

## 피부색상과 복합 특징을 이용한 유해영상 인식

장석우\*, 최형일\*\*, 김계영\*\*\*

### Adult Image Detection Using Skin Color and Multiple Features

Seok-Woo Jang\*, Hyung-Il Choi\*\*, Gye-Young Kim\*\*\*

#### 요약

유해영상 검출에서 정확하게 피부 색상 영역을 획득하는 것은 매우 중요하다. 그러나 기존의 방법들은 서로 다른 인종, 조명, 화장, 사용된 카메라 등과 같은 여러 원인으로 인해 피부 색상 추출에 여전히 문제를 가지고 있으며, 사전에 미리 정해진 피부 색상 분포 모델을 이용하여 영상에서 피부 영역을 검출한다. 이러한 문제를 해결하기 위해 본 논문에서는 눈 주변 영역에서 샘플을 추출하여 입력 영상에 최적으로 적합된 피부 색상 분포 모델을 생성하여 피부 영역을 강건하게 분할하고, 분할된 피부 영역에서 성인 영상을 대표할 수 있는 특징을 추출한 후, 분할된 피부 영역이 나체의 몸체를 포함하고 있는지를 뉴럴 네트워크 다층 퍼셉트론을 통해 여러 대표적인 특징들을 통합하면서 추론하는 새로운 방법을 제안하다. 본 논문의 실험에서는 피부 색상 영역 분할과 성인영상 검출의 두 가지 성능 측면에서 제안한 방법의 성능이 기존의 방법에 비해 보다 우수함을 보인다. 본 논문에서 제안한 강건한 유해영상 검출 기법은 얼굴 검출, 성인영상 필터링 등과 같은 관련된 여러 응용 분야에서 유용하게 활용될 것이라 기대된다.

#### Abstract

Extracting skin color is significant in adult image detection. However, conventional methods still have essential problems in extracting skin color. That is, colors of human skins are basically not the same because of individual skin difference or difference races. Moreover, skin regions of images may not have identical color due to makeup, different cameras used, etc. Therefore, most of the existing methods use predefined skin color models. To resolve these problems, in this paper, we propose a new adult image detection method that robustly segments skin areas with an input image-adapted skin color distribution model, and verifies if the segmented skin regions contain naked bodies by fusing several representative features through a neural network scheme. Experimental results show that our method outperforms others through various experiments. We expect that the suggested method will be useful in many applications such as face detection and objectionable image filtering.

▶ Keyword : 성인 영상(Adult image), 피부 색상(Skin color), 다중 특징(Multiple features), 뉴럴 네트워크(Neural network)

• 제1저자 : 장석우    교신저자 : 김계영

• 투고일 : 2010. 10. 04, 심사일 : 2010. 10. 18, 게재확정일 : 2010. 10. 21.

\* 안양대학교 디지털미디어학과 교수    \*\* 숭실대학교 글로벌미디어학부 교수    \*\*\* 숭실대학교 컴퓨터학부 교수

## 1. 서론

인터넷과 정보 기술의 빠른 발달로 그림, 영화, 음악 파일 등과 같은 다양한 멀티미디어 콘텐츠를 접근하고 브라우징하기가 매우 용이해졌다. 이에 반해, 포르노와 같은 성인 영상물이 인터넷 서핑과 이메일 첨부과 같은 경로를 통해서 가정과 학교로 쉽게 전송될 수 있게 되어 청소년이나 심지어 어린이까지도 이런 음란물로 인한 정신적인 피해가 우려되고 있는 실정이다. 따라서 이런 종류의 성인 영상물을 효과적으로 필터링하는 연구가 최근 들어 많은 관심의 대상이 되고 있다.

일반적으로, 기존의 성인 영상 검출 방법은 크게 IP(Internet protocol) 기반의 블랙리스트(blacklist) 블로킹, 문자 내용 기반의 필터링, 그리고 시각적 내용(visual content) 기반의 필터링의 3가지 부류로 분류될 수 있다[1]. IP 기반의 방법은 유해한 웹 사이트의 URL 목록을 저장하여 추후에 이 웹 사이트에 대한 접근을 차단하는 방법이다. 그러나 이 방법은 인터넷 콘텐츠의 내용이 매우 동적으로 변화하기 때문에 모든 유해한 웹 사이트의 URL 목록을 항상 보유하기가 어렵다는 단점이 있다. 문자 내용 기반의 방법은 웹 사이트가 포함한 문자를 분석해서 해당 사이트가 유해한 키워드를 포함하고 있는지를 판단하는 방법이다. 이 방법은 건강이나 성에 관한 일반적인 사이트를 유해물 사이트라고 오판하여 차단하는 오류를 가지고 있다. 이와 같이 위의 두 방법이 성인 영상을 차단하기에는 여러 가지 단점들이 있으므로 최근에는 많은 연구자들이 영상의 내용을 기반으로 성인영상을 검출하고 있다.

영상의 내용을 분석하여 성인영상을 검출하는 방법은 관련 문헌에서 많이 발견할 수 있다. Forth는 피부 필터와 인간 모양 그룹기를 이용하여 영상에서 나체를 검출하는 시스템을 제안하였다[2]. 칼라와 텍스처 특징이 피부 영역을 추출하기 위해 적용되었고, 추출된 영역은 인간 신체의 기하학적인 구조를 이용하는 그룹기에 입력되어 나체 영상을 포함하고 있는지를 판단했다. Bossom은 양자화된 칼라 공간의 유사도를 계산하고 이미지의 블롭(blob) 특징을 추출한 후, 이를 뉴럴 네트워크에 적용하여 포르노 영상을 분류하는 방법을 제안했다[3]. Zheng은 먼저 피부 색상 모델을 피부 영역을 추출하기 위해서 적용했고, 추출된 피부 색상 영역에서 획득된 여러 가지 특징들을 뉴럴 네트워크에 적용하여 유해 영상물을 검출하였다[4]. Hammami는 피부 색상에 기반한 시각적인 분석을 이용해 성인 영상을 검출하는 Webguard 시스템을 개발하였다[5]. 웹

페이지 내에 피부 색상 픽셀의 포함 정도를 나타내는 피부 색상과 관련된 시각적인 특징들을 사용했다. Lee는 YCbCr 공간에서 특별한 조명 효과에 기인한 색상 치우침을 견딜 수 있는 피부 색상 모델들을 이용해 피부 색상을 분할했다[6]. 그런 다음, 여러 가지 특징들을 이용해 분할된 피부 영역의 진위 여부를 파악했다. 이 외에도 다른 많은 방법들이 제안되고 있다 [7-8].

이런 기존의 방법에서 확인할 수 있듯이 입력된 영상으로부터 정확하게 피부 색상 영역을 획득하는 것은 성인 영상 검출에 매우 중요하나, 기존의 방법들은 피부 색상 추출에 여전히 문제를 가지고 있다. 즉, 인간의 피부 색상은 개인 간의 차이나 인종의 차이 등으로 인해 피부 색상이 기본적으로 동일하지 않다. 뿐만 아니라, 화장, 사용된 카메라, 조명 등의 여러 가지 다른 조건으로 인해 촬영된 영상 내에 존재하는 피부 색상은 조금씩 다르다. 따라서 사전에 정의된 피부색상 모델을 이용하는 기존의 알고리즘들은 위에서 기술한 문제들을 극복하기가 어렵다. 따라서 이런 문제를 해결하기 위한 최적의 솔루션은 입력 영상으로부터 피부 샘플을 신뢰성 있게 선택하고 입력된 영상 자체에 적합한 피부색상 모델을 적응적으로 생성하는 것이다.

따라서 본 논문에서는 입력 영상에 최적으로 적합된 피부 색상 분포 모델을 생성하여 피부 영역을 강건하게 분할하고, 분할된 피부 영역에서 성인 영상을 대표할 수 있는 특징을 추출한 후, 분할된 피부 영역이 나체의 몸체를 포함하고 있는지를 뉴럴 네트워크 다층 퍼셉트론을 통해 여러 대표적인 특징들을 통합하면서 추론하는 새로운 방법을 제안한다. 그림 1은 본 논문에서 제안한 성인영상 인식 시스템의 전체 개요도를 보여준다.

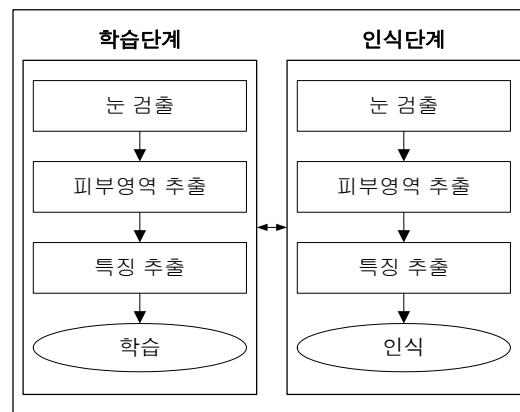


그림 1. 전체 개요도  
Fig. 1. Overall Flowchart

그림 1과 같이 제안된 성인영상 검출 방법은 크게 학습 단계와 인식 단계로 구성된다. 학습 단계에서는 피부 영역을 획득한 후 획득된 피부 영역으로부터 성인 영상을 특징지를 수 있는 여러 가지 시각적인 특징들을 추출한다. 그런 다음, 다층 퍼셉트론 뉴럴 네트워크를 이용하여 특징을 학습하여 저장한다. 인식 단계에서는 학습 단계와 유사하게 테스트 영상으로부터 먼저 피부 색상 영역을 추출한 후 추출된 영역으로부터 여러 가지 특징을 추출한다. 그런 다음, 뉴럴 네트워크를 이용하여 추출된 여러 특징들을 효과적으로 통합하면서 최종적으로 성인 영상의 존재 유무를 판단한다.

1장에서는 본 연구를 수행하게 된 동기 및 배경, 그리고 전체적인 개요를 기술하였다. 2장에서는 입력영상에서 피부 색상 영역을 획득하는 방법에 대해 기술하고, 3장에서는 획득된 피부색상 영역으로부터 성인영상을 나타내는 특징을 추출하는 기법을 소개하며, 4장에서는 여러 특징을 통합하면서 성인영상을 최종적으로 검출하는 방법을 기술한다. 그리고 5장에서는 본 논문에서 수행한 실험결과를 보이며, 6장에서는 결론 및 향후 연구방향을 제시한다.

## II. 피부 영역 검출

이 단계에서는 영상을 받아들여 눈을 검출하고, 눈 주위의 피부색상을 추출하여 피부색상 모델을 생성한다[9]. 일반적으로 눈은 얼굴 검출 및 인식에 있어 가장 신뢰성 있는 특징 중의 하나라고 알려져 있다. 본 논문에서는 눈을 검출하기 위해 Hsu가 제안한 아이 맵(Eye Map)을 이용한다[10]. 이를 위해, 먼저 입력영상을 YCbCr 칼라 공간으로 변환한 후 색상과 명암값을 이용해서 두 개의 아이 맵을 구한다. 식 (1)은 색상 아이 맵을 추출하는 식이다.

$$EyeMapC = \frac{C_b^2 + (255 - C_r)^2 + (C_b/C_r)}{3} \dots\dots\dots (1)$$

식 (1)에서  $C_b$ 와  $C_r$ 은 YCbCr 칼라 공간에서의 Cb와 Cr 값이며,  $C_b^2$ ,  $(255 - C_r)^2$ , 그리고  $C_b/C_r$ 은 0에서 255 사이의 값으로 정규화된다. 일반적으로 눈 주변의 피부 색상의 값은 Cr이 높고, Cb가 낮다는 관찰에 근거한 방식이다.

식 (2)는 명암 값을 이용하여 팽창과 침식 연산을 하여 명암대비 아이 맵을 구하는 식을 나타낸다. 식 (2)에서  $\oplus$ 는 팽창, 그리고  $\ominus$ 는 침식을 수행하는 것이고,  $g(x,y)$ 는 입력 영상의 명암 값을 나타낸다. 눈 영역에 있는 색상 값을 보면 어두운 부분과 밝은 부분을 모두 포함하고 있다. 따라서 영상 명암

값에 팽창과 침식을 적용하여 명암대비를 이용한 아이 맵을 추출한다.

$$EyeMapL = \frac{Y(x,y) \oplus g_\sigma(x,y)}{Y(x,y) \ominus g_\sigma(x,y) + 1} \dots\dots\dots (2)$$

식 (1)과 (2)에서 구한 두 개의 아이 맵을 식 (3)과 같이 통합하여 최종적인 아이 맵을 추출한다.

$$EyeMap = (EyeMapC) \text{ and } (EyeMapL) \dots\dots\dots (3)$$

그런 다음, 아이 맵을 임계값을 사용하여 이진화하여 눈의 후보 영역을 추출한다. 그리고 추출된 눈의 후보 영역들 중에서 길이, 밀집도, 가로 및 세로 비율 등 얼굴 안에 위치한 눈의 기하학적인 구조를 이용해서 눈의 진위 여부를 판단한다. 본 논문에서는 아이 맵을 이용해서 눈을 검출했으나 정확도만 보장된다면 다른 방법을 이용해서 눈을 검출해도 상관은 없다. 예를 들어, 얼굴의 구성요소인 입이나 코 또는 귀를 찾아서 수행할 수도 있을 것이며, 이 방법과 다른 기법을 혼합하여 사용할 수도 있을 것이다.

본 논문에서는 눈을 검출한 후 눈에 해당하는 최소 포함 사각형의 5배 확장된 사각형 영역에서 피부 샘플을 추출한다. 일반적으로 눈 주위의 선택된 영역에 피부색상이 분포할 확률이 높으나 안경이나 다른 배경 영역이 존재할 경우에는 선택된 영역을 피부 영역과 비피부 영역으로 구분해야 한다. 이를 해결하기 위해, 제안된 방법에서는 선택된 피부 샘플과 평균 피부 색상 사이의 거리를 식 (4)와 같이 계산하여 피부 맵 (skin map)을 생성한다.

$$SkinMap = 255 - \frac{|\hat{C}_r - C_r| + |\hat{C}_b - C_b|}{2} \dots\dots\dots (4)$$

식 (4)에서  $\hat{C}_r$ 과  $\hat{C}_b$ 는 각각 일반적인 피부 색상의 Cr과 Cb값을 나타낸다. 피부 맵은 0에서 255사이의 값의 범위를 가지며, 선택된 샘플이 평균적인 피부 색상에 근접할수록 255와 가까운 값을 가진다.

피부 맵에서 피부 샘플 최소만을 선택하기 위해 본 논문에서는 Otsu에 의해 제안된 히스토그램 이진화 방법을 사용한다[11]. Otsu 방법은 명암 히스토그램을 이진화하는 최적의 임계치를 사전지식 없이 통계적으로 선택하며, 히스토그램이 두 개의 확률밀도를 가질 때 성능이 우수하다고 알려져 있다. 그림 2는 입력영상으로부터 피부 맵을 생성한 후 이를 이진화하는 예를 보여주고 있다.

피부 색상을 모델링하는 것은 적합한 칼라 공간과 이 공간

에서 피부 색상과 관련된 클러스터를 확인해야 한다. 일반적으로, CbCr 공간에서 피부색상 샘플의 분포는 타원의 형태를 취하지만 테스트 영상마다 해당되는 타원의 위치는 다르므로 제안된 방법은 각 입력영상으로부터 직접 피부 샘플을 선택하여 영상 적응적인 피부 색상 모델을 생성한다. 제안된 피부색상 모델은 눈 주위에서 선택된 피부 샘플들을 사용해 생성되며, 타원형 클러스터의 중심, 그리고 얼마나 넓게 분포하는지가 지정된다. 본 논문의 타원형 피부 색상 모델은 식 (5)와 같이 정의된다.

$$\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} = 1 \dots\dots\dots (5)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b - C'_b \\ C_r - C'_r \end{bmatrix}$$

$$C'_b = \frac{1}{K} \sum_{t=0}^{K-1} C_b(t), \quad C'_r = \frac{1}{K} \sum_{t=0}^{K-1} C_r(t)$$

식 (5)에서 (Cb,Cr)은 테스트 영상의 Cb와 Cr 값을 나타내고,  $\theta$ 는 타원의 회전각을 나타낸다. 그리고 (a,b)는 타원의 장축과 단축을 의미하고, (ec<sub>x</sub>,ec<sub>y</sub>)는 타원의 회전 에러를 보상하는 값이다. (C<sub>b</sub>',C<sub>r</sub>')은 피부 모델의 중심을 표시한다. t는 각 피부 샘플을 표시하는 인덱스이며, K는 선택된 피부 색상의 개수를 의미하는 인수이다.

이와 같이, 피부 색상 분포에 대한 타원 모델을 생성한 후에는 이 모델을 이용해 입력영상에 대한 피부 분할을 수행한다. 본 논문에서는 사전에 이미 정의된 피부 색상 모델을 사용하는 대신 각각의 테스트 영상으로부터 영상 적응적인 피부 색상 모델을 생성하여 사용하므로 다양한 주변 조건에 관계없이 강건한 피부영역 분할이 가능하다.

### III. 특징 추출

피부색상 분포 모델을 이용하여 영상으로부터 피부 영역을 강건하게 추출한 후에는 추출된 피부 영역이 성인 영상을 포함하고 있는지의 여부를 판단하기 위해 추출된 피부 영역으로부터 성인 영상을 특징지을 수 있는 여러 가지 특징을 획득한다. 일반적으로, 주어진 영상 내에 나무, 사막, 음식, 동물의 가족이나 피부 등과 같이 피부 색상과 유사한 많은 영역들이 존재할 수 있으므로 성인 영상을 특징짓는 주요한 특징들만을 효과적으로 선택하는 것은 매우 중요한 일이다. 이를 위해, 본 논문에서는 위치(location), 크기(size), 연장도(elongatedness), 밀집도(compactness), 그리고 텍스처(texture)의 5 가지 특징을 활용한다.

위치 특징으로는 피부 영역의 무게중심과 영상의 중심 사이로부터 정규화된 유클리디안(Euclidean) 거리를 계산하여 사용한다. 이 특징은 분할된 피부 영역이 영상의 중심으로부터 얼마나 가까이에 위치해 있는지를 측정하는 척도이며, 일반적으로 나체나 성인 영상은 관찰자들의 시선을 집중시키기 위해 영상의 중심 부분에 위치한다는 많은 관찰에 근거한다. 본 논문에서 사용하는 위치 특징은 식 (6)과 같이 정의된다.

$$F_{loc}^i = \frac{1}{\sqrt{2}} \times \sqrt{\left(\frac{|CG_x^i - IC_x|}{W}\right)^2 + \left(\frac{|CG_y^i - IC_y|}{H}\right)^2} \dots\dots\dots (6)$$

식 (6)에서 i는 분할된 피부 영역의 순서를 나타내는 인덱스이고, W와 H는 영상의 너비와 높이를 나타낸다. CG<sub>x</sub><sup>i</sup>와

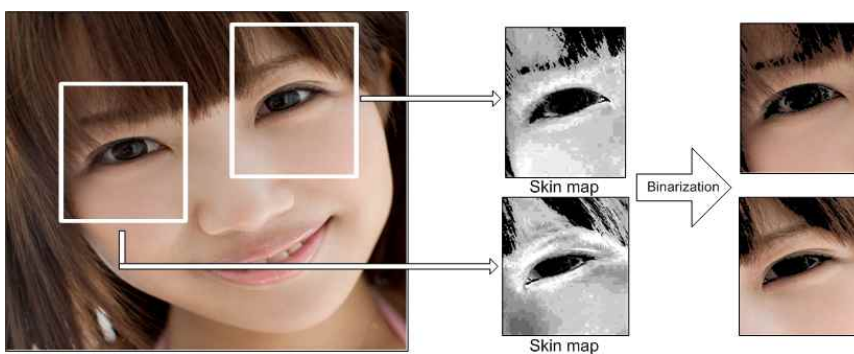


그림 2. 피부 맵 생성 및 이진화  
Fig. 2. Generating and Binarizing a Skin Map

$CG_y^i$ 는  $i$ 번째 피부 영역의 무게중심을 나타내고,  $IC_x$ 와  $IC_y$ 는 입력 영상의 중심위치를 나타낸다. 위치 특징  $F_{loc}^i$ 는 0에서 1 사이의 값을 가지며, 0에 가까울수록 해당되는 피부 영역은 영상의 중심에 위치한다.

크기 특징으로는 전체 영상에 대한 피부 영역의 크기 비율을 사용한다. 보통, 나체 영상의 크기는 보는 사람의 시각적 효과를 높이기 위해 영상에서 일정 비율 이상을 차지한다. 따라서 식 (7)에서 정의되는 크기 특징이 일정한 임계치 이상이어야 분할된 피부 영역이 성인영상을 포함할 가능성이 높다. 식 (7)에서  $N_i$ 는 해당되는 피부 영역 화소의 개수를 의미한다.

$$F_{size}^i = \frac{N_i}{W \times H} \dots\dots\dots (7)$$

연장도 특징으로는 분할된 피부 영역의 가로와 세로의 비율을 계산한다. 이 특징은 분할된 피부 영역의 프로파일이 얼마나 나체와 유사한 지를 측정하는데, 식 (8)과 같이 정의된다.

$$F_{elon}^i = \frac{L_{hor}(LER^i)}{L_{ver}(LER^i)} \dots\dots\dots (8)$$

식 (8)에서  $LER^i$ 는 분할된  $i$ 번째 피부 영역의 최소 포함 사각형을 나타내고,  $L_{hor}(LER^i)$ 와  $L_{ver}(LER^i)$ 는 최소 포함 사각형의 가로와 세로의 길이를 각각 나타낸다.

밀집도 특징으로는 분할된 피부 영역과 이 영역의 최소 포함 사각형의 비율을 계산한다. 밀집도는 영역의 밀도가 얼마나 높은지를 나타낸다. 일반적으로, 나체 영역은 피부 색상의 화소들이 군집을 형성하므로 밀집도가 높은 편이다. 이 특징은 나체 영상과 비 나체 영상을 구분하는 매우 좋은 특징 중의 하나인데, 식 (9)와 같이 정의된다. 식 (9)에서 특징 값이 1인 경우에는 피부 영역의 밀집도가 가장 높은 것을 나타낸다.

$$F_{comp}^i = \frac{N_i}{L_{hor}(LER^i) \times L_{ver}(LER^i)} \dots\dots\dots (9)$$

텍스처 특징으로는 분할된 피부 영역의 에지 분포가 얼마나 균일하게 존재하는 지를 측정한다. 보통, 비 성인영상은 에지가 날카롭게 분포되어 있으며, 성인 영상들은 에지가 균일하게 분포하고 있는 편이다. 본 논문에서 제안된 방법은 먼저 분할되어 있는 피부 영역에 라플라시안(Laplacian) 에지 연산자를 적용하여 에지의 크기를 추출한다. 일반적으로, 라플라시안 연산자는 영상의 2차 미분값을 측정하기 위해 사용되는데, 연산이 빠르고 모든 방향에서 에지를 효과적으로 검출할 수 있다고 널리 알려져 있다[12]. 본 논문에서는 라플라시안 에지를 구하기 위해 그림 3과 같은 마스크를 이용한다.

-1	-1	-1
-1	8	-1
-1	-1	-1

그림 3. 라플라시안 마스크  
Fig. 3. Laplacian Mask

그리고 추출된 에지의 크기가 전체 피부 영역의 평균 에지 크기와 유사한 화소들의 개수를 계산하는데, 식 (10)과 같이 정의된다. 식 (10)에서  $E(x_j^i, y_j^i)$ 는  $i$ 번째 피부 영역의  $(x, y)$  위치에서의 에지의 크기를 나타내고,  $\hat{E}^i$ 는  $i$ 번째 피부 영역의 평균 에지의 크기를 나타내며,  $E_{th}$ 는 실험에 의해 결정되는 임계치이다. 텍스처 특징은 1에 가까운 값을 가질수록 피부 영역의 에지가 균일하다.

$$F_{text}^i = \frac{1}{N_i} \times \sum_{j=1}^{N_i} \Phi(E(x_j^i, y_j^i)) \dots\dots\dots (10)$$

$$\Phi(E(x_j^i, y_j^i)) = \begin{cases} 1 & \text{if } \hat{E}^i - E_{th} \leq E(x_j^i, y_j^i) \leq \hat{E}^i + E_{th} \\ 0 & \text{else} \end{cases}$$

각각의 분할된 피부 영역을 특징지우는 위에서 계산된 5가지 타입의 정규화된 특징들은 특징 벡터의 형태로 구성될 수 있으며, 이는 성인영상을 최종적으로 분류하는 추론 모듈의 입력으로 전달된다.

#### IV. 유해영상 인식

본 논문에서 제안된 성인영상 검출 방법에서는 특징 추출 단계에서 생성한 다섯 가지의 특징들을 효과적으로 통합하면서 성인 영상과 비 성인 영상을 구분하기 위해서 다층 퍼셉트론(multi-layer perceptron) 뉴럴 네트워크를 적용한다 [13]. 다층 퍼셉트론은 분류(classification) 및 회귀(regression) 문제에서 유용하게 사용되는 전방향(feed-forward) 뉴럴 네트워크로 잘 알려져 있다. 다층 퍼셉트론은 일반화된 선형 모델, k-이웃(k-nearest neighborhood) 탐색기, SVM(support vector machine) 등과 같은 다른 분류 방법보다 통계적으로 좋은 성능을 보인다고 알려져 있다. 제안된 방법에서 사용하는 다층 퍼셉트론은 그림 4에서 보이듯이 하나의 중간 층을 가지는 3층의 전방향 뉴럴 네트워크이다.

각각의 피부 색상 영역에서 추출된 5가지 종류의 정규화된 시각적 특징은 입력 층으로 전해지며, 시그모이드 함수가 비선형 활성화 함수(activation function)로 사용된다. 다층 퍼셉트론 뉴럴 네트워크의 출력은 0에서 1사이의 값인데, 이 값은 분할된 피부 색상 영역이 나체 영역를 포함하고 있는지의 가능성 정도를 표현한다. 다시 말해서, 출력 값이 1에 가까울수록 입력 영상이 성인 영상을 포함할 가능성이 높은 것이다. 그리고 본 논문에서는 이진 결정(binary decision)을 위해 임계치 T를 설정한다. 본 논문에서 T는 테스트 영상이 성인 영상으로 인식되는 정도를 적응적으로 조절하는 역할을 한다. 예를 들어, 어린 아이들이 웹 페이지를 브라우징할 때에는 T 값은 대부분의 성인 영상들이 필터링되어 제거될 수 있도록 증가될 수 있다. 그렇지 않으면, T는 성인 영상으로 오 검출되는 비성인 영상의 개수가 최소화되도록 적절히 조절될 수 있다. 만일 테스트 영상에 하나 이상의 피부 영역이 존재할 경우에는 최대 출력 값을 가지는 영역을 선택하여 사용할 수 있다.

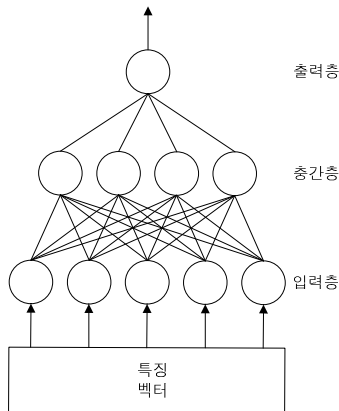


그림 4. 다층 퍼셉트론 구조  
Fig. 4. Structure of Multi-layer Perceptron

본 논문에서 제안된 성인영상 검출 방법이 입력 영상에 적용된 피부 색상 분포 모델과 5가지의 시각적인 특징을 이용하여 성인영상 검출을 시도하나 오 인식 가능성은 존재할 수 있다. 예를 들어, 상반신 영상(mug shot)의 경우에는 영상 내에 상당한 부분의 피부 영역을 포함하므로 이 영상은 종종 비성인 영상이 아니라 성인영상으로 분류될 수 있다. 이런 오 인식을 효과적으로 배제하기 위해서 대부분의 기존의 성인영상 검출 방법들은 보통 얼굴 인식기를 별도로 추가한다. 일반적으로, 상반신 영상은 얼굴 영역이 피부 영역의 상당한 부분을 차지한다는 특성을 가지므로 전체 피부 영역에 대한 얼굴 영역이 차지하는 비율을 활용하면 오 인식을 상당히 제거할 수 있다. 그러나 얼굴 검출 모듈은 시스템의 시간 계산상의 복잡도를 증가시킨다. 반면에, 제안된 성인영상 인식 방법은 피부 색상 분포 모델을 생성할 때 이미 눈을 검출하였으므로 검출한 눈과 얼굴 영역 사이의 기하학적인 관계를 이용해서 상반신 영상과 성인영상을 효과적으로 구분할 수 있다.

### V. 실험결과

본 논문의 실험을 위하여 사용한 컴퓨터는 인텔 Pentium Core 2 Duo의 2.66GHz CPU와 4GB의 메모리를 사용하였다. 그리고 운영체제로는 마이크로소프트사의 Windows XP Professional을 사용하였다. 그리고 제안된 성인영상 검출 소프트웨어 구현을 위한 컴파일러로는 마이크로소프트사의 Visual C++ 2003을 이용하여 제안된 성인영상 인식 알고리즘을 구현하였다. 또한, 실험에서 사용할 영상 데이터베이스 구축을 위하여 조명이나 다른 특정한 제약조건이 주어지지 않은 일반적인 자연환경에서 촬영한 다양한 종류의 성인영상과 비성인영상을 수집하여 활용하였다.



(a) 입력영상 (b) Hsu 방법 (c) Lee의 방법 (d) 제안 방법

그림 5. 피부 색상 영역 추출  
Fig. 5. Extracting Skin Color Regions

전체 2,400개의 영상을 이용하여 영상 데이터베이스를 생성하였는데, 이 중에는 1,200개의 나체영상, 400개의 비키니 영상, 400개의 인물영상, 그리고 400개의 기타영상이 있다. 나체영상은 상반신 나체영상과 전신 나체영상을 포함하고 있으며, 비키니 영상은 수영복이나 비키니 등과 같이 노출이 심한 옷을 입고 있는 비 성인영상을 포함하며, 인물영상은 배경과 조명 환경이 다른 정면에서 촬영한 전신 인물 영상과 상반신 비 성인영상을 의미한다. 그리고 기타 영상은 복잡한 배경이나 하나 이상의 얼굴을 포함하고 있는 다양한 영상을 포함한다. 전체 2,400 개의 영상 중에서 각 그룹에 속하는 영상의 1/2에 해당하는 1,200개의 영상은 학습을 위해 사용되고, 나머지 1,200 개의 영상은 테스트를 위해 사용되었다.

본 논문에서는 제안된 방법의 성능을 피부색상 분할과 성인영상 인식의 두 가지 측면에서 분석한다. 먼저, 본 논문에서는 피부 영역 분할을 위해 가장 대표적인 방법이라고 알려져 있는 Hsu[10]의 방법과 Lee[6]의 방법도 구현하였다. 그림 5는 제안된 방법과 기존의 방법을 이용하여 나체 영상으로부터 피부색상을 추출한 예를 보여준다. 그림 5의 피부색상 추출 결과 영상은 피부 색상, 그리고 검정색으로 표시된 비피부 색상으로 구성되어 있다. 그림 5에서 시각적으로 확인할 수 있듯이 기존의 방법들은 피부 색상 영역들의 경계를 정확하게 획득할 수 없는 것처럼 보인다. 그리고 배경이 피부색상과 유사한 영역을 포함하고 있으므로 많은 불필요한 영역을 피부 영역으로 오검출하는 경향이 있다. 이에 비해, 제안된 방법은 비교적 성공적으로 피부영역을 분할하는 것을 확인할 수 있다.

비키니 영상에 대한 피부 영역 검출 결과는 대체적으로 성인영상의 결과와 유사하다. 그 이유는 나체 영상과 비키니 영상은 모두 영상 전체에서 피부영역이 큰 부분을 차지한다는 공통점이 있기 때문인 것으로 추정된다. 인물 영상의 경우에는 영상에서 피부영역이 차지하는 부분이 상대적으로 크고, 비교적 좋은 실내외 환경에서 촬영되므로 기존의 방법과 제안된 방법 모두 좋은 피부색상 분할 결과를 산출하였다. 그리고 복잡한 주변 환경 하에서 촬영된 기타 영상의 경우에는 전체적으로 피부 분할 결과가 좋지 않지만 제안된 방법이 기존의 방법보다는 우수함을 시각적으로 확인할 수 있었다.

피부 색상 검출을 정량적으로 평가하기 위해 본 논문에서는 식 (11)과 같은 척도를 사용한다[14].

$$RMSE = \sqrt{\frac{1}{M \times N} \sum_{i=m}^M \sum_{j=n}^N |OB(i,j) - RB(i,j)|^2} \dots\dots\dots (11)$$

식 (11)에서 M과 N은 영상의 너비와 높이를 나타내고,

(i,j)는 영상의 위치를 나타내는 수평과 수직 인덱스이다. OB(i,j)는 그라운드 투루수(ground truth) 이진영상을 나타내고, RB(i,j)는 피부 영역 검출의 결과 이진 영상을 나타낸다. 그림 6은 기존의 방법과 제안된 피부 영역 검출 방법들에 대한 RMSE(root mean square error) 결과를 보여준다. 그림 6에서 확인할 수 있듯이 제안된 방법이 기존의 방법보다 좋은 결과를 산출함을 확인할 수 있다.

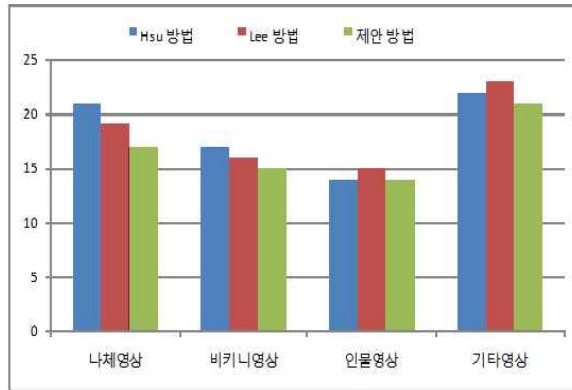


그림 6. 루트 평균 제곱 에러  
Fig. 6. Root Mean Square Error

표 1은 제안된 성인영상 검출 방법의 혼란 행렬(confusion matrix)을 보여준다. 성인영상과 비 성인영상의 인식률은 각각 94.33%와 92.68%이다.

표 1. 제안 방법의 혼란 행렬  
Table 1. Confusion Matrix of the Suggested Method

분류결과	테스트 영상	
	성인영상	비 성인영상
성인영상	566(94.33%)	44(7.32%)
비 성인영상	34(5.56%)	556(92.68%)

본 논문에서 제안된 성인영상 검출 알고리즘의 성능을 비교하기 위해 기존의 방법 중 Yang이 제안한 방법도 구현하였다[15]. 이 방법은 비교적 최근에 제안되었으며, 성능도 좋은 것으로 알려져 있다. 제안된 방법과 Yang의 방법의 성능을 비교 평가하기 위해 동일한 테스트 영상들이 Yang의 방법에 적용되었다. 이 방법에 대한 혼란 행렬은 표 2에 나와 있다. 표 2에서 확인할 수 있듯이 제안된 방법이 기존의 방법보다 성인영상을 보다 정확하게 검출하는 것을 확인할 수 있다.

표 2. Yang 방법의 혼란 행렬



Table 2. Confusion Matrix of Yang's Method

분류결과	테스트 영상	
	성인영상	비 성인영상
성인영상	512(85.33%)	79(13.17%)
비 성인영상	88(14.67%)	521(86.83%)

## V. 결론 및 향후연구

최근 들어 인터넷과 정보통신 기술의 빠른 발달로 인해 그림, 영화, UCC(user created contents), 음악 파일 등과 같은 다양한 멀티미디어 콘텐츠를 접근하고 브라우징하기가 매우 용이해졌다. 이에 반해, 포르노와 같은 성인 영상물들이 인터넷 서핑과 이메일 첨부과 같은 비교적 쉬운 경로를 통해서 가정과 학교로 쉽게 전송될 수 있게 되어 청소년이나 심지어 어린이까지도 이런 음란물로 인한 정신적인 피해가 우려되고 있는 실정이다. 따라서 이런 종류의 성인 영상물을 효과적으로 차단하고 필터링하는 연구가 최근 들어 많은 관심의 대상이 되고 있다.

본 논문에서는 강인한 피부색상 분할을 이용하는 새로운 성인영상 검출 방법을 제안하였다. 본 논문에서 제안된 방법에서는 먼저 얼굴에서 가장 안정적이고 중요한 구성요소라고 알려진 눈을 검출하고, 검출된 눈 주위로부터 신뢰성 있는 피부 샘플을 추출하여 입력된 영상에 적합한 피부 색상 분포 모델을 적용적으로 생성하였다. 그리고 생성된 피부 색상 모델을 이용하여 영상 전체에서 피부색상 영역을 효과적으로 분할하였다. 본 논문에서는 제안된 적용적인 피부 색상 분포 모델의 사용으로 주위 환경에 강인한 피부 색상 분할이 가능하였다. 그런 다음, 크기, 위치, 연장도, 밀집도, 텍스처 등의 성인 영상을 특징지을 수 있는 5가지의 기하학적 특징들이 분할된 피부 영역으로부터 추출되고, 마지막으로 5가지의 특징들을 다중 퍼셉트론 뉴럴 네트워크에 입력하여 효과적으로 다중의 특징들을 통합하면서 피부 영역이 성인영상을 포함하고 있는지를 판단하였다.

향후에 수행할 연구로는 여성의 가슴이나 엉덩이 등과 같이 성인영상을 좀 더 명확하게 특징지우는 인간의 신체 구성요소를 이용하여 성인영상의 검출 작업을 보다 크게 개선할 예정이다. 그리고 보다 복잡한 자연 환경에서 촬영한 영상들을 가지고 다양하게 실험하여 본 논문에서 제안된 성인영상 인식 방법의 실제 응용 프로그램으로의 적용 가능성을 보다 구체적으로 타진할 예정이다.

## 참고문헌

- [1] Jau-Ling Shih, Chang-Hsing Lee, and Chang-Shen Yang, "An Adult Images Identification System Employing Image Retrieval Technique," Pattern Recognition Letters, Vol. 28, No. 16, pp. 2367-2374, 2007.
- [2] D. A. Forsyth and M. M. Fleck, "Automatic Detection of Human Nudes," International Journal of Computer Vision, Vol. 32, No. 1, pp. 63-77, 1999.
- [3] A. Bosson, G. C. Cawley, Y. Chian, R. Harvey, "Non-Retrieval: Blocking Pornographic Images," In Proceedings of the International Conference on the Challenge of Image and Video Retrieval, Vol. 2383, pp. 50-60, 2002.
- [4] Huicheng Zheng, Mohamed Daoudi, and Bruno Jedynak, "Blocking Adult Images Based on Statistical Skin," Electronic Letters on Computer Vision and Image Analysis, Vol. 4, No. 2, pp. 1-14, 2004.
- [5] Mohamed Hammami, Youssef Chahir, and Liming Chen, "WebGuard: a Web Filtering Engine Combining Textual, Structural, and Visual Content-based Analysis," IEEE Transactions on Knowledge and Data Engineering, Vol. 18, No. 2, pp. 272-284, 2006.
- [6] Jiann-Shu Lee, Yung-Ming Kuo, Pau-Choo Chung, and E-Liang Chen, "Naked Image Detection Based on Adaptive and Extensible Skin Color Model," Pattern Recognition, Vol. 40, No. 8, pp. 2261-2270, 2007.
- [7] Y. H. Kuan and C. H. Hsieh, "Content-based Pornography Image Detection," In Proceedings of the International Conference on Imaging Science, Systems, and Technology, 2004.
- [8] Weiming Hu, Ou Wu, Zhouyao Chen, Zhouyu Fu, and Steve Maybank, "Recognition of Pornographic Web Pages by Classifying Texts and Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 6, pp.



1019-1034, 2007.

- [9] 박영재, 장석우, 김계영, “눈 주위의 피부색을 이용한 피부영역검출과 입술검출에 관한 연구,” 한국컴퓨터정보학회논문지, 제 14권, 제 4호, 19-30쪽, 2009년.
- [10] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, and Anil K. Jain, “Face Detection in Color Images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, pp. 696-706, 2002.
- [11] N. Otsu, “A Threshold Selection Method from Gray-Level Histogram,” IEEE Transactions on Systems, Man and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.
- [12] Hui Zhao, Qi Li, Huajun Feng, “Multi-focus Color Image Fusion in the HSI Space Using the Sum-Modified-Laplacian and a Coarse Edge Map,” Image and Vision Computing, Vol. 26, No. 9, pp. 1285-1295, 2008.
- [13] Baopu Li, Max Q.-H. Meng, “Texture Analysis for Ulcer Detection in Capsule Images,” Image and Vision Computing, Vol. 27, No. 9, pp. 1336-1342, 2009.
- [14] GuoJun Liu, XiangLong Tang, H. D. Cheng, JianHua Huang, and JiaFeng Liu, “A Novel Approach for Tracking High Speed Skaters in Sports Using a Panning Camera,” Pattern Recognition, Vol. 42, No. 11, pp. 2922-2935, 2009.
- [15] Jinfeng Yang, Zhouyu Fu, Tieniu Tan, Weiming Hu, “A Novel Approach to Detecting Adult Images” In Proceedings of the International Conference on Pattern Recognition, Vol. 4, pp. 479-482, 2004.

**저 자 소개**



**장 석 우**  
 2000년 8월 : 숭실대학교대학원 컴  
 퓨터학과 (공학박사)  
 2009년 3월 - 현재 : 안양대학교 디  
 지털미디어학과 교수  
 관심분야 : 로봇비전, 증강현실,  
 HCI, 게임, 비디오 색인  
 및 검색, 이러닝 등



**최 형 일**  
 1987년 6월 : 미시간대학교 전자공학  
 과 (공학박사)  
 1987년 9월 - 현재 : 숭실대학교 글로  
 벌미디어학부 교수  
 관심분야 : 컴퓨터 비전, 형태인식, 퍼  
 지어론, 비디오 검색, 인터  
 페이스 에이전트 등



**김 계 영**  
 1996년 2월 : 숭실대학교대학원 컴  
 퓨터학과 (공학박사)  
 2001년 3월 - 현재 : 숭실대학교 컴  
 퓨터학과 교수  
 관심분야 : 컴퓨터 비전, 형태인식,  
 생체인식, 증강현실, 신  
 호처리 등