

# 공간과 시간적 특징 융합 기반 유해 비디오 분류에 관한 연구

전 재 현<sup>†</sup> · 김 세 민<sup>\*\*</sup> · 한 승 완<sup>\*\*\*</sup> · 노 용 만<sup>\*\*\*\*</sup>

## 요 약

최근 인터넷, IPTV/SMART TV, 소셜 네트워크 (social network)와 같은 정보 유통 채널의 다양화로 유해 비디오 분류 및 차단 기술 연구에 대한 요구가 높아지고 있으나, 현재까지는 비디오에 대한 유해성을 판단하는 연구는 부족한 실정이다. 기존 유해 이미지 분류 연구에서는 이미지에서의 피부 영역의 비율이나 Bag of Visual Words (BoVW)와 같은 공간적 특징들 (spatial features)을 이용하고 있다. 그러나, 비디오에서는 공간적 특징 이외에도 모션 반복성 특징이나 시간적 상관성 (temporal correlation)과 같은 시간적 특징들 (temporal features)을 추가적으로 이용하여 유해성을 판단할 수 있다. 기존의 유해 비디오 분류 연구에서는 공간적 특징과 시간적 특징들에서 하나의 특징만을 사용하거나 두 개의 특징들을 단순히 결정 단계에서 데이터 융합하여 사용하고 있다. 일반적으로 결정 단계 데이터 융합 방법은 특징 단계 데이터 융합 방법보다 높은 성능을 가지지 못한다. 본 논문에서는 기존의 유해 비디오 분류 연구에서 사용되고 있는 공간적 특징과 시간적 특징들을 특징 단계 융합 방법을 이용하여 융합하여 유해 비디오를 분류하는 방법을 제안한다. 실험에서는 사용되는 특징이 늘어남에 따른 분류 성능 변화와 데이터 융합 방법의 변화에 따른 분류 성능 변화를 보였다. 공간적 특징만을 이용하였을 때에는 92.25%의 유해 비디오 분류 성능을 보이는데 반해, 모션 반복성 특징을 이용하고 특징 단계 데이터 융합 방법을 이용하게 되면 96%의 향상된 분류 성능을 보였다.

**키워드 :** 유해 비디오, 공간적 특징, 시간적 특징, 데이터 융합

## Using the fusion of spatial and temporal features for malicious video classification

Jae Hyun Jeon<sup>†</sup> · Semin Kim<sup>\*\*</sup> · Seungwan Han<sup>\*\*\*</sup> · Yong Man Ro<sup>\*\*\*\*</sup>

## ABSTRACT

Recently, malicious video classification and filtering techniques are of practical interest as ones can easily access to malicious multimedia contents through the Internet, IPTV, online social network, and etc. Considerable research efforts have been made to developing malicious video classification and filtering systems. However, the malicious video classification and filtering is not still being from mature in terms of reliable classification/filtering performance. In particular, the most of conventional approaches have been limited to using only the spatial features (such as a ratio of skin regions and bag of visual words) for the purpose of malicious image classification. Hence, previous approaches have been restricted to achieving acceptable classification and filtering performance. In order to overcome the aforementioned limitation, we propose new malicious video classification framework that takes advantage of using both the spatial and temporal features that are readily extracted from a sequence of video frames. In particular, we develop the effective temporal features based on the motion periodicity feature and temporal correlation. In addition, to exploit the best data fusion approach aiming to combine the spatial and temporal features, the representative data fusion approaches are applied to the proposed framework. To demonstrate the effectiveness of our method, we collect 200 sexual intercourse videos and 200 non-sexual intercourse videos. Experimental results show that the proposed method increases 3.75% (from 92.25% to 96%) for classification of sexual intercourse video in terms of accuracy. Further, based on our experimental results, feature-level fusion approach (for fusing spatial and temporal features) is found to achieve the best classification accuracy.

**Keywords :** Malicious Video, Spatial Feature, Temporal Feature, Data Fusion

※ 본 연구는 방송통신위원회의 정보보호원천기술개발사업의 연구결과로 수행되었음(KCA-2011-09914-06003).

† 준 회 원 : 한국과학기술원 전기 및 전자공학과 석박사통합과정

\*\* 준 회 원 : 한국과학기술원 정보통신공학과 박사과정

\*\*\* 정 회 원 : 한국전자통신연구원 선임연구원

\*\*\*\* 중 심 회 원 : 한국과학기술원 전기 및 전자공학과 교수

논문접수 : 2011년 1월 31일

수정일 : 1차 2011년 5월 31일, 2차 2011년 7월 4일

심사완료 : 2011년 7월 5일

## 1. 서 론

최근 인터넷, IPTV/SMART TV, 소셜 네트워크 (social network)와 같은 정보 유통 채널의 다양화로 멀티미디어 콘텐츠 (contents)의 유통 및 소비가 급격히 증가하고 있다. 이에 따라 유해한 이미지 및 비디오 등의 콘텐츠들을 손쉽게 인터넷을 통해 공유하고 접할 수 있게 되었다. 특히, 어린이들과 청소년들이 유해 콘텐츠를 접하기 쉬운 환경이 되었으며, 이를 막기 위한 유해 콘텐츠의 분류 및 차단 기술이 주목 받고 있다. 그러나, 현재 유해 이미지 분류 및 차단 기술은 많이 연구되는데 반해, 유해 비디오 분류 및 차단 기술에 대한 연구는 활발하지 않다.

기존의 유해 이미지 분류 연구에서는 이미지로부터 피부 영역의 비율이나 Bag of Visual Word (BoVW) 와 같은 공간적 특징 (spatial feature)을 추출하여 유해 유/무를 판단한다[참고문헌 추가]. 이 연구결과와 비디오가 연속된 프레임들로 구성된다는 점을 이용하여 [2-4]에서는 비디오를 구성하는 프레임들로부터 유해 이미지 분류 연구에서 이용되던 공간적 특징들을 추출하여 프레임들의 유/무해를 판단하고, 판단된 유해 프레임의 수가 특정 개수 이상이면 비디오를 유해 비디오로 판단한다.

또 다른 유해 비디오 분류 연구에서는 비디오의 프레임에서 추출할 수 있는 공간적 특징 이외에 시간적 특징 (모션의 반복성, 프레임 그룹의 특성)들을 이용하여, 유해 비디오의 모션으로부터 모션의 반복성을 검출하여 유해 비디오 판별에 이용하고 있다[5-7]. 모션의 반복성 검출을 위해 각 프레임에서 모션을 추출하여 모션 히스토그램을 구성하거나 수직 방향과 수평 방향으로 평균적인 모션 값을 얻어 모션의 반복성을 검출하고 유해 비디오를 분류한다. [9-11]에서는 하나의 프레임에서만 특징을 추출하지 않고, 비디오의 프레임들이 시간적 상관성 (temporal correlation)을 가지는 특징을 활용하여, 비디오를 특정 크기의 프레임 그룹들로 나누고 나누어진 프레임 그룹에서 특징을 추출한다. 하나의 프레임에서 추출되는 정보의 부정확함을 프레임 그룹에서 추출되는 특징을 이용함으로써 정보의 정확성을 높여서 유해 비디오 분류 성능을 향상시킨다.

최근의 유해 비디오 분류에서는 앞에서 언급된 공간적 특징과 시간적 특징 (모션의 반복성, 시간적 상관성)들을 같이 사용하여 유해 비디오 분류 성능을 향상시키는 연구들이 있다. [8]에서는 공간적 특징과 모션 반복성을 같이 사용하고 있으며, [9-11]에서는 공간적 특징과 시간적 상관성을 같이 이용하는 방법을 연구하고 있다. 그러나 현재까지의 유해 비디오 분류 연구에서는 특징 단계 융합(feature level fusion) 방법이 아닌 결정 단계 융합 (decision level fusion) 방법을 주로 사용하고 있어[8] 성능에 한계를 보이고 있다. 데이터 융합과 관련된 Data mining 연구들에서는 두 가지 이상의 특징들을 데이터 융합하여 사용하는 경우, 특징 단계 융합 방법이 결정 단계 융합 방법보다 더 높은 성능을 보이는 것으로 알려져 있다[14-16]. 그 이유로는 한 프레임

에서 얻어지는 공간적 특징과 비디오 시간적 방향에서 얻어지는 모션 반복성 특징을 결정 단계에서 융합하기에는 특징들의 영역 (domain)이 맞지 않기 때문이다 [15-16].

본 논문에서는 기존의 유해 비디오 분류 연구에서 사용되고 있는 공간적 특징과 시간적 특징들을 특징 단계 융합 방법을 이용하여 융합하여 유해 비디오를 분류하는 방법을 제안한다. 또한, 사용되는 특징들의 증가와 특징 융합 방법들에 따른 유해 비디오 분류 성능 변화에 대한 실험을 수행하고 그 결과를 분석한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 시간적 상관성을 고려하여 공간적 특징과 모션 반복성 특징을 융합하는 방법들을 서술한다. 그리고 3장에서는 사용되는 공간적 특징과 모션 반복성 특징의 타당성 (feasibility)를 보이고, 데이터 융합에 따른 유해 비디오 분류 성능 변화를 실험으로 비교한다. 마지막으로 4장에서는 본 논문에서 서술된 공간적 특징과 시간적 특징의 융합 방법에 따른 유해 비디오 분류 성능 변화와 앞으로의 연구에 대해서 서술한다.

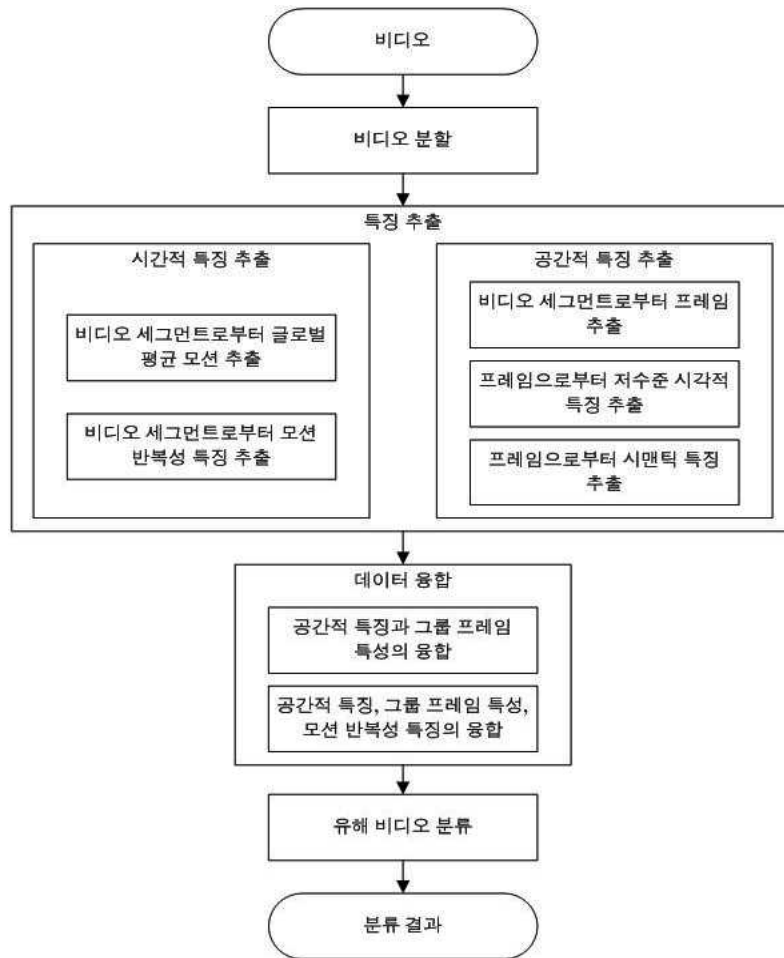
## 2. 공간적 특징과 시간적 특징의 융합에 기반한 유해 비디오 분류 시스템

본 장에서는 공간적 특징과 시간적 특징들의 데이터 융합에 기반한 유해 비디오 분류 시스템에 대하여 설명한다. 제안된 유해 비디오 분류 시스템은 (그림 1) 과 같으며, 다음과 같이 크게 네 부분으로 구성된다: 1) 비디오를 분할하는 부분, 2) 분할된 비디오에서 특징들을 추출하는 부분, 3) 추출된 특징들을 데이터 융합하는 부분, 4) 데이터 융합된 특징을 이용하여 유해 비디오를 분류하는 부분.

### 2.1 비디오 분할

비디오는 연속된 프레임들로 이루어지고 일반적으로 이 프레임들은 시간적으로 상관성을 가진다. 그러므로, 하나의 프레임에서 얻어지는 정보의 부정확함을 여러 개의 프레임으로부터 얻어지는 정보를 활용하면 비디오를 분류하는 성능의 정확성을 높일 수 있다. [9-11]에서와 같이 프레임 그룹의 시간적 상관성을 이용하기 위해, 본 논문에서 제안된 유해 비디오 분류 방법에서는 판별 대상이 되는 비디오  $V$ 가 입력되면 일정 크기  $N^{window}$ 의 비디오 세그먼트 (video segment)들  $\{V_1, V_2, \dots, V_s\}$ 로 분할한다. 그리고 분할된 비디오 세그먼트 별로 유/무해를 판단한 후, 비디오 세그먼트들의 유/무해 판단 결과를 이용하여 전체 비디오의 유/무해를 최종적으로 판단한다. 여기에서  $N^{window}$ 는 모션의 반복성이 가장 잘 관찰되는 프레임의 수를 나타낸다. 본 논문에서는 비디오를 구성하는 프레임의 공간적 특징만을 이용하는 것이 아니라, 모션의 반복성 특징 또한 이용하므로 비디오 세그먼트는 모션의 반복성이 잘 나타나는  $N^{window}$  크기를 가지도록 설정된다.

분할된 비디오 세그먼트  $V_s$ 에서의 공간적 특징과 모션의 반복성 특징을 추출하고, 추출된 공간적 특징들과 모션 반



(그림 1) 공간적 특징과 시간적 특징들의 융합에 기반한 유해 비디오 분류 시스템

복성 특징들을 데이터 융합하는 과정은 다음 절들에서 자세히 서술한다.

2.2 특징 추출

2.2.1 모션 반복성 특징 추출

본 논문에서 유해비디오는 남녀의 성교 (sexual intercourse) 가 출현하는 비디오들로 한정하여 연구한다. 남녀의 성교가

출현하는 장면에서는 동작의 반복성을 관찰 할 수 있는데, 이러한 반복성은 비디오 세그먼트에서 얻을 수 있는 모션 반복성 특징으로부터 측정될 수 있다. 본 절에서는 성행위 동작의 반복성을 측정하기 위해 모션 반복성 특징을 추출하는 과정과 모션 반복성 특징으로부터 모션 반복성 특징을 추출하는 과정을 설명한다.

s번째 비디오 세그먼트  $V_s$  는 연속된 프레임들  $V_s = \{F_{s,1}, F_{s,2}, \dots, F_{s,N^{window}}\}$ 로 표현할 수 있다.  $V_s$ 의 t번째 프레임  $F_{s,t}$ 의 매크로 블록 (macro block)들과 각 매크로 블록들에서의 2 차원 모션 벡터들 (motion vectors)은 (그림 2)와 같이 표현될 수 있다.  $m_{s,t}(i)$ 는  $m_{s,t}(i) = (m_{s,t}^x(i), m_{s,t}^y(i))$ 로 표현될 수 있다. 그리고 (그림 2)에서  $N_{s,t}^{blocks}$  는  $F_{s,t}$ 의 매크로 블록의 수를 나타낸다.

모션 정보에 기반을 둔 동작 반복성 판별 방법을 이용하면 동작이 반복적으로 나타나는 구간을 찾을 수 있다. 앞에서 언급된 모션 벡터에 Auto-Correlation Function (ACF) 이나 Discrete Cosine Transform (DCT)를 수행하면 동작의 반복성을 판별할 수 있는데 반복성을 검출하는 방법들 중 ACF를 이용한 방법을 살펴보면 다음과 같다.

$m_{s,t}(1)$	...	...	...	...
...	...	...	...	...
...	...	$m_{s,t}(i)$	...	...
...	...	...	...	...
...	...	...	...	$m_{s,t}(N_{s,t}^{blocks})$

(그림 2) s 번째 비디오 세그먼트의 프레임 t 에서의 매크로 블록들

$$\hat{R}_s^x(\tau) = \frac{1}{N^{window}} \sum_{t=1}^{N^{window}} m_{s,t}^x \cdot m_{s,t+\tau}^x, \tag{1}$$

$$\hat{R}_s^y(\tau) = \frac{1}{N^{window}} \sum_{t=1}^{N^{window}} m_{s,t}^y \cdot m_{s,t+\tau}^y,$$

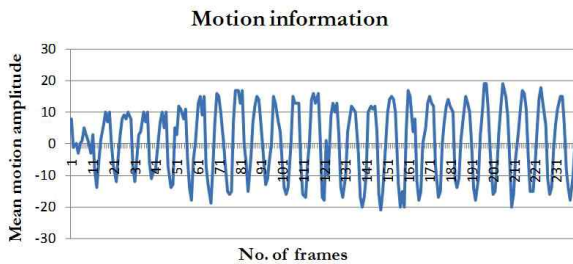
여기서  $m_{s,t}^x$  와  $m_{s,t}^y$  는 다음과 같이 정의된다.

$$m_{s,t}^x = \frac{1}{N_{s,t}^{blocks}} \sum_{i=1}^{N_{s,t}^{blocks}} m_{s,t}^x(i), \tag{2}$$

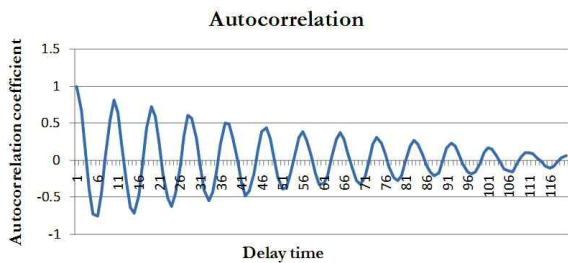
$$m_{s,t}^y = \frac{1}{N_{s,t}^{blocks}} \sum_{i=1}^{N_{s,t}^{blocks}} m_{s,t}^y(i).$$

비디오에서 모션은 크게 전체적 모션 (global motion)과 물체의 모션 (object motion)으로 나눌 수 있는데, 본 논문에서는 식 (2)과 같이 정의된 전체적 모션을 사용한다. 전체적 모션을 사용할 경우, 모션이 발생하는 영역이 전체 프레임에서 작으면 평균 모션의 값이 작고, 모션이 발생하는 영역이 크면 평균 모션의 값이 크게 된다. 즉, 비디오에서 줌인 (zoom in)과 줌아웃 (zoom out)에 따라 평균 모션의 값이 달라진다. 그러나, 이러한 부분은 모션의 반복성 (motion periodicity)을 찾는 데 크게 문제되지 않는다. 이에 대한 증명은 다음 장에서 실험을 통해 보이도록 한다.

(그림 3) (a)는 x축 방향으로 반복성이 존재하는 비디오 세그먼트에서의 x축 방향의 평균 모션의 변화를 나타낸다. 그리고 (그림 3) (b)는 (a)에서 얻어진 비디오 세그먼트의 x



(a) x축 방향의 평균 모션



(b) ACF 수행 결과

(그림 3) x축 방향으로 반복성이 존재하는 비디오의 특징 구간에서의 x축 방향의 평균 모션과 ACF수행 결과

축 방향의 평균 모션 값들에 대하여 ACF를 수행한 결과를 나타낸다. ACF 수행 결과를 보면, 지연 시간 (delay time)이 커질수록 자기상관 계수(Autocorrelation coefficient)의 값이 sine 함수와 같은 반복성을 보이는 것을 알 수 있다.

일반적으로 ACF를 적용 할 때에는 적용될 신호의 길이가 중요하다. 반복성이 나타나지 않을 정도로 신호의 길이가 짧으면, ACF를 적용하더라도 반복성을 관찰할 수 없다. 또한 반복성이 나타나는 길이보다 신호의 길이가 길면 ACF를 적용하여도 반복성이 뚜렷하게 나타나지 않게 된다. 그러므로 ACF를 이용하여 신호의 반복성을 검출하기 위해서는 신호의 길이를 정하는 식 (2)의  $N^{window}$ 의 크기 즉, 비디오 세그먼트의 크기가 매우 중요하다.  $N^{window}$ 의 크기 변화에 따른 반복성 검출의 성능 변화 실험은 다음 장에서 보이도록 한다.

반복성이 관찰되는 상교 비디오 세그먼트의 ACF 수행 결과들로부터 추출되는 모션의 반복성 특징을 다음과 같이 나타낼 수 있다.

$$\mathbf{X}_s^{per} = (x_s^{per}(1), x_s^{per}(2), \dots, x_s^{per}(N^{per}))^T. \tag{3}$$

모션의 반복성이 있는 유해 비디오 세그먼트들의 ACF 수행 결과를 보면, (그림 3) (b) 와 같이 ACF 결과의 로컬 미니멈 (local minimum)들과 로컬 맥시멈 (local maximum)들의 계수 값 차이가 크면서 일정크기의 지연 시간 (delay time)을 가지는 것을 볼 수 있다. 실험을 통한 관찰 결과 반복적인 동작이 존재하는 비디오 세그먼트에서 로컬 맥시멈과 로컬 미니멈의 지연시간 차이가 6~15 정도의 크기를 가진다. 이때, 총 10 개의 로컬 맥시멈과 로컬 미니멈들 간의 지연시간 차이들의 합은 60~150 정도의 범위를 가지게 되는데, 이 범위는 ACF의 전체 지연시간인  $N^{window}/2$  와 비슷하다. 따라서 본 논문에서는 모션의 반복성 특징을 표현하기 위해 20차원의 벡터로 정의하였다. 20차원의 벡터의 구성을 살펴보면, 첫 번째와 두 번째 성분은 첫 번째 로컬 맥시멈과 첫 번째 로컬 미니멈의 자기상관 계수 값 차이와 지연 시간 차이로 구성된다. 그리고 벡터의 세 번째와 네 번째 성분은 첫 번째 로컬 미니멈과 두 번째 로컬 맥시멈의 자기상관 계수 값 차이와 지연 시간 차이로 구성된다. 벡터의 나머지 성분들 또한 연속적으로 존재하는 로컬 미니멈과 로컬 맥시멈들의 자기상관 계수 값 차이와 지연 시간 차이로 구성된다.

### 2.2.2 공간적 특징 추출

비디오 세그먼트로부터 추출된 프레임에서 주로 사용되는 시각적 특징 (visual feature)으로는 MPEG-7, Scale-Invariant Feature Transform (SIFT), 피부 영역 (skin region) 검출 결과, Bag of Visual Word (BoVW)와 시맨틱 개념 검출 (semantic concept detection) 결과 등이 있다[참고 문헌]. 기존의 연구들에서는 시각적 특징으로 주로 피부 영역 검출이나 BoVW를 사용하고 있는데, 본 논문에서는

유해 이미지 분류에서 높은 성능을 보인 시맨틱 개념 검출 방법 [12-13]를 이용한다. [12-13]에서의 시맨틱 개념 검출 방법은 입력 이미지에 여성의 'breast', 'genital', 'bottom' 등이 존재하면 유해 이미지로 분류한다. [12][13]에서는 유해 이미지 분류를 위해 'naked body', 'bottom', 'breast', 'genital', 'dress', 'others' 등의 시맨틱 개념을 정의하여 사용하고 있다. 이러한 시맨틱 개념들은 유해 비디오 분류에도 이용 가능하다. 본 논문에서는 [12-13]에서 정의한 6개의 시맨틱 개념에 추가적으로 여성의 성기와 남성의 성기가 겹쳐지는 'sexual intercourse'를 정의하여 총 7 개의 시맨틱 개념들을 사용한다.

시맨틱 개념들 ( $C=\{c_1, c_2, \dots, c_{N^{sem}}\}$ ,  $N^{sem}$ 는 정의된 시맨틱 개념들의 수)은 이미지에서 얻어지는 MPEG-7 descriptor과 같은 저수준 시각적 특징(low-level visual feature)이 SVM의 학습을 통해 생성된 시맨틱 개념 검출기들을 통과하여 검출된다 [12][13]. 시맨틱 개념 검출기를 학습하는 과정은 다음과 같다. 우선 이미지에서 시맨틱 개념을 포함하는 부분을 잘라내고, 잘려진 부분에서 추출되는 저수준 시각적 특징을 이용하여 시맨틱 개념의 검출기를 학습시킨다.

이를 수식으로 정형화하여 표현하면 다음과 같다.

$s$ 번째 비디오 세그먼트에서 얻어지는  $t$ 번째 프레임  $F_{s,t}$ 로부터 추출되는 저수준 시각적 특징은 다음과 같이 표현할 수 있다.

$$\mathbf{X}_{s,t}^{low} = \left( x_{s,t}^{low}(1), x_{s,t}^{low}(2), \dots, x_{s,t}^{low}(N^{low}) \right)^T, \quad (4)$$

여기서,  $N^{low}$ 는 저수준 특징 벡터의 차원을 나타낸다.

$\mathbf{X}_{s,t}^{low}$ 로부터 [12][13]에서와 같이 SVM을 이용한  $n$ 번째 시맨틱 개념  $c_n$ 의 검출 결과  $X_{s,t}^{sem}(n)$ 을 다음과 같이 표현할 수 있다.

$$x_{s,t}^{sem}(n) = \sum_u \left\{ \mathbf{w}_n(u) \cdot \mathbf{z}_n(u) \cdot K(\mathbf{g}_n(u), \mathbf{X}_{s,t}^{low}) \right\} + a_n, \quad (5)$$

여기서,  $K$ 는 커널 함수(kernel function)을 나타내며,  $\mathbf{g}_n(u)$ 는 시맨틱 개념  $c_n$ 에 대한 초평면(hyperplane)의  $u$ 번째 지원 벡터(support vector)를 나타낸다. 그리고  $\mathbf{w}_n$ 은 지원 벡터에 상응하는 가중치 벡터(weighting vector)이며,  $\mathbf{z}_n$ 은 지원 벡터에 상응하는 클래스 벡터(class vector)를 나타낸다.  $a_n$ 은 시맨틱 개념  $c_n$ 에 대한 최적화된 임계 값(threshold value)을 나타낸다.

식 (5)와 같이  $F_{s,t}$ 으로부터  $N^{sem}$ 개의 시맨틱 개념 검출 결과를 얻으면, 검출 결과들을 직렬화(concatenation)하여 시맨틱 특징 벡터를 다음과 같이 구성할 수 있다

$$\mathbf{X}_{s,t}^{sem} = \left( x_{s,t}^{sem}(1), x_{s,t}^{sem}(2), \dots, x_{s,t}^{sem}(N^{sem}) \right)^T. \quad (6)$$

## 2.3 데이터 융합

### 2.3.1 공간적 특징과 시간적 상관성의 데이터 융합

$s$ 번째 비디오 세그먼트  $V_s$ 의 의미를 나타내는 시맨틱 특징 벡터  $\mathbf{X}_s^{sem}$ 은  $V_s$ 로부터 추출된  $N^{sample}$ 개의 프레임들 각각에 대해 식 (6)을 적용해서 얻어진 시맨틱 특징 벡터들을 데이터 융합하여 다음과 같이 얻을 수 있다.

$$\mathbf{X}_s^{sem} = \left( x_s^{sem}(1), x_s^{sem}(2), \dots, x_s^{sem}(N^{sem}) \right)^T. \quad (7)$$

$\mathbf{X}_s^{sem}$ 를 얻기 위한 데이터 융합 방법으로는 최대화(maximization) 방법과 평균화(averaging)방법을 적용해 볼 수 있다. 최대화 방법을 적용하면,  $V_s$ 의 시맨틱 특징 벡터  $\mathbf{X}_{s,t}^{sem}$ 의  $i$ 번째 특징 값은 다음과 같이 얻을 수 있다.

$$x_s^{sem}(i) = \frac{1}{N^{sample}} \max_{t \in \text{sampled-frames}} x_{s,t}^{sem}(i). \quad (8)$$

그리고 평균화 방법을 적용하면,  $\mathbf{X}_s^{sem}$ 의  $i$ 번째 특징 값은 다음과 같이 얻을 수 있다.

$$x_s^{sem}(i) = \frac{1}{N^{sample}} \sum_{t \in \text{sampled-frames}} x_{s,t}^{sem}(i). \quad (9)$$

$V_s$ 에서 추출되는 프레임의 수  $N^{sample}$ 의 변화와 데이터 융합 방법에 대한 유해 검출 결과는 3장의 실험에서 보이도록 한다.

### 2.3.2 공간적 특징, 시간적 상관성 그리고 모션 반복성 특징의 데이터 융합

$s$ 번째 비디오 세그먼트  $V_s$ 로부터 추출된 시맨틱 특징 벡터  $\mathbf{X}_s^{sem}$ 와 모션의 반복성 특징  $\mathbf{X}_s^{per}$ 은 데이터 융합(data fusion)을 통하여 유해 비디오 세그먼트를 검출하는데 이용된다. 여기서, 시맨틱 특징 벡터  $\mathbf{X}_s^{sem}$ 를 이용하여 SVM 기반으로  $V_s$ 의 유/무해 판별 신뢰 값을 다음과 같이 얻을 수 있다.

$$\Phi^{sem}(\mathbf{X}_s^{sem}) = \sum_u \left\{ \mathbf{w}_{sem}(u) \cdot \mathbf{z}_{sem}(u) \cdot K(\mathbf{g}_{sem}(u), \mathbf{X}_s^{sem}) \right\} + a_{sem}. \quad (10)$$

여기서,  $K$ 는 커널 함수(kernel function)을 나타내며,  $\mathbf{g}_{sem}(u)$ 는  $\mathbf{X}_s^{sem}$ 를 이용한 유/무해 판별에 대한 초평면(hyperplane)의  $u$ 번째 지원 벡터(support vector)를 나타낸다. 그리고  $\mathbf{w}_{sem}$ 은 지원 벡터에 상응하는 가중치 벡터(weighting vector)이며,  $\mathbf{z}_{sem}$ 은 지원 벡터에 상응하는 클래스

벡터 (class vector)를 나타낸다.  $a_{sem}$ 은 유/무해 판별에 대한 최적화된 문턱 값 (threshold value)을 나타낸다.

그리고 모션의 반복성 특징  $X_s^{per}$ 를 SVM을 기반으로 유/무해 판별 신뢰 값은 다음과 같이 얻어 질 수 있다.

$$\Phi^{per}(\mathbf{X}_s^{per}) = \sum_u \{ \mathbf{w}_{per}(u) \cdot \mathbf{z}_{per}(u) \cdot K(\mathbf{g}_{per}(u), \mathbf{X}_s^{per}) \} + a_{per} \quad (11)$$

그리고 얻어진 신뢰 값  $\Phi^{sem}(\mathbf{X}_s^{sem}), \Phi^{per}(\mathbf{X}_s^{per})$  은  $-\infty$  에서  $\infty$ 의 범위를 가지므로 시그모이드 적합화 (sigmoid fitting) 과정을 통해서 다음과 같이 0에서 1사이의 확률 값으로 변환 될 수 있다.

$$P(m | \mathbf{X}_s^{sem}) \cong \frac{1}{1 + \exp(A \cdot \Phi^{sem}(\mathbf{X}_s^{sem}) + B)},$$

$$P(m | \mathbf{X}_s^{per}) \cong \frac{1}{1 + \exp(A \cdot \Phi^{per}(\mathbf{X}_s^{per}) + B)} \quad (12)$$

여기서,  $A$  과  $B$  은 유/무해 판별에 대한 시그모이드 (sigmoid) 모델의 모양을 조정하기 위한 파라미터 (parameter)들이며, Regularized Maximum Likelihood problem [17]을 풀어냄으로써 파라미터 값들을 얻을 수 있다.

데이터 융합은 크게 결정 단계 융합 (decision-level fusion)과 특징 단계 융합 (feature-level fusion)으로 나뉘어진다 [16]. 결정 단계 융합은 다양한 특징 집합들이 있으면, 각각의 특징 집합에 대해 독립적으로  $V_s$ 의 유/무해 판별을 수행하여 얻어지는 신뢰 값 (confidence value)들을 최종 유/무해 판별에 이용한다.

결정 단계 데이터 융합 방법 중, 첫 번째 방법 가중치를 준 합 방법 (weighted sum)방법[16]이다. 이 방법은  $P(m | X_s^{sem})$ 와  $P(m | X_s^{per})$ 를 이용하여  $V_s$ 의 유해 확률 값  $P(m | V_s)$ 을 다음과 같이 구할 수 있다.

$$P(m | V_s) = \alpha \cdot P(m | X_s^{sem}) + \beta \cdot P(m | X_s^{per}), \quad (13)$$

여기서  $\alpha$ 와  $\beta$ 는 각각  $P(m | X_s^{sem})$ 와  $P(m | X_s^{per})$ 에 곱해지는 가중치 파라미터이고, 조건  $\alpha + \beta = 1$ 을 만족한다.

결정 단계 데이터 융합 방법 중, 두 번째 방법은 곱 (product) 방법[16]이다. 이 방법으로는 확률 값  $P(m | V_s)$ 를 다음과 같이 구할 수 있다.

$$P(m | V_s) = P(m | X_s^{sem}) \cdot P(m | X_s^{per}), \quad (14)$$

그리고 세 번째 방법으로 최대값 찾기 방법 [16] 있는데, 이 방법으로는 확률 값  $P(m | V_s)$ 를 다음과 같이 얻을 수 있다.

$$P(m | V_s) = \max(P(m | X_s^{sem}), P(m | X_s^{per})). \quad (15)$$

결정단계 융합 방법들에서의 마지막 네 번째 방법은  $V_s$ 의 유/무해 판별을 위해 융합된 특징 벡터 ( $X_s^{fused}$ )를 식 (13)에서 얻어진  $P(s | X_s^{sem})$ 와  $P(s | X_s^{per})$ 를 이용하여 다음과 같이 직렬화 [16] 하여 얻는다.

$$\mathbf{X}_s^{fused} = (P(m | \mathbf{X}_s^{sem}), P(m | \mathbf{X}_s^{per}))^T. \quad (16)$$

$V_s$ 로부터 얻어진  $X_s^{fused}$ 는 SVM을 기반으로 유/무해 판별 신뢰 값  $\Phi^{fused}(\mathbf{X}_s^{fused})$ 를 다음과 같이 얻을 수 있다.

$$\Phi^{fused}(\mathbf{X}_s^{fused}) = \sum_u \{ \mathbf{w}_{fused}(u) \cdot \mathbf{z}_{fused}(u) \cdot K(\mathbf{g}_{fused}(u), \mathbf{X}_s^{fused}) \} + a_m \quad (17)$$

그리고  $V_s$ 의 유/무해 판별 확률 값  $P(m | V_s)$ 는 식 (12)와 유사하게 신뢰 값을 시그모이드 적합화 과정 통해 다음과 같이 구해 질 수 있다.

$$P(m | V_s) \cong \frac{1}{1 + \exp(A \cdot \Phi^{fused}(\mathbf{X}_s^{fused}) + B)}. \quad (18)$$

특징 단계 융합 방법은 결정 단계 융합 방법에서 네 번째 방법 식 (16)과 유사한데,  $X_s^{fused}$ 를 얻는 과정이 다르다.  $X_s^{fused}$ 를 시맨틱 특징 벡터  $X_s^{sem}$ 와 모션의 반복성 특징  $X_s^{per}$ 들을 다음과 같이 직렬화시키는 방법으로 얻어진다.

$$\mathbf{X}_s^{fused} = [\mathbf{X}_s^{sem^T}, \mathbf{X}_s^{per^T}]^T. \quad (19)$$

얻어진  $X_s^{fused}$ 는 식 (17)과 (18)의 과정을 거쳐 최종적으로  $V_s$ 의 유/무해 판별 확률 값  $P(m | V_s)$ 을 얻을 수 있게 된다.

#### 2.4 비디오 세그먼트들의 유해 확률 값들을 이용한 유해 비디오 판별

비디오로부터 얻어진 비디오 세그먼트들  $V_s$ 는 각각 유/무해 판별을 거치게 된다. 그리고 전체 비디오에 대한 유/무해

판별 확률 값  $P(m|V)$  은 각각의 세그먼트들이 유해로 판단될 확률 값을 이용하여 다음과 같이 얻어진다.

$$P(m|V) = \max_s P(m|V_s). \quad (20)$$

### 3. 실험

본 장에서는 먼저 공간적 특징과 모션 반복성 특징의 검출 성능에 대한 실험 결과를 보인다. 그리고 공간적 특징, 모션 반복성 특징, 시간적 상관성 특징들의 사용에 따른 분류 성능 변화와 데이터 융합 방법에 따른 유해 비디오 세그먼트 분류 성능 변화에 대한 실험 결과를 기술한다. 마지막으로 유/무해 비디오 분류 실험에 대한 결과를 보인다.

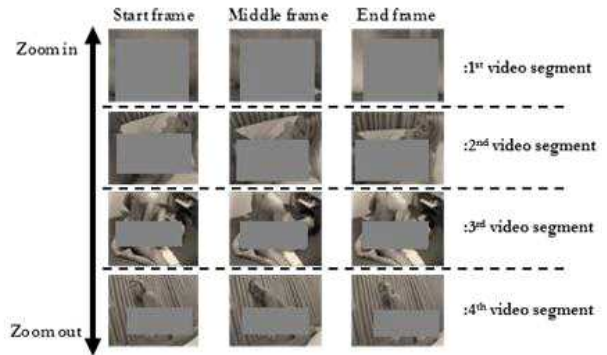
#### 3.1 공간적 특징과 모션 반복성 특징의 검출 실험

##### 3.1.1 모션 반복성 특징의 검출 성능

본 절에서는 줌-인과 줌-아웃이 모션의 반복성 검출 결과에 영향을 미치지 않는 것을 ACF 수행 결과를 통해 증명하고, 비디오 세그먼트의 크기  $N^{window}$  에 따른 모션 반복성 검출 성능 변화를 실험하고 분석한다.

##### (1) 줌-인 과 줌-아웃에 따른 모션 반복성 검출

비디오에서 줌-인과 줌-아웃에 따라 평균 모션의 값이 달라지게 되는데, 평균 모션에 ACF를 수행하게 되면 평균 모션의 값이 크고 작은 것에 영향을 받지 않고 모션의 반복성을 찾을 수 있다. 이에 대한 증명을 본 절에서 실험으로 증명한다.

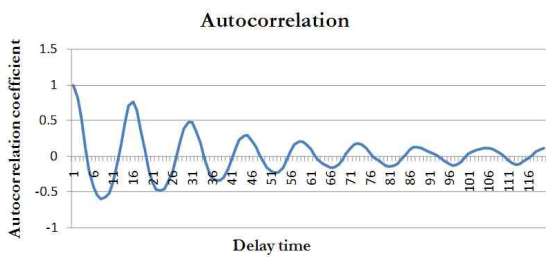


(그림 4) 줌-인된 비디오 세그먼트들과 줌-아웃된 비디오 세그먼트들의 예

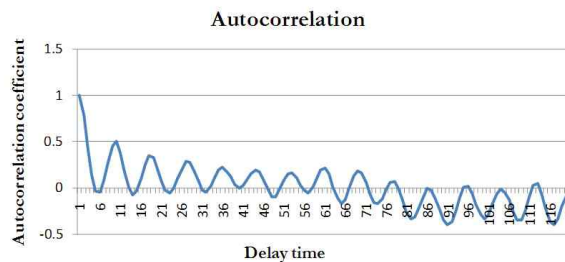
(그림 4)는 사람의 신체가 줌-인된 비디오 세그먼트와 줌-아웃된 비디오 세그먼트의 예를 나타낸다. 여기에서  $1^{st}$  비디오 세그먼트는 줌-인이 가장 많이 된 비디오로 평균 모션의 크기가 가장 크다. 그리고  $4^{th}$  비디오 세그먼트는 가장 줌-아웃된 비디오로 평균 모션의 크기가 가장 작다. (그림 5)의 각 비디오 세그먼트들의 ACF 수행 결과를 살펴보면, 비디오 세그먼트 별로 평균 모션의 크기는 다르지만 ACF의 수행 결과 모든 비디오 세그먼트들에서 평균 모션의 크기에 상관없이 비슷한 형태의 반복성을 보여주고 있다.

##### (2) $N^{window}$ 의 크기 변화에 따른 모션 반복성 검출

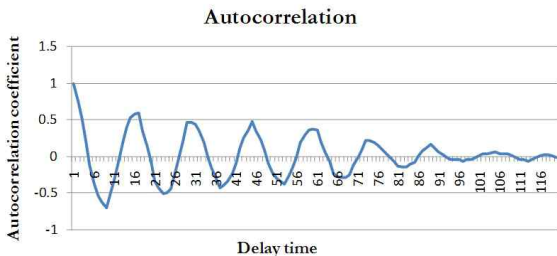
일반적으로 ACF를 적용 할 때 적용될 신호가 반복성이 나타나지 않을 정도로 짧으면, ACF를 적용하여도 반복성을 관찰할 수 없고, 반복성이 나타나는 길이보다 신호의 길이



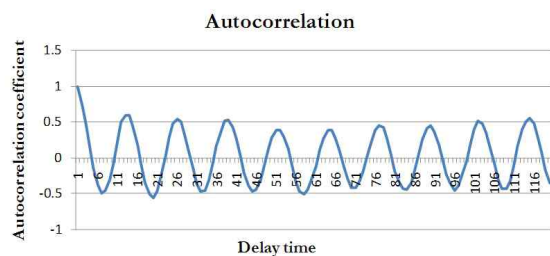
(a) 1st 비디오 세그먼트의 ACF 수행 결과



(b) 2nd 비디오 세그먼트의 ACF 수행 결과



(c) 3rd 비디오 세그먼트의 ACF 수행 결과

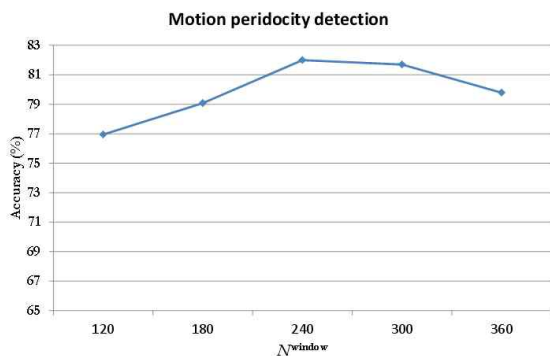


(d) 4th 비디오 세그먼트의 ACF 수행 결과

(그림 5) (그림 4)에서 예로 나오는 4개의 비디오 세그먼트들의 ACF 수행 결과

가 길면 ACF를 적용하여도 반복성이 정확히 나타나지 않게 된다. 그러므로 신호의 길이를 정하는 식 (2)의  $N^{window}$ 의 크기가 중요하다. 본 절에서는  $N^{window}$ 의 크기 변화에 따른 모션 반복성 검출 성능의 변화를 관찰하여 최적의  $N^{window}$ 을 제시한다.

$N^{window}$ 의 크기에 대한 실험은 120프레임, 180프레임, 240프레임, 300프레임, 360프레임으로 60프레임 간격으로 변화를 주면서 수행하였다. 사용된 특징은 2.2.2절에서 설명한 모션의 반복성 특징을 사용하였으며, 사용된 분류기(classifier)는 SVM을 이용하였다. 사용된 학습(learning) (DB)는 x축 방향 반복성 검출기와 y축 방향 반복성 검출기를 훈련시킬 때 각각 포지티브 데이터(positive data)로 98개의 비디오 세그먼트, 네거티브 데이터(negative data)로 98개의 비디오 세그먼트들을 이용하였다. 각 비디오 세그먼트들의 길이는  $N^{window}$ 의 크기 변화에 따라 각기 120프레임부터 360프레임까지 변화하도록 구성하였다. 그리고 테스트(test) DB는 x축과 y축 방향 반복성 검출기들 각각 포지티브 데이터로 136개의 비디오 세그먼트, 네거티브 데이터로 136개의 비디오 세그먼트들을 이용하였다. 포지티브 데이터들은 성교와 구강성교와 같이 모션의 반복성이 관찰되는 비디오 세그먼트들로 구성하였고, 네거티브 데이터들은 일반 드라마, 다큐멘터리, 모션 반복성이 관찰되지 않는 스포츠 등을 포함하는 비디오 세그먼트들로 구성하였다. 그리고 성능 측정은 (그림 6)과 같이 정확도(accuracy)로 x축 및 y축 반복성 검출기들의 성능을 평균적으로 측정하였다.



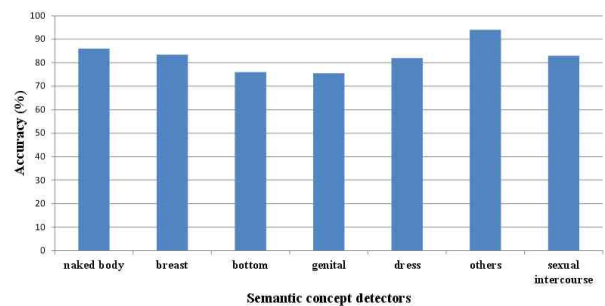
(그림 6)  $N^{window}$ 의 변화에 따른 모션 반복성 검출 성능 변화

(그림 6)의 실험 결과는  $N^{window}$ 의 값이 120프레임부터 360프레임까지 변화할 때, 분류 성능이 지속적으로 증가하는 것이 아니라 240프레임에서 최고의 성능을 보이고

$N^{window}$ 이 240프레임을 넘어가면 서서히 분류 성능이 감소하는 것을 나타낸다. 이를 통해 유추해 볼 수 있는 것은 성교나 구강성교 같은 동작을 포함하고 있는 유해 비디오들에서 관찰되는 모션의 반복성은 평균적으로 240프레임의 길이를 가진다는 것을 알 수 있다.

### 3.1.2 공간적 특징에 대한 타당성

본 논문에서 사용할 시맨틱 개념들은 [12][13]에서 사용되는 naked body, bottom, breast, genital, dress, others 이외에 추가적으로 'sexual intercourse'를 정의하여 유해 프레임 판별에 이용한다. 정의된 7개의 시맨틱 개념 검출기들은 SVM을 이용하여 학습된다. 학습을 위해 각 시맨틱 개념 검출기(detector)의 포지티브 데이터와 자신을 제외한 다른 시맨틱 개념들을 포함하는 네거티브 데이터를 동일한 비율(각 200장)로 모아서 구성한다. 그리고 성능 테스트를 위해 각 검출기마다 포지티브 데이터 100장의 이미지와 네거티브 데이터 100장의 이미지를 구성하였다. 사용된 시맨틱 개념 검출기들의 성능은 (그림 7)과 같다.



(그림 7) 시맨틱 개념 검출기들의 검출 성능

## 3.2 공간적 특징과 시간적 특징의 융합 방법 변화에 따른 분류 성능 변화

### 3.2.1 공간적 특징과 시간적 상관성의 데이터 융합 방법에 따른 분류 성능 변화

본 절에서는 비디오 세그먼트들의 유/무해 판별을 위해 공간적 특징만을 이용하여 SVM 기반으로 학습시켰을 때의 분류 성능에 대하여 서술한다. 그리고 유/무해 판별을 위한 비디오 세그먼트의 길이는 3.1.2절의 실험에서 가장 높은 모션 반복성 검출 성능을 보인  $N^{window}$  값인 240프레임으로 설정하였다. 비디오 세그먼트의 유/무해 분류기를 위한 학습 및 테스트 DB는 <표 1>과 같이 구성하였다.

<표 1> 유해 비디오 세그먼트 분류기의 학습 및 테스트 DB

Classifier	Positive DB	Negative DB							
		Drama	Documentary	Baseball	Martial arts	Boxing	Swimming	Hockey	Tennis
Sexual intercourse	200	22	22	22	23	23	22	22	22



비디오 세그먼트로부터 하나의 프레임만을 추출하여 시맨틱 특징정보를 얻었을 경우, 그리고 비디오 세그먼트로부터 일정 간격으로 3개, 5개의 프레임들을 추출하여 시맨틱 특징 정보들을 2.2.2절에서 언급한 최대화와 평균화 방법을 이용하여 데이터 융합을 하였을 경우의 유/무해 판별 성능 변화는 <표 2>와 같다.

<표 2> 사용되는 프레임의 수와 데이터 융합 방법에 따른 유해 판별 성능 변화

Accuracy		No. of used frames		
		1	3	5
Data fusion	Maximization	92.25	91.5	90
	Averaging		93.5	93.5

<표 2>의 실험 결과를 보면 하나의 비디오 세그먼트에서 이용되는 프레임의 수가 늘어남에 따라 유/무해 판별 성능이 증가하고 있다. 그리고 프레임으로부터 얻어진 시맨틱 특징 정보를 융합할 때, 평균화 방법을 이용하였을 경우의 유/무해 판별 성능이 최대화 방법을 이용하였을 경우보다 높은 성능을 보이고 있다. 이러한 결과는 비디오 세그먼트로부터 여러 프레임을 추출하고, 추출된 프레임들에서 얻어지는 시맨틱 특징정보들을 평균화 방법을 이용하면 높은 유/무해 판별 성능을 얻을 수 있다는 것을 의미한다.

3.2.2 공간적 특징, 시간적 상관성 그리고 모션 반복성 특징의 데이터 융합 방법에 따른 분류 성능 변화

본 절에서는 시맨틱 특징과 시간적 특징을 융합하여 사용할 때의 유해 검출의 성능 변화를 실험으로 보이도록 한다. 실험에 사용되는 DB의 구성은 3.3.1절의 실험에서 사용된 DB를 사용한다. 3.3.1절에서의 실험 결과를 살펴보면, 비디오 세그먼트로부터 5개의 프레임을 추출하여 얻어진 시맨틱 특징들을 평균화 시킨 특징을 이용하였을 때에 가장 높은 유해 검출 성능을 보이므로 향후 특징 융합 실험에서는 이것을 사용한다. <표 3>은 시맨틱 특징에 2.1절에서 정의된 모션 반복성 특징을 2.3.2절에서 언급된 여러 특징 단계 융합 및 결정 단계 융합 방법들로 융합하였을 경우들의 유해 판별 성능 변화를 나타낸다.

<표 3>의 실험 결과를 살펴보면, 특징 단계 데이터 융합을 하여 유해 판별을 수행 하였을 경우 공간적 특징만을 사용하였을 경우보다는 2.5% 그리고 다른 결정 단계 융합 보다는 2% 정도의 분류 성능이 향상되었음을 알 수 있다.

<표 3> 사용되는 특징들과 데이터 융합 방법들의 변화에 따른 유해 판별 성능 변화

Accuracy	Motion periodicity feature is only used	Spatial features are only used	Feature level fusion		Decision level fusion		
			Concatenation	Product	Sum	Max	Concatenation
Sexual intercourse	86.5	93.5	96	94.25	94.00	94.00	93.75

실험 결과를 정리해 보면 공간적 특징만을 사용하고 주위의 다른 프레임의 정보를 사용하지 않았을 경우의 유해 판별 성능은 <표 2>에서와 같이 92.25%였다. 그러나 공간적 특징과 다른 프레임의 정보들을 사용하고 추가적으로 모션 정보를 특징 단계 융합 방법으로 융합하여 유해 판별을 수행할 경우 분류 성능이 96%로 3.75%의 높은 성능 향상을 보였다.

4. 결 론

본 논문에서는 유해 비디오를 분류하기 위해 공간적 특징, 모션 반복성 특징, 프레임 그룹의 특성을 모두 이용하는 새로운 특징 추출 및 특징 융합 방법을 제안하고, 사용되는 특징들의 늘어남과 특징 융합 방법의 변화에 따른 유해 비디오 분류 성능 변화를 체계적인 실험을 통해 제시하였다. 실험 결과 공간적 특징만을 이용할 경우 유해 비디오의 분류 성능은 92.25%을 나타낸 반면, 공간적 특징과 모션 반복성 특징을 특징 단계 융합 방법으로 같이 사용할 경우 보다 3.75% 향상된 성능인 96%를 보였다.

본 논문에서는 유해 비디오 중에서 성교를 포함하는 비디오에 한정하여 유해성 분석 및 분류를 수행하였으나, 향후 연구에서는 성교뿐 아니라 노출, 자위 등과 같은 다양한 형태의 유해 행위를 포함하는 비디오에 대한 분석 및 분류 방법에 대한 연구가 진행되어야 할 것이다.

참 고 문 헌

- [1] N. Rea, G. Lacey, C. Lambe, and R. Dahyot, "Multimodal periodicity analysis for illicit content detection in videos," The 3rd European Conference on Visual Media Production (CVMP 2006), pp.106-114, 2006.
- [2] C. Y. Kim, O. J. Kwon, W. G. Kim, and S. R. Choi, "Automatic system for filtering obscene video," The 10th International Conference on Advanced Communication Technology (ICACT 2008), pp.1435-1438, 2008.
- [3] Z. Qu, L. Ren, A. Guo, and J. Yu, "Implementation of pornographic videos detection system," 2nd International Congress on Image and Signal Processing (CISP 2009), pp.1-4, 2009.
- [4] A. P. B. Lopes, S. E. F. de Avila, A. N. A. Peixoto, R. S. Oliveira, M. de M. Coelho, and A. de A. Araujo, "Nude detection in video using bag-of-visual-features," XXII Brazilian Symposium on Computer Graphics and Image Processing, pp.224-231, 2009.

[5] Z. Y. Qu, Y. Liu, Y. M. Liu, and L. N. Zhang, "A pornographic videos detection method based on optical flow direction's statistical histogram," International Symposium on Computer Network and Multimedia Technology (CNMT 2009), pp.1-4, 2009.

[6] Q. Zhiyi, L. Yanmin, L. Ying, J. Kang, and C. Yong, "A method for reciprocating motion detection in porn video based on motion features," 2nd IEEE International Conference on Broadband Network & Multimedia Technology (IC-BNMT '09), pp.183-187, 2009.

[7] Z. QU, Y. Liu, Y. Liu, K. Jiu, and Yo. Chen, "A porn video detection method based on motion features using HMM," Second International Symposium on computational Intelligence and Design (ISCID '09), pp.461-464, 2009.

[8] C. Jansohn, A. Ulges, and T. M. Breuel, "Detecting pornographic video content by combining image features with motion information," ACM Multimedia 2009, pp.601-604, 2009.

[9] S. M. Lee, H. G. Lee, and T. K. Nam, "A comparative study of the objectionable video classification approaches using single and group frame features," The 16th International Conference on Artificial Neural Networks (ICANN 2006), pp.613-623, 2006.

[10] H. G. Lee, S. M. Lee and T. K. Nam, "Implementation of high performance objectionable video classification system," International Conference on Advanced Communication Technology (ICACT 2006), pp.959-962, 2006.

[11] S. M. Lee, W. C. Shim, and S. H. Kim, "Hierarchical system for objectionable video detection," IEEE Transactions on Consumer Electronics, Vol.55, No.2, pp.677-684, May, 2009.

[12] S. M. Kim, H. S. Min, J. H. Jeon, Y. M. Ro, and S. W. Han, "Malicious content filtering based on semantic features," The ACM International Conference Proceeding 2009, pp.802-806, 2009.

[13] J. H. Jeon, S. M. Kim, J. Y. Choi, H. S. Min, and Y. M. Ro, "Semantic detection of adult image using semantic features," The 4th International Conference on Multimedia and Ubiquitous Engineering (MUE 2010), pp.1-4, 2010.

[14] J. Yang, J. Y. Yang, D. Zhang, and J. F. Lu, "Feature fusion: parallel strategy vs. serial strategy," Pattern Recognition, Vol.36, Issue 6, pp.1369-1381, June, 2003.

[15] X. Zhou, and B. Bhanu, "Integrating face and gait for human recognition at a distance in video," IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, Vol.37, No.5, pp.1119-1137, Oct., 2007.

[16] X. Zhou, and B. Bhanu, "Feature fusion of side face and gait for video-based human identification," Pattern Recognition Vol.41, Issue 3, pp.778-795, Mar., 2008.

[17] H. T. Lin, C. J. Lin, and R. C. Weng, "A note on platt's probabilistic outputs for support vector machines," Dept. Comp. Sci., National Taiwan Univ., 2003 ([online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/plattprob.ps>, Tech. Rep)

[18] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," IEEE Tran. On Pattern Analysis and Machine Intelligence, Vol.22, No.12, pp.1349-1380, Dec., 2000.

[19] S. J. Yang, S. K. Kim, and Y. M. Ro, "Semantic home photo categorization," IEEE Transactions on Circuits and System for Video Technology, Vol.17, No.3, pp.324-335, Mar., 2007.

[20] B. Li, J. H. Errico, H. Pan, I. Sezan, "Bridging the semantic gap in sports video retrieval and summarization," Journal of Visual Communication and Image Representation, Vol.15, Issue 3, pp.393-424, Sep., 2004.



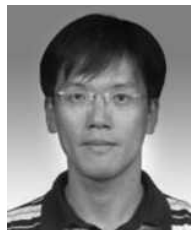
**전 재 현**

e-mail : sunguy@kaist.ac.kr  
 2008년 경북대학교 전기전자컴퓨터학부 (공학사)  
 2008년~현 재 한국과학기술원 전기 및 전자공학과 석박사통합과정  
 관심분야: 이미지/비디오 처리, 패턴 인식, 의료영상처리



**김 세 민**

e-mail : resemim@kaist.ac.kr  
 2006년 안동대학교 멀티미디어학과(공학사)  
 2008년 경북대학교 컴퓨터공학과(공학석사)  
 2008년~현 재 한국과학기술원 정보통신공학과 박사과정  
 관심분야: 이미지/비디오 처리, 패턴 인식



**한 승 완**

e-mail : hansw@etri.re.kr  
 1994년 전남대학교 전산학과(이학사)  
 1996년 전남대학교 전산계산학과(이학석사)  
 2001년 전남대학교 전산학과(이학박사)  
 2001년~현 재 한국전자통신연구원 선임연구원  
 관심분야: 정보보호, 계산이론, 영상 처리, 패턴 인식



**노 용 만**

e-mail : ymro@ee.kaist.ac.kr  
 1985년 연세대학교 전자공학과(공학사)  
 1987년 한국과학기술원 전기 및 전자공학과 (공학석사)  
 1992년 한국과학기술원 전기 및 전자공학과 (공학박사)  
 1993년~1995년 University of California Irvine 방문연구원

1996년~1997년 University of California Berkeley 연구원  
 2006년~2007년 University of Toronto 방문교수  
 1997년~2009년 한국과학기술원 ICC 공학부 교수  
 2009년~현 재 한국과학기술원 전기 및 전자공학과 교수  
 관심분야: 영상 및 비디오 신호처리, MPEG-7, feature recognition, image/video indexing, spectral analysis of image signal