

인터넷에 기반한 내용기반 검색 시스템 설계

Content-Based Retrieval System Design over the Internet

김 영 호, 강 대 성*

(Young Ho Kim and Dae-Seong Kang)

Abstract : Recently, development of digital technology is occupying a large part of multimedia information like character, voice, image, video, etc. Research about video indexing and retrieval progresses especially in research relative to video. This paper proposes the novel notation in order to retrieve MPEG video in the international standards of moving picture encoding. For realizing the retrieval-system, we detect DCT DC coefficient, and then we obtain shot to apply MVC(Mean Value Comparative) notation to image constructed DC coefficient. We choose the key frame for start-frame of a shot, and we have the codebook index generating it using feature of DC image and applying PCA(Principal Component Analysis) to the key frame. Also, we realize the retrieval-system through similarity after indexing. We could reduce error detection due to distinguish shot from conventional shot detection algorithm. In the mean time, speed of indexing is faster by PCA due to perform it in the compressed domain, and it has an advantage which is to generate codebook due to use statistical features. Finally, we could realize efficient retrieval-system using MVC and PCA to shot detection and indexing which is important step of retrieval-system, and we using retrieval-system over the internet.

Keywords : MVC(Mean Value Comparative), PCA(Principal Component Analysis), retrieval-system

I. 서론

최근 인터넷과 디지털 기술의 발달로 인터넷상에서 많은 정보를 습득하고 있다. 디지털 기술의 발달은 영상, 비디오, 음성 등과 같은 멀티미디어 정보가 매우 큰 비중을 차지하고 있다. 이러한 이유로 인터넷 사용자들은 인터넷상에서 원하는 데이터를 쉽고 정확하게 검색하고자 하는 색인 및 검색에 관한 연구들이 진행되고 있다[1,2]. 본 논문에서는 MPEG 비디오의 색인 및 검색을 위한 비디오의 최소 단위인 프레임들을 Shot 단위로 분류하기 위해 MVC(Mean Value Comparative)기법을 제안하였다. 기존의 압축 영역에서 이용한 DCT의 DC를 이용하여 DC영상을 구한 후, pixel간의 차분 값, 히스토그램에 대한 chi-square 값, 히스토그램의 분산의 차분 값, 양자화한 영상의 히스토그램의 각 bin 값들의 열과 행의 위치에 대한 분산의 chi-square 값을 이용하여 구해진 Shot 검출을 MVC 기법을 통하여 한번더 판별하였다. Shot의 시작 프레임을 대표 프레임으로 선정하고 색인을 위해 PCA(Principal Component Analysis)기법을 적용하였다. 멀티미디어검색 방법은 크게 텍스트 기반 검색(text-based retrieval)과 내용 기반 검색 방법(content-based retrieval)으로 나눌 수 있다. 초기의 영상 데이터 베이스 시스템에서는 텍스트 기반 검색으로, 이 방법은 각 영상에 대해 그 영상을 나타내는 주석 정보를 기술한 뒤, 이 정보를 이용하여 영상을 검색하게 된다. 하지만 멀티미디어 정보는 그 특성상 텍스트 주석 정보로는 표현하기에는 주관적이고, 질의 영상과 유사한 영상을 찾을 수 없다는 단점이 있다. 이에 반해 내용 기반 검색 방법은 물체의 형상, 색, 질감 등을 이용하여

검색할 수 있어, 자동적으로 영상의 특징을 추출하고 정보를 저장할 수 있는 장점이 있다. 내용 기반 검색 방법에는 여러 가지 방법들이 제시되었으며, 그 중 컬러 히스토그램을 사용하는 방법과 투영 기법을 이용하는 방법 등이 있다. 컬러 히스토그램을 사용한 방법은 영상의 전체 특성을 잘 나타내는 반면 영상의 위치정보의 손실로 인해 다른 영상을 같은 영상으로 오인식할 수 있다. 투영 기법을 사용한 방법은 영상의 크기에 따라 생성되는 데이터가 달라지기 때문에 다양한 형태의 데이터를 검색하는 데 문제가 발생하게 된다. 따라서 본 논문에서는 PCA기법을 멀티미디어 데이터에 적용한 방법을 제안한다. DC영상을 이차원 벡터화 시킨 후 초기 설정한 코드북 크기에 따라 특징추출을 수행한다. 멀티미디어 정보에 대한 색인 시 속도가 빠르다는 장점과 데이터의 통계적인 특성을 반영하기 때문에 최적의 코드북을 생성 가능한 장점이 있다. 색인 후 유사도를 통한 검색 시스템을 구현하였다.

2장에서는 전체알고리즘과 Shot 경계 프레임 검출, MVC 기법, 대표 프레임 색인 및 검색에 관해 서술하였고, 3장에서는 Shot 경계 프레임 검출 결과와 색인 및 검색에 대한 결과를 서술하였고, 4장에서 결론을 맺었다.

II. 멀티미디어 검색 시스템 설계

1. 전체 알고리즘

본 논문에서 사용된 전체 알고리즘은 다음과 같다.

- STEP 1. 비디오 시퀀스에서 I picture의 DC영상을 검출한다.
- STEP 2. 검출되어진 DC영상으로부터 4개의 feature를 구한다.
- STEP 3. MVC 기법을 적용 Shot 경계 프레임을 구한다.
- STEP 4. PCA 기법을 적용 대표 프레임 색인을 한다.
- STEP 5. 사용자의 질의영상에 대한 유사도를 검사하여

* 책임저자(Corresponding Author)

논문접수 : 2004. 10. 7., 채택확정 : 2005. 4. 1.

김영호 : 동아대학교 전자공학과(newkyh@hotmail.com)

강대성 : 동아대학교 전자컴퓨터공학부(dskang@daunet.donga.ac.kr)

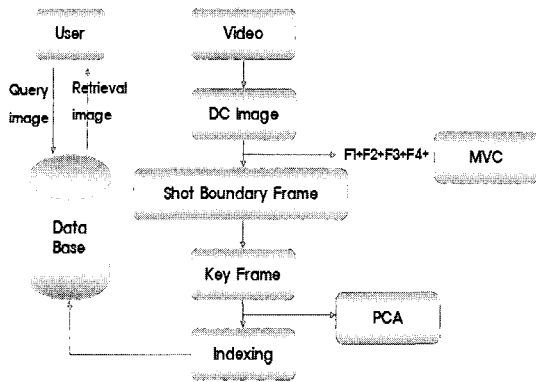


그림 1. 전체 알고리즘.

Fig. 1. A general algorithm.

응답 영상을 출력한다.

위의 그림 1은 전체 알고리즘의 구성도이다.

2. Shot 경계 프레임 검출

본 논문에서는 Shot을 검출하기 위하여 MPEG에서 I picture의 DCT변환의 DC계수만으로 영상을 구성한 DC영상만을 사용한다. Video 시퀀스의 decoding 과정을 없애므로써 처리시간을 대폭 감소시켜 효율적인 Shot 경계 프레임 검출이 가능하게 하였다. Shot 경계를 검출하기 위한 feature로서 4개의 feature를 조합하여 사용한다. feature를 구하는 수식은 다음과 같다[3].

$$Diffimg_i = \frac{\sum_{x=0, y=0}^{m-1, n-1} |img_{i-1}(x, y) - img_i(x, y)|}{NM} \quad (1)$$

$$X_i^2 = \sum_{k=0}^{n-1} \frac{(H_{i-1}(k) - H_i(k))^2}{H_i(k)}, \quad DiffD_i^2 = \frac{(D_{i-1} - D_i)^2}{D_i} \quad (2)$$

$$PX_i^2 = \sum_{k=0}^{n-1} \frac{(\rho X_{i-1}(k) - \rho X_i(k))^2}{\rho X_i(k)} \quad (3)$$

$$PY_i^2 = \sum_{k=0}^{n-1} \frac{(\rho Y_{i-1}(k) - \rho Y_i(k))^2}{\rho Y_i(k)} \quad (4)$$

$$\rho X(k)^2 = E(|BX_i(n) - \overline{BX_i(n)}|^2) \quad (5)$$

$$\rho Y(k)^2 = E(|BY_i(n) - \overline{BY_i(n)}|^2) \quad (6)$$

이상과 같이 구해진 DC영상의 feature들로부터 아래와 같은 단계로 Shot 경계 프레임을 검출한다.

1) 전체 프레임에 대한 각 파라미터의 평균을 구한다.

$$\overline{Diffimg}, \overline{X}, \overline{DiffD}, \overline{PX}, \overline{PY}$$

2) 각 feature 값이 아래의 조건식에서 condition 1과 2를 동시에 만족하면서 condition 3 혹은 4가 만족할 경우 Shot 경계 프레임으로 검출한다.

condition 1. $Diffimg_i > \alpha \overline{Diffimg}$

condition 2. $X_i > \beta \overline{X}$

condition 3. $DiffD_i > \gamma \overline{DiffD}$

condition 4. $PX_i > \delta \overline{PX}$ and $PY_i > \delta \overline{PY}$

3. MVC(Mean Value Comparative) 기법

I picture로부터 DC영상을 각각 추출한다. 기존의 4개의 feature를 순차적으로 비교하여 검출한 두 DC영상이 Shot 경계인지 판별해야 한다. 이는 Shot 경계 지점의 이웃 프레임들의 평균값들을 비교함으로써 이루어진다. 기존에 Shot 경계 지점으로 판별된 지점이 실제의 Shot 경계 지점인지 아닌지 이웃 프레임의 DC영상과 비교하여 재 판별한다. 실제 Shot 경계가 일어나는 지점을 기준으로 4개의 프레임의 평균값을 비교한다[4].

I-1, I, I+1, I+2 개의 프레임을 I, I+1 부분에서 Shot 경계 지점일 경우 I, I+1의 평균값과 I-1, I+1 과 I, I+2의 평균값을 비교하여 I, I+1 평균값이 큰 경우는 Shot 경계 지점이고 작은 경우 실제 Shot 경계가 일어나지 않고 프레임 내 최소값의 변화에 따른 오검출로 판별하였다.

4. 대표 프레임 색인 및 검색

대표 프레임 검출은 구해진 Shot 프레임들 중 그 비디오 시퀀스를 가장 잘 표현할 수 있는 대표 프레임을 찾는 과정이다. 본 논문에서는 분류 되어진 Shot의 시작 프레임을 대표 프레임으로 선정한다. 대표 프레임의 색인을 위해서 PCA 기법을 적용한다. 검출된 대표 프레임들을 2차원 입력 벡터로 나타내어 변동량이 가장 많은 주축을 구한다. 주축은 입력 벡터의 최대 고유치의 고유 벡터의 방향과 일치함으로써 데이터의 중복성을 제거할 수 있다. 다음으로 영역 분할 알고리즘을 이용해서 주축과 수직을 이루면서 노드의 평균을 지나는 지점을 기준으로 영역을 2개로 분할한다. 각 영역의 평균 오차를 계산하여 평균 오차가 가장 큰 영역을 선별한 후 선택된 영역에 영역 분할 알고리즘을 재 적용한다[5].

1) 주축을 구하는 알고리즘

① 입력벡터에 대한 기하학적 평균 M을 구한다.

$$m = \begin{bmatrix} M_x \\ M_y \end{bmatrix}, \quad M_x = \frac{1}{q} \sum_{i=0}^{q-1} I_x, \quad M_y = \frac{1}{q} \sum_{i=0}^{q-1} I_y, \quad (7)$$

여기서, q는 총 입력벡터의 개수이고, I_x, I_y 는 입력벡터의 x축, y축 I_x, I_y 의 i번째 요소를 나타낸다.

② 입력 벡터에 대한 공분산 행렬 C를 구한다.

$$C = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix}, \quad \text{where } C_1 = \frac{1}{q} \sum_{i=0}^{q-1} (x_i^2 - M_x^2),$$

$$C_2 = C_3 = \frac{1}{q} \sum_{i=0}^{q-1} (x_i y_i - M_x M_y), \quad C_4 = \frac{1}{q} \sum_{i=0}^{q-1} (y_i^2 - M_y^2) \quad (8)$$

③ 가장 큰 고유치(the largest eigenvalue)를 구한다.

$$A = C_1 + C_4$$

$$B = A^2 - 4(C_1 C_4 - C_2 C_3), \quad \lambda_{\max} = \frac{A + \sqrt{B}}{2} \quad (9)$$

④ 주축의 각도(θ)를 계산한다.

$$\theta = a \tan \left[\frac{\lambda_{\max} - C_1}{2} \right] \quad (10)$$

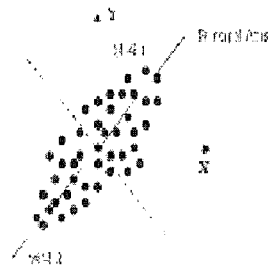


그림 2. 주축에 의한 영역 분할.
Fig. 2. Domain division for the principal axis.

2) 영역 분할 알고리즘

- ① 같은 영역에 할당된 입력 벡터의 평균 좌표를 계산한다.
- ② 주축을 구하는 알고리즘을 이용하여 주축과 수직이면서 노드의 평균 좌표를 지나는 직선과의 각도(ϕ)를 계산한다.
- ③ 다음의 식에 의해서 해당 영역의 각 데이터는 두 영역으로 분할된다.

If ($\cos(\phi) \leq 0$) then 영역 1
else 영역 2

그림 2는 주축을 구하는 알고리즘과 영역 분할 알고리즘을 적용하여 전체 입력 벡터를 두 부분으로 나누는 과정을 나타낸 것이다.

이상의 알고리즘에 의해 전체 입력 벡터는 여러 개의 영역으로 분할되고 각 영역의 평균값을 코드워드로 사용한다. 그리고 추출된 대표 프레임은 각 코드워드에 할당된 값을 카운트하여 코드북을 만든다. 이 카운트 값은 MPEG 비디오 시퀀스의 색인을 위한 파라미터로 사용된다. 색인이 끝난 후 검색을 위해 유사도를 평가한다. 검색을 위한 순서도는 다음과 같다.

1) 질의 영상과 응답영상의 값의 차이가 적은 순으로 유사도 평가를 한다.

$$Sim_i = \frac{\sum_{x=0, y=0}^{M-1, N-1} |img_{query}(x-y) - img_i(x-y)|}{NM} \quad (11)$$

- 2) Sim_i 의 값이 일정 임계치(S) 이하가 될 때까지 Step 1을 반복한다.
- 3) Sim_i 값이 작은 순으로 10개의 응답영상을 출력한다.

III. 실험 및 고찰

1. Shot 경계 프레임 검출

실험에서 사용된 대상 영상은 카메라의 후레쉬의 섬광 효과가 많은 영상(“Respect the Power of Love”)을 사용하였다. 프레임들은 320 × 240 사이즈에 8비트로 구성되어 있었다. 본 논문에서의 실험은 MPEG 비디오의 시퀀스로부터 DC 영상을 검출하여, 일반적인 Shot 검출에 사용되는 feature로 얻은 결과를 MVC 기법을 통해 한번 더 판별함으로써 개선된 결과를 보였다. 4분15초짜리 실험 MPEG 비디오의 DC영상은 511로 검출할 수 있었다. 그림 3은 실제로 Shot 경계프레임 영상들이다. 그림 4는 기존 4개의 feature를 이용해서 검출된 Shot 경계 프레임들이다. 0, 21, 52, 58, 78,

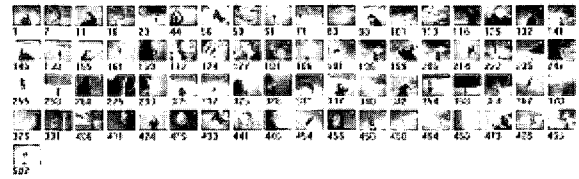


그림 3. 실제 Shot 경계 프레임들.
Fig. 3. The real shot boundary frames.

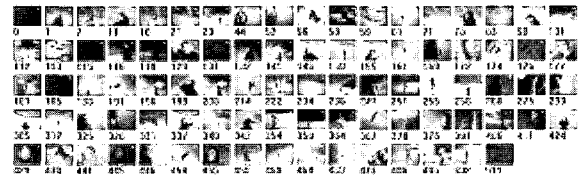


그림 4. 4개의 feature를 이용한 Shot 경계 프레임들.
Fig. 4. Shot boundary frames using 4 features.

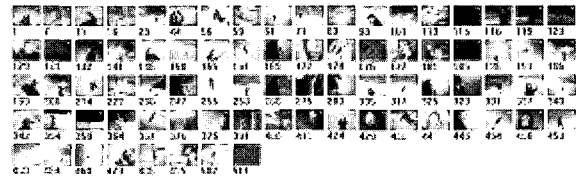


그림 5. MVC 기법을 이용한 Shot 경계 프레임.
Fig. 5. Shot boundary frames using MVC.

표 1. Shot 경계 프레임 검출 결과.

Table 1. Experimental results of shot boundary frames.

실험영상	Shot 경계 프레임 총 개수	검출개수	오검출	오검출율 (%)
f1, f2, f3, f4	73	92	19	26%
MVC	73	80	7	9.6%

83, 112, 115, 119, 131, 168, 176, 181, 185, 234, 251, 370, 445, 511 프레임이 오검출되었다. 그림 5는 기존 4개의 feature에 MVC 기법을 적용하여 검출한 Shot 경계 프레임들로서 115, 119, 120, 131, 176, 185, 511 프레임이 오검출되었다. 기존의 feature로 검출된 Shot 경계 프레임들 보다 오검출이 줄어들었음을 알 수 있다.

표 1은 Shot 경계 프레임 검출 결과를 나타내었다. 기존 4개의 feature로 검출했을 경우 오검출율이 26%, MVC 기법은 9.6%로 줄일 수 있었다.

2. 대표 프레임 검출 및 색인 검색

Shot 경계 프레임을 검출하고 나서 그 비디오의 시퀀스의 대표 프레임을 찾는 과정을 수행한다. 본 논문에서 MVC 기법으로 구해진 Shot 경계 프레임을 바탕으로 Shot을 분류하고 그 시작 프레임을 대표 프레임으로 선정한다. 선정되어진 대표 프레임을 색인하기 위해 PCA 기법을 적용하였다.

최종 영역의 평균값을 코드워드로 저장하고, 각 대표 프레임에 대해 코드워드에 할당된 수를 카운트한 값을 검색을 위한 색인 파라미터로 사용한다.

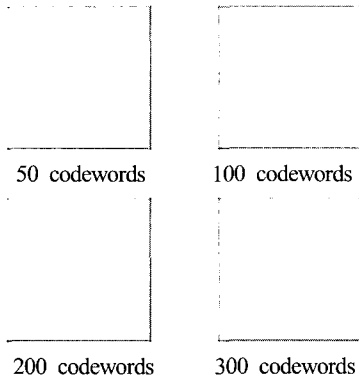
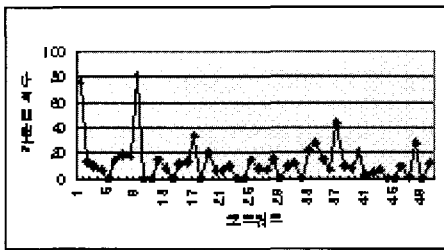


그림 6. 영역 분할에 의해 생성된 코드워드.
Fig. 6. Codeword for domain division.



(a) Key frame



(b) Codebook for figure (a)

그림 7. 코드북 생성 결과.
Fig. 7. Results of codebook.

표 2. 영상의 크기와 군집화에 소요되는 시간.
Table 2. Execution time for clustering and size of Image.

Size	# of Vectors	Execution time for clustering
320×240	300	4.878065 sec
	200	3.527954 sec
	100	2.278065 sec
	50	1.85175 sec

그림 6은 최종 영역의 분할 개수를 50, 100, 200, 300개로 주었을 때 생성된 코드워드를 보여준다. 그림 7의 (b)는 생성된 50개의 코드워드에서 대표 프레임(a)에 할당된 카운트 값을 나타낸다. x축은 코드워드를 나타내고 y축은 각 코드 워드에 할당되는 카운트 회수를 나타낸다.

표 2는 대표 벡터로 군집화 하는데 필요한 시간을 나타 내었다. 본 실험에서 군집화는 최종 대표 벡터의 수에 도달 할 때까지 반복한다. 같은 크기의 영상이라도 하나의 영 상을 군집화 하는데 걸리는 시간이 시스템의 상태에 따라서 조금씩 변하므로 같은 크기의 영상을 군집화 하는데 5회에 소요되는 시간의 평균값을 기록하였다.

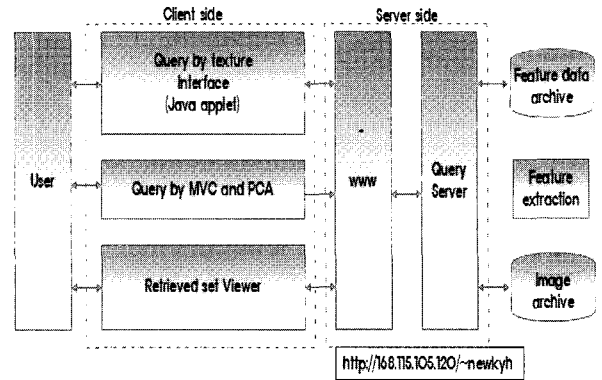


그림 8. 멀티미디어 검색 시스템.
Fig. 8. Multimedia retrieval system.

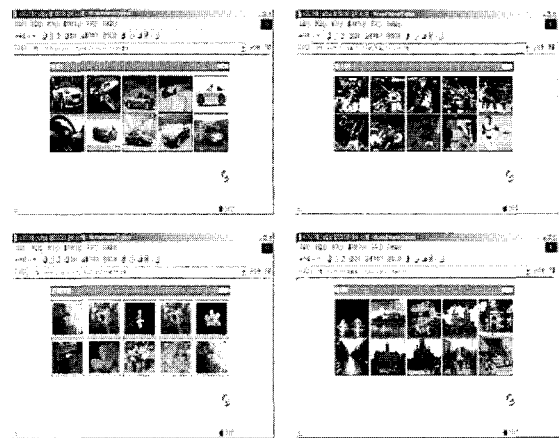
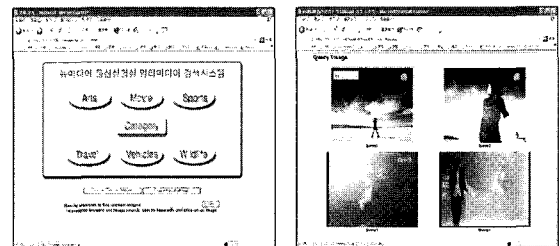
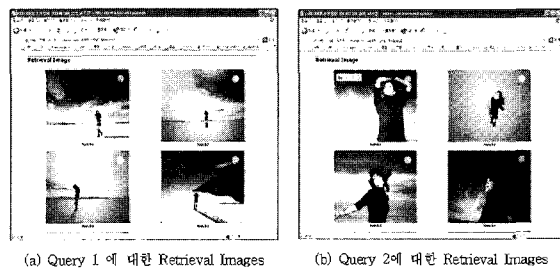


그림 9. 카테고리에 따른 출력 검색 영상.
Fig. 9. The result retrieval images for category.



(a) Beginning page (b) Query page

그림 10. 검색 시스템의 시작 화면과 질의 화면.
Fig. 10. The beginning page and query page of retrieval system.



(a) Query 1에 대한 Retrieval Images (b) Query 2에 대한 Retrieval Images

그림 11. 질의에 따른 결과 화면.
Fig. 11. The result images about the query images.

3. 멀티미디어 검색 시스템 구현

본 논문에서 제안한 MVC 기법과 PCA 기법 적용하여 실시간 데이터베이스 검색 시스템을 구축하였다. (<http://168.115.105.120/~newkyh>) 데이터베이스 상에는 대표 프레임과 특징 값들이 저장되는 공간과 질의 영상들과 특징 값들이 저장되는 공간, 두 가지 저장 공간으로 분류된다. 이 시스템은 사용자가 선택한 질의 영상에 대하여 matching 과정을 통해 그에 따른 결과 영상들을 matching 순서대로 화면에 출력한다. 본 검색 시스템은 카테고리 선택 부분과 내용 기반 검색 부분으로 나뉘어져 있다. 그림 8은 멀티미디어 검색 시스템의 전체 구성도이며, 그림 9는 카테고리에 따른 출력 검색 영상을 나타내었고, 그림 10은 본 멀티미디어 검색 시스템의 시작 화면과 질의 영상 화면을 보여주고 있고, 그림 11은 질의 영상 선택 후 화면에 출력된 matching 결과 영상들이다.

IV. 결론

본 논문에서는 MPEG 비디오의 색인 및 검색을 위한 비디오의 최소 단위인 프레임들을 Shot 단위로 분류하기 위해 MVC 기법을 제안하였다. 기존의 압축 영역에서 이용한 DCT의 DC를 이용하여 DC영상을 구한 후, pixel간의 차분 값, 히스토그램에 대한 chi-square 값, 히스토그램의 분산의 차분값, 양자화한 영상의 히스토그램의 각 bin 값들의 열과 행의 위치에 대한 분산의 chi-square 값을 이용하여 구해진 Shot 검출을 MVC 기법을 통하여 한번 더 판별하였다. Shot의 시작 프레임을 대표 프레임으로 선정하고 색인을 위해 PCA 기법을 적용하였다. DC영상을 이차원 벡터화 시킨 후 초기 설정한 코드북 크기에 따라 특징추출을 수행한다. 멀티미디어 정보에 대한 색인 시 속도가 빠르다는 장점과 데

이터의 통계적인 특성을 반영하기 때문에 최적의 코드북을 생성 가능한 장점이 있다. 제안한 MVC 기법과 PCA 기법 적용하여 실시간 데이터베이스 검색 시스템을 구축하였다. 데이터베이스 상에는 대표 프레임과 특징 값들이 저장되는 공간과 질의 영상들과 특징 값들이 저장되는 공간, 두 가지 저장 공간으로 분류된다. 이 시스템은 사용자가 선택한 질의 영상에 대하여 matching 과정을 통해 그에 따른 결과 영상들을 matching 순서대로 화면에 출력 한다. 앞으로는 여러 실험 영상을 통해 좀 더 많은 결과를 바탕으로, 보다 정확한 인터넷을 통한 내용기반 검색을 구현해야 할 것 이다.

참고문헌

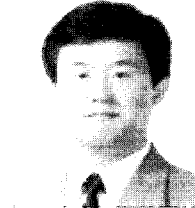
- [1] J. Fan, H. Luo, "Concept-oriented indexing of video databases: toward semantic sensitive retrieval and browsing", *IEEE TRANSACTION ON IMAGE PROCESSING*, OL. 13, NO. 7, JULY 2004.
- [2] K. R. Rao, J. J. Hwang. *Techniques and standards for image · video and audio coding*, Prentice Hall. 1999.
- [3] Jerry D. Gibson, *Image and video processing*, ACADEMIC PRESS pp. 687-704, 2000.
- [4] 김영호, 강대성, "프레임 특징비교를 이용한 압축비디오 분할," 한국 신호처리 시스템학회 논문지, pp. 25-29 2003.
- [5] Y. S. Avrithis, A. D. Doulamis, N. D. Doulamis and S. D. Kollias, "A stochastic framework for optimal key frame extraction from MPEG video databases," *Computer Vision and Image Understanding*, vol. 75, Nos. 1/2, July/August, pp. 3-24, 1999.

김영호



1999년 동아대학교 전자공학과 졸업.
2002년 동아대학교 전자공학과 석사.
현재 동아대학교 전자공학과 박사수료.
관심분야는 영상처리, 영상분할, 패턴인식, 멀티미디어 색인 및 검색.

강대성



1984년 경북대학교 전자공학과 졸업.
1991년 Texas A&M 대학교 전자공학과 석사.
1994년 Texas A&M 대학교 전자공학과 박사.
1983년 3월~1989년 6월 국방과학연구소. 1995년 9월~현재 동아대학교 전자컴퓨터 공학부 부교수. 관심분야는 영상처리, 패턴인식, 영상코딩, 멀티미디어 색인 및 검색.