

음란콘텐츠에 기반한 유해 음란 사이트의 차단

정회원 조 동 옥*

Blocking of Internet Harmful Pornographic Sites by Contents-based Method

Dong-Uk Cho* *Regular Member*

요 약

본 논문에서는 인터넷 역기능에 있어 가장 큰 문제로 인식되고 있는 유해 음란 사이트를 기술적으로 차단하기 위한 방법론을 제안하고자 한다. 현재까지의 기술적 차단 방법이 주로목록 기반과 단어 기반 방법이었기 때문에 새로이 생겨나는 사이트나 단어를 교묘히 변경한 경우 음란사이트 차단이 제대로 되지 않았다. 이를 위해 본 논문에서는 음란물의 내용이 될수 있는 즉, 음란물에서 성행위시 나오는 신음소리와 음란 영상을 추출하여 음란 사이트를 차단하고자 한다. 이를 위해 질감 분석, 곡선 적합등을 통해 영상에서 성적 주요 부위를 추출하였으며 프라트를 이용하여 음향 신호를 분석한후 자기 상관 함수로 음란 음향 신호를 차단하였다. 최종적으로 실험에 의해 본 논문의 유용성을 입증하고자 한다.

Key words : pornographic sites, texture, curve fitting, pratt

ABSTRACT

This paper proposes on the technical blocking method of Internet harmful pornographic sites which is the most Internet negative-function. At Present, most technical methods based on web sites back-searching or words filtering for blocking the pornographic Internet sites have limitations. For this, this paper proposes the acoustic and image based blocking method for filtering harmful Internet sites. For this, sexual main body parts are extracted by texture analysis and curve fitting. Also acoustic signals are analyzed using pratt tool and auto-correlation function is adopted for matching between prototype signals and test signals. Finally, the effectiveness of this paper is demonstrated by several experiments.

I. 서론

제3세대 IT기술의 핵심 인터넷은 우리의 생활 패러다임을 변화 시킬 정도로 순기능이 있는 반면 그 역기능 또한 기존에 오프라인상에서는 상상도 할 수 없을 정도여서 이에 대한 대책이 시급한 실정이다[1]~[3]. 주부 탈선과 원조 교제의 장이 되고 있는 채팅에서부터 시작해서 스팸메일, 전자메일 바이러스, 웹 자유 게시판에서 행해지고 있는 부적절

한 행동 그리고 70만개에 육박하는 전 세계 음란, 도박, 자살사이트 등 유해 정보 콘텐츠의 무분별한 유포 등이 인터넷 역기능의 주요 사례가 될 수 있다. 이 중 웹 자유게시판에서 행해지는 부적절한 행동은 유해단어를 필터링[4]하거나 비속어 처리 프로그램을 개발[5]하는 기술적 접근 외에 실명제의 도입, 회원 삼진아웃제 도입[6]등과 같은 정책적, 운영적 기법의 도입까지 행해져 그 역기능 해소에 주력하고 있다. 그러나 가장 큰 사회적 문제는 음란사이

* 충북과학대학 정보통신학과

논문번호 : 030582-1226, 접수일자 : 2003년 12월 26일

트의 무분별한 배포 행위와 채팅에 관한 문제 일 것이다. 이중 채팅은 기술적으로 차단 단어 설정과 차단 메뉴 리스트를 만들어 처리해야 할 것으로 여겨 지지만 음란 유해 사이트는 기술적으로 어떠한 방법으로 해결하는 것이 효과적인지 많은 논란이 되고 있다. 특히 현재 한글로 된 음란사이트가 세계 2 위이며 그 증가율은 세계 1위여서 그 심각성이 날로 더 커져가고 있으며 심지어 초등학교생에게도 음란사이트를 소개하는 메일이 전달되고 있는 실정이다. 이를 위해 정부에서는 e-메일 주소 추출기에 대한 법적 규제강화, 해외에 서버를 빌려 한글로 영업하는 불법 음란사이트를 인터넷서비스제공업체(ISP)의 협조를 받아 국제 관문국 단계에서 국내 유입을 차단하는 등의 정책을 시행하고 있다. 또한 ‘주니어 e-메일’ 계정 서비스를 보급하고 음란사이트의 카드 결제를 막는 등의 방법으로 인터넷 역기능을 해결하고자 하고 있다.

정책적인 방법외에 기술적으로는 KT등이 월정액 제로 음란사이트를 차단하는 서비스를 시행하고 있는데 대부분 기술적으로 음란사이트를 차단하는 방법은 목록기반과 단어기반의 방법을 채택하고 있다. 그러나 현재의 실정으로 목록기반과 단어기반만으로는 음란사이트에 대한 차단이 완벽하게 이루어지지 않아 이에 대한 기술적 방법론의 개발이 시급한 실정이다. 이를 위해 본 논문에서는 내용에 기반한 방법을 제안하고자 한다. 즉, 음란물의 내용이 될 수 있는 성교시 나는 신음 소리를 통해 음란사이트를 차단한다든지 또는 음란 영상을 검출하여 차단하는 방법을 제안하고자 한다. 이를 통해 기존에 목록기반과 단어기반에서 필터링이 안되는 음란사이트를 추가로 차단하고자 한다. 즉, 기존 방법은 목록에 기초한 경우 새로이 생겨 나는 음란 사이트를 일일이 추적하는것이 쉽지 않았고 단어 기반의 경우도 단어들을 교묘히 변형하여 보냄으로 음란물 차단이 안되는 경우가 많이 존재했었다. 이를 위해 본 논문에서는 내용 기반 음란물 차단방법을 제안하고자 하며 실험에 의해 제안한 방법의 유용성을 입증하고자 한다.

II. 음란사이트를 차단키 위한 기존의 기술적 방법론에 대한 고찰

현재 음란물을 인터넷을 통해 유통시키는 것은 과거 오프라인상에서 출판물이나 비디오등을 통해 유통시키는것과는 비교도 안될 정도로 그 전파력이

엄청난다. 실제로 근래 “XX양...”에 관한 성행위 콘텐츠들은 인터넷을 통해 접해 보지 않은 사람들이 없을 정도이다. 아래 표 1에 음란 사이트 실태에 대한 표를 나타내었다[7].

이같이 청소년들에게 막대한 피해를 주는 음란 유해사이트의 차단은 대단히 중요한 과제가 아닐 수 없다. 이를 해결키 위해서는 정책적 차단법이 규정되었으나 사이트 운영자가 해외에 서버를 두고 있는 관계로 법집행이 어려운 실정이다. 결국은 기술적으로 차단해야 할 수밖에 없는 실정이다.

표 1. 음란 유해 사이트에 대한 실태

분석 항목	내 용
언어별 분류	영어사용(56만 4천개, 83.6%), 한글사용(6만 4천개, 9.5%) 일어사용(1만 5천개, 2.2%), 독일어(1.3%), 프랑스어(0.6%)
유형별 분류	음란사이트(66만 8천개, 98.9%) , 도박사이트(6천 900개, 1.0%) 엽기, 마약, 폭력, 자살사이트(0.1%)
요일별 유해 정보사이트 이용량 분석	월(14.8%), 화(11.5%), 수(9.8%), 목(11.8%), 금(11.9%), 토(20.5%), 일(19.8%)
최근 한글 유해정보 사이트 발현률	하루 평균 268개씩 생겨나고 있으며 이는 같은 기간 전 세계 출현 (하루 평균 58개)의 45%에 해당됨

결국은 기술적으로 차단해야 할 수밖에 없는 실정이다. 이를 위해 목록기반의 차단기술이 발전하게 되었다[8]. 그러나 이는 새로이 생겨 나는 음란사이트를 목록 DB에 수시로 업데이트 하는 것이 문제가 된다. 또한 단어를 통해 필터링하는 방법[9]은 음란 서버에서 보내는 단어가 주기적으로 교묘히 변형되고 또한 단어 자체적으로는 음란성을 나타내지 않는 등의 방법으로 단어 필터링을 피해 가는 상황이기 때문에 이를 추적하는 것도 어려운 실정이다. 결국 목록기반과 단어기반 필터링의 문제점을 보완 내지는 해결하기 위해서는 내용 기반의 필터링 방법이 강구되어야만 한다. 이를 위해 본 논문에서는 음향과 영상기반의 음란 유해사이트 차단방법을 제안하고자 한다. 아래 표 2에 현존하고 있는 음란 유해 사이트 차단 프로그램에 대한 비교분석을 나타내었다.

III. 음란 콘텐츠에 기반한 차단

본 논문에서는 음란 콘텐츠 즉, 음란물의 내용에 기반한 음란사이트 차단방법을 개발하고자 한다. 이는

기술적으로는 음향과 영상처리에 기반하여 음란 유희사이트를 차단하고자 하는 방법이다.

1. 영상 콘텐츠의 처리

1.1 영상분할

영상분할을 위해 텍스처 분석을 행한다. 텍스처 분석을 행하는 방법은 크게 구조적 방법(10), (11), 신호처리 방법(12)과 통계적 방법(13), (14)으로 구분된다. 본 논문에서는 텍스처 분석을 통계적 방법인 GLCM 방법중 변화에 민감하게 반응하는 모멘트와 동일 성질에 반응하는 동차성을 조합하는 방법

$$H = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{255 - |i-j|}{255} G[i, j] \quad (2)$$

이를 4x4 윈도우 크기로 연속 수행하여 두 개의 배열 [A]와 [B]에 각각 저장한다. 이때 배열 [A]와 배열 [B]에 대해 특정한 임계치의 값으로 임계치 작업을 행하며 이를 통해 배열 [A]와 배열 [B]에 대해 OR 작업을 수행한다.

이의 결과는 신체의 주요 성적 부위에 해당하는 영역이 되며 이를 Y축으로 프로젝션을 행하여 차

표 2. 유희 사이트 차단 국내 프로그램들의 비교분석

제품구분	지키미 2.31	파로스	i-boho	수호친사 2000	컴지기 2.0
업체이름	인터피아 월드	아이탑	일레아트	플러스 기술	인터넷보
배포형태	유료	유료	유료	유료	유료
운영체제	Win96, XP	Win98, XP	Win96, 2000, XP	Win98, 2000, XP	Win98, XP
차단방법	목록, 단어	목록, 단어	목록, 단어	목록, 단어	목록, 등급, 이미지
차단대상	음란, 폭력, 마약	음란, 폭력	음란, 폭력, 도박	음란, 폭력, 마약	음란, 폭력, 도박
자동연결	음란, 폭력, 마약	음란, 폭력	음란, 폭력, 도박	음란, 폭력, 마약	음란, 폭력, 도박
목록갱신	음란, 폭력, 마약	음란, 폭력	음란, 폭력, 도박	음란, 폭력, 마약	음란, 폭력, 도박
추가삭제	음란, 폭력, 마약	음란, 폭력	음란, 폭력, 도박	음란, 폭력, 마약	음란, 폭력, 도박
내역조회	음란, 폭력, 마약	음란, 폭력	음란, 폭력, 도박	음란, 폭력, 마약	음란, 폭력, 도박
시간설정	음란, 폭력, 마약	음란, 폭력	음란, 폭력, 도박	음란, 폭력, 마약	음란, 폭력, 도박

을 참조하여 빈도수를 통한 GLCM내의 값들의 정규화와 동차성의 정규화를 행하는 방법을 제안하여 신체의 주요 성적 부분의 영역을 추출하고자 한다. 우선 GLCM(Gray Level Co-occurrence Matrix) $G[i, j]$ 는 변위벡터 $d = (dx, dy)$ 를 결정하고 명도 i 와 j 를 갖고 d 만큼 떨어져 있는 화소들의 쌍을 계산하여 생성한다. 이때 전체 빈도수를 GLCM의 각 원소값들에 나누어줌으로써 $\sum G[i, j] = 1$ 이 되도록 조정한다. 따라서 이를 이용하여 모멘트 M 은 하식 (1)과 같이 구한다.

$$M = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 G[i, j] \quad (1)$$

또한 정규화된 동차성 H 는 하식 (2)와 같이 새로이 정의하여 그 값을 구한다.

후 형성된 히스토그램에 대한 분석을 행한다.

1.2 히스토그램 분석과 곡선 적합

이제 앞절에서 구한 히스토그램을 분석하여 곡선 적합을 수행하고자 한다. 곡선 적합시 여성과 남성의 신체의 주요 성적 부위는 2차곡선이나 일량분포로 나타낼 수 있으므로 이에 대한 곡선적합을 수행한다. 우선 $(M+1)$ 개의 자료 $\{(x_i, y_i) | i = 0, 1, \dots, M\}$ 를 $n (< M)$ 차의 다항식 $P_n(x) = \sum_{k=0}^n a_k x^k$ 에 근사시키는 곡선적합은 아래와 같다.

$$E = \sum_{i=0}^M (y_i - P(x_i))^2 \quad (3)$$

위 식을 최소화하는 a_0, a_1, \dots, a_n 을 구하기 위

해서는 (3)식에서

$$\begin{aligned}
 E &= \sum_{i=0}^M y_i^2 - 2 \sum_{i=0}^M P(x_i) y_i + \sum_{i=0}^M (P(x_i))^2 \\
 &= \sum_{i=0}^M y_i^2 - 2 \sum_{i=0}^M \left(\sum_{j=0}^n a_j x_i^j \right) y_i \\
 &\quad + \sum_{i=0}^M \left(\sum_{j=0}^n a_j x_i^j \right)^2 \tag{4} \\
 &= \sum_{i=0}^M y_i^2 - 2 \sum_{j=0}^n a_j \left(\sum_{i=0}^M y_i x_i^j \right) \\
 &\quad + \sum_{j=0}^n \sum_{k=0}^n a_j a_k \left(\sum_{i=0}^M x_i^{j+k} \right)
 \end{aligned}$$

(4)식을 최소화하려면 $j = 0, 1, \dots, n$ 에 대하여

$$\begin{aligned}
 \frac{\partial E}{\partial a_j} &= \left(-2 \sum_{i=0}^M y_i x_i^j \right) \\
 &\quad + 2 \sum_{k=0}^n a_k \sum_{i=0}^M x_i^{j+k} = 0 \tag{5}
 \end{aligned}$$

따라서 $(n+1)$ 개의 미지수 a_j 에 대하여

$$\sum_{k=0}^n a_k \sum_{i=0}^M x_i^{j+k} = \sum_{i=0}^M y_i x_i^j \tag{6}$$

이차곡선에 대해서는

$$\begin{aligned}
 a_0 \sum_{i=0}^M x_i^0 + a_1 \sum_{i=0}^M x_i^1 + a_2 \sum_{i=0}^M x_i^2 \\
 = \sum_{i=0}^M y_i x_i^0 \tag{7}
 \end{aligned}$$

$$\begin{aligned}
 a_0 \sum_{i=0}^M x_i^1 + a_1 \sum_{i=0}^M x_i^2 + a_2 \sum_{i=0}^M x_i^3 \\
 = \sum_{i=0}^M y_i x_i^1 \tag{8}
 \end{aligned}$$

$$\begin{aligned}
 a_0 \sum_{i=0}^M x_i^2 + a_1 \sum_{i=0}^M x_i^3 + a_2 \sum_{i=0}^M x_i^4 \\
 = \sum_{i=0}^M y_i x_i^2 \tag{9}
 \end{aligned}$$

따라서 a_0 는 아래와 같이 구할 수 있다.

$$a_0 = \frac{\begin{vmatrix} \sum y_i x_i^0 & \sum x_i^1 & \sum x_i^2 \\ \sum y_i x_i^1 & \sum x_i^2 & \sum x_i^3 \\ \sum y_i x_i^2 & \sum x_i^3 & \sum x_i^4 \end{vmatrix}}{\begin{vmatrix} \sum x_i^0 & \sum x_i^1 & \sum x_i^2 \\ \sum x_i^1 & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{vmatrix}} = \frac{|D_0|}{|D|} \tag{10}$$

마찬가지로 a_1, a_2 로 (11), (12)것과 같이 하여 구할 수 있게 된다.

$$a_1 = \frac{|D_1|}{|D|} \tag{11}$$

$$a_2 = \frac{|D_2|}{|D|} \tag{12}$$

또한 일량분포에 대한 곡선적합은 $y = a + bx$ 에 대해 아래식과 같이 정리되어 일량분포에 대한 곡선적합을 수행할 수 있게 된다.

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{M \sum x_k^2 - (\sum x_k)^2} \begin{bmatrix} \sum y_k \\ \sum x_k y_k \end{bmatrix} \tag{13}$$

최종적으로 곡선적합 수식의 적용을 위해 아래와 같은 단계를 거친다. 우선 이차곡선 $f(x) = a_0 x^2 + a_1 x + a_2$ 에 대해 다음과 같은 작업을 수행한다. 이때 $\frac{df(x)}{dx} = 0$ 이 되는

x 의 값 $-\frac{a_1}{2a_0}$ 가 중앙에 있는지 여부와 기울기의 완만한 정도가 주요 평가 척도가 된다. 아래에 $x = -\frac{a_1}{2a_0}$ 이 중앙여부를 결정하는 함수식을 나타내었다.

$$\begin{aligned}
 \mu_p(x) &= 1, \frac{4}{5} * \left(\frac{\alpha + \beta}{2} \right) \leq x \leq \frac{6}{5} * \left(\frac{\alpha + \beta}{2} \right) \\
 &= \frac{x - \alpha}{\alpha' - \alpha}, \quad \alpha \leq x \leq \alpha' \\
 &= \frac{\beta - x}{\beta - \beta'}, \quad \beta' \leq x \leq \beta \tag{14}
 \end{aligned}$$

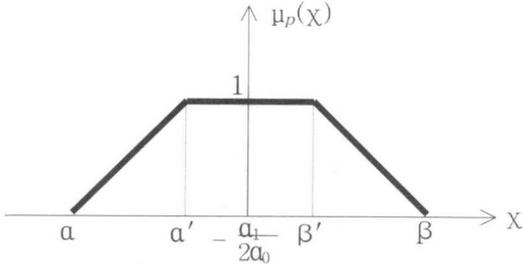


그림 1. 중앙여부를 결정하는 함수식

여기서 a 와 β 는 이차곡선의 시작과 끝점을 의미한다.

2. 음향 콘텐츠의 처리

이제 음란 콘텐츠에서 성행위시 발생하는 음향 콘텐츠를 처리하는 방법에 대해 제안하고자 한다. 이는 우선 음란사이트에서 나오는 성교시 신음 음향을 시계열 데이터로 변환하고 이를 2차원 스펙트럼으로 나타낸다. 차후 2차원 스펙트럼으로부터 특징을 추출하여 이를 표준 음향 신호와 상관 계수를 계산함으로써 음란 콘텐츠에서 음향 신호에 의한 음란사이트 차단을 행하고자 한다. 아래 그림 2에 음향신호 처리에 대한 전체 흐름도를 나타내었다.

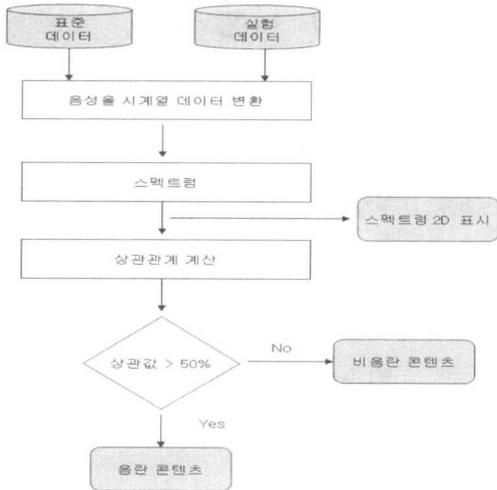


그림 2. 음향신호처리에 의한 음란 사이트 차단의 전체 흐름도

2.1 음향 신호 처리

주어진 음란 콘텐츠에서 음향신호를 처리하기 위해서는 표준이나 실험 음향 신호로부터 이의 신호

를 시계열 데이터로 변환하여 음향신호의 지문이라 할수 있는 스펙트럼을 생성해야 한다. 이를 위해 프라트 [15] 라 불리우는 음향 신호처리 툴을 적용하고자 한다. 프라트는 암스테르담대학의 Paul Boersma와 David Weenink가 만든 음성 분석 및 변형 프로그램 패키지인데 이것은 인터넷 홈페이지 [15] 에서 무료로 다운받을수 있다. 작업 환경은 프라트 객체창과 프라트 그림창으로 나뉘어 있는데 기본적으로 프라트 객체창은 모든 분석 작업을 통제하는 통제실과 같은 역할을 하고, 프라트 그림창은 분석 결과를 그림으로 그리고 동시에 필요한 문자와 화살표, 선등으로 그려 넣은뒤 그림 파일을 실을수 있는 기능을 가지고 있다. 아래 그림 3에 프라트의 설치 예를, 그리고 프라트의 read기능을 이용하여 샘플을 로드한 결과를 아래 그림 4에 나타내었다. 또한 draw기능을 이용하면 음성의 파형을 구할수 있으며 하단의 스펙트럼 기능을 이용하여 음성의 파형과 음향을 시계열데이터로 변환하는 것이 가능해 진다. 이를 아래 그림 5에 나타내었다.



그림 3. 프라트의 설치

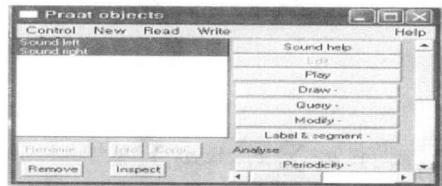


그림 4. read 수행의 예

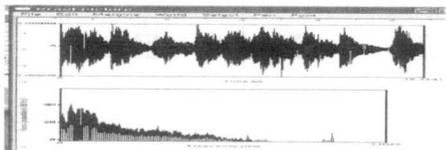


그림 5. draw기능 수행의 예

2.2 표준 스펙트럼신호와 실험 대상 스펙트럼 신호와의 상관 계수 계산

표준 스펙트럼 신호의 특징값과 실험 대상 스펙트럼 신호의 특징값의 상관 계수를 계산해야 해당 사이트가 음란 콘텐츠인지 아닌지를 계산할수 있다. 이를 위해 상관 값을 계산해 주어야 하는데 예를 들어 A신호와 B신호사이의 상관계수 값을 하식과 같이하여 구한다.

$$r = \frac{\sum_{m=1}^n A_m \bar{A} (B_m - \bar{B})}{\sqrt{(\sum_{m=1}^n A_m \bar{A}) (\sum_{m=1}^n (B_m - \bar{B})^2)}} \quad (15)$$

본 논문에서는 표준이 되는 음란 콘텐츠의 성교시 음향 신호를 우선적으로 10개를 정해 이를 표준 음란 음향신호로 정해 실험을 수행하고자 하며 차후 보다 다양한 형태의 음향 신호를 선정하고 분석하는 작업을 수행하고자 한다.

IV. 실험 및 고찰

1. 영상 콘텐츠에 대한 실험 결과 및 고찰

영상 콘텐츠에 대한 실험은 IBM-PC 상에서 C언어를 이용하여 행하였다. 우선 그림 6이 여성의 성적 부위 그리고 그림 7이 영역 분할 결과, 그림 8이 히스토그램 분석 결과이다. 마찬가지로 그림 9, 그림 12가 여성의 성적 부위 그리고 이에 대한 영역 분할 결과가 그림 10과 그림 13이며 히스토그램 분석 결과가 각각 그림 11과 그림 14이다. 남성의 성적 부위에 대한 실험 결과는 그림 15~그림 17에 나타내었다. 아울러 표 4에 각각의 신체 부위에 대한 곡선적합을 수행한 결과를 나타내었다. 실험 결과에서 알 수 있듯이 본 논문에서 제안한 방법이 남, 여 주요 성적 부위에 대한 영역 분할과 히스토그램 분석 그리고 이에 따른 곡선적합을 올바르게 행할 수 있음을 확인할 수 있다. 차후 보다 다양한 성행위 영상의 체위 형태 그리고 화상 채팅시의 성적 부위 노출과 자위 모습등 여러 유형의 음란 영상 콘텐츠에 대한 실험등도 추가로 행해지고 이에 대한 비교, 고찰등이 심도 있게 행해져야 하리라 여겨진다.

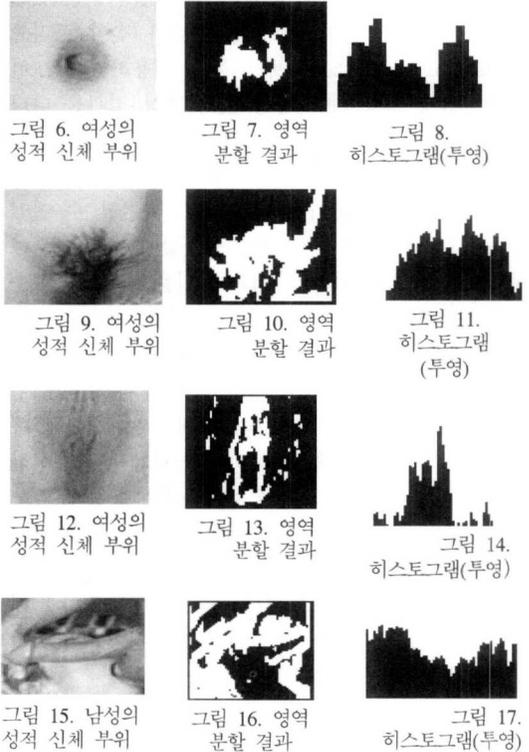


표 4. 곡선적합 결과

대상 영상	곡선 적합식
그림 6(여성)	$f(x) = -10.9975x^2 + 1.7047x - 0.0260$
그림 9 (여성)	$f(x) = -9.4905x^2 + 2.0014x - 0.0244$
그림 12 (여성)	$f(x) = 15.4450 - 0.1126x$
그림 15(남성)	$f(x) = 10.3231 - 0.1361x$

2. 음향 콘텐츠에 대한 실험 및 고찰

음향 콘텐츠에 대한 실험은 IBM-PC 상에서 C++을 사용하여 행하였다. 우선 인터넷 음란사이트에서 주로 많이 나오는 10개의 표준이 되는 실험샘플들을 wav 파일 포맷으로 변환하였다. 이때 10개의 표준 샘플은 성행위시 여성들이 흥분해서 내는 신음소리로 선정하였다. 남성보다 여성을 택한 이유는 남성들은 성교시 내는 신음소리가 주로 사정시로 한정되어 있는 반면에 여성들의 신음소리는 남성들보다 나오는 시간이 많기 때문에 이를 표준 음향신호로 택해야 차단이 용이하기 때문이다. 이때 각 샘플들은 미국, 일본, 한국 등 전 세계 포르노 관련 시장의 대부분을 차지하는 포르노 사이트를 대상으로 하여 보다 객관성있는 표준 샘플을 선정하고자 하였다. 아래 그림 18 ~그림 27에 10개의 샘플에

대한 음향신호와 이에 대한 스펙트럼을 나타내었으며 스펙트럼 생성은 앞서 언급한바와 같이 프라트(Pratt)라고 하는 소프트웨어 툴을 이용하였다 또한 아래 표5에 이에 대한 상관 값의 계산 결과를 나타내었다. 실험결과에서 알 수 있듯이 서로 다른 음향 신호 특성을 가지는 샘플에 대한 상관계수 값이어서 그 상관 값이 크지 않다는 것을 확인할 수 있었다. 이는 표준 샘플들이 서로 상관값이 작은 것, 다시 말해 서로 연관성이 적은 신음소리들을 표준 샘플로 삼았다는 것을 말해주며 실험 결과도 정확히 이를 입증해 주고 있다. 그러나 아래 표 6, 표 7, 표 8에서 알 수 있듯이 표준 샘플과 유사한 실험 대상(샘플 6과 샘플 7, 그리고 샘플 3)에 대해 음란콘텐츠의 음향신호를 분석하여 상관계수를 계산한 결과는 상관 계수값이 높게 나와 이는 음란 포르노 사이트라는 것을 정확히 지적해 주는 실험 결과임을 알 수 있었다. 향후 보다 다양한 표준 샘플을 선정하여 각종 음란사이트에서 나오는 음향신호에 대해 효율적으로 차단이 이루어지도록 하기 위한 연구가 지속되어야 하리라 여겨진다.

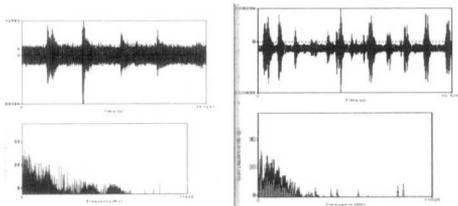


그림 18. 샘플1

그림 19. 샘플2

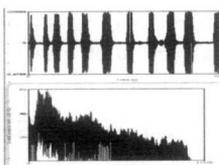


그림 20. 샘플3

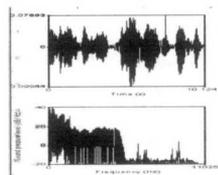


그림 21. 샘플4

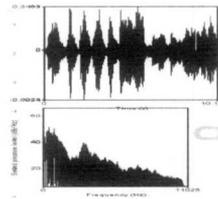


그림 22. 샘플5

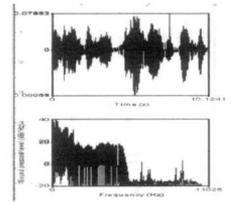


그림 23. 샘플6

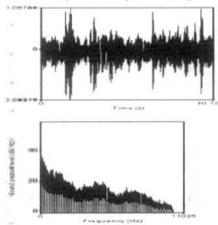


그림 24. 샘플7

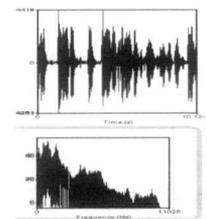


그림 25. 샘플8

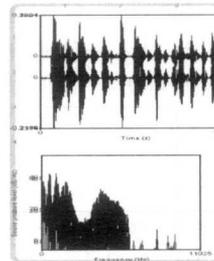


그림 26. 샘플9

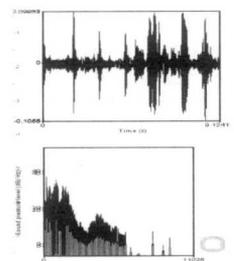


그림 27. 샘플10

표 4. 표준 샘플들 간의 상관 값

	샘플1	샘플2	샘플3	샘플4	샘플5	샘플6	샘플7	샘플8	샘플9	샘플10
샘플1	1.0000	0.0083	-0.0005	-0.0022	0.0037	-0.0018	0.0052	-0.0019	0.0094	-0.0052
샘플2	0.0083	1.0000	0.0005	-0.0024	-0.0056	0.0018	-0.0040	-0.0035	0.0041	-0.0033
샘플3	-0.0005	0.0005	1.0000	-0.0011	0.0001	-0.0029	0.0019	0.0016	-0.0017	-0.0010
샘플4	-0.0022	-0.0024	-0.0011	1.0000	0.0023	-0.0049	0.0007	0.0000	0.0033	-0.0015
샘플5	0.0037	-0.0056	0.0001	0.0023	1.0000	-0.0065	-0.0027	0.0030	-0.0004	0.0029
샘플6	-0.0018	0.0018	-0.0029	-0.0049	-0.0065	1.0000	0.0009	0.0030	0.0017	-0.0012
샘플7	0.0052	-0.0040	0.0019	0.0007	-0.0027	0.0009	1.0000	0.0045	0.0006	0.0008
샘플8	-0.0019	-0.0035	0.0016	0.0000	0.0030	0.0030	0.0045	1.0000	0.0072	-0.0009
샘플9	0.0094	0.0041	-0.0017	0.0033	-0.0004	0.0017	0.0006	0.0072	1.0000	0.0100
샘플10	-0.0052	-0.0033	-0.0010	-0.0015	0.0029	-0.0012	-0.0009	-0.0009	0.0100	1.0000

표 5. 실험결과 (샘플 6과 실험대상 A)

	샘플6	실험대상A
샘플6	1.0000	0.5236
실험대상A	0.5236	1.0000

표 6. 실험결과 (샘플 7과 실험대상 B)

	샘플7	실험대상B
샘플7	1.0000	0.5660
실험대상B	0.5660	1.0000

표 7. 실험결과 (샘플 3과 실험대상 C)

	샘플3	실험대상C
샘플3	1.0000	0.9285
실험대상C	0.9285	1.0000

V. 결론

본 논문에서는 인터넷 역기능에 있어 가장 큰 사회적 문제가 있는 음란사이트 차단 방법에 대해 제안하였다. 기존의 방법들이 목록과 단어에 기반하여 음란 유해 사이트를 차단하였는바 이는 새로운 내용을 목록기반 DB에 수시로 업데이트 해야 하는 문제점과 유해 단어가 주기적으로 교묘히 변경되는 문제로 말미암아 필터링이 올바르게 수행되지 않는 경우가 많았다. 본 논문에서는 이를 보완하기 위해 음란물의 콘텐츠에 기반한 방법을 제안하였다. 즉, 음란콘텐츠의 음향과 영상에 기반하여 음란 유해사이트를 차단하는 방법에 대해 제안하였다. 이를 위해 영상 기반의 경우 텍스춰 분석에 의한 영상분할, 히스토그램 분석, 곡선적합 등을 적용하는 방법을 제안하였으며 음향 기반의 경우 시계열 데이터 분석과 스펙트럼 생성 그리고 이에 따른 음란 콘텐츠 상관 계수값의 계산등에 대해 다루었다. 차후 다양한 실험 샘플에 대한 실험을 수행하여 제안한 방법의 상용화에 대한 연구가 지속적으로 행해져야 하

리라 여겨진다. 또한 이에 대한 병렬 처리 기법의 개발과 더불어 이를 확장하여 폰팅, 음란 화상 채팅, 음란 전화 차단등도 행할수 있는 시스템으로의 확장을 기해야 할 것으로 여겨진다.

참고 문헌

- [1] 추병완, “사이버 윤리의 정립과 방안”, 청소년보호 정책토론 자료집, 2001
- [2] 한국교육학술정보원, 교육기관 정보화 역기능 방지에 관한 연구, 방문사, 2000
- [3] 이재선, 전용희, “인터넷 등급 서비스를 이용한 효과적인 유해사이트 선별 기술에 관한 연구”, 한국정보처리학회 추계종합학술대회 논문집, 제9권, 제2호, 2002
- [4] 김치민, 김응근, “인터넷 게시판에서 정보통신윤리교육을 위한 유해 단어 필터링 시스템의 설계와 구현”, 한국정보처리학회 추계종합학술대회 논문집, 제9권, 제2호, 2002
- [5] 조아영, “웹 게시판 비속어 처리 프로그램의 설계 및 구현”, 한국컴퓨터산업교육학회 논문지, Vol. 2, No. 10, 2001
- [6] 중앙일보, 비방하는 글 몰살. 청와대 홈페이지 삼진아웃제 도입, 2003년 7월 5일자, 1면
- [7] 이상경, “국내 인터넷 이용 행태 및 조사 방법”, Telecommunication Review, Vol. 13, No. 3, 2003
- [8] 김재천, “인터넷 유해 사이트 차단 프로그램 분석 및 활용 방안”, 홍익대학교, 2001
- [9] 장운정 외 3인, “유해 문자 가중치를 이용한 유해 사이트 차단 방법”, 한국정보처리학회 춘계종합학술대회 논문집, 제10권, 제1호, 2003
- [10] R. C Gonzalez and R. E. Woods, Digital Image Processing, Addison Wesley, 1992
- [11] N. Ahuja and A. Rosenfeld, “Mosaic Models for Textures”, IEEE Trans. on PAMI, 1981
- [12] V. Alvestad, “Unsupervised Image Texture Segmentation with Optimized Filters”, Master’s thesis, stavanger, Norway, 1995
- [13] J. R. Parker, Algorithms for Image Processing and Computer Vision, John Wiley & Sons, pp. 150-175, 1997
- [14] H. C. Lin et al, “Extracting Periodicity of a Regular Texture Based on Autocorrelation Functions”, Pattern Recognition letters, 1997
- [15] http://www.fon.hum.uva.nl/pratt/download_win.html

조 동 욱(Dong-Uk Cho)

정회원



1983년 2월: 한양대학교

전자공학과(공학사)

1985년 8월: 한양대학교

전자공학과(공학석사)

1989년 2월 : 한양대학교

전자통신공학과(공학박사)

1991년 3월~2000년 2월 : 서원대학교정보통신과
부교수

2000년 3월~현재 : 충북과학대학 정보통신학과
교수

2001년 4월 : 충청북도지사 표창

2001년 10월: 한국정보처리학회 우수논문상

2002년 12월: 한국콘텐츠학회 학술상

2004년 5월: 한국정보처리학회 우수논문상

<관심분야> 인터넷 역기능의 기술적 해결, 영상 이
해, 정보 보호