Journal of Korean Institute of Intelligent Systems, Vol. 23, No. 2, April 2013, pp. 184-189
http://dx.doi.org/10.5391/JKIIS.2013.23.2.184

Gaussian Mixture Model과 프레임 단위 유사도 추정을 이용한 유해동영상 필터링 시스템 구현

A Realization of Injurious moving picture filtering system with Gaussian Mixture Model and Frame-level Likelihood Estimation

김민정* · 정종혁*†

Min-Joung Kim, Jong-Hyeog Jeong†

*경운대학교 항공정보통신공학과

† Department of Aviations Information & Communication Engineering, Kyungwoon University

요 약

본 논문에서는 인터넷 및 인터넷 저장 공간에 제한없이 유통되고 있는 유해동영상을 필터링하기 위해 유해동영상에 포함된 특정 소리를 이용한 유해 동영상 필터링 시스템을 제안한다. 이를 위하여 소리의 특성을 잘 표현할 수 있는 Gaussian Mixture Model을 이용하였으며, 필터링 대상 데이터와 소리모델과의 유사도를 계산하기위해 프레임단위 유사도 추정을 이용하였다. 또, 실시간 처리를 위하여 비교대상 데이터의 수를 줄임으로서 실시간 처리가 가능한 프루닝 방법을 적용하였으며, 고정도의 구별 성능을 위하여 기존 화자식별에서 우수한 성능을 보였던 MWMR 방법을 적용하였다. 식별실험결과, 일반 영상과 유해 영상의 기준인 전체프레임 대비 유사도 높은 프레임의 비를 50%로 설정한 경우, 판별 오류율은 6.06%였으며, 프레임 비의 기준이 60%인 경우, 오류율은 3.03%를 나타내어 소리를 이용한 유해동영상 필터링 시스템이 효과적으로 일반영상과 유해영상을 구별할 수 있는 것을 확인하였다.

키워드 : 필터링, 유해동영상, 화자식별, 유사도, 소리식별

Abstract

In this paper, we propose the injurious moving picture filtering system using certain sounds contained in the injurious moving picture to filter injurious moving picture which is distributed without limitation in internet and internet storage space. For this purpose, the Gaussian Mixture Model which can well represent the characteristics of the sound, is used and frame level likelihood estimation is used to calculate the likelihood between filtering target data and the sound models . Also, the pruning method which can real-time proceed by reducing the comparing number of data, is applied for real-time processing, and MWMR method which showed good performance from existing speaker identification, is applied for the distinguish performance of high precision. In the identification experiment result, in case of the frame rate which is the proportion of total frame to high likelihood frame, is set to 50%, identification error rate is 6.06%, and in case of frame rate is set to 60%, error rate is 3.03%. As the result, the proposed system can distinguish between general and injurious moving picture effectively.

Key Words : Filtering, Injurious Moving Picture, Speaker Identification, Likelihood, Sound Identification.

1. 서 론

최근 정보통신의 급격한 발달과 더불어 수많은 영상 데이터들이 인터넷 가상공간 및 인터넷 저장 공간에 활발히 전송되어지고 있다. 이러한 데이터들 중에는 어린이나 청소년들이 접하지 않아야할 불법 성적(性的)영상물이나 성인영상물이 일반 영상물과 함께 존재한다. 이러한 유해영상물은 일반 영상물과 구분되어 엄격한 관리가 필요하지만, 하루에도 수십만 건의 데이터들이 전송되어지는 실정이며, 이러한 데이터들에서 유해영상물을 구별한다는 것은 매우 어려운 일다. 이를 위하여 웹하드 업체 등에서는 표본 검사 및 영상물의 제목 등으로 유해영상물을 판별하고 있지만, 이러한 방법들은 수많은 데이터들 중에서 유해영상물을 구분하여 효과

접수일자: 2012년 11월 30일
심사(수정)일자: 2013년 3월 27일
게재확정일자 : 2013년 4월 4일
† Corresponding author

적으로 관리하기위한 방법으로는 매우 부족하다. 따라서, 유해영상물을 효과적으로 구별하기 위한 방법으로 본 논문에서는 유해영상물에 포함된 성적(性的)소리 데이터를 이용하여 유해영상물을 판별해 내는 필터링 시스템을 제안한다.

유해영상물에는 일반영상물에는 거의 존재하지 않는 남녀의 신음소리, 괴성 등의 성적(性的)소리 데이터가 포함되어 있다. 영상물에서 이러한 소리들을 찾아낼 수 있거나 또는 이러한 소리가 많이 포함되어 있는 영상물만을 구별할 수 있다면 유해영상물의 관리는 한층 체계적으로 가능할 것이다.

유해영상물에 포함된 소리를 찾아내는 방법으로는 화자식별에서 사용되는 방법을 적용할 수 있다. 화자식별이란 발성 화자가 등록된 화자인지 아닌지를 판별하는 것이다. 이를 위해 사전에 등록화자의 음성특징을 추출하여 화자의 음성학적 특징을 모델로 만들어 등록시켜 놓고, 입력되는 화자 발성과 사전에 등록된 화자 모델과의 유사도를 계산하여 등록된 화자인지 아닌지를 판별하는 기술이다.

이 방법을 유해영상물 필터링에 적용하여, 사전에 유해영상물에 포함되어 있는 성적(性的)소리 데이터들을 채집하여 음성학적 소리모델로 등록시켜 놓고, 인터넷상에 존재하는 영상물에서 소리데이터만을 실시간 추출하여 사전에 등록된 성적(性的) 소리모델과 비교하는 것이다. 이렇게 비교한 결과값이 특정 문턱치를 넘으면 해당 영상물은 유해영상물로 간주하고 해당 영상물만을 전수 검사하는 것이다.

이를 위하여 성적(性的)소리 데이터의 음성학적 특징을 표현하기 위하여 화자식별에서 특정소리 패턴을 표현하는데 좋은 성능을 보인 Gaussian Mixture Model[1]을 이용하고, 실시간 처리를 위하여 대상 영상데이터에서 실시간으로 소리데이터만을 추출하고, 소리모델과 추출되어진 소리데이터의 유사도 측정을 위하여 프레임단위 유사도 측정 방법[2],[3],[4]과 비교 모델 프루닝(Pruning) 방법[5]을 적용한다.

본 논문의 구성은 다음과 같다. 2장에서 음성학적 특징을 잘 표현할 수 있는 것으로 알려진 Gaussian Mixure Model과 유사도 추정 방법에 대해서 설명하고, 3장에서 유해동영상을 판별하기 위한 방법 및 실시간 처리를 위한 방법과 본 논문에서 제안하는 시스템을 소개한다. 4장에서 실험결과를 평가하고, 마지막으로 5장에서 결론을 맺는다.

2. 소리 모델과 유사도 측정 방법

2.1 Gaussian mixture model[1]

GMM(Gaussian mixture model)은 출력확률밀도함수가 가우시안 밀도 혼합(Gaussian density mixture)인 1개의 상태만으로 구성된 CHMM(Continuous Hidden Markov Model)의 한 형태이다.

특정소리인식에 GMM을 사용하는 이유로, 첫째, GMM은 음향학적 클래스(Acoustic class) 집합을 모델링 할 수 있다는 것이다. 특정소리에 대응되는 음향 공간은 모음이나 비음, 파찰음과 같은 음소를 표현하는 음향학적 클래스의 집합으로 표현될 수 있는데, 이러한 음향학적 클래스는 특정소리를 구별하는데 이용되는 특정소리에 대한 정보를 가지고 있다[6]. i번째 음향학적 클래스의 스펙트럼 형태는 i번째 component 밀도의 평균 μ_i으로 표현되고, 평균 스펙트럼형태의 변화는 공분산행렬 $\sum_i$로 표현된다. 모든 학습 및 테스트 소리는 레이블되지 않기 때문에, 음향학적 클래스는 hidden으로 볼 수 있다. 독립특징벡터를 가정하면, 이러한 hidden 음향학적 클래스로부터 추출된 특징벡터의 관측밀도가 Gaussian mixture이다.

둘째, Gaussian basis 함수의 선형조합은 샘플분포(Sample distribution)의 클래스를 표현할 수 있다는 것이다[7]. GMM의 성질 중 하나가 임의의 형태를 가지는 밀도를 부드러운 형태로 근사시키는 것이다. unimodal 가우시안 소리모델은 평균벡터(Mean vector)와 공분산(Covariance)으로 소리의 특징분포를 표현하고, VQ-distortion 모델은 특징벡터의 이산집합으로 소리분포를 표현한다. 이와 같은 점을 고려하여 구성된 GMM은 가우시안 함수의 이산집합을 사용하고, 각각의 평균과 공분산을 가지게 함으로써 이들 두 모델의 특징을 혼합한 형태이다[8].

가우시안 혼합 밀도는 M component 밀도의 가중합계이며, 다음의 식에 의해 얻어진다[1].

$$p(x|\lambda)=\sum_{i=1}^{M}c_i N(x;\mu_i,\sum_i) \tag{1}$$

여기서, x는 d-차원 랜덤 백터이며, $c_i, i=1,\ldots,M$은 mixture weight이고, 각 component 밀도는 평균 μ_i과 공분산 $\sum_i$을 가지는 d-variate Gaussian 함수이다.

$$N(x;\mu_i,\sum_i)=\frac{1}{(2\pi)^{\frac{d}{2}}|\sum_i|^{\frac{1}{2}}}exp\left\{-\frac{1}{2}(x-\mu_i)^t\sum_i^{-1}(x-\mu_i)\right\} \tag{2}$$

여기에서, mixture weight는

$$\sum_{i=1}^{M}c_i=1 \tag{3}$$

로 제한한다.

Gaussian mixture density는 모든 component밀도의 mixture weight와 공분산행렬, 평균벡터로 구성된다.

$$\lambda=\{c_i,\mu_i,\sum_i\} \qquad i=1,\ldots,M \tag{4}$$

특정소리모델 학습은 주어진 학습소리로부터 학습특징벡터의 분포와 가장 잘 맞는 GMM, λ 파라미터를 추정하는 것이다. GMM의 파라미터를 추정하는 방법에는 여러 가지가 있으나, 가장 잘 알려진 방법으로는

MLE(maximum likelihood estimation)가 있다. MLE는 주어진 학습데이터에서 GMM의 유사도를 최대화하는 모델 파라미터를 찾는데 사용된다.

T 학습벡터 $X=x_1, x_2, \ldots, x_T$의 열에서, GMM 유사도는 식(5)와 같고,

$$P(X|\lambda)=\prod_{t=1}^{T} p(x_t|\lambda) \tag{5}$$

이를 로그영역에서 표현하면 식(6) 같다.

$$L(X|\lambda)=\sum_{t=1}^{T} \log p(x_t|\lambda) \tag{6}$$

2.2 프레임 단위 최대 유사도(Frame Level Maximum Likelihood)방법

일반적인 소리식별 방법은 Bayes의 정리[7]에 따라 식(7)에서 N개의 소리 중 사후확률 $P(\lambda_i|X)$, $1\le i\le N$를 최대로 하는 모델 λ_{i^*}의 소리 i^*를 찾는 것이다.

$$P(\lambda_i|X)=\frac{p(X|\lambda_i)P(\lambda_i)}{p(X)} \tag{7}$$

여기서, 사전정보가 없기 때문에 소리모델들은 동일하다고 가정한다. 즉, 사전확률 $P(\lambda_i)$는 식 (8)와 같다.

$$P(\lambda_i)=\frac{1}{N},\ 1\le i\le N \tag{8}$$

식 (7)의 분모인 $p(X)$는 발성 X의 빈도에 대한 무조건적인 우도를 나타내며 모든 소리에 대해 동일한 값을 가진다. 따라서 식 (7)에서 식별소리는 $p(X|\lambda_i)$의 사후확률이 최대가 될 때의 소리가 되며 다음과 같이 결정된다.

$$i^*=\arg\max_i p(X|\lambda_i) \tag{9}$$

일반적인 소리식별 시스템에서는 식별소리를 결정하는데 있어서 사용된 소리전체로부터 유사도를 계산한다. 그러나 이 경우 발성문장의 내용이 달라질 경우 문제가 있다. 이를 개선하기 위해 프레임단위를 이용한 소리식별의 경우에는 백그라운드 소리모델에 의한 유사도 정규화를 통해 소리문장의 내용변화에 따른 특징변화를 최소화 할 수 있기 때문에 시스템의 성능을 향상시킬 수 있었다[2][3][4]. 프레임단위 유사도를 식 (7)에 적용하면 식 (10)와 같이 된다.

$$p_{norm}(x_t|\lambda_i)=\frac{p(x_t|\lambda_i)}{\frac{1}{B}\sum_{b=1}^{B} p(x_t|\lambda_b)} \tag{10}$$

여기에서, $p(x_t|\lambda_b)$는 t 프레임에서의 b 백그라운드 소리모델의 유사도이며, $p_{norm}(x_t|\lambda_t)$는 t 프레임에서 백그라운드 소리모델에 의해 정규화 된 i번째 소리의 유사도를 나타낸다.

식 (10)를 이용하여 각 소리의 입력 x_t, $(t=1,2,\cdots,T)$의 각 프레임별 유사도를 계산하여 합한 각 소리 모델 i에 대한 점수로부터 식별소리는 식 (11)를 이용하여 결정한다.

$$Sc_i(X|\lambda_i)=\frac{1}{T}\sum_{t=1}^{T} \log p_{norm}(x_t|\lambda_i) \tag{11}$$

3. 유해동영상 필터링 방법 및 시스템

3.1 수정된 가중모델순위(Modified Weighting Model Rank;MWMR)방법[9][10]

가중모델 순위(Weighting Model Rank; WMR) 방법[11]은 식별소리를 결정하는 점수를 계산할 때 테스트 소리데이터와 소리모델들과의 프레임 유사도를 그대로 사용하지 않고, 계산된 유사도들의 상대적 순위에 따라 정해진 가중치를 스코어 계산에 사용함으로서 소리데이터들 간의 변별력을 제고하는 방법이다. 이 방법은 프레임 단위에서 소리모델들 사이의 변별력을 크게 할 수 있어 최대유사도(Maximum Likelihood; ML) 방법보다 식별성능을 향상시킬 수 있지만, 각 프레임의 유사도에 의한 변별력은 고려되지 않아 소리의 음성학적 특성을 잘 표현하지 못하는 프레임의 경우에도 유사도의 순위만 높다면 큰 가중치를 부여함으로써 결과적으로는 소리의 변별력을 감소시킬 수 있는 단점이 있다. 즉, i_{th}프레임의 유사도 순위가 $i-1_{th}$프레임의 유사도 순위와 상대적으로 동일한 위치를 차지할 경우 $i-1_{th}$ 프레임과 동일한 가중치를 갖게 된다. 그러므로 전체 프레임에서 순위의 합계가 동일한 두 소리모델이 존재할 경우 동일한 가중치 합을 출력하는 단점이 있다.

이를 개선할 수 있는 한 방법으로 가중치를 결정하는데 있어서 프레임 유사도의 크기까지 고려할 경우 소리모델들 사이의 변별력을 좀 더 크게 할 수 있다. 즉, 프레임 단위 유사도의 상대적 위치에 따라 결정된 가중치와 프레임단위 유사도를 곱한 값을 식별소리를 결정하는 스코어 계산에 이용한다. 이러한 방법을 수정된 가중모델순위(Modified WMR; MWMR) 방법이라 한다.

MWMR 방법을 간략히 설명하면 다음과 같다.

첫 번째 단계에서는 각 테스트 벡터 x_t, $t=1,2,\cdots,T$에서 프레임 유사도 $p(x_t|\lambda_i)$, $i=1,2,\cdots,N$을 계산하고 이를 오름차순으로 정렬한다. 즉, 가장 큰 유사도를 가지는 소리모델은 최상위에 위치시키고, 가장 낮은 유사도를 가지는 소리모델은 최하위에 위치시킨다.

두 번째 단계에서는 소리모델의 각 순위에 따라 가중치 $w(r)$을 결정한다. 이때, 가중치는 식 (12)과 같

은 지수함수를 이용하는 방법이 사전실험에서 우수한 성능을 나타내었다[6][10].

$$w(r_\lambda) = \exp(\alpha - \beta r_\lambda), \quad r_\lambda = 1, \cdots, N \tag{12}$$

여기서, α와 β는 순위의 확률밀도함수에 따라 결정되는 가중치 요소로서 소리 데이터 수에 밀접한 관계가 있으므로, 소리 데이터 수에 따라 새롭게 계산되어야 한다. 그리고 여기서 $w(1) \approx N$이 되도록 설정한다[6].

세 번째 단계에서는 각 모델 λ_i의 순위에 해당하는 가중치 $w_t(r_{\lambda_i})$와 유사도 $p(x_t|\lambda_i)$의 곱을 이용하여 전체 스코어 $Sc(X|\lambda_i)$를 계산한다.

$$\log Sc(X|\lambda_i) = \sum_{t=1}^{T} p(x_t|\lambda_i) w_t(r_{\lambda_i}) \tag{13}$$

여기서, $w_t(r_{\lambda_i})$와 $p(x_t|\lambda_i)$는 각각 시간 t에서 순위가 r_{λ_i}인 모델 i의 가중치와 프레임 유사도를 나타낸다.

그림 1에 MWMR 방법의 흐름도를 나타내었다.

이 MWMR 방법은 이전 연구[9][10]에서 WMR 방법보다 향상된 화자식별 성능을 나타내어 그 유효성을 확인할 수 있어 본 논문에서는 유해동영상 필터링을 위한 방법으로 MWMR을 적용하였다.

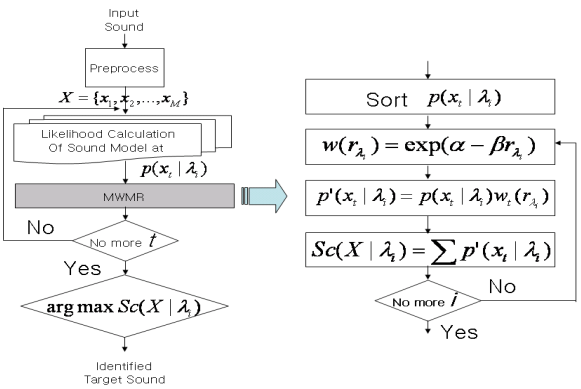

그림 1. Modified Weighting Model Rank 방법 흐름도

Fig. 1. Flow-chart of Modified Weighting Model Rank Method

3.2 실시간 필터링을 위한 비교 데이터 프루닝

일반적인 특정 소리식별 방법에서는 식별대상의 데이터 수가 많을수록 계산량이 증가하며 이는 식별 시간의 증가를 가져오게 되어 실시간 특정소리 식별시스템 구현이 어렵게 된다. 계산량을 줄이기 위한 방법으로서는 비교대상이 되는 입력 프레임의 수를 줄이는 방법이나 비교 프레임 수를 줄이는 방법을 고려할 수 있을 것이다. 입력 프레임수를 줄이는 방법으로서는 대상소리의 특성이 충분히 포함된 입력 프레임만을 사용하는 방법[5]등이 제시되었으나, 이 경우, 특정소리의 특징이 포함된 데이터의 감소로 인한 식별성능의 저하가 발생한다. 이를 해결하고 전체적인 계산량을 감소시키기 위하여 비교대상 프레임 수를 줄이는 방법을 고려할 수 있다. 한편, 고정도의 식별성능을 위해서는 기존에 많이 이용되고 있는 ML 방법보다 향상된 식별성능을 나타내는 WMR 방법이나 MWMR 방법 등의 적용을 고려할 수 있다. 하지만 WMR 방법이나 MWMR 방법의 경우, 대상소리 데이터 수가 달라질 때 마다 가중치 요소가 새롭게 계산되어야 하는 단점이 있으므로 이를 함께 고려하여 주어야만 한다. 따라서 본 논문에서는 고정도의 특정소리 식별성능을 유지하기 위하여 MWMR 방법을 적용하면서도 식별시간의 감소를 위하여 추가적인 가중치 요소의 계산없이 전체적인 계산량을 줄일 수 있는 비교 모델 프루닝(Pruning)[5] 방법을 적용한다. 이 방법은 소리 선택단과 소리 식별단으로 구성되며, 각 단에서의 처리순서는 다음과 같다.

소리선택단

a. 입력프레임의 일부와 전체 등록 소리모델들과의 프레임유사도 계산.
b. 각 소리모델별 누적 프레임 유사도 값의 크기에 따라 각 소리모델을 내림차순으로 정렬
c. 상위 소리모델들의 일부만을 선택하여 소리 식별단에서 사용하고 나머지 소리모델은 프루닝.

소리식별단

a. MWMR을 적용하여 선택된 상위 순위의 소리 모델들과 입력소리데이터의 전체 프레임을 이용하여 프레임유사도 계산
b. 각 소리별 누적 프레임 유사도를 구한 후 가장 큰 누적 프레임 유사도를 가지는 소리를 식별소리로 결정

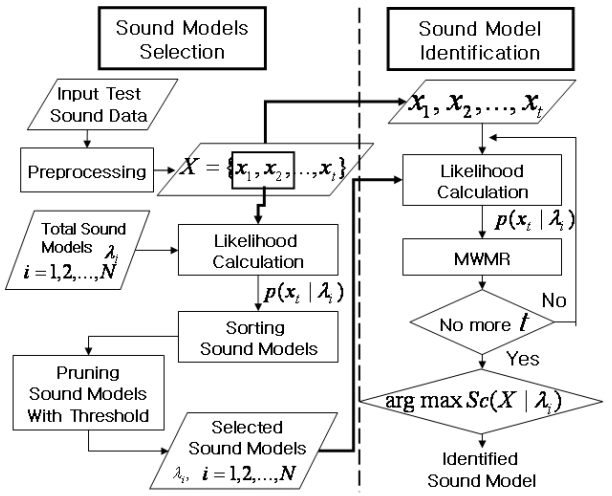

그림 2. 유해동영상 필터링 시스템의 흐름도

Fig. 2. Flow-chart of injurious Moving Picture Filtering System

3.3 유해동영상 필터링 시스템

그림 2에 유해동영상 필터링 시스템의 흐름도를 소리선택단과 소리 식별단으로 구분하여 나타내었다. 소리 식

별단에서는 소리 선택단에서 선택된 소리모델들 중에서 식별소리모델을 결정하므로 식별성능의 저하없이 전체적인 계산량을 감소시킬 수 있다. 또, MWMR 방법을 적용하므로 비교대상 소리모델들을 사전에 설정된 가중치 요소에 맞는 소리 데이터 수까지 프루닝할 수 있어 특정소리의 식별 성능 향상도 기대할 수 있다.

4. 실험 및 결과

4.1 실험데이터 및 실험 방법

유해소리 패턴은 동양, 서양, 남성, 여성, 성인, 기타 행위 소리 등의 조합으로 구성할 수 있으나, 본 논문에서는 제안한 시스템의 유효성을 확인하기 위하여 동양-성인-여성 음성의 1가지 패턴으로 제한하였다. 영상데이터의 종류로는 드라마, 오락, 다큐 등의 일반 영상물과 유해영상물, 두 종류로 선택하였다. 비교소리모델 작성을 위하여 다수의 일반 영상으로부터 2.7GByte의 소리데이터를 추출하여 일반소리모델을 작성하였으며, 유해영상의 소리모델 작성을 위해서 다수의 유해영상물로부터 200MByte의 성적(性的)소리 데이터를 수작업으로 분류 후 유해소리모델을 작성하였다. 이때, 비교모델 작성을 위한 소리데이터의 한 프레임 길이는 2초로 하였다.

테스트를 위한 영상데이터로는 비교모델작성에 사용하지 않은 50편의 일반 영상과 50편의 유해영상을 이용하였으며, 테스트 소리데이터의 한 프레임 길이는 60초로 하였다. 실험은 작성된 일반 및 유해소리모델과 입력데이터의 각 프레임과의 유사도를 측정하였다. 각 데이터의 유해 및 일반영상의 판별을 위하여 두 소리모델과 입력 데이터의 각 프레임 유사도 추정 결과로부터 일반영상과 유사도가 높은 프레임의 수가 입력데이터 전체프레임의 50%이상이면 일반영상으로 판단하고, 유해영상과 유사도가 높은 프레임이 전체 프레임의 50%이상이면 유해영상으로 판단하였다.

소리모델 작성 및 실험에서 소리데이터의 분석조건으로 GMM 혼합수는 식별률 및 계산량을 고려하여 16으로 고정하였으며, 특징 파라미터는 켑스트럼 계수 10차와 회귀계수 10차만을 사용하였다. 음성 특징 파라미터의 분석조건을 표 1에 나타내었다.

표 1. 전처리를 위한 분석조건
Table 1. Analysis condition for pre-processing

Sampling Rate/Resolution	16 kHz/16bits
Pre-emphasis coefficient	0.98
Hamming Windows	yes
Frame length	320 points
Frame Shift	160 points
Cepstrum vector dimension	10

4.2 실험결과 및 고찰

그림 3에 유해동영상 식별 시스템의 실험결과를 나타낸다. 그림 3에서 세로축은 유해 영상과 일반 영상의 기준이 되는 유사도 높은 프레임과 전체프레임의 비를 나타내고, 가로축은 100개의 영상물의 순번을 나타내며, 중앙의 굵은 선은 유해 영상의 판별 기준인 프레임비의 문턱치를 나타낸 것이다. 그림 3에서 프레임비가 높을수록 일반영상과 유사도가 높은 프레임이 많은 것이며, 프레임비가 낮을수록 유해영상과 유사도가 높은 프레임이 많은 경우이므로 판별기준이 되는 프레임 비의 문턱치를 조절함으로서 판별 오류율을 조절할 수 있다.

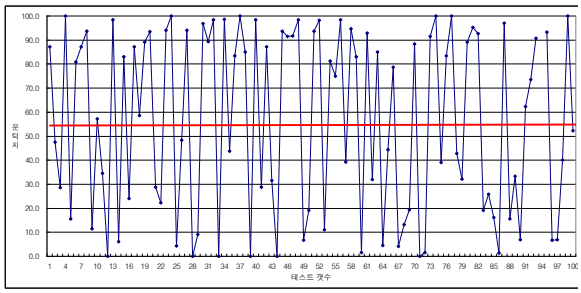

그림 3. 유해동영상 식별 실험결과
FIg. 3. Identification Restlt of Injurious Moving Picture

식별실험 결과, 유사도 높은 프레임과 전체 프레임의 프레임비가 50%인 경우 유해영상과 일반영상의 판별 오류율은 6.06%였으며, 60%인 경우 판별 오류율은 3.03%였다.

오류가 발생하는 경우를 분석한 결과, 테스트 데이터에 배경 잡음이나 배경 음악이 존재하거나 음성보다 큰 경우 발생한 것으로 분석되었다.

5. 결론 및 향후 연구

본 논문에서는 인터넷 상에 존재하는 유해동영상을 필터링하기 위한 유해 동영상 필터링 시스템을 제안하였다. 이를 위하여 소리데이터의 음성학적 특징을 잘 표현할 수 있는 Gaussian Mixture Model을 이용하여 소리모델을 작성하였으며, 기존 화자식별에서 우수한 성능을 보였던 MWMR 방법을 적용하여 유사도 측정을 수행하였다. 또, 실시간 필터링을 위하여 비교대상 소리데이터의 수를 줄임으로서 실시간 처리가 가능한 프루닝 방법을 적용하였다. 식별실험결과, 유사도 높은 프레임과 전체프레임의 비를 50%로 정한 경우 구별 오류율은 6.06%였으며, 프레임 비가 60%인 경우, 오류율은 3.03%를 나타내어 영상에 포함된 소리데이터를 이용한 유해동영상 필터링 시스템의 유효성을 확인할 수 있었다. 유해영상과 일반영상의 구별 오류의 원인으로는 유해 소리와 배경잡음이 함께 존재하는 경우나 배경 잡음이 유해 소리보다 큰 경우인 것으로 분석되었다. 향후, 오류율을 줄이기 위해 이러한 배경 잡음 처리 방법이 보완되어야 할 것이고, 다양한 유해소리패턴을 추가하여 세밀한 필터링이 가능하도록 해야 할 것이다.

References

[1] D. A. Reynolds and R. C. Rose, "Robust Text - Independent Speaker Identification using Gaussian Mixture Speaker Models," *IEEE Trans. on SAP*, Vol. 3, No. 1, pp. 72-83, 1995.

[2] D.A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models", *Speech Communication* , Vol. 17, No.1-2,pp.91-108, 1995.

[3] A. Rosenberg, J. DeLong, C.Lee, B.Juang and F. Soong, "The use of cohort normalized scores for speaker verification", *proc. ICSLP*, pp.599-602, 1992.

[4] T. Matsui and S. Furui, ""Likelihood normalization for speaker verification using a phoneme- and speaker-independent model," *Speech Communication*, Vol. 17, pp. 109-116, Aug. 1995.

[5] M. J. Kim, S. J. Oh, H. Y. Jung, S. Y. Suk, H. Y. Chung and H. Y. Chung, "Modified Weighting Model Rank Method for Improving the Performance of Real-Time Text-Independent Speaker Recognition System," *Journal of the Acoustical Society of Korea*, Vol. 21, No. 1(s), pp. 107-110, 2002.

[6] H. Matsumoto and H. Wakita, "Vowel normalization by frequency warped spectral matching," *Speech Communication,* Vol. 5, No. 2, pp. 239-251, 1986.

[7] K. Fukunaga, Introduction to Statistical Pattern Recognition. Academic Press, Inc., second ed., 1990.

[8] H. Gish and M. Schmidt, "Text-independent speaker identification," *IEEE Signal Processing Magazine*, pp. 18-32, Oct. 1994.

[9] M. J. Kim, S. J. Oh, H. Y. Jung, and H. Y. Chung, "Frame Selection, Hybrid, Modified Weighting Model Rank Method for Robust Text-Independent Speaker Identification," *Journal of the Acoustical Society of Korea*, Vol. 21, No. 8, pp. 735-743, 2002.

[10] M. J. Kim, S. J. Oh, S. Y. Suk, H. Y. Jung, and H. Y. Chung, "Modified Weighting Model Rank Method for Improving the performance of real-time text-independent speaker recognition system," *Proc., Acous. Soc. Korea,* pp. 107-110, July 2002.

[11] K. Markov and S. Nakagawa, "Text-independent speaker identification on TIMIT database," *Proc. Acoust. Soc. Jap.*, pp. 83-84, March 1995.

저 자 소 개

김민정(Min-Joung Kim)

1997년 : 경일대학교 전자공학과 공학사
1999년 : 영남대학교 멀티미디어 통신공학과 공학석사
2003년 : 영남대학교 정보통신공학과 공학박사
2010년～현재 : 경운대학교 항공정보통신공학과 교수

관심분야 : 음성신호처리, 패턴매칭, 화자인식
Phone : +82-53-479-1315
E-Mail : manjukz@naver.com

정종혁(Jong-Hyeog Jeong)

1992년 : 부경대학교 전자공학과 공학사
1994년 : 동아대학교 전자공학과 공학석사
1999년 : 한국해양대학교 전자통신공학과 공학박사
2000년～현재 : 경운대학교 항공정보통신공학과 교수

관심분야 : 신호처리, USN, 전파통신
Phone : +82-54-479-1314
E-Mail : jhjeong@ikw.ac.kr