論文
3-5-9

2단계 분류기법을 이용한 영상분류기 개발
(A Study on development for image detection tool using two layer voting method)

김 명 관*
(Myung-Gwan Kim)

요 약

영상물에 대한 학습과 분류를 위해 단순 베이지안, N-Nearest 방법 등이 사용된다. 이 방법들은 단순하면서 높은 정확도를 갖는다. 본 논문에서는 2단계 투표를 통해 이들 방법들을 조합하여 사용하였다. 유해 영상물들을 대상으로 학습 및 분류를 실험하였다. 결과로 색상분포에 따른 영상 분류가 실시간 처리 및 유해 영상 인식에 효과적임을 보였다. 또한 2단계 투표 방식의 알고리즘으로 약 2000장 이상의 사진을 가지고 학습 및 분류를 시행했으며 결과 80%에 가까운 높은 정확도와 대상 사진에 영향 받지 않는 안정도를 보였다.

키워드 : 유해사진 탐지, 자동 분류, 영상 분류

ABSTRACT

In this paper, we propose a Internet filtering tool which allows parents to manage their children's Internet access, block access to Internet sites they deem inappropriate. The other filtering tools which like Cyber Patrol, NCA Patrol, Argus, Netfilter are oriented only URL filtering or keyword detection methods. Thease methods are used on limited fields application. But our approach is focus on image color space model. First we convert RGB color space to HLS(Hue Luminance Saturation). Next, this HLS histogram learned by our classification method tools which include cohesion factor, naive baysian, N-nearest neighbor. Then we use voting for result from various classification methods. Using 2,000 picture, we prove that 2-layer voting result have better accuracy than other methods.

Keyword : Machine classification, injurious Image Detection, Voting method

1. 서 론

인터넷의 확산은 전 세계를 하나의 문화와 정치, 생활권역으로 만들어 가고 있다. 인터넷을 통한 비즈니스는 엄청난 시장 규모를 기록해 가고 있으며 앞으로 더욱 확대될 추세이다. 현재 인터넷 비즈니스의 절반 가량이 유해영상을 게시하고 회원제로 돈

* 정회원 : 서울보건대학 전산정보처리과

논문접수 : 2002. 4. 3
심사완료 : 2002. 4. 26

을 받는 사이트들이다. 이들 사이트들은 대부분 미국과 일본, 유럽에 위치하고 있으며 경제성으로 인해 해마다 급증하는 추세이다. 세계적으로 이런 사이트들에 대한 청소년들의 해악을 방지하기 위해서 각종 정부규제나 인터넷 내용 등급제 등을 실시하고 있으며 각종 음란 유해정보 차단 소프트웨어들도 상용화되어 있다. 특히 1997년 7월에 독일 본에서 개최된 광역정보망(Global Information Network) 장관회의에서 29개 유럽국가 장관들이 내용등급에 관한 원칙에 합의한 상태이다[1].

기존의 유해 정보 차단 기법에는 사이트 등급(PICS, Platform for Internet Content Selection)에 의한 방법, 블랙 리스트 기법, 화이트 리스트 기법, 해당 사이트의 키워드에 대한 분석기법 등이 있다. 그러나 이들 기법들은 많은 제약을 가지고 있다.

본 논문에서는 기존의 기법들과는 다르게 유해 이미지 자체를 분석하므로 국적이나 언어 등과 독립된 방지 효과를 얻을 수 있도록 하였다. 또한 통계학에 기준 한 상관계수법, 확률론에 따른 베이지안, 거리에 따른 분류인 K-nearest neighbor 분류기를 사용하여 각각의 분류 정확도를 살펴보았다. 실험 결과 앞의 분류기(classifier)들을 1 단계 분류기로 놓아 복합적인 수치를 만들고 2단계에서 투표(voting)케 함으로서 분류 정확도를 높일 수 있었다. 2장에서는 기존의 유해 정보 차단 기법들과 문제점을 살펴보고 3장에서 유해 영상들의 특성에 대하여 기술하였다. 4장에서 각종 분류기의 내용과 실험결과들을 기술한다.

2. 기존 유해 정보 방지 기법들

기존 유해 정보 차단 기법들 중 대표적인 방법은 PICS(Platform for Internet Content Selection)이다. PICS는 웹사이트 내용에 대해 선택적으로 접근하도록 해주는 기반 구조로서 필터링 소프트웨어와 등급 서비스간의 원활한 연동을 도와주는 기술 규격이다. 이를 구현한 시스템으로는 RSACi(Recreational Software Advisory Council on the Internet)과 세이프 서프가 대표적이다. RSACi는 게임 소프트웨어의 등급을 정하기 위해 출범된 모임으로서 4개 분야(섹스, 신체노출, 어투, 폭력)에 대한 기준을 정하고 있다. 세이프 서프는 95년 개발된 것으로 부모나 교사들이 어린이들에게 안전한 인터넷 사용을 만들어주기 위해 마약 도박 등 11개 분야에 대해서 1에서 9까지 등급을 매겼으며 상당히 세밀하게 구성되어 있다. 현재 마이크로소프트의 익스플로러는 이 PICS와 호환된다. 이 밖에 유럽의 INCORE, 독일의 ECO 포럼과 차일드 넷 인터네셔날(Childnet International), 오스트리아 ABA 등이 국제 내용등급 작업그룹(International working group on content rating)을 결성하고 있다[1]. 그러나 이런 등급 제에 의한 방지 기법들은 나라별 사정에 따라 다르며 강제적인 규제를 할 수 없는 사이버 공간에서 완전히 유해 사진을 방지하기에는 역부족이다[17].

다음의 방법으로 '유해물 방지 프로그램' 이 있다. 이들 프로그램으로는 컴퓨서브사의 '아동을 위한 인터네트 박스, 서브 워치, 넷 내니 같은 것들이 있다. 이 프로그램들은 크게 두 종류로 나눠지는데 하나는 유해 단어나 텍스트를 미리 입력해 두었다가 청소년들이 데이터를 검색할 때 이들 단어가 나오면 인터네트 접속이 자동으로 차단되도록 한 것들이고, 다른 하나는 미리 유해 사이트의 리스트를 구해 이들 사이트에 원천적으로 들어갈 수 없도록 하거나 들어갈 수 있는 사이트의 리스트를 운용하며 이 사이트들 이외에는 못 들어가도록 하는 것이다. 그러나 이 소프트웨어들에는 한계가 있다. 새로 생긴 사이트나 신종 유행어 따위를 계속적 모니터하고 감시할 수는 없기 때문이다. 또한 단어에 의한 차단은 다양한 국가들의 언어로 유해 사이트들이 만들어지고 있으므로 한계를 지닌다. 여기다가 뉴스 그룹이나 FTP 사이트 등 설명하는 문장 없이 오직 사진만으로 이루어진 경우 막을 방법이 없다.

이밖에 변칙적인 방법으로 사진을 계속 다운로드 받으면 일단 의심하고 차단하는 소프트웨어들이 있다. 또한 넷 필터(www.netfilter.net)처럼 기존의 웹브라우저를 프록시로 설정해놓고 모든 URL요청에 대하여 프록시 서버인 넷 필터에 의해 유해 정보를 차단하는 방법이 있다. 이들 방법들도 개인의 웹 사용을 제약하므로 바람직한 방법으로 볼 수 없으며 역시 모든 유해 사이트들을 모니터 할 수 없다는 한계를 지니고 있다. 국내에 나와있는 제품들의 성격과 내용은 다음 표와 같다.

<표 1> 국내 유해정보 차단 제품들

<Table 1> Injurious site detection S/W

	언어기반	DB 기반	복합방식
판매 중인 제품명	액티브웹-캐시	수호천사, 넷피아 브라우저, Spoon, 웹 매니저, 지킴이, ContentsFilter	안티 엑스, Web Keeper, 맘씨, 녹스2, 웹 몬스터, 컴지기, 아이보안관
문제점	그림 중심의 사이트는 탐지가 어렵다.	모든 유해 사이트 주소를 파악할 수는 없다. 특히 미국, 일본이 아닌 도메인은 탐지가 어렵다.	인터넷 및 컴퓨터 사용에 제약을 가해 사용에 어려움이 있다.

위와 같이 내용을 분석하여 키워드 추출 중심의 언어기반, 유해사이트 목록 또는 합법사이트 목록을 유지하는 DB 기반의 기법, 사용시간이나 시간대, 키워드, DB 등을 복합적으로 사용하는 제품들이 있다. 그러나 이들 제품들은 영상 위주의 사이트에 취약하거나(언어기반), 계속적으로 DB를 업그레이드해 주어야 하거나(DB 기반), 사용상의 제약이 있거나(복합 방식)하는 약점을 가지고 있다. 따라서 색상을 기반으로 유해 영상을 차단하는 기법은 위의 단점에 좋은 대안이 될 수 있다.

3. 색상을 기반으로 한 유해 사진 분석

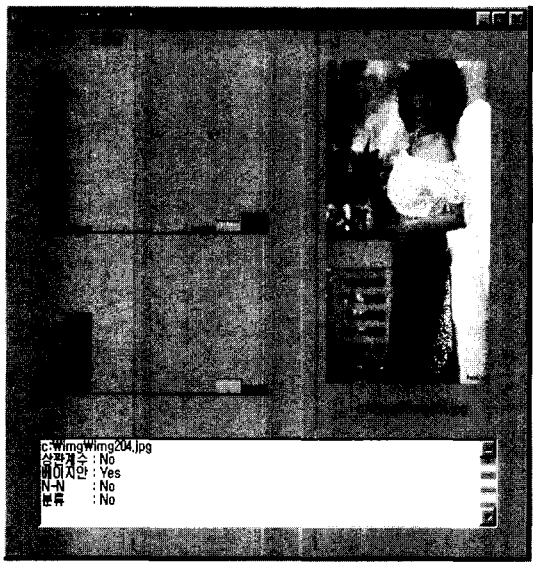
최근 영상 인식 분야에서 색상을 기반으로 하는 인식 알고리즘의 시도가 많은 장점을 갖는 것으로 알려지고 있다. 기존의 방법에서는 명암으로 대상 영상을 변화시키고 이중 에지를 추출하여 대상 영상을 인식하였다. 그러나 이 에지 추출 시 잡음에 민감하고 처리 과정의 계산이 시간을 많이 소모하여 실시간 처리에 어려움이 있다. 반면 색상을 기반으로 하는 영상 인식은 주로 얼굴을 추출하는 응용에 많이 사용되고 있다[3][4]. 이 경우 기존에 방법에 비해 에지 추출 등의 과정이 없고 단순한 색상 히스토그램을 사용하므로 실시간 처리에 강하다. 또한 색상 과학(Color Science)로 부터 인체 부위에 대한 인식은 기존의 방법들 보다 잡음에 강하다는 실험 결과가 있다[2]. 본 연구는 주로 여성의 누드 사진이 대상이므로 이와 같은 색상 분포에 따른 응용이 적합한 것으로 보인다.

실험 영상은 RGB로 구성되며 이 RGB 값을 HSL(Hue, Saturation, Luminance)로 바꾸었다. 다음은 RGB에서 HSL을 구하기 위한 간단한 알고리즘이다[4][5].

```
max = R;
if(max<G) max = G; if(max<B) max = B; min
= R;
if(min>G) min = G; if(min>B) min = B;
L = (max + min)/2.0;
if( max == min) { S=0; H=UNDEFINED; }
else {
   delta = max - min;
   if( L < 0.5) S= delta / (max+min);
        else S=delta / (0.2 - max - min);
   if(R == max) H = (G-B) / delta;
        else { if(G==max) H = 2.0 + (B-R) / delta;
           else { if( B==max ) H=4 + (R-G) / delta; }
        }
   H *= 60.0; if (H < 0.0) H += 360.0;
}
```

여기에서 L 값 즉, 밝기는 사진마다 촬영시의 밝기가 다르고 대상 사진 인식과 관련이 없으므로 사용치 않는다[11]. 학습 대상 사진들의 특성이 HSL로 변화한 후 두드러지게 나타났다. 분포도를 이용한 결과 채도인 S 값은 유해 사진을 구별하는데 별로 영향을 주지 않는 것을 알 수 있었다. 즉 색상 Hue에 분포가 유해 사진들과 비유해 사진들에 있어서 많이 다름을 알 수 있었다. [그림 1]에서는 구현된 유해 사진 분류기의 화면 모습을 보여주고 있다. 화면에 보이는 사례는 대표적인 유해 사진과 비유해 사진의 색상 분포를 그래프로 보여주고 있다. 색상의 구분은 360 가지로 나누었으며 그림의 왼쪽 위의 히스토그램은 3000여장의 사진을 학습한 후의 분포 모습이며 왼쪽 아래 히스토그램은 오른편 모델의 전

신사진을 사용하여 얻은 색상 분포 그래프이다.(옷을 입은)



[그림 1] 유해 영상 분류기 모습

[Fig. 1] Display for injurious picture classification

따라서 본 연구에서는 True color인 대상 사진을 360 레벨의 색상 분포만으로 분류 작업을 할 수 있었다. 아래 문자 상자에 나타난 값은 상관계수, 베이지안, N-N 분류 기법으로부터 얻은 (No, Yes, No) 값이며 결과로 'NO'가 나왔다. 따라서 위 그림은 유해 영상이 아닌 것으로 판단한다.

4. 분류 기법 및 실험 결과

본 논문에서는 유해 사진을 인식하기 위하여 인터넷 상에서 가장 많이 사용하는 JPEG 파일들을 기반으로 하였다. 이 실험을 위해 모두 5000여장의 사진 데이터베이스를 구축하였다. 이중3000장을 학습용 데이터로 사용하고 1000여장을 분류 확인용으로 사용하였다. 또한 유해성이 없는 사진을 1000여장 준비하였다. 유해성이 없는 사진은' 30%가 탤런트 및 모델들의 사진으로 30%는 운동 경기중의 운동선수들의 모습으로 30%는 산, 바다, 비행기, 군함 등 전혀 연관이 없을 사진으로 구성하였다.

본 실험에서 사용한 분류 기법(classification method)은 상관계수법[6], 베이지안[7], Nearest-Neighbor[8] 등 3가지이다. 상관계수는 통계학에서 사용하는 두 표본 사이의 상관도를 구하는 방법이다. 상관계수(Correlation)는 다음과 같이 구한다. 여기에서 posi(i)는 실험 사진의 색상 분포를 img_freq(i)는 학습되어 있는 유해 사진들의 색상 분포를 나타낸다.

```
For i = 0 To max_arry - 1
  sum_xy = sum_xy + img_freq(i) * posi(i)
  sum_x = sum_x + img_freq(i)
  sum_y = sum_y + posi(i)
  exp_x = exp_x + img_freq(i) ^ 2
  exp_y = exp_y + posi(i) ^ 2
Next i
  Crr=((max_arry*sum_xy) - (sum_x * sum_y))
/          (Sqr(max_arry*exp_x-sum_x          ^
2)*Sqr(max_arry* exp_y - sum_y ^ 2))
```

베이지안은 대표적인 사전확률과 사후확률을 이용한 분류 기법[13]으로서 본 연구에서는 단순 베이지안 분류기법(naive bayesian classification method)을 사용하였다[9]. 이 방식은 실험 사진에 있는 색상 각각에다가 학습된 해당 색상의 확률들을 계속 곱해나가는 것이다.

```
For j = 0 To posi(i)
  prob = prob * (img_freq(i) / tot_count)
Next j
```

3번째 방법은 k-nearest neighbor 방식으로서 학습되어 있는 각각의 그림들을 하나의 벡터로 보고 유클리드 기하학에서의 거리를 구하여 가장 가까이 있는 벡터의 분류에 따라 분류를 결정하는 방식이다[9]. 다음과 같이 구현하였다. scale1과 scale2는 학습된 데이터와 실험 데이터의 규모를 맞추어주는 변수이다.

```
For i = 0 To max_arry - 1
  k_near = k_near + (img_freq(i) / scale1 -
posi(i) / scale2) ^ 2
Next i
  k_near = Sqr(k_near)
```

<표 2> 각 방법에 따른 정확도

<Table 2> Accuracy for each methods

	학습된 유해 영상	학습 안된 유해 영상	비유해 영상	평균
상관 계수법	30%	40%	92%	54%
Naive Bayesian	78%	78%	56%	71%
Nearest Neighbor	94%	94%	65%	84%
2-Layer Voting	78%	76%	80%	78%

실험결과는 학습 대상 유해 영상과 학습하지 않은 유해 영상, 비유해 영상을 사용하여 표 2.과 같이 얻었다.

<표 2>의 결과를 보면 상관계수법, Naive Bayesian, Nearest Neighbor 등은 대상 영상에 따라 정확성이 상당히 차이가 나는 것을 알 수 있다. 그러나 본 논문에서 사용한 2-Layer Voting 방식은 대상 영상과 상관없이 안정된 정확도를 보이고 있다.

5. 결론 및 향후 연구 방향

본 논문에서는 유해 영상을 구별하기 위한 유해 사진 분류기 개발에 관한 내용을 다루었다. 인터넷의 확산과 무분별한 인터넷 비즈니스의 성행에 의한 유해 정보는 아무 제약 없이 청소년들에게 노출되어 있는 현실이다. 이를 방지하기 위하여 블랙 리스트를 관리하는 기법, 키워드를 사용한 필터링 기법 등이 제안되고 있으나 역부족인 현실이다. 본 논문에서는 유해 영상을 대상으로 한 직접적인 분류 기법을 다루었다. 즉, 대상 영상을 RGB 색상 공간에서 HLS(Hue Luminance Saturation)으로 바꾸어 필요 없는 밝기 정보 등을 제거하였다. 실험 결과 색상정보인 Hue 만 가지고 유해영상 분류를 성공적으로 할 수 있음을 보였다.

실시간 처리에 유리하며 정확도가 높은 상관계수법, Naive Bayesian, Nearest Neighbor 기법 등을 사용하였다. 이와 같이 3가지 방법을 동시에 테스트하여 얻은 결과를 분석하였으며 각 방법의 추정 결과를 하나의 단계를 더 두어서 모아진 결과를 가지고 투표하는(Voting) 기법을 사용하여 보았다. 이 방법을 2-Layer Voting 이라고 부르며 이 기법으로 실험해본 결과 대상 영상과 상관없이 안정된 정확도를 보여 주었다.

이와 같은 다층 분류기법은 다양한 분류에 응용될 수 있으며 특히 멀티미디어 데이터의 학습 및 분류에 적합할 것으로 보인다. 앞으로 좀더 다양한 학습 기법들을 실험해 보고 대상 데이터에 대해 적합한 학습 기법을 찾아볼 필요가 있다. 유해 정보 차단을 위해서는 키워드 기법, 블랙리스트 기법 등과 상호 보조한다면 더 나은 정확도를 보일 수 있을 것이다. 또한 비디오 영상과 같이 프레임이 다수인 경우 위와 같은 Voting 기법이 더 효율적일 것이다. 본 결과물은 (주) 에이전텍과 산학협력으로 개발하였으며 상품화 가능성이 높다는 것을 실제 실험을 거쳐 입증하였다.

※ 참고문헌

[1] 홍성명, "인터넷 유해 정보 방패막 세우기", 월간 인터넷, 1997, 11

[2] 유태웅, 오일석, "색채 분포 정보에 기반 한 얼굴 영역 추출", 정보과학회 논문지(B) 제 24권 제 2호, 1997, 2

[3] R, Baeza-Yates, "Modern Information Retrieval", 345-363, Addison Wesley, 1999

[4] Gose, "Pattern Recognition and Image Analysis", PTR, 1996

[5] I. Pitas, "Digital Image Processing Algorithm", Prentice Hall, pp 2-40, 1995

[6] David B. Skalak, "Prototype Selection Composite Nearest Neighbor Classifiers", Ph.D Thesis of UNION College, 1997

[7] I. Kononenko, "Comparison of inductive and naive Bayesian learning approaches to automatic knowledge acquisition" Current trends in knowledge acquisition, Amsterdam IOS Press, 1990

[8] T. Mitchell, "Machine Learning", McGraw-Hill, 1997

[9] M. Fayyad, "Advances in Knowledge Discovery and Data Mining", MIT Press, 1996

[10] Sestito, Dillon, "Automatic Knowledge Acquisition", Prentice Hall, 1994

[11] Y. Gong and M. Sakauchi, "Detection of regions matching specified chromatic features", Computer vision and Image Understanding, Vol 61, No 2, pp 263-269, 1995

[12] E. Alpaydin, "GAL : Networks that Grow When They Learn and Shrink When They Forget", International Computer Science Institute, Berkeley: CA, TR-91-032

[13] D. Fisher, "Knowledge Acquisition via Incremental Conceptual Clustering", Machine Learning, 2, 139-172, 1987

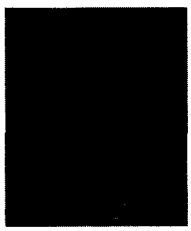
[14] D. W. Aha, "Instance-Based Learning Methods", Machine Learning, 6, 37-66, 1991

[15] C. D. Elliott, "A Process model of emotions in a multi-agent system", Ph.D thesis, north-west Univ., 1992

[16] 장동혁, "디지털 영상처리의 구현", 95-124, 정보게이트, 2001

[17] 정보통신윤리위원회, "인터넷 내용등급 서비스", http://www.icec.or.kr/, 2002

김 명 관



1985. 숭실대학교 전자계산학과 졸업(학사)

1987. 숭실대 대학원 전자계산학과졸업(석사)

1989-1993 한국전자통신연구소 인공지능연구실 연구원

1996 - 숭실대학원 박사과정 수료

1993 - 현재 서울보건대학 전산정보처리과 조교수

관심분야 - 에이전트, 기계학습, 자연어처리