

템플릿 매칭기반 음란 이미지 탐지

신형섭*, 김현정**, 원일용*

*서울호서전문학교 사이버해킹보안과

**건국대학교 컴퓨터공학과

e-mail: sprtms121@naver.com, nygirl@konkuk.ac.kr, clcccc@shoseo.ac.kr

Template matching-based Pornographic Image Detection

Hyung-Sub, Shin*, Hyun-jung, Kim**, Il-Young, Won*

*Cyber Hacking Security Seoul Hoseo Technical College,

**Dept. of Computer Science Konkuk University

요 약

인터넷 공간에서 무한한 자료의 공유는 손쉽게 유용한 정보를 얻는다는 순기능도 있지만, 청소년들이 음란물을 쉽게 접할 수 있다는 역기능도 제공된다. 음란물은 신체적으로나 정신적으로 미성숙한 상태에 있는 청소년들의 정서와 태도형성, 그리고 행위까지도 영향을 미칠 수 있다[8]. 따라서 방대한 인터넷 공간에서 음란 이미지를 자동으로 감지하는 연구는 청소년들을 보호하기 위해 필요하다.

본 연구에서 음란 이미지를 판단하는 알고리즘을 제안하고, 실험으로 제안된 알고리즘의 성능을 평가했다. 음란 이미지를 판단하는 핵심 알고리즘은 템플릿 매칭 기법이며, 실험결과 대체적으로 안정된 성능을 확인할 수 있었다.

1. 서론

정보통신기술의 발달로 인터넷을 통해 무한한 정보공유가 가능해지면서 사람들은 손쉽게 유용한 정보를 획득할 수가 있다. 하지만 이러한 순기능이 있는 반면 여러 가지 역기능도 등장했다. 특히 급속도로 퍼져 나가는 음란물은 초·중·고 학생들에게 부정적인 영향을 미치고 있어, 실질적인 음란물 제어가 필요하다. 통계에 따르면 청소년들의 91.4%가 음란물은 본 적이 있다고 답변하고 있다[1].

인터넷에 유포되고 있는 음란물을 자동으로 감지하여 청소년을 보호하기 위해 다양한 기술이 개발되고 있다. 대부분의 상업용 제품들이 적용하고 있는 음란물 감지기법은 다음과 같다. 먼저 기존에 알려진 모든 음란파일들의 개별적 특징을 추출하여 음란물 DB에 저장해 두고, 실시간으로 의심되는 파일을 이 DB에 있는 특징과 비교하여 음란물 여부를 판단하는 방법이다. 그러나 이러한 기존의 방법은 알려지지 않은 음란물은 탐지하지 못하며, 음란물 파일이 조금만 변경되어도 탐지할 수 없다는 문제점을 가지고 있다. 특히 기하급수적으로 증가하고 있는 음란물을 모두 DB화하는 것은 지속적인 노력과 비용이 필요하다[2][3].

본 연구는 기존의 방법과는 다르게, 등록되어있지 않은 음란물에 대해서도 음란물 여부를 판단할 수 있는 패턴매칭기반의 음란 이미지 탐지기법에 관한 것이다. 2장에서는 관련연구를 언급하고, 3장에서는 새로운 음란 이미지 탐지기법을 제안한다.

4장에서는 제안된 시스템의 유용성을 검증하기 위한 실험 및 결과를 기술한다. 마지막으로 5장에서 결론 및 향후연구과제에 대하여 언급한다.

2. 관련연구

2.1 AdaBoost - Face Detection

최근 기계학습 분야에서 커널머신을 이용한 대표적 분류기로 AdaBoost가 주목받고 있으며 다양한 패턴인식 문제에 적용되고 있다[3][9]. 부스팅(boosting)은 구별 기능이 약한 여러 개의 검출기(classifier)를 조합하여 강한 검출기를 만드는 학습 방법이다. 많은 검출기들 중에서 찾고자 하는 대상과 적합한 패턴을 정해진 개수만큼 찾아내고 이들 사이에서 선형적인 관계를 찾는 학습 방법이다. 선형적인 관계는 찾아진 검출기들의 대한 각각의 계수(coefficients)를 찾아내고 이들을 선형적으로 융합하는 것을 의미한다[10].

$$F_{strong}(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t F_t(\mathbf{x}) > \lambda_t \\ -1 & \text{otherwise} \end{cases}$$

T 는 사용자가 실험을 통하여 적당한 값을 선정, 몇 개의 패턴으로 원하는 물체를 표현할지 나타낸다. F_{strong} 은 찾아진 검출기 F_t 와 계수 α 에 대한 선형 조합의 형태를 갖는다. 여기서 F_{strong} 이 -1이면 틀린 경우, 1이면 맞는 경우를 의미한다. \mathbf{x} 는 일정한 크기를 가지는 임의의 입력 벡터이다[10].

2.2 Template Matching

템플릿 매칭방법은 구현하기 쉬운 장점을 가지고 있다. 보통 템플릿으로 표준 패턴을 수동으로 미리 작성하거나 함수 파라미터 형식으로 정의하고 주어진 이미지 위에서 이동하면서 템플릿인 표준 패턴과 상관관계를 계산하여 대응하는 위치를 찾는다. 그 중에 매칭방법은 상관계수 (Correlation Coefficient)를 이용한다[4].

$$r = \frac{\sum_{x,y} [f(x,y) - \bar{f}][t(x,y) - \bar{t}]}{\sqrt{\sum_{x,y} [f(x,y) - \bar{f}]^2 \times \sum_{x,y} [t(x,y) - \bar{t}]^2}}$$

f(x,y)는 추출된 샘플이미지에서 픽셀(x,y)의 명암 값이고, \bar{f} 는 추출된 샘플이미지 에 평균 명암값이다. 또한 t(x,y)는 템플릿 이미지 에서 픽셀(x,y)의 명암값이고 \bar{t} 는 템플릿이미지의 평균 명암값이다. r은 상관계수의 결과값으로 범 위는 $-1 \leq r \leq 1$ 를 갖으며, 1이면 템플릿과 샘플이 정확히 일치하고, 1에 가까울수록 정합률이 좋다[5].

2.3 피부색 검출

색상은 이미지에서 객체를 식별하게 하는 중요한 요소이며, 이미지 검색에 가장 널리 사용되는 시각 특징 중에 하나이다[7]. 본 연구에서 사용한 피부색 검출 기법은 YCrCb 색공간의 색차성분 (CrCb) 부분의 범위 [77<Cb<127 and 133<Cr<173] 를 피부색 범위로 사용한다[6].

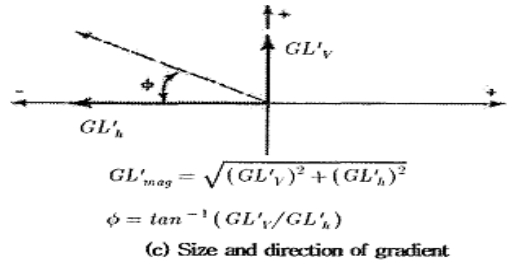
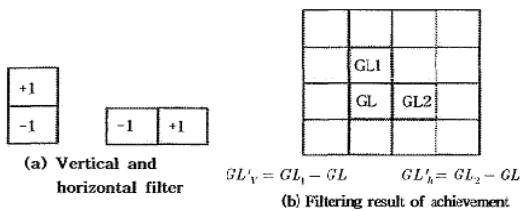
$$Y < -0.299 * R + 0.587 * G + 0.114 * B$$

$$Cr < - (R - Y) * 0.713 + \text{delta}$$

$$Cb < - (B - Y) * 0.564 + \text{delta}$$

2.4 질감 검사

질감은 이미지의 표면 특성으로 표현되는 화소간의 밝기 차이를 나타낸 것으로서 인간의 시각 패턴에 있어 중요한 요소로 작용한다. 질감의 구조가 옷감의 프린트 패턴이나 벽돌무늬처럼 규칙적으로 반복되는 기하학적인 도형의 배열을 분석하는데 주로 이용된다. 영상처리의 일반적인 문제는 화소와 화소사이에서 갑작스럽게 나타나는 밝기값의 변화이다. 이러한 변화는 현격이 다른 두 대상물 사이에 경계인 외곽선을 나타낸다. 하나의 필터에 의해 생성되는 강화영상은 한 방향에 대해서만 우월적으로 나타내며, 수평과 수직방향으로 영상필터링을 수행하여 (그림 1) 과 같은 벡터 값을 산출함으로써 얻어질 수 있다[7][11].

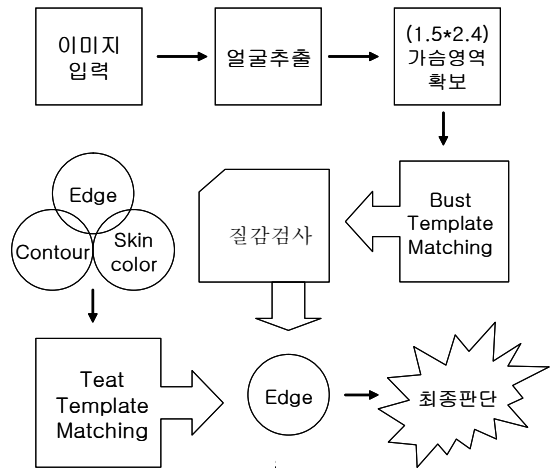


(그림 1) Vector Calculation of image gradient

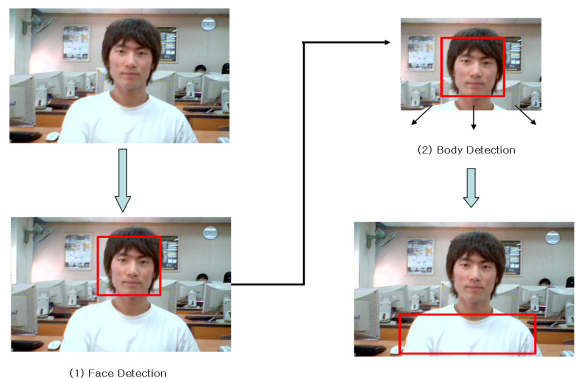
3. 음란이미지 검출

3.1 검출 알고리즘

음란이미지 검출은 (그림2) 와 같은 과정을 통해 검출 한다. 첫 번째, 입력 이미지에서 학습된 데이터 파일을 통해 얼굴검출을 실시하고 음란이라고 판단되는 가슴영역을 확보한다. 즉, 가슴영역은 (그림3) 과 같이 검출된 얼굴사이즈를 통해 가슴영역을 확보한다.



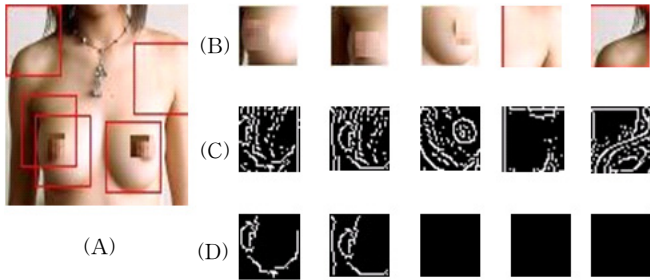
(그림 2) 검출 알고리즘 흐름도



(그림 3) 음란이미지 중 가슴을 찾는 과정

확보된 가슴 영역에서 미리 확보한 가슴 이미지와 템플릿 매칭을 시도한다. 템플릿의 사이즈는 얼굴사이즈에 적당한 크기로 설정 했고, 매칭된 영역에서 오검출 영역을 줄이고, 정확한 가슴영역을

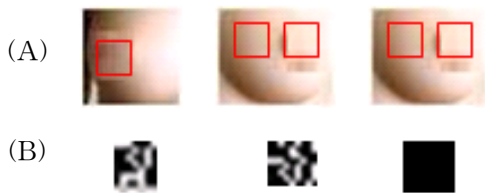
확보하기위해 (그림4)와 같이 질감 검사를 실시한다. 매칭된 가슴영역은 에지로 변환하여 외곽선 검출을 실시한다. 이 때 픽셀값이 정해놓은 Threshold 값을 넘어간다면 가슴영역이라 판단한다.



(A)가슴영역 (B)템플릿매칭 결과 (C)캐니엣지 변환 (D)외곽선 검출

(그림 4) 질감 검사

질감 검사와 피부색 필터링이 된 가슴 영역에서 최종적으로 미리 확보한 유두(nipple) 이미지를 템플릿 매칭한다.(그림5) 와 같이 매칭된 유두영역을 캐니에지로 변환한 뒤, 정해놓은 임계치 값보다 픽셀의 평균값이 높다면 최종적으로 음란이라고 판단한다.



(A) 매칭된 유두 이미지 (B) 에지로 변환된 이미지
(그림 5) nipple Template Matching

4. 실험 및 분석

4.1 실험 및 결과

인터넷을 통해 음란이미지와 비음란 이미지 총 200여장으로 테스트했다. 음란이미지는 얼굴과 가슴이 노출된 사진으로 구성했고, 비음란 이미지는 스포츠, 연예, 일반인 등 각종 음란과 관련 없는 사진, 그리고 음란 이미지와 유사한 이미지(비키니, 운동선수)등으로 구성하였다.

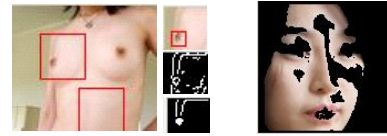
<표 1> 실험 상세 결과표

Data set	Bust size	skin color	Threshold	canny	contour	비고
1	0.7	2.0	0.85	50(60)	50	○
2	0.7	2.0	0.85	50(60)	50	○
3			0.85			○

실험결과 음란 이미지는 80개의 검출과 20개의 오검출로 80% 정도의 확률이며 정상 이미지는 92개의 검출과 8개의 오검출로

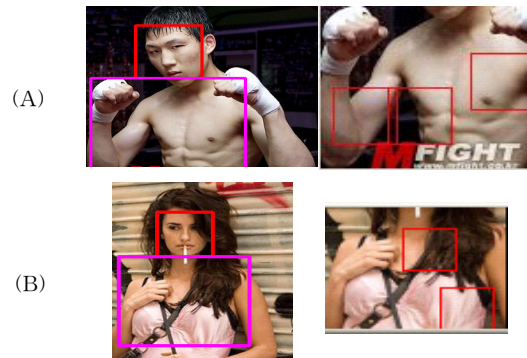
92%의 확률 결과가 나왔다.

(그림 6)에 (A)와 같이 윤곽선이 흐릿한 사진은 좀 더 높은 threshold를 설정해야 탐지가 되고, (B)는 조명으로 인해 피부색이 변질되어 정해놓은 피부색 수치를 넘지 못할 경우 얼굴이라고 판단하지 않는 오류를 발생한다.



(A) 35번 Image (B) 54번 Image
(그림 6) 검출 실패 이미지

(그림 7)은 정상 이미지에서 오검출한 이미지이다. (A) 사진은 여성의 가슴 이미지를 템플릿 매칭 하였지만, 피부색, 근육으로 인한 가슴윤곽, 꼭지 등으로 조건이 맞아 가슴으로 판별되었다. (B) 사진 또한 피부색, 머리카락으로 인한 윤곽 등으로 인해 오검출한 경우이다.



(그림 7) 오검출 이미지

실험 결과의 정확도는 이미지(정상, 음란)/ 검출이미지 * 100으로 86 %이며, 에러율은 14% 로 나왔다. 14% 중에 음란사진을 정상사진으로 찾은 경우는 10%, 정상사진을 음란사진으로 찾은 경우는 4%였다.

<표 2> 음란 이미지 탐지 실험 결과

원본 \ 결과	단위 : 장	
	정상	음란
정상	92	8
음란	20	80

$$\text{정확도} = \frac{92+80}{92+80+20+8} * 100 = 86\%$$

5. 결과

본 연구에서 음란 이미지를 자동으로 인식하는 기법에 대하여 언급하였다. 본 논문에서 제안한 방법은 기존의 방법들처럼 알려진 음란이미지의 시그니처를 이용하는 방법이 아니고, 알려지지 않은 음란 이미지를 탐지하는 방법이다. 제안된 알고리즘의 핵심은 패턴매칭 기반의 음란이미지 탐지이다. 제안된 기법의 유용성을 검증하기 위해 음란물 여러 개를 무작위로 인터넷에서 수집하여 실험을 실시하였으며, 제안된 알고리즘의 실험결과는 대체적으로 우수한 성능을 보였다.

그러나 제안된 알고리즘은 다음과 같은 상황에서는 음란물을 판단하는데 어려움이 있다. 첫째 얼굴이 없는 음란 이미지는 인식하는 경우, 둘째, 가슴부위가 가려진 음란이미지 인식에도 어려움이 있다. 셋째, 확보한 음란이미지의 질이 나쁜 경우 추후 개선해야하는 연구 방향은 이러한 문제를 해결할 수 있도록 알고리즘을 확장하는 것이다.

참고문헌

- [1] “인터넷 음란물 접촉과 청소년 성의식간의 관계”, 최정순, 2009
- [2] “내용기반 유해정보 방지기술 개발에 관한 연구”, ETRI, 2005
- [3] “피부색상을 이용한 유해 영상 분류기 개발”
윤진성, 숭실대 컴퓨터학과, 2009
- [4] “Adaboost face Detection system Based on
skin-color Filtering and Normalization Methods
for Face Recognition”, 김중규, 2005
- [5] “템플릿 매칭에 기반한 잡음에 강인한 악보인식”, 유재명, 2009
- [6] “Skin Segmentation Using Color Pixel Classification:
Analysis and Comparison”, Son Lam Phung, 2005
- [7] “색상 특징 정보와 질감 정보를 이용한 이미지 검색방안”
이인갑, 청운대학교 정보산업대학원, 2005
- [8] “청소년의 인터넷 음란물 이용 행태와 대처방안에 관한 연구”,
김덕모, 언론과학연구, 2003
- [9] “아이디부스트(Adaboost)와 원형기반함수를 이용한 다중표적 분류기법”,
김재협 외 3, 대한전자공학회, 2010.5, pp. 22-28
- [10] “얼굴 특징점 검출 및 응용프로그램”, 신길수, 세종대학원, 2007
- [11] “지형공간정보를 이용한 해안선 변화의 자동추출”
김용석, 동아대학원, 2007