論文2001-38CI-5-6

MPEG 비디오의 통계적 특성을 이용한 검색 시스템

(Retrieval System Adopting Statistical Feature of MPEG Video)

劉 玲 達 *, 姜 大 星 *, 金 大 鎭 **

(Young-Dal Yu, Dae-Seong Kang, and DaiJin Kim)

요 약

현재 많은 정보들이 비디오 데이터로 전송 또는 저장되고 있으며 고성능 PC의 보급과 internet과 같은 통신망의 대중화로 이런 비디오 데이터는 급속도로 증가하고 있다. 본 논문에서는 이런 비디오 데이터의 검색을 위하여 비디오 스트림을 분석하여 shot을 찾아내고 이들 중 key frame을 찾는 방법에 대하여 연구하고 이로서 사용자의 질의에 부합하는 비디오를 검색한다. 본 논문에서는 shot 경계 검출을 위해 객체의 움직임에 강인하면서 shot 내에서의 칼라의 변화에 둔감한 새로운 feature를 제안하고, shot frame에서 구한 각 feature들의 통계적 특성을 이용하여 스트림의 특징에 따라 weight를 부가하여 구해진 characterizing value의 시간 변화량을 구한다. 구해진 변화량의 local maxima와 local minima는 비디오 스트림에서 각각 가장 특징적인 frame과 평균적인 frame을 나타낸다. 이 순간의 shot frame을 구함으로서 효과적이고 빠른 시간 내에 key frame을 추출한다. 추출되어진 key frame에 대하여 원 영상을 복원한 후, 색인을 위하여 다수의 parameter를 구하고, 사용자가 질의한 영상에 대해서 이들 parameter를 구하여 key frame들과 가장 유사한 대표영상들을 검색한다. 실험결과 일반적인 방법보다 더 나은 결과를 보였고, 높은 검색율을 보였다.

Abstract

Recently many informations are transmitted and stored as video data, and they are on the rapid increase because of popularization of high performance computer and internet. In this paper, to retrieve video data, shots are found through analysis of video stream and the method of detection of key frame is studied. Finally users can retrieve the video efficiently. This Paper suggests a new feature that is robust to object movement in a shot and is not sensitive to change of color in boundary detection of shots, and proposes the characterizing value that reflects the characteristic of kind of video (movie, drama, news, music video etc.). The key frames are pulled out from many frames by using the local minima and maxima of differential of the value. After original frame(not dc image) are reconstructed for key frame, indexing process is performed through computing parameters. Key frames that are similar to user's query image are retrieved through computing parameters. It is proved that the proposed methods are better than conventional method from experiments. The retrieval accuracy rate is so high in experiments.

* 正會員, 東亞大學校 電氣電子컴퓨터工學部
(School of Electrical, Electronic, and Computer Eng., Dong-A University)
** 正會員, 浦港工科大學校 컴퓨터工學科
(Dept. of Computer Eng., POSTECH)
※ 본 연구는 한국과학재단(KOSEF: 1999-2-302-011-2)의 지원으로 수행되었음.
接受日字:2000年7月21日, 수정완료일:2001年7月11日

Ⅰ. 서 론

현재 많은 정보들이 비디오 데이터로 전송 또는 저장되고 있으며 고성능 PC의 보급과 internet과 같은 통신망의 대중화로 이런 비디오 데이터는 급속도로 증가하고 있다. 이에 많은 비디오 스트림들의 database화를 위해 고용량의 비디오 데이터를 효과적으로 색인하고 검색할 수 있는 기술이 연구되어 지고 있다[1-6]. 검색은 구문기반검색과 내용기반검색으로 나눌 수 있으며, 모든 과정을 인간의 주관적인 관점이 배제되고 객관적으로 처리되는 자동화된 내용기반검색[2-6]이 주로 연구되어 지고 있다. 본 논문에서는 이런 내용기반검색을 위하여 비디오 스트림을 분석하여 shot을 찾아내고 이들 중 key frame을 찾아 색인 하여 사용자의 질의에 부합하는 비디오를 검색한다. 본 논문에서는 shot 경계 검출을 위해 객체의 움직임에 강인하면서 shot 내에서의 칼라의 변화에 둔감한 새로운 feature를 제안하고, shot frame에서 구한 각 feature들의 통계적 특성을 이용하여 스트림의 특징에 따라 weight를 부가하여 구해진 characterizing value의 시간 변화량을 구한다. 구해진 변화량의 local maxima와 local minima는 비디오 스트림에서 각각 가장 특징적인 frame과 평균적인 frame을 나타낸다. 이 순간의 shot frame을 구함으로서 효과적이고 빠른 시간 내에 key frame을 추출한다. 추출되어진 key frame에 대하여 원 영상을 복원한 후, 색인을 위하여 다수의 parameter를 구하고, 사용자가 질의한 영상에 대해서 이들 parameter를 구하여 key frame들과 가장 유사한 대표영상들을 검색한다.

key frame을 추출하기 위하여 비디오 스트림은 여러 개의 shot들로 나뉘어져야 한다. shot의 boundary를 찾고 frame의 색인을 위하여 많은 feature들이 제안되어지고 분석되어 졌다. 일반적으로 frame의 feature는 두 가지 타입으로 분류되어진다. 첫 번째는 motion vector를 이용함으로서 칼라의 변화에 강인한 타입이다. 그러나 이런 타입의 feature는 만약 객체의 움직임이 하나의 shot 내에서 급격히 발생할 경우 그 shot을 별개의 두 개의 shot으로 간주하는 단점이 있다. 두 번째 타입은 히스토그램이나 chi-square를 사용함으로써 객체의 움직임에 강인한 feature이다. 그러나 이런 타입의 feature가 사용될 때, 하나의 shot에서 칼라의 큰 변화는 feature의 차분을 매우 크게 하는 단점이 있다. 이런 상반된 특징으로 인하여 두 가지 타입의 feature를 조합하는데 많은 어려움이 있다. 그래서 key frame의 최적화를 위하여 frame들 간의 관계를 정의하기 위한 새로운 feature의 연구가 요구되어 진다.

본 논문에서는 비디오 데이터로서 현재 가장 많이 이용되어지고 있고, 또한 국제 표준안인 MPEG 비디오를 이용하여 실험하였다. 그리고 비디오 스트림의 I-picture의 dc image만을 shot 검출에 이용함으로서 비디오 데이터의 복원과정을 거치지 않음으로서 처리시간을 단축시킨다. key frame 추출 후 색인 과정을 위해 key frame으로 추출된 frame에 대해서는 보다 정확한 색인을 위하여 원 영상을 복원하여 처리함으로서 보다 효율적인 색인과 검색을 수행한다.

Ⅱ. MPEG 비디오검색을 위한 key frame검출

MPEG video의 검색 시스템을 구현하기 위하여 검색에 적합한 database를 구현하여야 한다. database를 구현하기 위해서는 video 스트림의 분석과 각 frame들의 분류가 선행되어야 한다. 이런 database를 구현하는 시스템은 다음과 같은 단계로 나뉘어진다.

① frame feature 검출
② shot boundary frame 추출
③ key frame 추출
④ key frame 색인

위와 같은 단계로서 database를 구축한 후, video의 검색은 사용자의 질의에 따라 유사도가 가장 높은 key frame을 순서대로 출력한다. frame의 feature를 검출하는 단계는 그 후의 단계들에 가장 큰 영향을 주며, feature가 적절하게 검출되지 않았을 경우에는 그 후의 모든 과정에서 잘못된 결과를 출력하게 된다.

일반적으로 하나의 video 스트림은 많은 수의 frame을 갖고 있다. 그러나 근접한 frame들은 유사한 내용을 갖고 있는 영상들로 구성되고, 이는 video 스트림의 모든 frame들에 대해 shot을 검출하고 key frame을 찾는 것은 비효율적인 과정임을 알 수 있다. 그러므로 본 논문에서는 shot을 검출하기 위하여 MPEG에서 I picture의 DCT변환의 dc계수만으로 영상을 구성한 dc image만을 사용함으로서 처리시간을 단축시킬 수 있다. 또한 video stream의 decoding과정이 없으므로 계산량을 대폭 감소시켰다. 검색 시 보다 정확한 질의영상과의 유

사도 측정을 위하여 key frame으로 추출된 영상에 대해서는 원 영상을 복원하여 효율적인 검색이 가능하도록 하였다.

1. 전체 알고리즘

본 논문의 전체 알고리즘은 다음과 같다.

Step 1. video 스트림에서 I-Picture의 dc image를 추출한다.

Step 2. 추출되어진 dc image로부터 4개의 feature를 구한다.

Step 3. 주어진 조건식에 따라 shot boundary frame을 구한다.

Step 4. 구해진 shot frame에서 5개의 parameter를 구한다.

Step 5. shot frame 전체의 각 parameter에 대한 통계적 분포에 따라 parameter에 대한 weight를 부가하여 characterizing value를 구한다.

Step 6. characterizing value의 시간에 대한 변화량을 구한다.

Step 7. key frame을 검출한다.

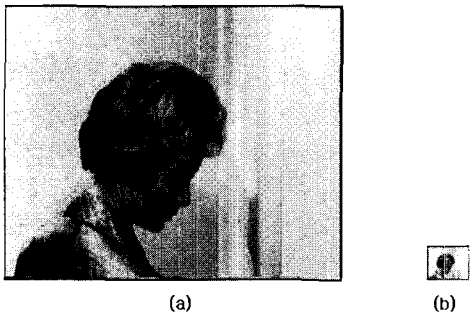
Step 8. 검출되어진 key frame의 원 영상을 복원한다.

Step 9. key frame을 색인 한다.

Step 10. 사용자의 질의영상에 대한 유사도를 검사하여 응답 영상을 출력한다.

2. Shot boundary frame 추출

본 연구에서는 shot을 검출하기 위하여 MPEG에서 I picture의 DCT변환의 dc계수만으로 영상을 구성한 dc image만을 사용한다. video 스트림의 decoding 과정을 없앰으로서 처리시간을 대폭 감소시켜 효율적인 shot boundary frame 추출이 가능하게 하였다. 그림 1는 원 영상과 dc image를 나타낸 것이다.



(a) (b)

그림 1. 실험영상(a)원 영상 b)dc image.
Fig. 1. Experiment image (a)original image (b)dc image.

본 연구에서는 shot boundary frame를 추출하기 위한 feature로서 4개의 feature를 조합하여 사용한다. 이 feature를 구하는 수식은 다음과 같다.

$$DiffImg_i = \frac{\sum_{x=0,y=0}^{M-1,N-1} |Img_{i-1}(x,y) - Img_i(x,y)|}{NM} \quad \text{(feature 1)}$$

$$X_i^2 = \sum_{k=0}^{n-1} \frac{(H_{i-1}(k) - H_i(k))^2}{H_i(k)} \quad \text{(feature 2)}$$

$$DiffD_i^2 = \frac{(D_{i-1} - D_i)^2}{D_i} \quad \text{(feature 3)}$$

$$PX_i^2 = \sum_{k=0}^{n-1} \frac{(\rho X_{i-1}(k) - \rho X_i(k))^2}{\rho X_i(k)} \quad \text{(feature 4)}$$

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

$$\rho X_i(n)^2 = E(|BX_i(n) - \overline{BX_i(n)}|^2)$$

$$\rho Y_i(n)^2 = E(|BY_i(n) - \overline{BY_i(n)}|^2)$$

이상과 같이 구해진 dc image의 feature들로부터 아래와 같은 단계로 shot boundary frame을 검출한다.

Step 1. 각 parameter들의 전체 프레임에 대한 평균을 구한다.

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

Step 2. dc image의 각 feature 값이 아래의 조건식에서 condition 1과 2를 동시에 만족하면서 condition 3 혹은 4가 만족할 경우 shot boundary frame으로 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 $PY_i > \delta\,\overline{PY}$

(※ 각 상수 $\alpha, \beta, \gamma, \delta$ 는 실험을 통해서 2.0, 2.0, 2.0, 2.0으로 검출)

3. Key frame 추출 알고리즘

key frame 검출은 구해진 shot frame들 중 그 video 스트림을 가장 잘 표현할 수 있는 대표 frame을 찾는 과정이다. 본 논문에서는 각 video 스트림마다 종류에 따라 다른 특징을 갖고 있고 key frame이 그 video 스트림의 특징에 따라 parameter에 대해 민감도가 다른 것을 반영하기 위하여 전체 shot frame의 각

parameter들의 통계적 특성에 따라 weight를 부가한 characterizing value를 구함으로서 보다 적합한 key frame 추출을 수행한다.

다음은 key frame 추출을 위한 shot frame의 parameter들이다.

① 휘도의 평균 : shot frame 전체 pixel에 대한 휘도의 평균이다.

$$f_1 = AveShot_i = \frac{\sum_{x=0,y=0}^{M-1,N-1} Shot_i(x,y)}{MN}$$

② 히스토그램의 분산값 : shot frame의 히스토그램의 분포 특성을 나타낸다.

$$f_2 = DisShot_i^2 = E(|H_i(n) - \overline{H_i(n)}|^2)$$

③ 양자화한 shot frame 히스토그램의 각 bin값들의 열과 행의 위치에 대한 분산의 평균 : 히스 토그램과 위치정보의 조합을 나타낸다.

$$\rho X_i(n)^2 = E(|BX_i(n) - \overline{BX_i(n)}|^2)$$

$$\rho Y_i(n)^2 = E(|BY_i(n) - \overline{BY_i(n)}|^2)$$

$$f_{3x} = AveDisX_i = \frac{\sum_{k=0}^{n_1-1} \rho X_i(k)}{n_1}$$

$$f_{3y} = AveDisY_i = \frac{\sum_{k=0}^{n_2-1} \rho Y_i(k)}{n_2}$$

④ 이전 shot frame과의 휘도 차 : 이전 shot frame과 휘도의 변화량을 나타낸다.

$$f_4 = DiffShot_i = \frac{\sum_{x=0,y=0}^{m-1,n-1} |Shot_i(x,y) - Shot_{i-1}(x,y)|}{MN}$$

⑤ 휘도 차의 누적에 대한 f_4의 비율 : 전체적인 휘도 변화율에 대한 상대값을 나타낸다.

$$f_5 = AccDiffShot_i = \frac{\sum_{k=0}^{i-1} DiffShot_k}{DiffShot_i}$$

위의 parameter들을 각 검출된 shot frame에 대해서 구하고 난 후 key frame을 검출하기 위하여 다음과 같은 단계를 거쳐 characterizing value를 구한다.

Step 1. shot으로 검출된 모든 프레임에 대한 각 feature들의 분산을 구한다.

$$\rho_{f_n}{}^2 = E(|F_n - \overline{F_n}|^2)$$

Step 2. 각 feature들의 전체 평균에 대한 차를 구한다.

$$f_n{}' = |f_n - \overline{f_n}|$$

Step 3. 각 feature들에 대해 분산에 대한 비율만큼 weight를 부가하여 각 프레임의 특징 값 (C_m)을 구한다.

$$C_m = \omega_1 f_1{}' + \omega_2 f_2{}' + \cdots + \omega_n f_n{}'$$

여기서 m은 shot으로 검출된 프레임의 개수, 가중치 ω는 아래와 같다.

$$\omega_1 = \frac{\rho_1}{\sum_{i=1}^{n} \rho_i}, \quad \omega_2 = \frac{\rho_2}{\sum_{i=1}^{n} \rho_i}, \quad \cdots, \quad \omega_n = \frac{\rho_n}{\sum_{i=1}^{n} \rho_i}$$

위의 단계를 거쳐 계산되어진 characterizing value의 시간에 대한 변화량을 구하여 local maxima와 local minima를 구한다. 각 local maxima와 local minima는 그 비디오 스트림의 가장 특징적인 frame과 평균적인 frame을 나타내게 되며 이 순간의 frame을 key frame으로 추출할 수 있다.

4. Key frame의 색인과 검색

보다 정확한 색인을 위하여 key frame으로 추출된 dc image를 원 영상으로 복원한다. 그리고 key frame의 색인을 위하여 parameter들을 구한다. 원 영상에 대한 parameter를 구함으로 보다 정확한 색인이 가능하고, 검색 시 입력되는 질의영상과 좀 더 객관적인 유사도 측정이 가능하다. 색인과 검색 알고리즘의 단계는 다음과 같다.

Step 1. key frame으로 추출 된 frame을 원 이미지로 복원한다.

Step 2. frame의 색인을 위한 parameter를 구한다.

① 영상의 휘도 평균(p_1)

② 히스토그램의 분산(p_2)

Step 3. Step2에서의 색인 값과 질의영상의 색인 parameter값의 차이가 적은 순으로 유사도 평 가를 한다.

$$Sim_i = \frac{\sum_{x=0,y=0}^{M-1,N-1} |Img_{query}(x,y) - Img_i(x,y)|}{NM}$$

Step 4. Sim_i의 값이 일정 임계치(S) 이하가 될 때까

지 Step 3을 반복한다.

Step 5. Sim_i값이 작은 순으로 10개의 응답영상을 출력한다.

Ⅲ.실험 및 고찰

1. Shot boundary frame 추출

test 영상으로는 영화 psycho의 일부분으로서 384 × 288 크기의 영상이고, 총 프레임 수는 1618개이다. 실험결과, 일반적인 shot 검출에 사용되는 pixel간 차분영상 feature와 히스토그램 chi-square feature만을 사용하여 shot 검출 시 보다 제안하는 히스토그램 분산과 히스토그램 bin의 위치에 따른 분산 값을 이용하여 검출했을 때, 더 개선된 결과를 보였다. shot 검출의 기준으로는 직접 사람의 지각으로 shot이 바뀌는 frame을 검출하여 비교 기준으로 사용하였다. 결과는 표 1과 같다

표 1. shot frame 검출 결과

Table 1. Experiment results of shot detection.

	correct	miss	extra detection	total
F1, F2	45/56 (80%)	11	12	57
F1~F4	51/56 (91%)	5	9	60

표 1에서 correct는 사람의 지각으로 뽑은 shot과 정확히 일치하는 shot을 검출한 경우이고, miss 항목은 사람이 shot으로 지각하였으나 검출하지 못한 경우이고, extra detection 항목은 사람이 shot으로 지각하지 않았으나, shot으로 검출한 경우이다. extra detection의 경우는 shot의 근처 frame을 검출하는 경우가 대부분이라 문제가 되지 않으나, miss와 같이 shot인 부분을 검출하지 못하는 것은 다음 과정인 key frame 추출에 잘못된 결과를 초래할 수도 있다.

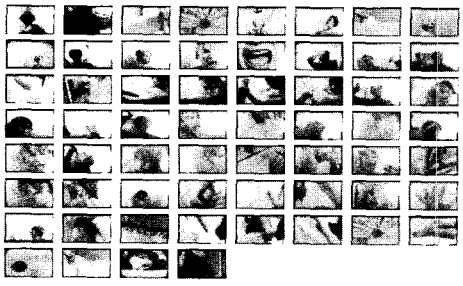


그림 2. psycho에서 검출된 shot frame들

Fig. 2. Shot frames of detected in psycho.

2. Key frame 추출

shot frame을 추출하고 나서 그 video 스트림의 대표 frame을 찾는 key frame 추출을 수행한다. 본 실험에서는 각 shot frame들 중에 key frame을 찾기 위해 5개의 parameter를 구하고 전체 shot frame의 각 parameter들의 통계적 특성에 따라 weight를 부가한 characterizing value를 구한다. 다음의 그림 3은 key frame을 구하기 위한 characterizing value를 그래프로 나타낸 것이다. 그림 4은 구해진 characterizing value의 시간에 대한 미분값이다. 계산되어진 characterizing value의 시간에 대한 미분값을 구하여 local maxima와 local minima를 구해보면, 각 local maxima와 local minima는 그 비디오 스트림의 가장 특징적인 frame과 평균적인 frame을 나타내고 있다. 이 순간의 frame을 key frame으로 추출하였다. 특히 28~31번까지의 shot frame의 시간에 대한 미분치가 거의 동일하다는 것을 볼 수 있을 것이다. 이것은 video 스트림의 중요한 부분에서 shot의 급격한 변환으로 유사한 여러 frame이 shot으로 검출됨으로써 발생한 것으로 실제 video 스트림에서 대부분 가장 대표적인 frame으로 검출될 수 있다. 그림 5는 psycho 데이터에서 key frame으로 추출된 dc image들이다.

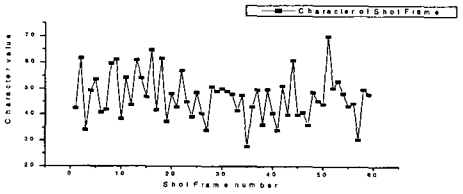


그림 3. characterizing value

Fig. 3. Characterizing value.

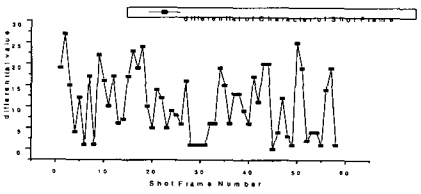


그림 4. characterizing value의 미분값

Fig. 4. Differential value of characterizing value.

3. 색인과 검색

제안한 방법으로 색인 하여 비디오 데이터에서 추출한 총 106개의 key frame으로 데이터베이스를 구성하여 각 비디오의 frame들 중 key frame이 아닌 frame 각 50개를 임의로 선정하여 질의영상으로 입력하여 검색한 결과는 다음의 표 2와 같다. 150개의 영상 중 1순위에 정확한 검색이 이루어진 경우는 92%로서 신뢰도 높은 검색이 가능하다.

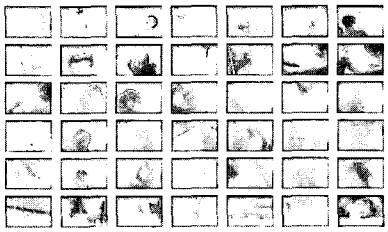
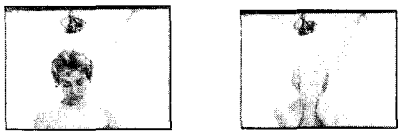


그림 5. key frame으로 추출된 dc image
Fig. 5. Key frames of detected.

표 2. 검색결과
Table 2. results of retrieval.

	1st out	2nd out	3rd out	others
accuracy	138/150	5/150	3/150	4



(a) (b)

그림 6. 검색결과 (a) 질의영상 (b) 응답영상
Fig. 6. results of retrieval.(a)Query image (b)Retrieval image.

그림 상과 질의영상에 대해 1순위로 출력된 retrieval 영상6은 검색시스템에 입력한 query 영이다. 사용자가 검색시스템을 이용할 때에 자기가 찾고자 하는 query 영상을 입력함으로서 출력된 retrieval 영상 중 사용자의 요구와 일치된 영상을 선택함으로서 원하는 비디오 데이터를 찾을 수 있다.

Ⅳ. 결 론

본 연구에서는 객체의 움직임에 강인하면서 shot내에서의 칼라의 변화에 둔감한 새로운 feature를 제안하고, 이를 이용하여 shot frame을 검출하였다. shot frame에서 구한 각 feature들의 통계적 특성을 이용하여 video 스트림의 특징에 따라 weight를 부가하는 characterizing value를 제안하고, 이를 적용하여 key frame을 추출하였다. 실험 결과 일반적으로 사용되어지는 feature와 함께 제안한 feature를 적용함으로서 보다 정확한 shot frame 추출이 가능하였다. 또한 제안된 characterizing value의 시간에 대한 미분의 local maxima와 local minima에서 video 스트림의 중요한 대표 frame을 key frame으로 추출하였고, 특히 가장 대표되는 스트림의 중요 부분에서 뚜렷한 변별력을 보였다. 색인과 검색에 있어서는 key frame을 색인하고 질의영상 입력 시 질의영상의 색인 parameter를 구해 parameter값에 따라 순차적으로 유사도를 비교함으로서 보다 빠르고 효율적인 검색이 가능하게 하였다.

참 고 문 헌

[1] K. R. Rao, J.J.Hwang. "Techniques and Standards for Image · Video and Audio Coding," Prentice Hall. 1996.

[2] Yunis S. Avrithis, Anastasios D. Doulamis, Nikolaos D. Doulamis and Stefanos 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.

[3] J. Meng, Y. Juan, and S.F. Chang, "Scene change detection in a mpeg compressed video sequence," in Proc. SPIE-Digital Video Compression : Algorithms and Tech.,(San Joce, CA), Feb. 1995.

[4] M. Abdel-Mottaleb and R. Desai, "Image Retrieval by Local Color Features," The 4th IEEE Symposium on Computers and Communications, Egypt, July 1999.

[5] D. Zhong, S. F. Chang, "Spatio-Temporal Video Search using the Object-Based Video Representation," Proc. ICIP'97, Vol1, pp. 21~24, Oct 1997, Santa Babara, CA.

[6] Wei Xiong and Chung-Mong Lee, "Efficient Scene Change Detection and Camera Motion Annotation for Video Classification," Computer Vision and Image Understanding Vol.71, Nos 2/2, August, pp. 166~181, 1998.

저 자 소 개

劉 玲 達(正會員)
1998년 : 동아대학교 전자공학과 학사. 2000년 동아대학교 전자공학과 석사. 2001년 : LG 디지털 TV 연구소 ASW팀 재직중

姜 大 星(正會員) 第 34券 C編 第 7號 參照
현재 동아대학교 전자공학과 조교수

金 大 鎭(正會員) 第 36券 S編 第 8號 參照
현재 포항공과대학교 컴퓨터공학과 부교수