

내용기반 영역에서의 효과적인 장면전환에 관한 연구

이 해 문[†] · 오 일 균^{††} · 이 재 연^{†††} · 배 영 래^{††††} · 장 종 환^{†††††}

요 약

동영상의 데이터베이스에서 원하는 영상검색을 위해 히스토그램을 이용한 알고리즘이 널리 사용되었다. 본 논문에서는 HVS(Human Visual System)의 특성을 이용한 히스토그램 알고리즘을 제안하였고 또한 기존의 알고리즘과 제안한 알고리즘의 성능을 분석하여 성능이 우수한 알고리즘을 제안한다.

A Study on the Scene Change Detection on the Content-Based Domain

Hae-Moon Lee[†] · Il-Kyun Oh^{††} · Jae-Yeon Lee^{†††} ·
Young-Rae Bae^{††††} · Jong-Whan Jang^{†††††}

ABSTRACT

Histogram-based methods have been used generally for retrieving an image to be searched in the video database. In this paper, we present an algorithm to retrieve an image incorporating the HVS. Comparing the performance of the previous algorithm with that of the proposed one, the new and better algorithm is proposed than the previous algorithm.

1. 서 론

최근 들어 대용량의 데이터를 저장할 수 있는 장비와 초고속 통신망의 발전으로 많은 양의 멀티미디어 데이터를 처리할 수 있는 환경으로 발전하고 있다. 멀티미디어 환경하에서의 중요한 문제의 하나는 얼마나 빠른 시간 내에 얼마나 정확한 정보를 찾아낼 수 있는 멀티미디어 정보검색과 데이터베이스 구축이 연구 중이며, 이에 대한 연구가 활발히 진행되고 있다.

멀티미디어 데이터 중 가장 효과적인 정보전달 매체는 동영상이다. 동영상 정보의 내용기반 정보검색을 하기 위한 중요한 기술의 하나는 video indexing이다 [1, 6-7]. Nagasaka[5]는 내용기반 정보검색을 위한 video indexing을 하기 위해 히스토그램 방법을 제안하였지만 히스토그램의 임계값이 영상의 값의 분포도에 따라 변경되는 문제점을 가지고 있다. 이 문제를 해결하기 위해 Hampapur[2] 및 Shahraray[9] 등은 모델기반 장면전환기법을 제안하였다. 그러나 이 방법들은 압축이 안된 영상에만 적용가능 하므로 MPEG으로 압축된 동영상 데이터에 이 방법을 적용하기 위해서는 먼저 디코딩 한 다음 이 방법을 적용함으로 데이터를 처리하여 시간이 오래 걸리고 메모리양이 많이 필요한 문제점을 나타내고 있다.

※ 본 논문은 1998년도 배재대학교 공학연구소에서 일부 지원하였음.

† 준 회원 : 한국전자통신연구원

†† 준 회원 : 배재대학교 대학원 정보통신공학과

††† 정 회원 : 한국전자통신연구원 책임연구원

†††† 정 회원 : 한국전자통신연구원 영상정보처리연구팀 팀장

††††† 종신회원 : 배재대학교 컴퓨터전자정보공학부 교수

논문접수 : 1999년 1월 22일, 심사완료 : 1999년 8월 21일

최근에는 이 문제를 해결하기 위해 압축된 비트스트림이 갖고 있는 내용을 직접 분석하여 장면전환에 필요한 의미 있는 특징 정보를 추출하여 동영상의 장면 전환 및 검색에 활용하려고 노력하고 있다[3, 4, 10]. 그리고 압축된 비트스트림 데이터의 DCT(Discrete Cosine Transform)계수를 사용하여 장면전환을 구하는 방법 뿐만 아니라, 장면전환에 중요한 정보를 나타내는 P 및 B 픽처의 Motion 벡터 정보를 사용하여 신뢰성 있는 장면전환 추출 알고리즘을 개발하려고 한다. 그러나 현재까지 제안된 방법들을 응용의 최종관찰자인 인간이 어떻게 보고, 무엇을 보는지를 고려하지 않고 장면전환 알고리즘을 설계하려고 노력하고 있다. 본 논문에서는 DC coefficient 값에 HVS(Human Visual System)의 특성을 이용한 새로운 알고리즘을 제안하여 내용기반 검색을 위한 장면전환 추출 알고리즘의 성능을 향상하였다.

2. 본 문

동영상의 영상검색을 위해 장면을 분할하기 위해서는 위에서 언급한 바와 같이 여러 가지 방법들이 제안되어 있다. 이들 중 구현하기 간단하고 널리 사용되는 프레임간의 히스토그램 차이를 이용하는 방법이 식 (1)에 주어진다.

$$sad(i, i+1) = \sum_{j=0}^{N-1} |F_i(j) - F_{i+1}(j)| > T \quad (1)$$

sad(sum of the absolute difference)은 i 번째 프레임과 $i+1$ 번째 프레임간 차의 절대값의 합을 나타내는 함수, N 은 인텐서티 값의 범위, $sad(i, i+1)$ 는 i 번째 프레임과 이웃하는 $i+1$ 번째 프레임과의 유사성을 측정한다. $F_i(j)$ 는 i 번째 프레임의 j 인텐서티 값을 갖는 히스토그램 함수이며 T 는 장면전환의 임계값이다. 유사한 프레임들이 많이 있는 곳에서는 $sad(i, i+1)$ 가 작은 값을 나타내지만 장면전환이 있는 곳에서는 $sad(i, i+1)$ 가 크게 나타나며, 이 값이 임계값 T 보다 크면 장면전환이 있는 것으로 결정한다. 본 논문에서는 기존에 제안한 방법에 성능을 개선하기 위해 HVS 특성을 고려한 새로운 알고리즘을 제안한다. HVS특성은 낮은 값 및 높은 값을 갖는 인텐서티 부근에서는 $sad(i, i+1)$ 은 큰 값을 갖지만 최종 관찰자인 인간은 낮은 값 및 높은 값의 인텐서티 부근에서는 HVS의 인

식률이 낮으므로 장면전환을 인식할 수 없는 경우도 있다[8]. 이러한 경우에는 장면전환으로 인식되지 않아야 하지만 $sad(i, i+1)$ 값이 임계값 보다 높으면 장면전환으로 인식되어 오류가 발생한다. 이러한 문제점을 개선한 알고리즘을 제안하여 기존의 방법과 성능을 분석하여 더 좋은 결과를 얻을 수 있다.

영상은 Y, C_b, C_r 성분으로 구성되어 있으므로 식 (1)을 식 (2)로 확장할 수 있다.

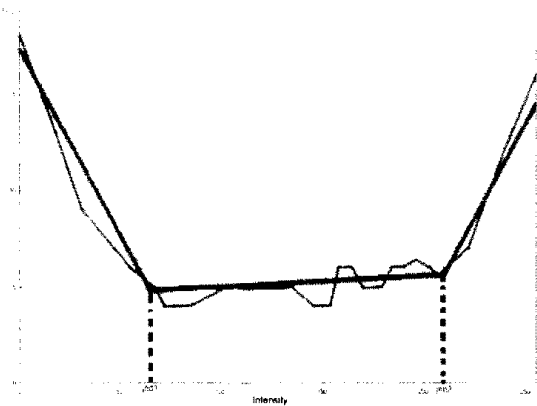
$$sad(i, i+1) = \sum_{j=0}^{N-1} \left(\begin{aligned} &|F_i^Y(j) - F_{i+1}^Y(j)| \\ &+ |F_i^{C_b}(j) - F_{i+1}^{C_b}(j)| \\ &+ |F_i^{C_r}(j) - F_{i+1}^{C_r}(j)| \end{aligned} \right) \quad (2)$$

본 논문에서는 Y, C_b, C_r 성분 중 정보의 대부분을 가지고 있는 휘도성분인 Y 의 HVS의 특성을 이용한다. Y 의 HVS의 중요한 특성 중에 하나는 JND(Just Noticeable Difference)이다. JND는 HVS특성중 인간의 contrast 인식률을 효과적으로 측정할 수 있는 방법이다. 본 논문에서는 JND를 측정하기 위해 split-merge 방법을 선택하였다. 실험을 하기 위해 측정하려는 정사각형의 영상 크기를 반으로 나눈다. 왼쪽에 있는 $\frac{1}{2}$ 크기의 영역을 상수의 인텐서티로 고정하고 오른쪽에 있는 $\frac{1}{2}$ 크기의 영역을 관찰자가 차이를 인식할 때 까지 인텐서티의 값을 선형적으로 증가한다. 차이를 처음 인식할 때 오른쪽과 왼쪽의 차이의 값이 왼쪽에 있는 영상의 인텐서티에 대한 JND값이다.

본 논문에서는 5명의 실험자가 인텐서티 0부터 255까지의 모든 인텐서티에 대한 JND 값을 측정하였다. 5명이 측정한 JND의 평균 곡선은 (그림 1)에 가는 선으로 주어진다. 실험적인 결과는 HVS의 contrast 인식률의 특성과 일치한다[6]. JND값은 낮고 높은 인텐서티 부근에서 가장 높았으며, 중간 부분인 인텐서티 부분에서는 JND값이 최소값을 가지므로 가장 민감한 반응을 나타냈으며 거의 상수 값을 갖는다.

식 (2)에 제안한 히스토그램 방법에 HVS의 특성을 고려한 가중치 계수를 계산하기 위해서는 측정한 JND 곡선에 식 (3)에 주어진 Linear Regression Method를 사용하여 1차 곡선으로 근사화 한다.

$$Min \left\{ \sum_{i=0}^{N-1} |a_k x + b_k - J_i|^2 \right\} \text{ for } a_k \text{ and } b_k \quad (3)$$



(그림 1) 5사람이 측정하여 평균한 JND 값
(굵은 선은 JND의 근사화한 곡선이다)

여기에서 a_k 및 b_k 는 1차 곡선의 계수이며 J_i 는 (그림 1)에 가는 선으로 나타난 JND 곡선의 인텐시티 i 의 JND값이다. 식 (3)에 최소값을 갖는 계수를 구하여 (그림 1)에 굵은 선으로 나타낸다. jnd_1 혹은 jnd_2 는 3개의 곡선을 경계하는 임계값이다. (그림 1)에서 jnd_1 과 jnd_2 는 각각 60와 210이다.

동영상의 히스토그램 분포가 대부분 0와 jnd_1 사이 또는 jnd_2 와 255 영역에 집중 분포되어 있는 프레임의 경우 인간은 장면전환 인식율이 저하되므로 가중치 계수 $W^Y(j)$ 값을 작게 하여 sad 값을 낮추어 장면전환 정확성을 높일 수 있다.

반면에 동영상의 히스토그램이 jnd_1 및 jnd_2 영역 사이에 대부분 존재하는 프레임에서는 인간의 인식률의 크게 됨으로 가중치계수 $W^Y(j)$ 을 크게 하여야 한다. 모든 영역에 걸쳐 고르게 분포하는 프레임에서는 평균적으로 변화가 없어야 함을 나타낸다. 제안한 알고리즘에 가중치 계수를 삽입하기 위해서는 HVS의 인식이 낮은 값에서는 가중치계수를 낮게 하고 인식이 높은 곳에서는 가중치계수를 높은 값으로 한다. 그렇게 하기 위해서는 근사화 한 3개의 곡선을 x축으로 회전하면 가중치계수를 계산할 수 있다. 회전한 3개의 곡선으로부터 얻은 함수는 $f_u(i)$ 라고 한다.

$f_u(i)$ 로부터 가중치 계수를 계산하는 식은 (4)에 주어진다.

$$K \cdot \frac{1}{N} \sum_{i=0}^{N-1} f_u(i) = 1 \quad (4)$$

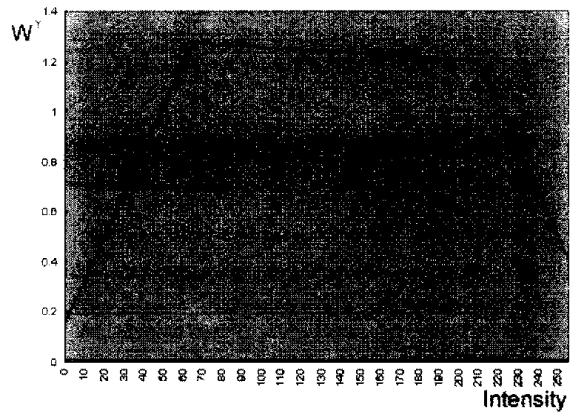
여기에서 $N=256$ 이다. 식 (4)로부터 가중치 계수 $W^Y(j)$ 는 식 (5)로 주어진다.

$$W^Y(j) = \frac{1}{K} \cdot f_u(j) \quad (5)$$

여기에서 $\frac{1}{N} \sum_{i=0}^{N-1} W^Y(i) = 1$ 이다.

식 (5)로 부터 얻은 JND 특성을 이용한 가중치 계수의 함수 $W^Y(j)$ 는 (그림 2)에 주어진다. $W^Y(j)$ 를 사용하면 식 (2)는 식 (6)과 같이 표현할 수 있다.

$$sad_w(i, i+1) + \sum_{j=0}^{N-1} \left(\begin{array}{l} W^Y(j) |F_i^Y(j) - F_{i+1}^Y(j)| \\ + |F_i^{C_1}(j) - F_{i+1}^{C_1}(j)| \\ + |F_i^{C_2}(j) - F_{i+1}^{C_2}(j)| \end{array} \right) \quad (6)$$



(그림 2) 제안한 알고리즘의 가중치 계수의 곡선

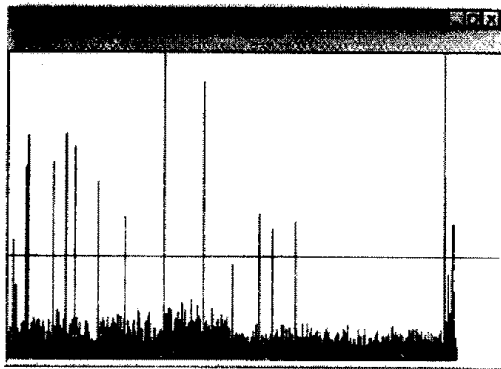
제안한 알고리즘을 요약하면 다음과 같다.

- ① 프레임을 조사하여 장면전환 임계값 T 를 계산한다.
- ② HVS의 특성을 고려하기 위해 Y성분의 JND값을 측정하여 가중치 계수 $W_i^Y(j)$ 을 계산한다.
- ③ 식 (6)에 주어진 제안한 함수값 $sad_w(i, i+1)$ 값을 계산한다.
- ④ $sad_w(i, i+1)$ 의 값이 임계값 T 보다 작으면 장면전환이 없고, 임계값 T 보다 크면 장면전환이 일어난 것으로 결정하여 키 프레임을 추출한다.
- ⑤ 최종 프레임에 도착할 때까지 반복 수행한다.

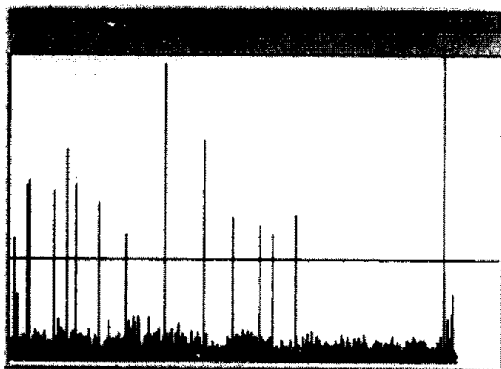
3. 성능분석 및 결과

일반적으로 사용하는 광고 및 뉴스의 두 종류 MPEG

샘플데이터를 사용하여 기존의 방법과 제안한 방법의 장면전환 성능분석을 하였다. MPEG 샘플데이터의 장면전환 검출 성능에는 데이터의 값의 분포 정도에 따라 차이가 날 것이다. HVS 특성을 이용하지 않는 기존 방법과 HVS의 특성을 이용한 제안한 방법의 히스토그램의 분포는 (그림 3)과 (그림 4)에 주어진다.

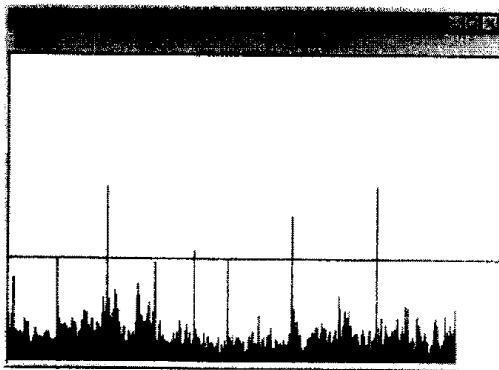


(a) 기존의 방법에 의한 sad 값 계산

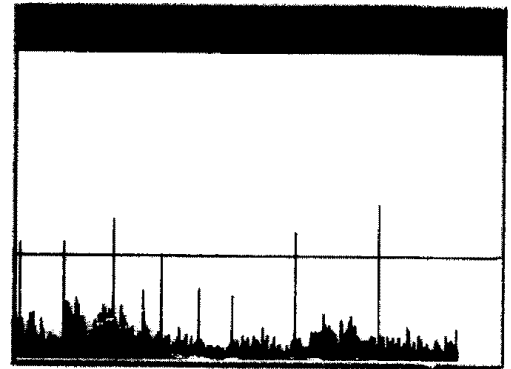


(a) 제안한 방법에 의한 sad_w 값 계산

(그림 3) 광고 MPEG 데이터를 사용한 기존의 방법과 제안한 방법과의 성능분석



(a) 기존의 방법에 의한 sad 값 계산



(b) 제안한 방법에 의한 sad_w 값 계산

(그림 4) 뉴스 MPEG 데이터를 사용한 기존의 방법과 제안한 방법과의 성능분석

2개의 MPEG 데이터의 성능분석은 <표 1>에 주어진다.

<표 1> 2개의 MPEG 샘플데이터로부터 얻은 실험결과

		경우 1	경우 2	경우 3
광고	기존의 방법	15	1	2
	제안한 방법	15	0	0
뉴스	기존의 방법	48	2	1
	제안한 방법	48	0	0

경우 1은 샘플이 갖고 있는 실제의 장면전환 개수이고, 경우 2는 장면전환을 인식하지 못한 개수이며, 경우 3은 장면전환이 아닌데 잘못 인식한 개수이다.

기존에 제안한 알고리즘의 장면전환 임계값은 반복 실험을 통해 장면전환 검출 실패율이 최소화하는 적절한 임계값으로 결정하였다. <표 1>에서 경우 2인 경우는 영상의 데이터가 대부분 jnd_1 과 jnd_2 사이에 분포되어 있어 인간의 인식률은 높는데 같은 임계값을 사용함으로 장면전환을 검출하지 못하였는데 제안한 알고리즘에서는 높은 가중치 계수를 곱하여 sad_w 값을 증가하여 임계값 보다 크므로 장면전환 검출하는데 성공하였다. 또한 경우 3인 경우는 영상의 데이터가 대부분 jnd_1 보다 작거나 jnd_2 보다 큰 값에 분포되어 있으므로 낮은 가중치 계수를 곱하여 잘못 인식한 장면전환을 인식하지 못하게 한다. 제안한 알고리즘은 기존에 광범위하게 사용한 히스토그램 방법에 HVS의 특성을 고려한 가중치 계수를 삽입하여 실패율이 감소하였다.

본 논문에서는 동영상데이터의 휘도성분의 HVS특

성을 이용한 알고리즘을 제안하여 기존의 방법보다 더 좋은 결과를 얻었다. 동영상 데이터는 휘도 신호인 Y 성분뿐만 아니라 색차 신호인 C_b 및 C_r 을 포함하고 있으므로 HVS을 색차 신호에 확장할 수 있는 알고리즘 연구가 진행 중이다. 또한 제안된 방법을 압축된 비트 스트림 동영상에 직접 적용할 수 있으면 상당히 많은 시간을 필요로 하는 디코딩 처리시간을 제거함으로써 일반적인 PC에서도 실시간으로 동작할 수 있다.

참 고 문 헌

[1] M. Flickner et al. Query by image and video content : The QBIB system, IEEE Computer, 28(9) : 23-32, September, 1995.

[2] A. Hampapur, R. Jain, and T. Weymouth, Production Model Based Digital Video Sequention, Multimedia Tools and Applications, Vol.1, No.1, pp.9-46, Mar. 1995.

[3] V. Kobla and D. Doerman, Compressed Domain Video Indexing Techniques Using DCT and Motion Vector Information in MPEG Video, SPIE, Vol.3022, pp.200-211, 1997.

[4] Jianhao Meng, Y. Juan, S. F. Chang, Scene Chang Detection in a MPEG Compressed Video Sequences, Proc. of SPIE, Vol.2419, pp.14-25, 1995.

[5] A. Nagasaka and Y. Tanaka, Automatic Video Indexing and Full Motion Search for Object Appearance, Proc. Of IFIP on Visual Database System, pp.113-127, Sep. 1991.

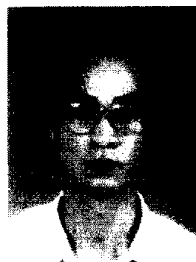
[6] V. Ogle and M. Stonebraker, Chabot: Retrieval from a relational database of images. IEEE Computer, 28(9) : 40-48, September 1995.

[7] A. Pentland, R. Picard, and S. Sclaroff, Photo-book : Content-based manipulation of image database, Internaltional Journal of Computer Vision, 18(3) : 233-254, June 1996.

[8] W. K. Pratt, Digital Image Processing, Wiley Interscience, 1998.

[9] B. Shabaray, Scene Change Detection and Content Based Sampling of Video Sequences, Proc. of SPIE, Vol.2419, pp.2-13, Feb. 1995.

[10] Boon Lock Yeo and Bede Liu, Rapid Scene Analysis on Compressed Video, IEEE Transaction on Circuit and System for Video Technology, Vol.5, No.6, pp.533-544, December 1995.



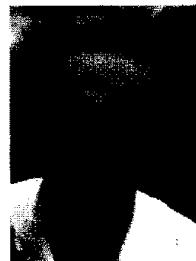
이 해 문

e-mail : hmlee@etri.re.kr
 1989년 건국대학교 전자공학과 졸업(학사)
 1989년~현재 한국전자통신연구원
 관심분야 : 멀티미디어, 비디오검색, 영상처리



오 일 군

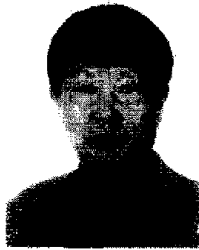
e-mail : oilvirus@icelab.paichai.ac.kr
 1999년 배재대학교 정보통신공학과 졸업(학사)
 1999년 3월~현재 배재대학교 정보통신공학과 석사과정
 관심분야 : 영상처리, 멀티미디어



이 재 연

e-mail : leeyj@etri.re.kr
 1984년 서울대학교 이학사(제어계측)
 1986년 한국과학기술원 공학석사(전기및전자)
 1992년 일본東海대학 공학박사(광공학)

1986년 3월~현재 한국전자통신연구원 컴퓨터·소프트웨어 기술 연구소 책임연구원 재직중
 관심분야 : 영상처리, 멀티미디어



배 영 래

e-mail : yljb@etri.re.kr
 1976년 서울대학교 이학사(해양물리), 1986년 한양대학교 공학석사(전자계산학)
 1992년 Univ. of Kent 공학석사(전자공학)

1995년 Univ. of Kent 공학박사(전자공학)
 1976년~1979년 해운장교
 1979년~1980년 삼성
 1980년 KIST 입소
 1999년 현재 한국전자통신연구원 컴퓨터·소프트웨어
 기술 연구소 영상정보처리연구팀장
 관심분야 : 영상처리, 컴퓨터 비전, 병렬처리



장 종 환

e-mail : jangjw@mail.paichai.ac.kr
 1979년 한양대학교 전자통신공학과 졸업(학사)
 1986년 North Carolina 주립대학, 전기 및 컴퓨터공학과(공학석사)

1990년 North Carolina 주립대학, 전기 및 컴퓨터공학과(공학박사)
 1990년~현재 배재대학교 컴퓨터전자정보공학부 부교수 겸 정보통신대학원 교학부장
 1992년~1996년 배재대학교 전자계산소 소장
 1994년~현재 대전광역시 지역정보화 추진협의회 전문위원
 1998년~현재 정보통신부 지정 정보통신창업지원센터 소장
 1995년~현재 (사) 한국정보처리학회 멀티미디어시스템 연구회 운영위원
 1995년~현재 (사) 한국통신학회 대전충남지부 이사
 1999년~현재 대전광역시 최고정보화 책임관(CIO)
 1999년~현재 대전광역시 과학기술위원회 전문위원
 1999년~현재 한국소프트웨어진흥원 대전S/W지원센터 운영위원
 관심분야 : 멀티미디어, 영상처리, 비디오검색