

論文2003-40SP-1-4

PCA에 기반한 압축영역에서의 MPEG Video 검색기법

(PCA-Based MPEG Video Retrieval in Compressed Domain)

李 暲 和 * , 姜 大 星 *

(Kyung-Hwa Lee and Dae-Seong Kang)

요 약

본 논문에서는 동영상 압축 부호화에 대한 표준안인 MPEG 기반의 압축 비디오 스트림으로부터 DCT DC 계수를 추출하고, 이들로 구성된 DC 영상으로부터 장면 전환 검출을 수행한 후 대표 프레임을 추출한다. 또한 추출된 대표 프레임을 PCA(Principal Component Analysis) 방법을 이용하여 데이터베이스의 색인 정보로 저장한 후 입력된 질의 영상과 가장 유사한 대표 영상들을 검색하는 방법에 대해 제안한다. 즉, 추출된 대표 프레임에 대해 주성분해석 기법을 적용하여 통계적인 특성을 가진 데이터를 특징으로 추출함으로써 코드워드의 수에 따른 코드북을 생성하고 이를 데이터베이스의 색인 정보로 저장한다. 실험 결과 제안된 방법이 검색에 있어 우수한 성능을 나타내고 또한 통계적인 데이터의 특성을 이용하기 때문에 처리 시간과 상당한 양의 메모리 공간을 줄일 수 있음을 확인하였다.

Abstract

This paper proposes a database index and retrieval method using the PCA(Principal Component Analysis). We perform a scene change detection and key frame extraction from the DC image constructed by DCT DC coefficients in the compressed video stream that is video compression standard such as MPEG. In the extracted key frame, we use the PCA, then we can make codebook that has a statistical data as a codeword, which is saved as a database index. We also provide retrieval image that are similar to user's query image in a video database. As a result of experiments, we confirmed that the proposed method clearly showed superior performance in video retrieval and reduced computation time and memory space.

Keyword : PCA(Principal Component Analysis), DCT, DC image, codebook, codeword

I. 서 론

영상을 포함한 멀티미디어 데이터 량이 폭발적으로 증가하면서, 이러한 방대한 양의 데이터들의 저장, 브라우징(browsing), 인덱싱(indexing), 검색(retrieval)을 위

해 많은 기법들이 개발되고 있으며^[1], 디지털 박물관, 인터넷 쇼핑, 의료영상 검색 시스템과 같은 많은 응용 분야의 핵심 기술로서 영상 데이터베이스의 인덱싱 및 검색에 대해 활발한 연구가 진행되고 있다. 특히 영상 데이터베이스로부터 원하는 것을 찾고자하는 영상 검색 방법은 기존의 텍스트 주석에 기반한 검색 방법의 한계를 극복하는 새로운 분야로 각광을 받고 있다^[2-4]. 본 논문에서는 장면 전환 검출 알고리즘과 각 셋을 대표하는 대표 프레임 추출에 대한 방법을 제안하고, 추출된 대표 프레임에 주성분 해석 기법을 적용하여 입력 벡터의 차원을 축소한다. 또한 입력 벡터의 통계적

* 正會員, 東亞大學校 工科大學 電子工學科
(Dept. of Electronic Eng., Dong-A University)
(Dept. of Electronics Engineering, Pusan Nat'l Univ)
接受日字:2001年6月13日, 수정완료일:2002年12月3日

인 특성을 데이터베이스의 색인 정보로 저장한 후 입력되는 질의 영상에 대해 원하는 유사한 영상을 비디오 전체 내용을 검색하지 않고 대표 프레임에서 제시하는 영상 검색 기법에 대해 제안한다.

본 논문의 구성은 다음과 같다. I장 서론에 이어 II장에서는 장면 전환 검출 및 대표 프레임 추출 알고리즘에 대해 기술하고 III장에서는 추출된 대표 프레임에 PCA 기법을 적용한 색인 및 검색 기법에 대해 설명한다. IV장에서는 제안한 방식을 적용한 실험결과 및 연구 내용을 검토하고, 마지막 V장에서는 결론 및 향후 연구 방향에 대해 언급한다.

II. 장면 전환 검출과 대표 프레임 추출

1. MPEG 비디오 스트림으로부터 DC 영상 추출

본 논문에서는 I-프레임에 대해서 DCT DC 계수를 구하고 이를 정규화시킨 후 DC 영상을 구성하였다. 그림 1은 MPEG 비디오 스트림으로부터 구성된 I-프레임에 대한 DC 영상을 나타낸 것이다.

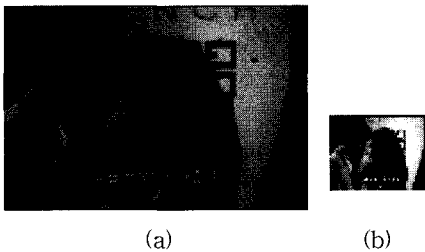


그림 1. MPEG 스트림으로부터 구성된 DC 영상
(a) 원 영상(I-프레임), (b) DC 영상
Fig. 1. DC image from MPEG video stream.
(a) I-type frame, (b) DC image

2. 장면 전환 검출 알고리즘

먼저 장면 전환 검출을 위한 이웃 프레임간의 DC 영상에 대해 다음의 파라미터들을 계산한다.

$$\textcircled{1} \text{DiffImg}_i = \frac{\sum_{x=0, y=0}^{x \leq M-1, y \leq N-1} |I_{i-1}(x, y) - I_i(x, y)|}{NM}$$

여기서, $I_i(x, y)$ 는 현재 DC 영상의 픽셀 값이다.

$$\textcircled{2} X_i^2 = \sum_{k=0}^{k \leq q} \frac{(H_{i-1}(k) - H_i(k))^2}{H_i(k)}$$

여기서, $H_i(k)$ 는 i 번째 DC 영상 히스토그램의 k 번째 bin의 값을 나타내고, $H_{i-1}(k)$ 은 그 이전 값이다. n 은 8bit 영상을 사용할 경우 256이다.

$$\textcircled{3} \text{DiffD}_i^2 = \frac{(D_{i-1} - D_i)^2}{D_i}$$

여기서, D_i 는 i 번째 DC 영상의 히스토그램 분산값이고, D_{i-1} 은 그 이전 값이다.

$$\textcircled{4} \text{PX}_i^2 = \sum_{k=0}^{k \leq q} \frac{(\rho X_{i-1}(k) - \rho X_i(k))^2}{\rho X_i(k)}$$

$$\text{PY}_i^2 = \sum_{k=0}^{k \leq q} \frac{(\rho Y_{i-1}(k) - \rho Y_i(k))^2}{\rho Y_i(k)}$$

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

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

여기서 M, N 은 각각 DC 영상의 가로, 세로 크기이며, DiffImg_i 는 DC 영상의 차분, X_i^2 은 DC 영상의 히스토그램에 대한 chi-square, DiffD_i^2 은 DC 영상 히스토그램 분산의 차분, $\text{PX}_i^2, \text{PY}_i^2$ 은 양자화된 영상의 히스토그램의 각 bin 값들의 열과 행의 위치에 대한 분산의 chi-square 값, $BX_i(n)$ 은 bin에 대한 열의 픽셀 값, $BY_i(n)$ 은 bin에 대한 행의 픽셀 값이고, $\rho X_i(n)^2, \rho Y_i(n)^2$ 은 i 번째 DC 영상의 n 번째 bin의 열과 행에 대한 분산이다. 이상과 같이 구해진 DC 영상의 파라미터들로부터 다음의 단계로 장면 전환 프레임을 추출한다.

STEP 1. 각 파라미터들의 전체 프레임에 대한 평균을 구한다.

STEP 2. DC 영상의 각 파라미터 값이 다음의 조건식에서 condition 1과 2를 동시에 만족하면서 condition 3 혹은 4가 만족할 경우 장면 전환 프레임으로 검출한다.

condition 1. $\text{DiffImg}_i > \alpha \overline{\text{DiffImg}}$

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

condition 3. $\text{DiffD}_i > \gamma \overline{\text{DiffD}}$

condition 4. $\text{PX}_i > \delta \overline{PX}$ and $\text{PY}_i > \delta \overline{PY}$

여기서 각 상수 $\alpha, \beta, \gamma, \delta$ 는 실험에 의해 얻어진 최적화된 상수를 나타낸다.

3. 대표 프레임 추출 알고리즘

대표 프레임 추출은 구해진 장면 전환 프레임들 중 그 비디오 스트림을 가장 잘 표현할 수 있는 대표 프레임을 찾는 과정이다. 본 논문에서는 대표프레임을 추출하기 위하여 장면전환 프레임에 5개의 파라미터를

조합하여 characterizing value를 구함으로서 보다 적합한 대표 프레임 추출을 수행한다. 다음은 장면 전환 프레임에 대해 다음의 파라미터들을 계산한다.

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

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

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

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

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

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

$$\textcircled{4} f_4 = DiffS_i = \frac{\sum_{x=0, y=0}^{x \leq M, y \leq N} |S_i(x, y) - S_{i-1}(x, y)|}{MN}$$

$$\textcircled{5} f_5 = AccDiffShot_i = \frac{\sum_{k=0}^{k \leq i} DiffShot_k}{DiffShot_i}$$

여기서 f_1 은 첫 프레임에 대한 명암도의 평균, f_2 는 첫 프레임에 대한 히스토그램 분산값, f_{3x}, f_{3y} 는 양자화한 첫 프레임 히스토그램의 각 bin값들의 열과 행의 위치에 대한 분산의 평균, f_4 는 이전 첫 프레임과의 명암도 차, f_5 는 명암도 차의 누적에 대한 f_4 의 비율이다. 위의 파라미터들을 검출된 첫 프레임에서 구하고 난 후 다음과 같은 단계를 거쳐 대표 프레임을 검출한다.

STEP 1. 첫으로 검출된 모든 프레임에 대한 각 파라미터들의 분산을 구한다.

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

STEP 2. 각 파라미터들의 전체 평균에 대한 차를 구한다.

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

STEP 3. 각 파라미터들에 대해 분산에 대한 비율만큼 weight를 부가하여 각 프레임의 특징 값(c_m)을 구한다.

$$C_m = \omega_1 f_1' + \omega_2 f_2' + \dots + \omega_n f_n'$$

여기서 m 은 첫으로 검출된 프레임의 개수, 가중치 ω 는

$$\omega_1 = \frac{\rho_1}{\sum_{i=1}^m \rho_i}, \omega_2 = \frac{\rho_2}{\sum_{i=1}^m \rho_i}, \dots, \omega_n = \frac{\rho_n}{\sum_{i=1}^m \rho_i}$$

STEP 4. 시간에 대한 특징값의 변화를 구한다.

STEP 4에서 구한 값들에 대해 시간의 도함수를 구하면, 각 local maxima는 첫으로 검출된 프레임 중 가장 특징적인 프레임이 나타나는 순간이고, 각 local minima는 그 구간에서 가장 평균적인 프레임이 나타나는 순간이다. 각 local maxima와 local minima에서 대표 프레임을 추출한다^[5].

III. PCA를 이용한 대표프레임 색인 및 검색기법

본 논문에서는 앞장에서 기술한 대표 프레임 추출 알고리즘에 의해 추출된 대표 프레임에 대한 영상 색인 및 검색 기법으로 주성분 해석을 이용하는 방식에 대해 제안한다. 주성분 해석 기법은 입력 공간내의 변동량이 많은 방향으로 주축을 발견하여 데이터의 중복성을 제거하는 기법으로 주성분 분석 기법이라고도 한다. 내용기반 검색은 영상 특징 정보의 차이값을 척도로 하여 데이터베이스 내의 모든 영상들과 검색 대상 영상을 비교하는 것으로 방대한 계산량을 요구한다. 따라서 주성분 해석을 이용한 고차원 데이터의 차원 감소는 영상 분석에 있어 유용하다. 즉, $n \times n$ 의 2차원 영상을 n 개의 1차원 데이터 요소로 축소함으로써 상당한 양의 처리 시간과 메모리 공간을 줄일 수 있다. 또한 주성분 해석에 의하여 계산된 주축은 입력 벡터의 최대 고유치의 고유 벡터의 방향과 일치하며 입력 벡터의 변동량이 가장 큰 방향이다. 본 논문에서는 추출된 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^[6].

IV. 실험결과

본 장에서는 제안한 대표 프레임 추출 및 영상 검색 기법의 성능을 평가하기 위해 테스트 영상으로 352 × 240 크기를 갖는 8318 프레임과 13062 프레임을 갖는 뮤직비디오 2편에 대해 256 그레이 레벨 해상도의 휘도 성분을 갖고 실험한 결과에 대해 기술한다.

뮤직비디오 2편 모두 전반적으로 많은 카메라 움직임을 포함하고 있으며, 물체의 움직임은 보통 정도이다. 첫 번째 뮤직비디오 영상은 panning이 영상 전반에 걸쳐 사용되었고, 2번의 fade-out/fade-in에 의한 장면 전환이 일어난다. 두 번째 뮤직비디오 영상은 후반부에는 flashlight 빛에 이은 급격한 장면 전환이 4번 반복되는 부분이 있고, close-up된 카메라 동작을 보여주는 부분이 있다. 본 논문에서 제안한 알고리즘에 의한 실험 결과 급격한 장면 전환은 거의 모두 검출할 수 있었으며, 두 개의 fade-out/fade-in도 검출되었다. <그림 2>는 제안한 알고리즘에 의해 검출된 장면 전환 프레임이다.



그림 2. 추출된 장면전환 프레임
Fig. 2. Detected shot frames by the proposed method.



그림 3. 추출된 대표 프레임
Fig. 3. Detected key frames by the proposed method.

<그림 3>은 구해진 장면전환 프레임 중에서 그 video 스트림을 가장 잘 표현할 수 있는 대표 프레임을 추출한 것이다.

<표 1>은 군집화하여 얻은 대표 벡터의 수 50, 100, 200, 300인 경우에 대해 기존의 알고리즘과 비교한 결과를 보여준다. 실험결과 군집의 수가 300, 200, 100인 경우 제안하는 기법이 Histogram Intersection 보다 효과적인 검색 결과로 나타났다.

표 1. 기존의 알고리즘과 군집화하여 얻은 대표벡터를 이용한 방법의 비교

Table 1. Comparison of histogram intersection and the proposed method.

Method	검색 결과
Histogram Intersection	29
vector 300	33
vector 200	33
vector 100	30
vector 50	28

이렇게 얻어진 대표 프레임을 히스토그램 평활화를 시키고, 전체를 2차원 입력 벡터로 나타낸다. 주성분 해석을 이용하여 전체 입력 벡터의 주축을 구하고, 주축과 직교하면서 입력 벡터의 중점을 지나는 직선에 의하여 두 영역으로 나눈다. 나누어진 각각의 영역에 대해서도 영역 분할 알고리즘을 통해 동일한 방법으로 재분할한다. 각각의 재분할된 영역도 초기의 입력 벡터와 마찬가지로 영역 분할이 이루어지며, 영역 분할은 초기에 설정한 최종 영역 분할의 개수가 될 때까지 반복된다. <표 2>는 영역 분할 실험중 초기의 영역 분할을 위한 정보를 나타낸다.

표 2. 초기 영역 분할을 위한 정보

Table 2. Domain division information of the first stage.

Covariance matrix	C_1	0.33
	C_2	0.32
	C_3	0.32
	C_4	0.34
λ_{max}		0.65
θ		0.159 (9.134 degree)

최종 영역의 평균값을 코드워드로 저장하고, 각 대표 프레임에 대해 코드워드에 할당된 수를 세어 카운트한 값을 검색을 위한 색인 파라미터로 사용한다. <그림 4>는 최종 영역의 분할 개수를 50, 100, 200, 300개로 주었을 때 생성된 코드워드를 보여준다. <그림 5>는 생성된 50개의 코드워드에서 대표 프레임에 할당된 카운트 값을 나타낸다. x축은 코드워드를 나타내고 y축은 각 코드워드에 할당되는 카운트 회수를 나타낸다.

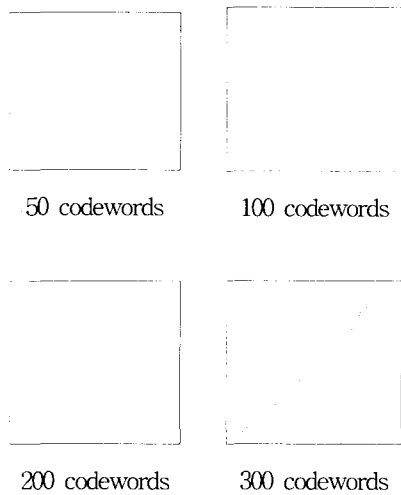


그림 4. 영역 분할에 의해 생성된 코드워드
Fig. 4. Codeword for domain division.

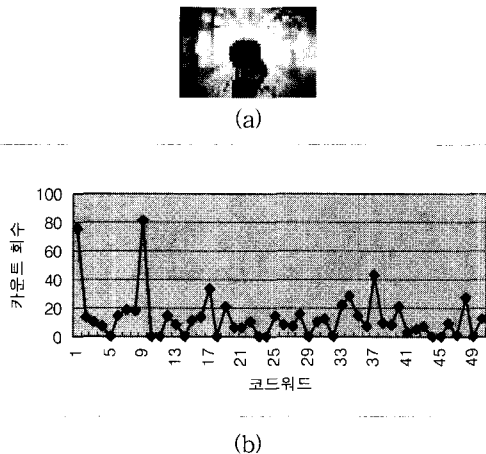


그림 5. 코드북 생성 결과
(a) 대표프레임, (b) 그림(a)대한 코드북 생성 결과

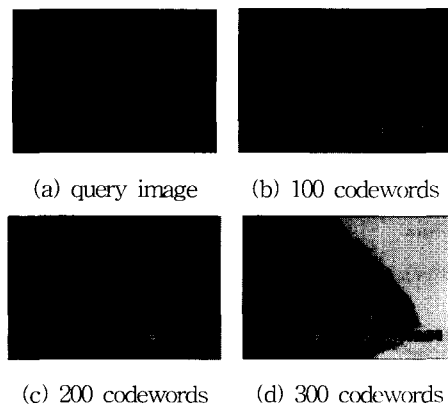
Fig. 5. Result of codebook.
(a) key frame, (b) codebook for picture (a).

<표 3>은 대표 벡터로 군집화 하는데 필요한 시간을 나타내었다. 본 실험에서 군집화는 최종 대표 벡터의 수에 도달할 때까지 반복한다. 같은 크기의 영상이라도 하나의 영상을 군집화 하는데 걸리는 시간이 시스템의 상태에 따라서 조금씩 변하므로 같은 크기의 영상을 군집화 하는데 5회에 소요되는 시간의 평균값을 기록하였다.(PentiumIII-650MHz사용)

표 3. 영상의 크기와 군집화에 소요되는 시간
Table 3. Execution time for clustering and size of Image.

Size	# of Vectors	Execution time for clustering
352×240	300	4.878065 sec
	200	3.527954 sec
	100	2.278065 sec
	50	1.85175 sec

<그림 6>은 제안한 주성분 해석 기법의 주축과 영역 분할 알고리즘을 이용한 검색 결과를 나타낸 것이다. 최종 영역의 평균값을 코드워드로 하여 코드워드 수에 따른 유사도가 가장 높은 검색 결과를 나타낸 것이다. 결과에서 보듯이 입의 질의 영상이 주어졌을 때 코드워드 수에 따른 검색 결과 코드워드의 개수가 많을 수록 좀 더 나은 검색 결과를 보이긴 했지만 적은 수의 코드워드로도 질의 영상과 아주 유사한 검색 결과를 보였다. 또한 검색된 대표 프레임들을 사용자에게 제공함으로써 비디오 전체 내용을 검색하지 않더라도 원하는 것을 찾아낼 수 있으므로 검색 효율 측면에서도 뛰어나다는 것을 알 수 있다. 이상의 실험에 근거해 볼 때 제안한 방법은 효율성 및 검색에 있어 타당하고 우수한 성능을 갖는다는 것을 알 수 있다.



(a) query image (b) 100 codewords
(c) 200 codewords (d) 300 codewords

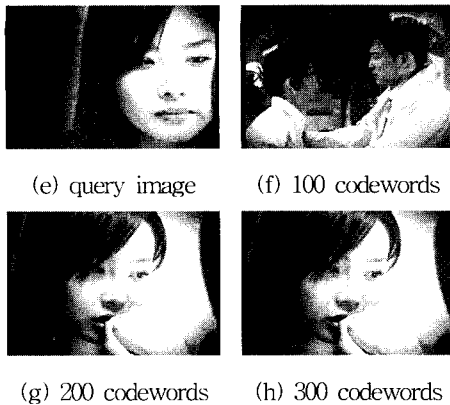


그림 6. 코드워드 수에 따른 검색결과의 예
(a),(e) 질의 영상, (b)-(d),(f)-(h) 각 코드워드 수에 따른 검색된 대표프레임

Fig. 6. Retrieved images for each codeword.
(a),(e) query image, (b)-(d),(f)-(h) the detected key frames for each codeword

V. 결 론

본 논문에서는 MPEG 비디오 스트림으로부터 DCT DC 계수를 추출하고, 이로 구성된 DC 영상에 대해 장면 전환 및 대표 프레임 추출하고, 추출된 대표 프레임을 2차원 입력 벡터로 나타낸 후 주성분 해석에 기반한 주축을 구하는 알고리즘과 영역 분할 알고리즘을 이용한 영상 검색 기법을 제시하였다. 실험 결과의 분석으로부터 제안한 방법은 입력 벡터의 국부적인 특성이 아닌 데이터의 통계적인 특성을 이용하고, 데이터의 중복성을 제거한 주성분 데이터를 특징으로 이용함으

로써 검색에 있어서도 우수한 성능을 보이며, 상당한 양의 처리 시간과 메모리 공간을 줄일 수 있는 장점이 있다. 향후 제안한 방법은 색상과 같은 색인정보와 모션 데이터를 이용할 경우, 보다 나은 내용기반에 충실한 색인과 검색 수단을 제공할 것이다.

참 고 문 헌

- [1] Myron Flickner et al, "Query by Image and Video Content: The QBIC System", IEEE Computer, Vol. 28, No. 9, pp. 23~32, 1995.
- [2] V. N. Gudivada and V. V. Raghvan, "Content-Based Image Retrieval Systems", IEEE Computer, Vol. 28, No. 9, pp. 18~22, 1995.
- [3] F. Idris and S. Panchanathan, "Review of Image and Video Indexing Techniques", Journal of Visual Communication and Image Representation, Vol. 8, No. 2, pp. 146~166, 1997.
- [4] Anil. K. Jain, Aditya Vailaya, "Image Retrieval using Color and Shape," Pattern Recognition, Vol. 29, No. 8, pp. 12~33, 1996.
- [5] 유영달, "Robust feature를 이용한 key frame 추출과 MPEG video 검색", 동아대학교 대학원 공학석사학위논문, 1999.
- [6] 이경화, "주성분 해석 기법을 이용한 멀티미디어 검색에 관한 연구", 동아대학교 대학원 공학석사학위논문, 2001.

저 자 소 개

李 暲 和(正會員)
2000년 동의대학교 컴퓨터공학과 학사. 2002년 동아대학교 전자공학과 석사

姜 大 星(正會員) 第 38卷 CI編 第 5號 參照
현재 동아대학교 전기전자컴퓨터공학부 부교수