

---

# 은닉 마르코브 모델을 이용한 비디오 요약 시스템

박호식\* · 배철수\*

## Video Summarization Using Hidden Markov Model

Ho-sik Park\* · Cheol-soo Bae\*

### 요 약

본 논문에서는 비디오 검색을 위한 비디오 사진 분류 시스템을 제안하였다. 제안된 시스템은 3개의 모듈인 특징 추출, 은닉 마르코브 모델 생성, 그리고 비디오 사진 분류로 구성되어 있다. 같은 등급에 속한 비디오 화면들이 반드시 유사하지 않으므로 견실한 Hidden Markov Model을 구성하기 위해서는 충분한 학습이 필요하였다. 제안된 시스템은 텔레비전 야구 중계 방송의 비디오 화면을 15가지 등급으로 분류하여 분석 및 요약하는 실험을 한 결과 평균 84.72%의 인식률을 얻을 수 있었다.

### ABSTRACT

This paper proposes a system to analyze and summarize the video shots of baseball game TV program into fifteen categories. Our System consists of three modules: feature extraction, Hidden Markov Model (HMM) training, and video shot categorization. Video Shots belongs to the same class are not necessarily similar, so we require that the training set is large enough to include video shot with all possible variations to create a robust Hidden Markov Model. In the experiments, we have illustrated that our system can recognize the 15 different shot classes with a success ratio of 84.72%

### 키워드

Feature Extraction, Hidden Markov Model, Video Shot categorization

### 1. 서 론

멀티미디어 데이터베이스 내에서 디지털 비디오의 양이 증가함에 따라 내용 기반에 의한 비디오 검색 기술을 필요로 하고 있다. 최근 들어 여러 가지 다양한 내용 기반의 비디오 분류 방법들[1-3]이 개발되고 있는데 크게 두 가지로 분류할 수가 있다. 첫째는 하나의 비디오 장면을 여러 사진들로 나누고 각 사진에서 핵심프레임을 추출하는 것이다. 핵심프레임은 그 비디오 장면에 대한 기준을 포함한 상태로 라이브러리에 저장된다. 이 장면들을 검색하기 위해서 사용자는 하나의 질의 이미지를 제시

하고 검색 시스템은 사전 정의된 특징들을 이용하여 유사한 핵심프레임들을 검색한다. 두 번째 방식은 정보 검색의 개념에 기초한 것으로, 비디오 장면들의 이미지에 따라서가 아닌 장면들의 의미에 따라 분류되어진다. 이 방식은 분석과 분류로 구성된다[3]. 분석은 비디오 스트림으로부터 비디오 사진을 추출하는 작업이며, 분류는 비디오 사진을 비디오 사진 등급들 중 하나로 분류하는 것이다.

근래 들어 이미지/비디오 데이터베이스의 논문들은[4-9] 이미지의 내용에 의한 이미지/비디오 분류에 초점을 맞추고 있다. 이 분류 작업은 자동 비디오 사진 분석 및 표현을 필요로 한다. 이러한 분

석과 표현은 사건을 탐색하고 동작들을 이해하고, 마지막으로 검색을 위해 비디오 사진을 요약하게 된다.

비디오 검색을 위해서 비디오 사진들을 분석하고 각 사진의 의미를 서술할 필요가 있기에, 각 사진의 의미에 근거하여 분류하는 비디오 사진 분류 시스템을 제안한다. 제안된 시스템은 3개의 모듈, 즉 특징 추출, 은닉 마르코브 모델(Hidden Markov Model - 이하 HMM) 생성, 그리고 비디오 사진 분류로 구성되어 있다. 제안된 시스템은 15개의 서로 다른 HMM들로 구성되어 있다. 이것들은 15개의 서로 다른 비디오 사진 등급들을 테스트하는데 사용되었다.

본 논문에서는 야구 경기의 장면을 가지고 15개의 내용 기반의 사진 등급을 다음과 같이 정의하였다.: (1) 왼손 타자의 전면, (2) 왼손 타자의 측면, (3) 투수와 타자, (4) 내야수, (5) 주자의 확대 화면, (6) 상공에서 본 그라운드, (7) 상공에서 본 타자, (8) 투수의 확대 화면, (9) 주자를 견제하는 투수, (10) 오른손 타자의 전면, (11) 오른손 타자의 측면, (12) 타자의 원거리 모습, (13) 포수와 투수, (14) 외야수, (15) 타격 슬로우 모션. 이 밖의 다른 형태의 비디오들은 실험에서 제외되었다. 제안된 시스템을 이용하여 실험한 결과 평균 인식률은 84.72%였다.

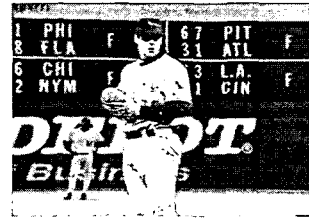
## II. 특징 벡터 생성

비디오 화면에서 움직이는 물체의 동작과 위치 정보는 비디오 요약에 있어서 필수적인 것이다. 여기서 동작이란 카메라의 이동과 확대에 의한 장면의 변화, 그리고 사물의 움직임을 포함하는 것으로 서로 다른 사진의 이미지 프레임들의 색상 분포 및 구조 정보 또한 중요한 작용을 한다. 본 논문에서는 비디오 화면에 움직이는 사물이 있다고 가정하였다.

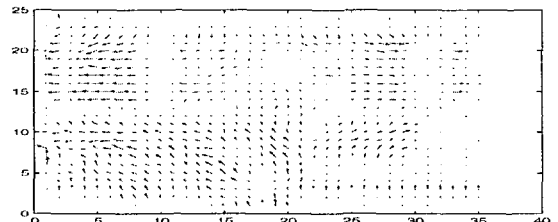
### 2.1 동작 추정

동작 추정은 광학 흐름 방정식[10]이라고 하는 시공간 이미지 강도 경사도를 기반으로 하였다. 광학 흐름 방정식은 적합한 공간과 시간에서 인접한 벡터 변화가 유연하게 하는 유연성 제약과 관련되어 있다. 동작 벡터의 크기와 위상은 움직이는 물체의 이동 속도와 방향을 나타내므로, 동작 벡터의 크기 및 위상의 막대그래프 분포에서 동작 특징으로서 추출하였다. 동작 벡터들의 크기 분포를 10개의 구간으로 나누고,  $Pe(i)$ 를 크기 구간  $i$ 에 속한 동작 벡터의 수라고 하면,  $fe(i)=Pe(i)/(\text{프레임의 총 픽셀 수})$ 라 할 수 있다. 여기서  $1 \leq i \leq 10$ 이며,  $fe(i)$ 는 동작 벡터의 크기로부터 추출된 특징들을 나타

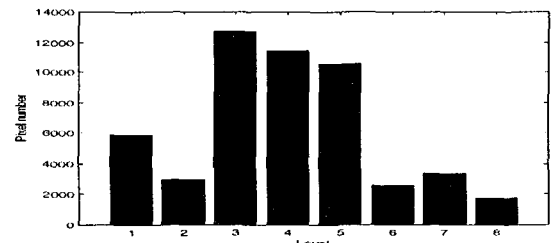
낸다. 또한 동작 벡터의 위상 분포를 8개의 구간으로 나누었으며,  $Pp(i)$ 를 위상 구간  $i$ 에 속한 동작 벡터들의 수라고 하면,  $fp(i)=Pp(i)/(\text{프레임의 총 픽셀 수})$ 라 할 수 있다. 여기서  $1 \leq i \leq 8$ 이며,  $fp(i)$ 는 동작 벡터의 위상으로부터 추출된 특징들을 나타낸다.



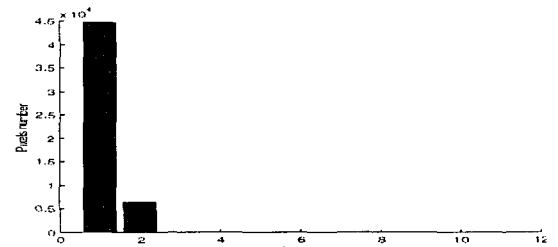
(a)



(b)



(c)



(d)

그림 1. (a) 투수의 확대 화면 (b) 동작 벡터 (c) 위상에 대한 막대그래프 (d) 크기에 대한 막대그래프

Fig 1. (a) close-up shot of the pitcher (b) Motion Vector Field (c) Histogram of phase (d) Histogram of Magnitude

예를 들어 그림 1(a)와 같은 투수의 확대 화면이 있다면, 이 비디오 사진에서 그림 1(b)의 동작 벡터들이 바깥쪽을 향하고 있음을 알 수 있다. 배경에 몇 개의 작은 물체들이 있을 수 있으며 따라서 주요 대상물의 동작 벡터들이 여러 방향을 가리킬 수 있다. 위상 막대그래프(그림 1(c))로부터 구간 3, 4, 5의 최고점들이 동작의 3가지 주요 방향들을 나타내고 있음을 알 수 있다. 또한 크기 막대그래프(그림 1(d))로부터 오직 첫 번째 구간의 최고점이 주요 동작 속도를 나타내고 있음을 알 수 있다.

동작의 방향과 크기 분포에 대한 위와 같은 분석 외에 동작 벡터의 분포와 관련된 다른 특징들을 알 수 있다. 동작 벡터가 2차원 공간  $X=(x, y)$ 에서의 동작의 강도라고 하면, 벡터  $v(x, y, t)$ 를 이용하여 동작의 분포를 특징지을 수 있다. 시점  $t$ 에서 동작 분포를 특징짓기 위하여 무게중심  $\vec{m}(t)^T = [m_x(t), m_y(t)]$ 을 사용하였다.

$$m_x(t) = \frac{\sum_{x,y} x \cdot v(x,y,t)}{\sum_{x,y} v(x,y,t)} \quad m_y(t) = \frac{\sum_{x,y} y \cdot v(x,y,t)}{\sum_{x,y} v(x,y,t)} \quad (1)$$

벡터  $\vec{m}(t)$ 는 이미지의 "동작의 중심"으로 해석될 수 있다.  $v(x, y, t)$ 는 시점  $t$ , 위치  $(x, y)$ 에서의 동작벡터의 크기를 나타낸다. 이 중심의 움직임에 대한 HMM의 모델링 용량을 증가시키기 위해서 우리는 동작의 중심에 대한 특징의 변화  $\Delta \vec{m}(t)$ 를 특징 벡터에 포함시킨다. 특징의 변화는  $\Delta m_x(t) = m_x(t) - m_x(t-1)$ 과  $\Delta m_y(t) = m_y(t) - m_y(t-1)$ 로 정의된다. 또 하나의 유용한 특징으로 동작의 중심에서 이미지들의 모든 지점들까지 동작의 평균 절대편차  $\sigma_x(t)' = [\sigma_x(t), \sigma_y(t)]$ 를 다음과 같이 정의한다.

$$\sigma_x(t) = \frac{\sum_{x,y} v(x,y,t) |x - m_x(t)|}{\sum_{x,y} v(x,y,t)} \quad \sigma_y(t) = \frac{\sum_{x,y} v(x,y,t) |y - m_y(t)|}{\sum_{x,y} v(x,y,t)} \quad (2)$$

### 2.2 색상 정보

YIQ 모델의 이점은 광도(Y)와 색상 정보(I와 Q)가 분리되었다는 점이다. 광도는 눈에 의해 지각되는 빛의 양과 비례하고 이미지의 광도 요소가 이미지의 색상 정보에 영향을 주지 않고 처리될 수 있다는 점이다. 예를 들어 단순히 Y 요소에 막대그래프를 평준화함으로써 YIQ 포맷으로 표현된 이미지 색상에 대해 막대그래프 평준화를 적용할 수 있다는 것이다. 이때 이미지에 있는 색상 정보는 이 평준화작업에 영향을 받지 않는다.

본 논문에서는 YIQ 색상 모델에 기초하여 색상

을 분석한다. 이미지의 각 픽셀의 Y, I, Q 값을 계산한 결과 Y의 값이 0에서 255 사이에, I의 값이 -45에서 105 사이에, 그리고 Q가 -95에서 95 사이에서 변화됨을 알 수가 있었다. 동작 벡터 분석과 마찬가지로, Y, I 또는 Q 막대그래프 분포를 10개의 구간으로 나누고,  $P_y(i)$ 를 Y의 값이 구간  $i$ 에 속하는 픽셀의 수,  $f_y(i) = P_y(i) / (\text{총 픽셀의 수})$ 라고 하였다. 여기서  $i$ 는 구간 번호( $1 \leq i \leq 10$ )를 나타낸다. 마찬가지로  $P_i(i)$ 를 I의 값이 구간  $i$ 에 속하는 픽셀의 수라고 하고,  $f_i(i) = P_i(i) / (\text{총 픽셀의 수})$ 라고 하였으며,  $P_q(i)$ 를 Q의 값이 구간  $i$ 에 속하는 픽셀의 수라고 하고,  $f_q(i) = P_q(i) / (\text{총 픽셀의 수})$ 라고 하였다.

### 2.3 구조 정보

이미지 내에서 보다 상세한 구조 정보를 얻기 위해서 본 논문에서는 하나의 이미지를 6개의 하위 이미지들로 나누어 그 이미지 내에 있는 물체들의 지역적 특징들을 나타내고자 하였다. 하나의 이미지 프레임을 6개의 하위 이미지들로 나누어 각 하위 이미지들의 구조를 분석했다. 이미지 프레임을 윗 부분과 아래 부분으로 나누고 각 부분을 왼쪽, 가운데 및 오른쪽 열로 나누었다. 따라서 이미지 프레임은 6개의 동일한 크기의 하위 프레임들로 나누어졌다. 6개의 하위 이미지들에 대한 구조 정보를 얻기 위해서 각 하위 이미지들에 대해 푸리에(Fourier) 분석을 실시하였다(그림 2 참조).

각 하위 이미지의 파워 스펙트럼 밀도 분포는 대칭을 이룸으로 절반만을 분석하면 된다. 하위 이미지의 주요 구조 방향을 찾기 위하여 이 분포를 8개의 구간으로 나누었다. PFP를 주파수 영역에서 이미지의 중심에 대한 그 픽셀의 위상이라 하고, 또한  $E_j(i)$ 를 하위 이미지  $j$ 의 위상  $i$ 의 에너지라고 하였다.  $E_j(i)$ 는 위상이 구간  $I$ 에 속한 모든 픽셀의 파워 밀도의 합계를 나타내고, 여기서  $1 \leq i \leq 8, 1 \leq j \leq 6, \theta_{jmax} = \text{Arg}(\max(E_j(i)))$ 이며,  $\theta_{jmax}$ 는 하위 이미지  $j$ 의 주요 구조 방향을 나타낸다.

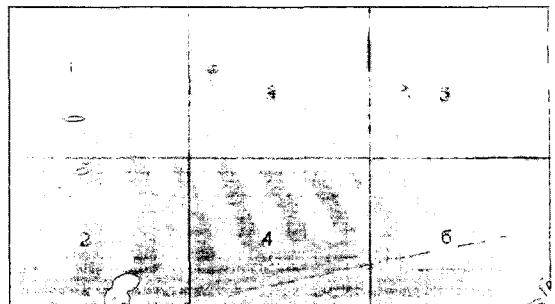


그림 2. 6개의 하위 이미지  
Fig 2. the six sub-images

또한, 하위 이미지의 주요 구조 방향과 더불어, 특징들로서 주요 방향의 에너지와 관련된 다른 방향에서의 에너지의 편차를 계산한다.

$$\sigma_f(i) = \frac{\sum_{j=1}^8 (E_j(\max) - E_j(i))}{8} \quad (3)$$

이것은 하위 이미지의 복잡성을 특징지을 수 있다. 본 논문에서는 24개의 동작 정보 특징들, 30개의 색상 정보 특징들, 그리고 12개의 구조 정보들을 추출했다. 하나의 이미지로부터 총 66개의 특징을 추출하여 그 이미지의 내용을 정확하고 효과적으로 서술하는 66 차원 특징 벡터를 생성하였다.

### III. 비디오 사진 분류를 위한 HMM

HMM은 비디오 화면의 동적 시간적 행동을 특징짓기 위하여 사용되었다. 본 논문에서는 비디오의 내용 분석을 위해 HMM을 두 가지 다른 방법으로 활용하고자 하였다: (i) 비디오 스트림으로부터 추출되어진 측정 화면으로 입력되는 데이터를 위한 하나의 모델의 가능성을 판단한다. 그리고 (ii) 주어진 측정 화면으로부터 해당되는 최적의 화면을 계산한다. 전자의 접근 방법은 주어진 비디오 스트림 유형의 내용을 파악하는 것, 즉 그 입력 데이터가 어떤 내용 등급에 속하는지를 파악하는 것이다. 한편, 후자의 접근 방법은 그 화면의 개요를 제공하며 그 사진들을 그 모델의 상태로 표현되는 범주들로 분류한다.

(i)의 성능은 (ii)에서 개발된 모델들의 효율성에 크게 영향을 받는다. 총체적으로, 이 두 프로세싱 모델들은 비디오 내용 특징화를 위한 유연하고 완전한 방법을 정의한다. 본 논문의 핵심은 (ii)를 이루기 위하여 다양한 내용 영역을 위해 적절한 HMM을 개발하는 것이다. HMM은 이중 확률 모델이다. 즉 각 상태에 대한 측정 부호 확률 분포는 측정 가능한 부호의 내적 상태 변화성을 포착하며, 상태 전환 확률은 측정 가능한 부호의 내적 동적 구조를 설명한다.

#### 3.1 벡터 양자화

본 논문에서는 벡터양자화를 이용하여 다차원 벡터 화면들을 유한 부호 화면으로 전환해야 한다 [11]. 벡터양자화는 학습을 위해 모든 k-차원 학습된 특성 벡터들을 M개의 클러스터들로 나누는 것

이다. 이 클러스터들의 중심 궤적은 k-차원 벡터  $C_i$  이며, 각 클러스터에는 부호화문자  $o_i$  라고 하는 하나의 양자화 값이 배정되어 있다. 벡터양자화 방법은 한 클러스터의 학습 화면들로부터 시작하여 그것들을 반복적으로  $M=2n$  클러스터로 나눈다. 각 클러스터의 중심 궤적은 K-means 클러스터링에 의해 정밀해진다. 실험 결과에 의해, 제안된 인식 시스템은 코드북의 크기  $M=256$  에서 높은 성능을 보였다. 일단 벡터 양자화 코드북이 확보되면, 코드북은 이어지는 HMM 작업에서 특정 벡터를 학습하고 테스트하는데 사용된다.

#### 3.2 Hidden Markov Model

HMM 모델은 시간적으로 다양한 화면들을 조사하는데 사용 할 수 있다. 다음과 같은 매개변수들을 규정함으로써 HMM의 요소들을 정의하였다. N: 모델에서의 상태의 수. 시점 t 에서 모델의 상태는  $qt$  이고  $1 \leq qt \leq N, 1 \leq t \leq T$  이다. 여기서 T는 출력되는 측정 가능 부호 화면의 길이이다. M: 상태 당 구분되는 측정 가능한 부호들의 수.  $ot$  가 시점 t 에서 각 상태에 대해 측정 가능성이 가장 높은 부호들 중 하나라고 가정할 때,  $0 \leq ot \leq M-1$  이다.  $\pi N$ : N-성분 벡터의 초기 상태 확률을 나타낸다.  $\pi N = \{\pi_i\}$ 이며, 여기서  $\pi_i = P(qt=i), 1 \leq i \leq N$  이다.  $AN \times N$ :  $N \times N$  행렬에서 상태가 상태 i에서 상태 j로 전환하는 상태 전환 확률을 나타낸다.  $AN \times N = \{a_{ij}\}$ 이며 여기서  $a_{ij} = P(qt=j | qt-1=i), 1 \leq i, j \leq N$  이다.  $BM \times N$ :  $M \times N$  행렬이 시스템이 상태 j와 시간 t 에서 측정 가능한 상징 ot를 생성할 것임을 나타낸다.  $BM \times N = \{b_j(ot)\}$ ,  $b_j(ot) = P(Ot=ot | qt=j), 0 \leq ot \leq M-1, b_j(ot) \geq 0$ , 그리고  $\sum_{ot=0}^{M-1} b_j(ot) = 1, 1 \leq j \leq N$  이다.

이산 HMM의 매개변수 집합  $\lambda$  는 하나의 벡터  $\pi$  와 두 행렬 A 및 B 에 의해 표현된다. 비디오 사진을 HMM을 이용하여 보다 정확하게 서술하기 위해서는 HMM 매개변수들을 적절하게 선택할 필요가 있다. 매개변수 선택 작업은 HMM 학습 작업이라고 부른다. 이 매개변수 집합  $\lambda$  는 출력 측정 가능 부호화면 O의 최대 성능을 측정하기 위한  $P(O|\lambda)$ 를 평가하는데 사용될 수 있다. 여기서 T는 각 이미지 화면에 대한 프레임의 수이다. 각  $P(O|\lambda)$ 를 평가하기 위해서 상태의 수 N, 측정 가능한 부호의 수 M (코드북의 크기)을 선택하고, 다음으로 한 집합의 해당 학습 데이터를 이용하여 각 HMM을 학습함으로써 확률 밀도 벡터  $\pi$  와 행렬 A 및 B의 결과를 계산한다. 초기 상태 분포는  $\pi = 1.0$  이다.

HMM 에는 3가지 기본 요소들이 있다: (1) 확률 평가. 하나의 HMM 매개변수 집합  $\lambda=(\pi, A, B)$ 과 출력 측정가능부호화면  $O=(o_1, o_2, \dots, o_T)$ 이 주어진 가운데,  $P(O|\lambda)$ 를 평가한다. (2) 최적 상태 화면. 측정가능부호화면  $O=(o_1, o_2, \dots, o_T)$ 와 HMM 매개변수 집합  $\lambda=(\pi, A, B)$ 가 주어진 가운데, 최적 상태 화면  $q=(q_1, q_2, \dots, q_T)$ 를 결정한다. (3) 매개변수 측정. 출력측정가능부호화면  $O=(o_1, o_2, \dots, o_T)$ 를 생성하는 출력 확률  $P(O|\lambda)$ 을 극대화하기 위해 HMM 매개변수 집합  $\lambda=(\pi, A, B)$ 를 규제한다.

HMM을 비디오 화면분류에 사용하기 위해서 출력 확률  $P(O|\lambda)$ 을 계산한다. 이것을 이용해 HMM 매개변수 집합  $\lambda=(\pi, A, B)$ 가 주어진 가운데 출력 측정가능 부호화면  $O=(o_1, o_2, \dots, o_T)$ 를 생성할 것이다. 이를 계산하는 가장 명료한 방법은 길이 T의 모든 가능한 상태 화면들을 열거하는 것이다. 그렇게 하면 NT개의 가능한 상태 화면 조합이 생기며 여기서 N은 상태의 총수이다. 하나의 상태 화면  $q=(q_1, q_2, \dots, q_T)$ 이 있다고 가정하자. 이 문제를 풀기 위해서는 보다 효율적인 방법이 필요함으로 전·후방향 작업[11]을 사용한다. 정방향 변수는  $a_i(i)=P(o_1, o_2, \dots, o_T, q_i=\lambda|\lambda)$ 로 정의된다. 이것은 HMM 매개변수 집합  $\lambda$ 에 의한 상태 i와 시간 t에서의 부분적 측정가능부호화면  $o_1, o_2, \dots, o_T$ 의 확률이다. 전방향 작업과 후방향 작업 이용하여  $a_i(i)$ 를 구할 수 있다.

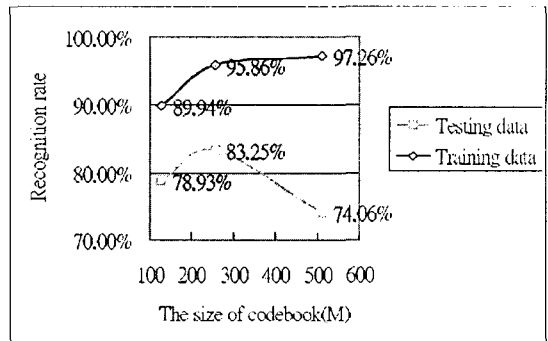
마지막으로, 모델 매개변수들을 조정하고 다른 화면의 측정 가능 부호들을 파악하거나 인식하는데 사용될 수 있는 신호 모델을 구축하는데 학습 측정 가능부호화면 집합이 사용된다. 그러나 부호화면의 확률을 보편적으로 극대화하는 모델 매개변수 집합을 최적화하는 효율적인 방법이 없으므로 반복 절차를 이용하여 가능성 함수  $P(O|\lambda)$ 이 지역적으로 극대화될 수 있도록 최대 확률 모델 매개변수 집합  $\lambda$ 를 선택하는데 Baum-Welch 방식[11]이 사용된다.

#### IV. 실험 결과 및 고찰

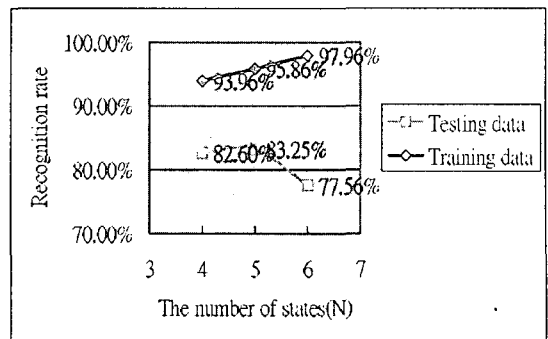
본 논문에서는 메이저리그의 6 경기, 한국 프로 야구의 6 경기에서 추출한 1800개의 비디오 사진들을 사용하였다. 제안된 시스템은 녹화 된 야구 경기기를 여러 다른 사진들로 나누었으며 각 비디오 사진은 다른 프레임 번호를 가질 수 있다고 가정했

다. 또한 동일한 등급에 속한 비디오 사진일지라도 구성 프레임의 수가 다를 수 있다.

각 비디오 사진은 다른 수의 이미지 화면들로 구성될 수 있고, 이미지 화면들의 일부는 학습을 위해 그리고 나머지는 테스트를 위해 사용되었다. 각 전체 칼라 이미지 프레임의 크기는 352x240이며, 프레임률은 초당 29.97프레임으로 하고, 각 비디오 화면의 길이는 1에서 2.5초로 하였다. TV로부터 녹화된 이미지 상에 너무 많은 노이즈와 번짐 효과 없이 빠르게 움직이는 물체를 포착하기 위해 mpeg-1 인코더 카드를 사용하였다.



(a)



(b)

그림 3 (a). 코드북 크기에 따른 인식률  
(b). 상태 수에 따른 인식률

Fig 3 (a). recognition of rate by codebook size  
(b). recognition of rate by state number

실험에서, 하나의 특성 추출 프로세스를 위한 이미지 프로세싱은 프레임 당 약 2초가 소요되었으며, 이 시간은 동작 측정, FFT 및 YIQ 색상 변환 등을 포함한 것이다. 전체적으로 인식 시스템은 약 1.5에서 2.5초가 소요되었다. 사진들의 내용을 추출하고 부호화하기 위해 약 1800개의 실험용 이미지

화면들이 사용하였는데 이들 중 55%는 학습 단계에서, 그리고 그 나머지는 테스트 단계에서 사용되었다. 추출된 정보는 벡터 화면으로 전환되어 다시 학습과 인식 프로세스를 위한 부호들로 양자화되었다.

동일한 사진 등급에 속한 비디오 사진들이라도 서로 다른 사람, 서로 다른 움직임, 그리고 카메라맨의 조작에 따라 다르게 보일 수 있으므로 보다 확실한 분류 시스템을 디자인하기 위해 우리는 각 사진 등급에 대해 가능한 많은 비디오 사진들을 학습 데이터로 선택하였다. 학습이나 인식 프로세스를 위해 HMM을 사용하기 전에, 하나의 특징 벡터가 하나의 측정 가능 부호로 할당되었고. 테스트를 위한 코드북 크기 M은 2의 제곱수로 선택되었다. 그림 3(a)에서 우리는 M=256일 때 최상의 성능을 나타냄을 볼 수 있다. 이러한 학습 부호 화면에 기초하여 사진들을 모델링하기 위한 1차 5 상태 HMM을 효과적으로 생성할 수 있었다. 다른 상태 수의 HMM들(4-상태, 5-상태, 6-상태)을 이용하여 제안된 시스템을 테스트했으며 5-상태 HMM이 최상의 성능을 보임을 확인했다 (그림 3(b)).

실험 결과로부터 우리는 사진 등급 1, 2, 10 그리고 11에서 오차율이 높음을 발견했다. 이것은 그것들의 내용이 서로 유사하기 때문이다. 그것들 중 하나가 그 나머지 다른 것들로 인식될 확률이 높았다. 또한 사진 등급 1과 3은 유사하여 서로를 오인할 확률이 매우 높았다. 동일한 등급에 속한 비디오 사진들이 반드시 유사하지는 않기 때문에 확실한 Hidden Markov Model을 만들기 위해서 이 학습 세트가 가능한 많은 변형들을 모두 포함하기에 충분한 만큼 클 것이 요구되어진다. 본 실험에서 제안된 시스템은 15가지 서로 다른 사진 등급들을 84.72%의 성공률로 인식함을 보였다.

## V. 결 론

본 논문에서는 비디오 검색을 위해서 비디오 사진들을 분석하고 각 사진의 의미를 서술할 필요가 있기에, 각 사진의 의미에 근거하여 분류하는 비디오 사진 분류 시스템을 제안하였다. 제안된 시스템은 3개의 모듈인 특징 추출, 은닉 마르코브 모델 생성, 그리고 비디오 사진 분류로 구성되었다. 제안된 시스템은 15개의 서로 다른 HMM들로 구성되었고, 이것들은 15개의 서로 다른 비디오 사진 등급들을 테스트하는데 사용되었다. 본 논문에서는 15가지 등급의 야구 경기의 장면을 분석 및 요

약하여 분류하는 실험을 한 결과 제안된 시스템은 84.72% 평균 인식률을 나타내었다.

## 참고문헌

- [1] E. Ardizzone, M. L. Casia, and D. Monlinelli. "Motion and color based Video Indexing". ICPR, Vienna, Austria, Aug. 1996.
- [2] R. W. Picard. "A Society of Models for Video and Image Libraries". Per. Comp. Sec Technical 1996.
- [3] S. W Smoliar and Il. Zhang. "Content-Based Video Indexing and Retrieval". IEEE Multimedia, 1(2): 62-72, 1994.
- [4] M. Das, E. Riseman, and B. Draper. "FORCUS: Searching for multi-colored objects in a Diverse Image Database," IEEE CVPR, pp.756-761, 1997.
- [5] Y. Deng, D. Mukherjee, "Object-based Video Representation," SPIE: Storage and Retrieval for Image and Video Databases VI, pp.203-213, 1998.
- [6] A. Ilampapur et al., "Mirage Video Engine," SPIE: Storage and Retrieval for Image and Video Databases V, pp.188-197, 1997.
- [7] J. Ilung, S. R. Kumar, M. mitra, W. Zhu, and R. Zabih, "Image Indexing Using Color Corriograms," CVPR, pp.762-768, 1997.
- [8] W. Niblack et al., "The QBIC Project: Querying Images By Content Using Color, Texture, and Shape," SPIE: Storage and Retrieval for Image and Video Databases, pp.173-181, 1993.
- [9] E. Petrakis and C. Faloutsos, "Similarity Searching in Medical Image Databases," IEEE Transactions on Knowledge and Data Engineering 9(3) pp.435-447, 1997
- [10] B. Ilorn and B. G. Shunck, "Determining optical flow", Artif. Intelli., vol.17, pp.185-203, 1981.
- [11] Rabiner, L. R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, vol.77, No.2, pp.257-285, February 1989.
- [12] M. M. Yeung and B. L. Yeo, "Time-

constrained Clustering for segmentation of Video into Story Units." IEEE ICPR, pp.375-380, 1996.

[13] M. M. Yeung and B. Liu. "Efficient matching and clustering of video shots." IEEE ICIP Vol. I, pp.338-341, 1995.

[14] Y. Linde, A. Buzo, and R. Gray, "An Algorithm for Vector Quantizer Design". IEEE Transaction on Communications, Vol. COM-28, No.1, pp.84-85, January, 1980.

저자소개

박호식(Ho-Sik Park)



1994년 2월 연세대학교 의용전자공학과 졸업(공학사)  
 2001년 2월 관동대학교 대학원 전자통신공학과 졸업(공학석사)  
 2001년 3월~현재 관동대학교 대학원 전자통신공학과 박사과정 재학중

※ 관심분야 : 영상처리, 신호처리시스템, 영상압축

배철수(Cheol-Soo Bae)



1979년 2월 명지대학교 전자공학과 졸업(공학사)  
 1981년 2월 명지대학교 대학원 전자공학과 졸업 (공학석사)  
 1988년 8월 명지대학교 대학원 전자공학과 졸업 (공학박사)  
 1991년 3월~현재 산업자원부 공업기술기반 평가위원 및 심의위원  
 1998년 3월~현재 강원도 정보화위원회 위원  
 1999년 3월~2001년 5월 관동대학교 공과대학 학장  
 2000년 3월~2002년 2월 관동대학교 양양캠퍼스 창업보육센터 소장  
 2001년 6월~2003년 8월 관동대학교 평생교육원장  
 2001년 3월~현재 해양정보통신학회 강원지부장  
 2001년 11월~현재 기초과학연구소 운영위원  
 2002년 5월~현재 중소기업 정보화 혁신 컨소시엄 추진위원  
 2003년 1월~현재 한국통신학회 국내저널 편집부 위원장  
 2003년 1월~현재 대한전자공학회 이사  
 1981년~현재 관동대학교 정보통신공학부 교수  
 ※ 관심분야 : 영상처리, 신호처리시스템, 영상압축