 분류를 위해 이용하는 이미지 특성 벡터를 제안한다. 마지막으로 본 논문에서 제안한 유해 이미지 등급 선별 기법을 적용한 실험 결과와 이미지의 유해 등급 분류에 대한 판별 성능을 평가한다.

키워드 : 유해 이미지, 이미지 컨텐츠 등급, 관심 영역, 피부색 검출, 다중 SVM

Abstract This paper presents the image feature extraction method and the image classification technique to select the harmful image flowed from the Internet by grade of image contents such as harmlessness, sex-appealing, harmfulness (nude), serious harmfulness (adult) by the characteristic of the image. In this paper, we suggest skin area detection technique to recognize whether an input image is harmful or not. We also propose the ROI detection algorithm that establishes region of interest to reduce some noise and extracts harmful degree effectively and defines the characteristics in the ROI area inside. And this paper suggests the multiple-SVM training method that creates the image classification model to select as 4 types of class defined above. This paper presents the multiple-SVM classification algorithm that categorizes harmful grade of input data with suggested classification model. We suggest the skin likelihood image made of the shape information of the skin area image and the color information of the skin ratio image specially. And we propose the image feature vector to use in the characteristic category at a course of training resizing the skin likelihood image. Finally, this paper presents the performance evaluation of experiment result, and proves the suitability of grading image using image feature classification algorithm.

Key words : harmful image, image content rate, region of interest, skin filtering, multiple support vector machine

1. 서 론

21세기를 살아가는 오늘날의 세계는 산업 사회에서 정보화 사회로 급속하게 패러다임의 이동이 진행되고 있다. 이 시대를 사는 우리는 인터넷을 통해 시간과 공간을 뛰어 넘어 언제, 어디서든 손쉽게 필요로 하는 정보를 획득할 수 있게 되었다. 하지만 인터넷은 유용한 정보와 손쉬운 활용이라는 순기능 뿐만 아니라 사회적으로 통제를 필요로 하는 유해한 정보 역시 인터넷을 이용하는 이용자들에게 무차별적으로 제공함으로써, 바람직스럽지 못한 역기능을 발생시키고 있다. 그 중에서 특히 유해 정보 (본 논문에서의 유해 정보는 음란성 정보를 말한다)는 사회적인 보호를 받아야 하는 청소년을

† 정 회 원 : 한국전자통신연구원 개인정보보호연구팀 연구원
tynam@etri.re.kr
iamready@etri.re.kr
†† 비회 원 : 한국외국어대학교 전자정보공학과 교수
cmhan@hufs.ac.kr
논문접수 : 2004년 11월 8일
심사완료 : 2005년 2월 17일

비롯하여 판단력과 절제력이 부족한 미성년 인터넷 이용자들에게 심각한 사회 문제를 야기시키고 있는 실정이다. 현재 유해 정보를 차단하기 위해서는 인터넷 상에 유해 정보를 가진 웹 사이트의 주소, 유해 단어 및 유해 단어를 대량으로 저장하고 있는 데이타베이스를 이용하여 청소년들의 유해 정보로의 접근을 차단하고 있다. 이렇게 고정된 데이타베이스를 이용하는 경우에는 새로운 유해 사이트, 유해 단어가 생겨 날 때 마다 데이타 베이스를 갱신해 주어야 한다. 하지만 인터넷은 동적인 공간으로서 하루에도 수많은 사이트들이 생성되고 사라지는 그 때마다 데이타 베이스를 갱신하는 것은 불가능한 일이며, 또한 최신 정보를 갖고 있는 데이타베이스를 실시간으로 일반 이용자 단말에서 이용하기에는 너무나 어려운 현실이다.

최근에는 이러한 유해 정보를 차단하기 위해서 웹사이트 검색의 주요 컨텐츠인 유해 이미지의 특성을 이용하는 방법이 대두되고 있다. 이 방법은 일반적인 이미지와는 특별하게 구별되는 유해성 이미지의 특징을 이용하여 유해 이미지 여부를 판별함으로써, 미성년자들이 유해 사이트나 웹 페이지로 접근하는 것을 차단할 수 있고 유해 스팸 메일 수신을 차단하는 연구가 이루어지고 있다. 현실 세계에서 웹사이트 내용의 유해 정도를 구분하는 최종 판단은 ICRA(RSACi)[1], Safenet[2]에서 명시하고 있는 인터넷 내용 등급의 기준에 따라 사람의 지각에 의해 판단한다. 그러므로 객관적이고 일관성이 있는 컨텐츠에 대한 유해 등급의 분류가 이루어지려면 컴퓨터의 기준에 적합한 기준을 정립하는 것이 중요하다. 그러나 컴퓨터에서 이미지를 분석하여 특성 차이에 따라 등급을 구분하는 것은 무척 어렵고 정확도에 있어서 신뢰할 수 있는 값을 얻어 낸다는 것은 매우 어려운 일이다. 특히 유해 이미지에 있어서 등급간의 구별 즉 무해, 선정, 유해(누드) 및 심한 유해(성인물) 등급간의 차이를 구분하기 위해서는 고난도의 이미지 처리 기법이 요구된다.

본 논문에서 다루고자 하는 연구 내용은 기존의 웹사이트, 스팸메일 등의 통신에서 이미지 컨텐츠의 유해 등급의 판별을 위해 사람의 지각에 의해 판단하는 문자적 등급 기준을 컴퓨터의 처리 기준에 적합한 이미지 컨텐츠의 유해 등급으로 분류하기 위한 이미지의 특징 추출과 분류 방법이다.

본 논문의 구성은 다음과 같다. 2장에서는 인터넷 내용 등급 표준인 PICS[3]와 RSACi의 등급 체계에 대해 살펴보고, 유해 이미지 처리에 연구되고 있는 이미지 기술을 간략하게 정리한다. 3장에서는 입력 이미지로부터 피부색을 검출하기 위한 방법을 제시하고, 유해 정보의 특징을 추출하기 위한 이미지 관심 영역 검출 기법을 제안한다. 4장에서 학습을 통해서 이미지를 4종류의 등급으로 선별하기 위해 이용될 유해 이미지 분류 모델을 생성하는 다중 SVM(Support Vector Machine) 학습 기법과 입력 이미지를 정해진 등급 중 하나로 분류하는 다중 SVM 분류 방법을 제시한다. 특히 피부색 영역 이미지와 피부색 확률 이미지의 정보를 합쳐서 만든 피부색 가능성 분포 이미지와 이를 축소하여 학습 과정에서 등급 분류에 이용할 이미지 특징 벡터를 제안한다. 5장에서는 유해 이미지 등급 선별 알고리즘을 유해 등급 선별 과정에 적용한 실험 결과와 이미지의 유해 등급 판별 성능을 평가한다.

2. 관련 연구

최근 인터넷의 역기능을 방지하려는 취지에서 유해 정보 방지 기술에 대한 연구 개발이 진행되어 왔다. 유해 정보 방지 기술은 내용 등급 분야로 대표되며, 내용 등급 표준인 PICS(Platform for Internet Content Selection)[3]을 기반으로 하는 연구 개발이 주를 이루고 있다.

PICS는 특정 웹 페이지의 내용에 관한 정보가 기술된 메타 정보를 컴퓨터의 소프트웨어를 통해 인식하고 선별할 수 있도록 하는 기술규격으로, 가장 일반적인 이용에서 볼 때, PICS는 주로 부모나 교사 등 청소년 보호자들이 미성년자의 인터넷 접속을 지도하고 통제하기 위한 자녀 통제 장치에 이용되어 왔다. RSACi, SafeSurf 등을 통해 잘 알려져 있는 인터넷 내용등급 시스템 이외에도, PICS는 디지털 라이브러리의 검색 시스템, 의료 정보등에 대한 평가 및 선별 시스템, 메타 검색엔진의 카테고리 분류 시스템 등에 이용되고 있다. 하지만, PICS의 아이디어가 구체적인 발전을 이루어낸 분야는 무엇보다 인터넷 내용등급 시스템이라고 할 수 있다. PICS가 동작되기 위해서는 특정한 등급 시스템이 만들어져야 하며, 이 등급 시스템은 특정 등급에 컨텐츠의 내용을 기술하기 위한 표준적인 분류어와 범주를 정의하고 있으며, 표1에 이러한 등급시스템의 표준이 되는 RSACi의 등급체계를 설명하고 있다.

컴퓨터로 유해 이미지 검출을 위한 연구가 진행되기 시작한 것은 웹이 일반화되기 시작한 90년대 중반부터이다. 이때에는 주로 이미지 내에 옷을 벗은 사람의 존재 여부에 대한 알고리즘들이 개발되었다[4,5]. 이러한 연구 내용은 먼저 피부 영역을 위한 효과적인 마스크를 색깔과 질감을 이용해서 만들고, 이 마스크를 통과한 피부 지역을 인간의 신체 모습에 부분적으로 비교하여 많은 부분이 매치가 되면 그만큼 노출 부분이 많다는 것을 유추함으로 유해 정도 즉, 옷을 벗은 정도를 측정한 것이다. 좀더 복잡한 유해 이미지 검출에 필요한 특징에

표 1 RSACi의 등급체계

범주 / 수준	신체노출(Nudity)	성행위(Sex)	폭력(Violence)	언어(Language)
Level 4	자극적인 정면 신체노출	노골적인 성행위 또는 성범죄	강간 또는 잔인하고 이유없는 폭력	노골적인 저속어 또는 극단적인 혐오성 언급
Level 3	정면 신체노출	노골적이지 않은 성행위	과격한 폭력 또는 인간의 죽음	폭언 또는 혐오성 언급
Level 2	부분적인 신체노출	착의 상태의 성적 접촉	실제 대상물에 대한 파괴	심하지 않은 비속어 또는 불경어
Level 1	노출 복장	격렬한 키스	인간에 대한 상해	가벼운 비속어
Level 0	위의 어느 것과도 관계없음	위의 어느 것과도 관계없음 또는 가벼운 키스(로맨스)	위의 어느 것과도 관계없음 또는 스포츠와 관련된 것	위의 어느 것과도 관계없음

대한 연구가 진행된 것은 칼라 히스토그램을 이용한 방법[6], MPEG7 기술자를 이용한 방법[7], 이미지 내용을 기반으로 하는 방법[8]과 같은 연구에서였다.

최근에는 인터넷의 급속한 발달에 따라 방대한 데이타를 손쉽게 얻을 수 있게 됨에 따라 통계적인 방법을 이용한 이미지 분류 방법이 많이 사용되기 시작했다. 특히 데이타 마이닝이라는 학문의 발달로 인해 이미지 마이닝이라는 새로운 분야가 생기면서 신경망을 이용하는 방법[9][10], 결정 트리 모델 방법[11], SVM을 이용한 방법[12]과 같은 이미지 분류에 대한 연구가 매우 활발히 진행되고 있다. 특히 SVM은 1979년 Vapnik이 두 가지 클래스를 가지는 객체를 구분하기 위해 제안된 후[13], 그 성능을 인정 받아서 숫자 인식, 물체 인식, 얼굴 검출, 유해 이미지의 분류[14] 등의 많은 분야에서 이용되고 있으나, 유해와 무해의 두 종류만을 판별하는 연구에 한정되어 있다.

본 논문에서는 표 1에서 정의된 RSACi의 등급 체계를 컴퓨터에서 이미지 컨텐츠를 자동 분류하기 위한 판단 기준인 4종류의 등급인 무해(0등급), 선정(1등급), 유해(2등급), 심한 유해(3등급)으로 정의하고자 한다. 무해는 자연물이나 정상적인 의복을 갖춘 상태를 말하며, 선정은 비키니, 속옷 등을 갖추었으나 노출 정도가 선정적으로 보이는 상태를 말하며, 유해는 상반신을 의복이 없이 노출 시킨 누드를 말하며, 심한 유해는 의복을 갖춘 상태를 구분하기 어려운 상태(하반신 노출, 전신 노출)과 성인물을 말한다.

3. 이미지의 피부 영역 검출 및 관심 영역 검출 기법

본 장에서는 이미지 컨텐츠의 등급 분류를 위한 전처리 단계로서, 피부색을 특징으로 하는 유해 이미지의 피부색 검출을 위한 모델을 정의하고 알고리즘을 제시하고자 한다. 피부색이라는 컬러 모델만을 사용하는 경우 판별 오류가 많이 발생할 수 있기 때문에 사람 피부의 균일한 성질을 이용하여 피부를 검출하기 위한 피부 영역 기반의 피부 검출 알고리즘을 제시하고, 피부 영역의 판별 오류를 최소화하기 위해 피부 영역을 제외한 배경 부분이나 노이즈를 제거함으로써 유해 이미지의 판별율을 높이는 관심 영역 검출 알고리즘을 제안한다.

이미지를 처리하여 정해진 등급으로 구분하기 위해서는 먼저 입력 이미지를 이미지 정규화, 노이즈를 제거하는 이미지 평활화(smoothing) 과정의 전처리하는 과정과 학습과정을 통해 생성한 이미지 특징 벡터를 가지고 전처리된 이미지를 분류하는 후처리 과정으로 나뉜다.

이미지의 유해 여부를 판단하기 위해서 입력 이미지가 들어오면 먼저 크기를 400x400으로 같게 해주는 이미지 정규화를 수행한다. 이 이미지 정규화 과정은 인터넷에 존재하는 이미지들의 크기가 매우 다양하기 때문에 입력 이미지의 크기가 커질수록 처리 시간이 늘어나는 것을 방지하기 위해 이용한다. 이 정규화 과정 후에는 이미지 상의 노이즈를 제거하기 위해 메디안 필터를 이용하여 이미지 평활화를 수행한 후, 피부색 검출을 수행한다.

3.1 영역 기반 피부색 검출 기법

이미지에서 피부색을 추출하기 위해서는 먼저 피부색을 찾을 컬러 스페이스와 피부색 모델을 정의해야 한다. 현재 피부색과 피부색이 아닌 경우를 분리하는 것은 컬러 스페이스와 무관하다고 알려져 있기 때문에 제안된 방법에서는 RGB 컬러 스페이스를 사용한다. 제안된 방법에서는 피부색과 피부색이 아닌 경우에 대해서도 가우시안 복합 모델을 적용한다. 피부색에 대해서만 가우시안 복합 모델을 이용하는 경우에는 사람의 피부색과 다른 경우에도 피부색으로 오검출하는 경우가 많이 발생하게 되므로, 피부색과 피부색이 아닌 경우의 확률 비율을 수식 (1)을 이용하여 피부색 여부를 판별한다.

$$P(c|skin) = \sum_{i=1}^{k} \pi_i \cdot p_i(c|skin)$$

$$where\ p_i(c|skin) = \frac{1}{2\pi|\sum_{si}|^{1/2}} \cdot \exp^{\frac{1}{2}(c-\mu_{si})\cdot\sum_{si}(c-\mu_{si})} \quad (1)$$

여기에서 c는 컬러 벡터를 의미하며, k는 가우시안 모델의 개수를 나타낸다. 제안된 방법에서는 16개의 가우시안 모델(k=16)을 사용한다. 가우시안 복합 모델에서 πi는 i번째 가우시안 모델의 가중치를 의미하며, μi, Σ i는 i번째 가우시안 모델의 평균과 공분산 행렬를 의미한다.

각 픽셀에서 피부색일 확률, 피부색일 아닐 확률을 가우시안 복합 모델을 통해서 계산 한 후, 수식 (2)로 피부색의 확률비를 계산 한다. 예를 들면 피부색 확률비를 1.0으로 할당하여 1.0 이상인 경우는 피부색일 확률이 크고, 1.0보다 작은 경우는 피부색이 아닐 확률이 높아지게 된다.

$$SkinRatio\quad(c) = \frac{p(c \mid skin)}{p(c \mid non - skin)} \qquad (2)$$

그림 1의 (b)는 입력 이미지 (a)를 고정 임계값을 이용하는 픽셀 기반 검출 방법의 결과를 보이고 있으며, 이미지의 상단과 하단 부분에 피부 영역이 아니지만 피부색으로 혼동할 수 있는 유사한 색 때문에 피부 영역으로 오인하여 검출된 것을 볼 수 있다. 이러한 원인은 피부색 확률비의 정보만을 이용하는 픽셀 기반 검출 방법의 단점으로, 이것을 해결하기 위해서는 주변 픽셀의 정보를 같이 이용함으로써 피부 영역의 균일한 특징을 반영하는 영역 기반의 검출 방법인 적응형 피부 분할 방법[15]을 적용한다. 적응형 피부 분할 방법은 피부색 확률비에 낮은 임계값을 적용하여 초기 피부색 영역을 추출 한다. 이 때 임계값이 낮기 때문에 대부분의 피부색 영역이 검출 되지만, 피부색이 아닌 영역도 다수 포함하게 된다. 사람의 피부색은 같은 색깔이 넓은 영역에 걸쳐서 존재하는 균일성을 가지고 있기 때문에, 피부색이 아닌 영역을 제거 하기 위해 추출된 피부색 영역에 대해서 균일성을 검사하게 된다. 이 때 균일성을 검사하는 기준으로는 수식 (3)과 같이 피부색 영역 내에서의 에지 픽셀의 개수, 영상 강도의 표준 편차(σ)가 이용한다.

$$(\sigma \le 40)\ \text{AND}\left\{\left(\frac{N_e}{N_d} \le 1.5\right)\text{OR}\left(\frac{N_e}{N_s} \le 0.02\right)\right\} \qquad (3)$$

N_e는 피부색 영역에서의 에지 픽셀의 개수, N_d는 피부색 영역을 포함하는 사각형의 최대 크기(높이 또는 너비), N_s는 피부색 영역의 픽셀 수를 의미한다. 피부색 후보 영역이 만약 균일하지 않다면, 피부색 후보 영역이 균일해 질 때까지 임계값을 높이는 방법이 이용한다. 그림 1의 (c)는 적응형 피부 분할 방법을 수행한 결과 피부색으로 오인된 이미지의 상단과 하단을 피부색이 아닌 것으로 표현된 것을 볼 수 있다.

3.2 관심 영역 검출 알고리즘

지금까지 유해 이미지 분류에 관한 연구는 이미지 전체 영역에서 유해 이미지의 특징을추출하였으나, 제안하는 방법에서는 관심 영역을 설정하고 관심 영역 안에서만 유해 이미지 분류 특징을 추출하는 방법을 제안한다.

(a) 입력 이미지

(b) 픽셀기반 피부색검출 (Sth=1.0)

(c) 적응형 피부 분할

그림 1 피부색 검출 결과

관심 영역을 추출하는 목적은 입력 이미지의 영역에서 피부색 영역이 얼마나 존재하는 지를 측정하여 판별 오류를 최소화 하고자 하는 것으로, 이 피부색 영역 부분만을 제외하고 배경 부분이나 노이즈 부분을 제거함으로써, 유해성 여부를 정확히 판별하고자 하는 것이다. 즉, 관심 영역 검출은 이미지의 특징을 추출할 영역을 설정하는 기능을 수행한다. 유해 이미지는 일반적으로 신체와 배경으로 구성되어 있으며, 유해 이미지의 전체 영역에서 특징을 추출하게 된다면 추출된 특징이 배경

(a) (b)

그림 2 관심 영역 검출 및 관심 영역 이미지

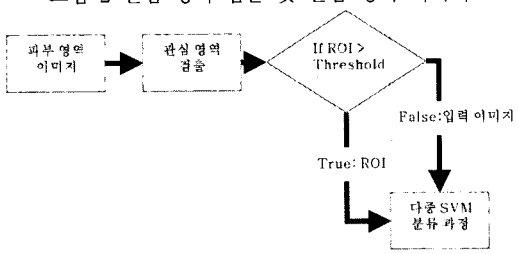


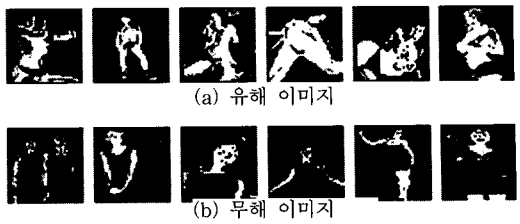
그림 3 관심 영역 검출 흐름도

영역의 영향을 받기 때문에 분류 성능을 떨어뜨리게 된다. 유해 이미지에서 중요한 것은 사람이 있는 영역이기 때문에 피부색 영역 이미지를 활용하여 피부색 영역을 포괄하는 최대 영역을 관심 영역으로 설정한다. 다음의 그림 2의 (a)에서 작은 상자는 각각의 피부색 영역을 의미하며, 가장 큰 상자는 각각의 피부색 영역을 모두 포함하는 최대 피부색 영역을 의미하며, 이 가장 큰 상자를 관심 영역으로 설정한다. 왼쪽 이미지의 경우 신체의 위쪽에 피부색으로 오인한 영역까지 포함하여 관심 영역이 설정되었으며, (b)의 경우 얼굴과 신체 두 부분의 피부색 영역을 포함하는 관심 영역이 설정된 것을 볼 수 있다.

이 때 전체 이미지에서 관심 영역이 차지하는 비율이 임계값(0.2)보다 적은 경우에는 관심 영역을 전체 이미지로 설정하게 하였다. 이것은 이미지에서 관심 영역이 너무 적은 경우에는 이미지의 특성을 제대로 반영하지 못하기 때문이다. 관심 영역에서 유해 이미지를 분류 할 수 있는 특징을 추출하게 되면 배경의 영향을 많이 받지 않기 때문에 유해 이미지의 특징을 잘 특성화 시킬 수 있다. 또한 이미지 전체 영역에 비해서 유해한 영역의 크기가 적을 경우에도 관심 영역에서 특징을 추출한다면 유해 이미지로 분류 할 수 있기 때문에 유해 이미지 분류 성능이 향상 된다. 관심 영역 검출 절차는 그림 3과 같이 구성된다.

4. 다중 SVM 기반 유해 이미지 등급 선별 방법

본 장에서는 유해 이미지가 어떤 등급 인지를 선별하는데 가장 중요한 요소인 이미지 특징 벡터를 생성하는 다중 SVM 학습 기법, SVM 분류 모델과 유해 이미지 컨텐츠 등급 선별 알고리즘을 제안한다.



(a) 유해 이미지

(b) 무해 이미지

그림 4 피부색 가능성 분포 이미지의 예

4.1 이미지 특성 벡터

지금까지 연구되어온 피부 영역에 대한 고차원의 특징들을 살펴보면 피부 영역의 개수, 이미지에서 피부 영역이 차지하는 비율 등 이었다는 것을 알 수 있다. 이러한 고차원의 특징들은 피부 영역에 대한 정보를 잘 이용할 수 있지만, 너무 단순하여 유해 이미지의 특징을 잘 반영하지 못하며, 피부색 정보에 많이 의존하고 있기 때문에 이미지에서 피부 영역이 잘못 검출된 경우와 피부 영역이 적게 존재하는 경우에 문제점이 발생한다.

이러한 문제점을 해결하기 위하여 유해 이미지의 형태를 이용할 수 있지만, 유해 이미지를 포함하는 복잡한 배경이 이미지 형태를 특성화 하는데 큰 장애 요인이 된다. 본 논문에서는 유해 이미지의 피부 영역 이미지들 간에는 형태의 유사성이 있고, 유해 이미지와 무해 이미지 간에는 형태의 피부영역 형태의 차이가 있음에 착안하여, 유해 이미지의 형태 정보를 특성화한 피부 가능성 분포 이미지를 생성한다. 그림 4는 피부색 가능성 분포 이미지의 예를 표현하고 있으며, (a)는 유해이미지를 (b)는 무해 이미지를 나타낸다. 그림 5의 (d)피부가능성 분포 이미지는 그림 5의 (b)피부색 분포 이미지와 그림 5의 (c) 피부 영역 이미지를 가지고 수식 (4)를 이용하여 구성하게 된다.

$$L(x,y)=\begin{cases}0, & \text{if } S(x,y)>0 \;\&\; R(x,y)\le r_0\\ \dfrac{255}{(r_1-r_0)}\cdot\big(R(x,y)-r_0\big), & \text{if } S(x,y)>0 \;\&\; r_0<R(x,y)<r_1\\ 255, & \text{if } S(x,y)>0,\ R(x,y)\ge r_1\end{cases} \tag{4}$$

수식 (4)에서 L(x,y)는 피부색 가능성 분포 이미지를 의미하고, S(x,y)는 피부색 영역 이미지를 의미하며 0 또는 255의 값을 가진다. R(x,y)는 피부색 분포 이미지를 의미하며, r_0와 r_1은 피부색 비율의 최대값과 최소값을 의미한다.

피부색 가능성 분포 이미지는 이미지에서 피부색 영역의 형태와 피부색일 확률 등의 정보를 가지고 있게 되지만, 차원이 너무 높기 때문에 크기를 줄여서 40x40

(a) 입력이미지 (b) 피부색분포이미지

(c) 피부영역이미지 (d) 피부색가능성분포이미지

그림 5 피부색 가능성 분포 이미지

의 형태로 만들었으며, 이를 이미지 특징 벡터라고 한다. 이미지 특징 벡터는 피부색에 대한 정보와 피부 영역의 형태에 대한 정보를 모두 가지고 있기 때문에 유해 이미지를 분류 하는 데 있어서 좋은 특징이 된다.

4.2 다중 SVM 학습 및 다중 SVM 분류

4.2.1 SVM 학습 방법

SVM은 범주를 알고 있는 n 개의 샘플이 주어질 때, (x_1,y_1), (x_2,y_2), $\cdots(x_n,y_n)$ $\in \chi \times \{\pm 1\}$, 샘플들을 분리할 수 있는 최적의 separating hyper-plane을 만들어 준다. 여기에서 x는 특성 벡터를 의미하며, y는 특성 벡터의 범주를 나타내며, 그 값은 +1 또는 -1값이다. SVM은 주어진 샘플들을 수식 (5)와 같은 매핑 함수를 이용하여 다음과 같은 특징 공간으로 매핑시키는 매핑 함수 Φ(xi)로 매핑 시킨다.

$$\Phi : \chi \rightarrow F, x \mapsto \Phi(x) \tag{5}$$

이때 특징공간 F는 고차원이기 때문에 특징 공간 상에서 범주를 분류해 주는 최적의 separating hyper-plane을 얻기 위한 비선형 결정 함수인 판별 함수를 수식 (6)과 같이 정의할 수 있다.

$$f(x)=\mathrm{sgn}\left(\sum_{i=1}^{n} y_i\alpha_i(\Phi(x)\bullet\Phi(x_i))+b\right)$$

$$=\mathrm{sgn}\left(\sum_{i=1}^{n} y_i\alpha_i k(x,x_i)+b\right) \tag{6}$$

매핑 함수의 $\Phi(x)\cdot\Phi(x')$는 계산량이 많기 때문에 특징 공간상에서 커널 트릭을 이용하여 간단하게 계산하며, SVM 커널 유형의 하나인 Radial Basis function을 이용하며, Radial Basis Function의 함수는 수식 (7)과 같다.

$$k(x,x')= \exp\{-g\cdot|x-x'|^2\} \tag{7}$$

그림 6은 특징 공간 상에서 데이타를 분류해주는 optimal separating hyper-plane예를 보이고 있다. 이때 두 개의 범주(class)의 각각의 경계선에 존재하는 샘플들을 support vector라고 한다.

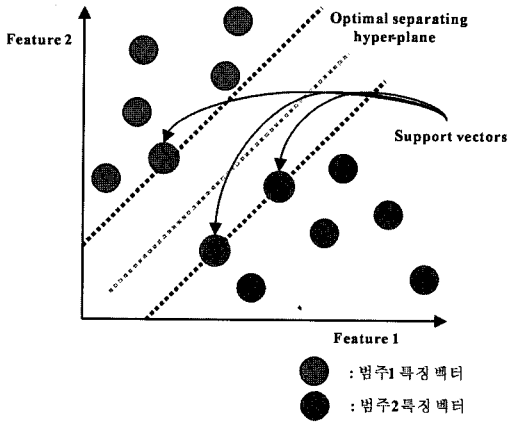


그림 6 특징 공간과 support vector

4.2.2 다중 SVM 분류 기법

one-against-one 방법을 사용하여 다중 SVM 분류 방법을 실현한다. One-against-one 방법은 각각 클래스와 클래스를 하나씩 비교하는 방법으로, C개의 클래스가 존재할 경우 학습 과정에서 C*(C-1)/2 개의 다음의 이진 분류기가 생성된다. 본 논문에서는 4종류의 등급을 대상으로 하게 되므로, 6개의 이진 분류기가 생성된다.

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \tag{8}$$

수식 (8)에서 상수 C는 0보다 커야 하며, 경험 오류와 복잡도 항의 트레이드 오프를 조절하는 역할을 한다. 학습 데이타의 i번째 등급과 j번째 등급에 대해서, 수식 (8)에서 정의한 식을 풀어서 수식 (9)와 같은 다중 SVM 분류 모델을 생성한다.

$$\min_{\omega^{ij},b^{ij},\xi^{ij}} \frac{1}{2}(\omega^{ij})^T\omega^{ij} + C\sum_t \xi_t^{ij}$$

$$(\omega^{ij})^T\phi(x_t)+b^{ij} \geq 1-\xi_t^{ij}, \ if\ y_t = i,$$

$$(\omega^{ij})^T\phi(x_t)+b^{ij} \leq -1+\xi_t^{ij}, \ if\ y_t = j,$$

$$\xi_t^{ij} \geq 0. \tag{9}$$

학습을 통해서 생성된 C*(C-1)/2 개의 분류기를 사용하여 분류를 할 때 제안하는 방법에서는 투표 방법을 사용한다. 투표는 만약 i 등급과 j등급을 분류하는 분류기에서 i번째 등급이라고 결정 한 경우, i번째 등급의 카운트를 하나 증가시키는 방법을 사용하였으며, 최종 분류 결과는 카운트가 가장 높은 등급으로 결정한다. 즉,

6개의 분류기 (0등급, 1등급), (0등급, 2등급), (0등급, 3등급), (1등급, 2등급), (1등급, 3등급), (2등급, 3등급)에서 분류기의 최대 카운트에 따라 등급을 결정하게 된다.

다중 SVM 분류기의 입력 벡터로는 SVM 학습에서 생성한 이미지 특징 벡터를 이용하며, 커널로는 radial basis function을 사용한다. 커널에 사용되는 파라미터들은 그리드 서치와 k-fold cross validation을 통하여 구한다.

4.2.3 그리드 서치를 이용한 다중 SVM 학습 기법

다중 SVM 학습에서는 학습에 이용하는 데이타의 한계를 극복하고 일반성을 확보하기 위해 k-fold cross-validation 기법을 이용한다. K-fold cross-validation 기법은 학습 과정에서 사용 가능한 데이타 수가 적을 경우 데이타를 k개의 서로소(disjoint) 집합으로 분할하고, 각각의 분할한 데이타의 집합에 대해 작업을 수행하는 것이다. 다중 SVM 학습 과정의 목적은 학습 데이타를 이용하여 구하고자 하는 등급 0부터 등급 3까지를 분류할 수 있는 최적의 분류 모델을 구하는 것으로, 그림 7에서 나타낸 것과 같이 그리드 서치 방법을 이용한다. 그리드 서치는 SVM 학습에서 최적의 파라미터를 찾기 위해서 이용하는 방법으로서, 파라미터의 값을 일정 간격으로 증가 또는 감소 시키면서 파라미터에 따른 분류 모델을 생성하고, 4.2.2의 다중 SVM 분류 모델과 같은 생성된 분류 모델을 사용하여 학습 데이타에 적용한 후, 정확도가 가장 높은 파라미터를 선택하는 방법이다.

상세한 그리드 서치의 과정은 다음과 같다.

- K-fold cross-validation 방법은 학습 데이타를 같은 크기를 가지는 k개로 나눈 후, 하나의 데이타 집합을 나머지 k-1개의 집합으로 학습시킨 분류 모델을 이용하여 오류를 측정하는 과정을 k번 반복한다.

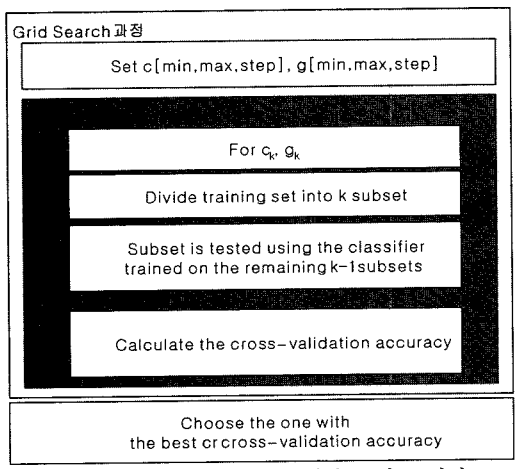


그림 7 다중 SVM학습을 위한 그리드 서치

- 그리드 서치와 K-fold cross-validation을 이용한 학습 파라미터 결정은 그리드 서치를 하는 중간 과정에서 분류 모델을 생성하여 정확도를 측정하는 과정을 K-fold cross-validation으로 대체한 것이다. 그리드 서치에서 최적값을 결정한 파라미터들의 최대, 최소값을 설정한 후, 비용(cost) 파라미터 c_k와 radial basis 커널인 g_k를 사용하여 k-fold cross-validation을 수행한다.
- 이 때 학습 데이타를 k개의 집합으로 나눈 다음, 하나의 집합을 제외하고 k-1개의 집합을 c_k, g_k 파라미터를 사용하여 학습한다.
- k-1개의 집합과 c_k, g_k 파라미터를 이용하여 학습을 하면 분류모델이 생성된다. 이 분류 모델을 이용하여 나머지 하나의 집합을 분류하여 정확도를 측정한다. 이런 과정을 k번 반복하면 c_k, g_k 파라미터를 가질 때 cross-validation accuracy를 측정된다.
- 특정 파라미터 쌍에 대해서 하나의 cross-validation accuracy가 나오고, 그리드 서치에서 정한 모든 파라미터 값에 대해서 cross-validation 정확도를 계산했을 때, cross-validation 정확도가 가장 높을 때의 파라미터 값을 선택하여 SVM 학습에 이용한다.

5. 다중 유해 등급 분류 실험 및 성능 분석

5.1 다중 등급 분류 실험

유해 이미지 등급 분류를 위한 시스템은 인텔 펜티엄 M 1.50 GHz, 512M, WinXP sp1 환경의 노트북을 이용하여, 유해 이미지를 구별하는 특징으로는 40x40의 이미지 특성 벡터를 이용한다. 다중 등급의 유해 이미지 분류를 위해서는 공개 소프트웨어인 libsvm-2.6 을 이용하며, SVM 커널의 최적 파라미터를 찾기 위해서 libsvm-2.6에서 제공하는 그리드 서치 프로그램(python)을 이용하여 수행한다. 실험 데이타는 인터넷의 웹사이트를 통해서 일반 영상과 유해 영상을 수집하여 이미지 컨텐츠 유해 정보 기준에 따라 무해, 선정, 유해, 심한 유해의 4종류의 등급으로 정의한 0등급~3등급까지의 4종류로 구성하고, 등급0~등급3에 각각 1,000장을 할당하여, 800장은 학습에 이용하고, 나머지 200장은 테스트 데이타로 사용한다. 상용 시스템의 경우 수만 여장 이상을 학습에 이용하고 있으나, 실험실 수준에서 이미지의 유해 정도를 분류하는 관련 연구[16]의 경우, 상용 환경의 이메일과 웹으로부터 유해 이미지 3,967장, 무해 이미지 7,038장의 총 11,005장의 이미지를 테스트 데이타로 이용하였고, [17]에서는 2,594장의 유해이미지와 일반 이미지 7,574장 총 10,168의 데이타를 이용하고 있다. 또한, 이미지의 학습과 분류의 비는 [5]에서는 8:2로, [18]에서는 3:1로 실험하였다. 본 논문에서는 실험 데이

표 2 유해이미지 등급분류 실험결과 (ROI 검출이 없는 경우) - 각 등급당 200장

등급	피부색 검출 과정[0등급]	다중SVM [0등급]	다중SVM [1등급]	다중SVM [2등급]	다중SVM [3등급]	처리시간 (초)
0등급	157	29	9	3	2	1.048
1등급	6	12	132	20	30	1.388
2등급	2	7	35	103	53	1.362
3등급	1	2	21	66	110	1.217
다중 SVM 분류 판별율		67.44%	68.04%	52.02%	55.27%	58.99%
전체 이미지등급 판별율		93.0%	66.0%	51.5%	55.0%	66.37%

표 3 유해이미지 등급분류 실험결과 (ROI 검출을 수행한 경우) - 각 등급당 200장

등급	피부색검출 과정[0등급]	다중SVM [0등급]	다중SVM [1등급]	다중SVM [2등급]	다중SVM [3등급]	처리시간 (초)
0등급	157	28	11	4	0	1.0042
1등급	6	9	148	13	24	1.398
2등급	2	6	28	126	38	1.369
3등급	1	4	21	41	133	1.217
다중 SVM 분류 판별율		65.11%	76.28%	63.63%	66.83%	68.61%
전체 이미지등급 판별율		92.5%	74.0%	63.0%	66.5%	74.0%

타 중에서 학습에 이용하는 이미지는 전체 이미지의 80%인 3,200장을 이용하고, 실험 데이타를 효율적으로 이용하기 위해 k-fold cross-validation 기법을 적용함으로써, 이용된 데이타의 수 배의 구성 효과를 보인다.

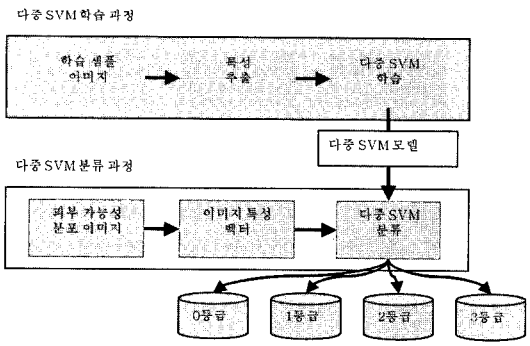


그림 8 유해 이미지 컨텐츠 등급 분류 시스템의 구성

5.2 성능 분석

다중 SVM을 이용한 유해 이미지 컨텐츠의 분류 성능은 다음과 같이 분석된다. 표2와 표3의 0등급부터 3등급 이미지 중에서 피부색 검출 과정에서 탐지된 166장 중에서 실제 0등급으로 검출한 157장의 탐지하여 94.57%의 정확률을 보이고 있다.

관심 영역 검출 기법의 활용에 따른 분석을 해보면, 관심 영역 검출을 수행한 전체 판별율 결과(74.0%)가 관심 영역 검출을 수행하지 않은 결과(66.37)보다 약 7.6%의 개선 효과를 가져왔다. 그러나 각 등급간의 차이를 분석해 보면 0등급의 차이는 거의 미미하지만, 관심 영역 검출을 수행한 결과가 관심 영역 검출 과정이 없는 분류 과정보다 1등급의 경우 8%, 2등급과 3등급에서 각각 11.5% 이상의 높은 검출율을 나타내었다. 따라서, 유해 이미지의 경우 배경이나 노이즈에 의한 판별 오류를 최소화하기 위한 관심 영역 검출이 이미지의 다중 분류의 정확성에 상당 부분 기여한다는 사실을 알수 있다.

표 2와 표 3에서 나타난 이미지 컨텐츠 등급을 분류하는데 있어서 등급간의 오판단을 하게 하는 이유는 본 논문이 피부색에 기반한 방법이기 때문에, 조명으로 인한 피부색의 변화, 복장을 구분하기 어려운 속옷 착용으로 인한 피부색 분포의 변화, 피부색과 유사한 배경이 있어 피부 분포의 구분이 어려운 이미지와 같은 경우 등급간을 구분하는 판별 성능을 떨어뜨리는 주요 요인이 된다는 것을 알 수 있다.

6. 결 론

본 논문에서는 기존의 웹사이트, 스팸메일 등 인터넷을 통해 유입되는 유해성 이미지를 컴퓨터의 처리 기준에 따라 유해 이미지의 컨텐츠 등급을 가려내기 위한 이미지의 특징 추출과 다중 SVM 분류 기법을 제안하였다. 본 논문에서 유해 이미지 컨텐츠를 분리하는데 있어서 가장 큰 주안점을 두었던 피부색 검출은 무해 이미지인 0등급 이미지의 분류에서 94.57%의 높은 판별율을 보임으로써, 유해 이미지 분류의 가장 좋은 특징을 갖는 것을 알 수 있다. 또한 다중 SVM 학습 및 다중 SVM 분류 알고리즘을 이용한 유해 이미지 컨텐츠 등급 분류 기법은 최근까지 연구[12,14,17,18]에서의 유해, 무해의 두 가지 유형의 90% 정도의 분류 판별율과 직

접적인 비교는 어렵지만, 무해 이미지인 0등급의 판별율이 92.5%이고, 다중 유해 등급을 63%~74% 판별하는 결과를 거둠으로써, 유해 이미지의 특성을 이용한 다중 이미지 컨텐츠 등급을 분류할 수 있음을 증명하였다. 본 논문에서 제안한 유해 이미지 분류 기법은 현재의 인터넷, 스팸 메일, P2P 뿐만 아니라 향후 고속 인터넷이나 방송 통신 융합망(BcN) 시대에 가장 각광 받을 것으로 예상되는 멀티미디어 스트림의 유해 정도를 실시간으로 판단하여 유해 정보를 선별적으로 차단하는 분야에 적용될 수 있을 것이다. 향후 본 논문에서 제안한 유해 이미지 분류 기법 및 알고리즘의 적용성을 넓히기 위해서는 데이타 시료를 확장 시킴으로써 유해 이미지의 등급별 선별율을 보다 높일 수 있을 것으로 보인다.

참 고 문 헌

[1] ICRA, http://www.icra.org/

[2] Safenet, http://www.safenet.ne.kr/

[3] PICS, http://www3.org/PICS

[4] Margaret Fleck, David Forsyth, and Chris Bregler. "Finding Naked People," ECCV96 (1996 European Conference on Computer Vision), Vol. II, pp. 592-602, 1996.

[5] David Forsyth and Margaret Fleck, "Identifying nude pictures," IEEE Workshop on the Applications of Computer Vision 1996, pp103-108, 1996.

[6] Michael J. Jones and James M. Rehg, "Statistical color models with application to skin detection," Technical Report Series, Cambridge Research Laboratory, December 1998.

[7] Sung-joon Yoo, "Intelligent multimedia information retrieval for identifying and rating adult images," KES 2004, Data Mining and Knowledge Discovery, 2004.

[8] Will Archer Arentz and Bjorn Olstad, "Classifying offensive sites based on image contents," Computer Vision and Image Understanding, Vol 94, pp. 293-310, 2004.

[9] F. Rosenblatt, "The Perceptron: A probabilistic model for information storage and organization in brain," Psychology Review 65, pp. 386-408.

[10] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. "Learning representations by back-propagating errors," Nature (London), Vol. 323, pp. 533-536.

[11] David Hand, Heikki Mannila, and Padhraic Smyth, "Principles of Data Mining," MIT Press, pp. 343-347, 2001.

[12] Feng Jiao, Wen Giao, Lijuan Duan, and Guoqin Cui, "Detecting adult image using multiple features," IEEE conference, 2001.

[13] V. Vapnik. Statistical Learning Theory. Wiley, New York, NY, 1998.

[14] Chiyoon Jeong, Chongsung Kim, and Ki-sang Hong, "Appearance-based Nude Image Detection," 2004, International Conf. On pattern Recognition Proceeding, 2004.

[15] S. Phung, A. Bouzerdoum, and D. Chai, "Adaptive skin segmentation in color images," In International Conf. On Acostics Speech and Signal Processing, 2003.

[16] Alison Bosson, Gavin C. Cawley, Yi Chan, Richard Harvey, "Non-retrieval: Blocking porno graphics images," CVIR02.

[17] Huicheng, Mohamad Daoudi, and Bruno Jednak, "Blocking Adult Images based on Statistical Skin detection," ELCVIA (Electronic letters on Compute Vision and Image Analysis) 0(0): 1-7, 2000.

[18] Seong-Joon Yoo, Min-ho Jung, Hee Beom Kang, Chee Sun Won, Soo-Mi Choi, "Composition of MPEG-7 Visual Descriptors for Detecting Adult Images on the Internet." HIS2003.

남 택 용

1987년 충남대학교 계산통계학과 이학사 1990년 충남대학교 계산통계학과 이학석사. 2005년 한국외국어대학교 전자정보공학과 공학박사. 1987년~현재 한국전자통신연구원 정보보호연구단 개인정보보호연구팀 팀장(책임연구원). 관심분야는 정보보호, 인터넷, 이미지마이닝 등

정 치 윤

2002년 포항공과대학교 전자전기공학과 공학사. 2004년 포항공과대학교 전자전기공학과 공학석사. 2004년~현재 한국전자통신연구원 정보보호연구단 개인정보보호연구팀 연구원. 관심분야는 컴퓨터 비젼, 네트워크보안 등

한 치 문

1977년 경북대학교 전자공학과 공학사 1983년 연세대학교 전자공학과 공학석사 1990년 일본 동경대학교 전자정보공학과 공학박사. 1977년~1983년 한국과학기술연구원(KIST) 연구원. 1983년~1997년 한국전자통신연구원(ETRI) 계통연구부장(책임연구원). 1997년~현재 한국외국어대학교 전자정보공학부 교수. 1999년~2001년 한국외국어대학교 전자계산소 소장. 2002년~2004년 한국외국어대학교 정보산업공과대학 학장. 관심분야는 차세대스위칭기술, 개방형네트워크기술, VoIP기술, 네트워크보안기술