

---

# 골프 동영상에서의 강건한 선수명 인식

## Robust Recognition of a Player Name in Golf Videos

정철곤, Cheolkon Jung\*, 김종규, Joongkyu Kim\*\*

---

**요약** 스포츠의 경기에서 비디오 문자는 득점이나 선수명과 같은 중요한 정보를 제공한다. 본 논문에서는 골프 동영상에서 선수명 정보를 강건하게 인식하는 방법을 제안한다. 골프 경기의 경우, 원하는 선수의 플레이 장면을 검색하고자 하는 요구가 많은 스포츠 종목이다. 이러한 기능을 구현하기 위해 골프 동영상에 포함된 문자 정보를 이용한다. OCR 에 의해 검출된 문자 정보를 인식한 후, 사전 등록된 선수명 DB 를 이용해 선수명 정보를 인식한다. 이렇게 획득된 선수명 정보를 이용해 원하는 선수의 플레이 장면을 검색할 수 있도록 하였다. 다양한 골프 동영상에 대하여 실험을 수행한 결과, 본 논문에서 제안한 방법이 강건하게 선수명을 인식하는 것을 확인하였다.

**Abstract** In sports videos, text provides valuable information about the game such as scores and information about the players. This paper proposed a robust recognition method of player name in golf videos. In golf, most of users want to search the scenes which contain the play shots of favorite players. We use text information in golf videos for robust extraction of player information, By using OCR, we have obtained the text information, and then recognized the player information from player name DB. We can search the scenes of favorite players by using this player information. By conducting experiments on several golf videos, we demonstrate that our method achieves impressive performance with respect to the robustness.

**핵심어:** *Text information, player name, key caption, golf navigation system*

---

\*주저자 : 성균관대학교 정보통신공학부 박사후연구원 e-mail: [ckjung@ece.skku.ac.kr](mailto:ckjung@ece.skku.ac.kr)

\*\*교신저자 :성균관대학교 정보통신공학부 교수; e-mail: [jkkim@skku.edu](mailto:jkkim@skku.edu)

## 1. 서론

근래 들어 사용자가 원하는 시간에 원하는 내용의 콘텐츠를 선택적으로 즐기고자 하는 욕구가 증대되고 있다. 이러한 요구에 부합하여 소니의 로케이션 프리, 슬링 미디어의 슬링 박스와 같은 유비쿼터스 TV 들이 등장하고 있다. 이러한 유비쿼터스 TV 들은 자동으로 스포츠의 하이라이트를 생성해준다든지, 원하는 장면을 쉽게 검색할 수 있도록 해주는 멀티미디어 콘텐츠 분석 기술들을 탑재하고 있다. 스포츠 동영상의 경우, 이러한 요약이나 하이라이트 생성, 검색 기술이 필요한 장르로서 활발하게 이에 대한 연구가 진행되어 오고 있다 [1-7].

Ekin [4] 은 스포츠 비디오에서 경기장의 색상을 이용하여 각 장면을 경기장면과 경기 외 장면으로 구분하는 방법을 제안하였으며, 또 이를 이용하여 축구와 농구 동영상을 요약하는 방법을 제안하였다. 이 방법은 경기 외 장면을 제외한 경기 장면으로 이루어진 요약 동영상을 생성하였으나, 선택된 경기 장면 가운데 어떤 장면이 중요한 장면인지에 대한 고려가 되어 있지 않다. Pan [5] 은 스포츠 동영상에서 중요장면은 슬로우 모션으로 반복하여 보여진다는 점을 이용하여 슬로우 모션을 검출하고 주변의 경기장면을 중요장면으로 검출하는 방법을 제안하였다. 그러나 이 방법에서 역시 중요한 이벤트의 검출에는 효과적이지만, 그 이벤트의 의미에 대해서는 아무런 정보도 제공하지 못한다. Zhang [2, 6] 은 문자정보를 이용해 야구 동영상을 요약하는 시스템을 개발하였다. 이 방법은 문자 정보를 추출하여 야구 경기의 요약에 활용하였다. 고속의 비디오 문자 정보 추출을 위해 비압축 영역의 특징치인 discrete cosine transform (DCT) 계수와 모션 벡터를 이용하였다. 이로 인해 득점이나 이닝 정보 등을 자동으로 추출함에 의해 장면에 대한 의미정보를 부여해 줄 수 있는 방법이다. 정(Jung) [7]은 득점 정보를 이용하여 농구 동영상을 요약하는 방법을 제안하였다. 농구 경기의 경우, 득점 장면이 많이 나타나기 때문에 득점자체로서는 하이라이트 장면이라 보기 어렵다. 이에 득점정보로부터 역전이나 추격, 점수 등의 장면에 중요도를 부여함에 의해 하이라이트를 구성하였다.

골프 경기의 경우, 원하는 선수의 플레이 장면을 검색하고자 하는 요구가 많은 스포츠이다. 예를 들어, 사용자들은 골프 경기에서 타이거 우즈, 아니카 소렌스탐, 박세리 선수들의 play 장면만을 선택해서 보고 싶어한다. 더 나아가, 특정 선수의 퍼팅 샷, 드라이버 샷, 티 샷으로 분류해서 보는 것을 사용자들은 원한다. 본 논문에서는 이러한 기능을 가능하게 하기 위해 스포츠 동영상 내에 포함된 문자 정보를 인식하여 원하는 선수의 play 장면을 검색할 수 있도록 하는 방법을 제안하였다.

## 2. 본론

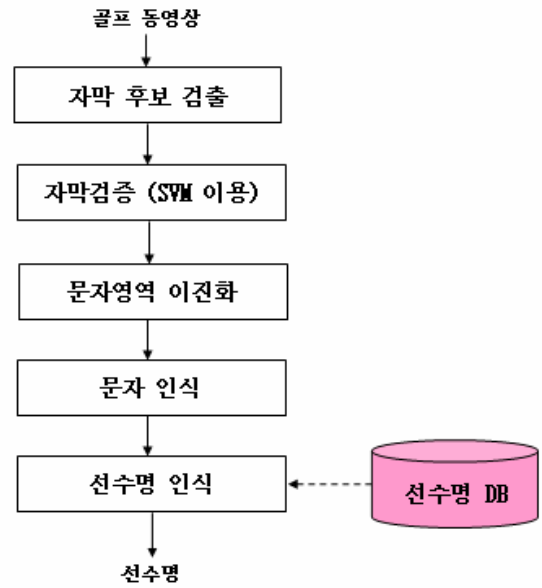


그림 1. 제안된 선수명 검출 방법

본 논문에서 제안된 방법은 장면들로부터 선수명을 강건하게 인식하기 위해 자막 검증 모듈과 자막 인식 결과를 이용하였다. 제안된 선수명 인식 방법은 그림 1 과 같이 자막후보검출단계, SVM 분류기에 의한 자막검증단계, 문자영역 이전화 단계, 문자인식단계, 선수명 인식단계로 구성된다. 여기서 입력은 골프 동영상이며, 출력 결과는 각 장면에서 인식된 선수명이다.

### 2.1 자막후보검출

입력 영상은 골프 동영상의 스트림 상에서 얻은 영상인데, 구체적으로 각 장면의 대표 영상 (key frame) 이다. 고속의 자막후보 검출을 위해 에지 정보를 이용하였다. 자막 후보 검출 방법은 에지 맵 구성 단계, 자막후보영역 스캐닝 단계, 자막후보영역 생성단계로 구성된다. 에지 맵은 소벨 에지 검출기에 구성되며, 구성된 에지 맵에 대해 고정된 사이즈의 윈도우(예, 8x16)를 이용해 자막후보영역을 스캐닝한다. 이렇게 스캐닝 된 영역에 대해 연결 요소 분석 (connected component analysis) 을 통해 자막후보영역을 생성하게 된다.

### 2.2 SVM 에 의한 자막검증

자막 검증 단계는 검출된 자막후보 영역이 자막영역인지 아닌지를 검증하는 단계이다. 자막후보에 대한 검증은 support vector machine (SVM) 분류기에 의해 수행된다. 이렇게 자막후보 검증과정을 둔 이유는 비 자막영역의 인식에 따른 처리시간을 최소화 하기 위함이다. 자막후보 영역에 대한 검증과정은 다음과 같다. 2.1 절의 출력결과인 자막후보영역은 윈도우의 크기에 의해 문자영역 이외에도 배경영역이 많이 포함되어 있다.

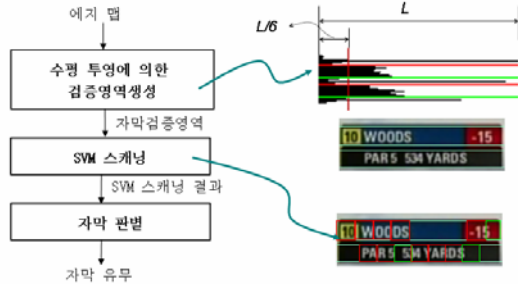


그림 2. SVM에 의한 자막 검증

따라서 이러한 문제점을 해결하기 위해 그림 2와 같이 검증영역 결정단계를 두어 에지의 수평 투영을 통해 검증 단계를 위한 검증영역을 생성한다. 검증은 SVM 분류기에 의한 스캐닝을 통해 수행된다. 이때, 입력 특징치는 정규화된 그레이 값이며, 자막영역 판별 정확도는 약 96.7%이다 [8]. 또한 그림에서와 같이수평 투영한 픽셀의 개수에 대한 최대값을  $L$  이라고 하면, 임계값은  $L/6$  로 설정한다.

이렇게 영역이 설정되면 SVM 스캐닝을 수행하게 되는데, 이때 높이를 15 pixel 로 정규화하고, 15x15 윈도우에 대하여 SVM 분류기에 의해 판정을 수행한다. SVM 스캐닝 시, 각 윈도우의 에지 평균이 큰 경우에 대해서만 SVM 판정을 수행한다. 그림 6의 우측 최하단의 그림은 SVM 스캐닝에 의한 결과로서, 적색 박스는 허용된 (accepted) 박스를 나타내고, 녹색 박스는 거절된 (rejected) 박스를 나타낸다 (박스의 색은 원본 파일 참조). 여기서 허용 (accepted)으로 판정된 윈도우의 개수가 5 개 이상인 경우, 자막영역으로 간주한다.

### 2.3 문자이진화

이렇게 자막의 유무가 판정되면 자막영역에 대하여 문자 이진화를 수행한다. 그런데 자막영역 내부에는 컬러 극성이 상반된 문자들이 함께 존재하고 있다. 이에 문자 이진화 과정은 그림 3과 같이 문자의 컬러극성이 다른 경우에도 이진화가 가능하도록 dual binarization 방법을 제안하였다. 그림에서 Otsu 방법 [9]에 의해 결정된 임계치보다 작은 경우 1을, 큰 경우는 0을 부여하여 첫 번째 이진화 영상을 생성한다. 반대로 임계치보다 큰 경우 1을, 작은 경우는 0을 부여하여 두 번째 이진화 영상을 생성한다. 그리고 두 이진화 영상에 대해 문자크기에 대한 제한 조건을 두어 잡음성 영역을 제거한 후 결합하여 최종 이진화 영역을 결정한다. 이 방법은 문자의 특징을 이용하여 제한 조건을 둬으로써 문자의 컬러 극성이 다른 경우에도 분할이 가능한 방법이다.

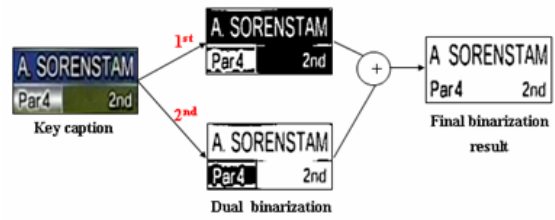


그림 3. Dual binarization 방법

### 2.4 문자인식

문자영역을 OCR (optical character recognizer)에 입력하기 위해서는 먼저 라인 단위의 문자영역을 생성해야 한다. 대체적으로 OCR은 라인단위의 문자영역을 입력으로 하기 때문이다. 기존에 사용되던 방법인 수평투영에 의한 방법은 문자라인이 서로 중첩된 경우, 라인분리를 할 수 없다. 이에 본 논문에서는 수평투영에 의한 방법을 사용하지 않고 CCA (연결 요소 분석, Connected Component Analysis)를 수행하여 라인 단위의 문자영역을 생성하여 문자 인식을 수행한다. 이렇게 OCR에 의해 인식결과를 얻게 되면 유사 단어 보상을 수행하게 된다. 실제로 인식된 결과를 보면 문자 'O'를 숫자 '0'으로, 문자 'l'을 숫자 '1'로, 문자 'g'를 숫자 '9'로 오인식 하는 경우가 자주 발생한다. 따라서 이에 대한 보상을 유사 단어 보상을 통해 해결한다.

### 2.5 선수명 인식

2.4 절의 인식 결과로부터 선수명을 인식하는 과정은 다음과 같다. 인식결과를 선수명 DB와 비교하여 가장 유사도가 높은 선수를 선택한다. 유사도의 기준은 워드 단위의 스트링 매칭 방법을 이용해 측정한다. 또한 선수명 DB는 EPG, 웹정보, 또는 콘텐츠에 포함된 리더보드(leader board) 자막을 이용해 구성된다. 워드 단위의 스트링 매칭은 full name matching과 family name matching의 순서로 수행한다.

## 3. 실험 결과

실험은 Intel Pentium Processor 2.40GHz와 512M RAM에서 수행되었으며, 실험 동영상은 스포츠 채널로부터 얻은 골프 경기의 mpeg2 동영상(총 6경기 9.5시간)을 사용하였다. 그림 4는 본 실험에 사용된 골프경기 요약 시스템의 유저 인터페이스이다. 그림에서 보는 바와 같이 좌측 상단은 동영상이 재생되는 곳이며, 우측의 'caption' 메뉴에서는 자막 판정에 의한 결과를, 'Name' 메뉴에서는 선수명 인식결과를 표시하였다.

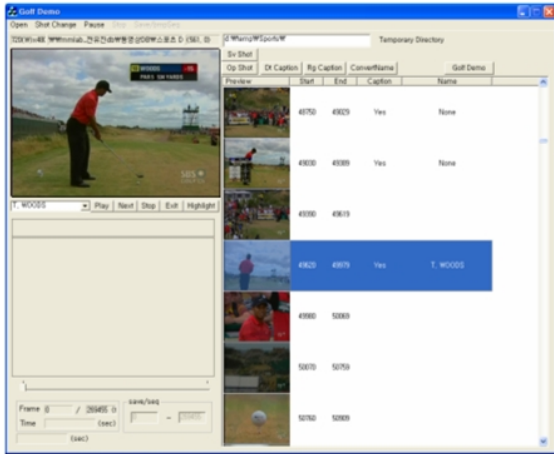


그림 4. 유저 인터페이스

표 1. 성능 평가

제목	재현율	정확율	처리시간	길이
LPGA 투어_041029	92/111	92/93	42.8s	1.5h
LPGA 투어_050105	54/70	54/54	36.5s	1h
CJ 나인브릿지 0401	60/66	60/60	30.2s	1h
브리티시_2005	40/61	40/42	51.7s	1.5h
브리티시_2006	65/76	65/71	94.4s	3h
CJ 나인브릿지 0402	104/108	104/104	64.2s	1.5h
<b>평균</b>	<b>84.3%</b>	<b>97.8%</b>	<b>33.7s</b>	<b>1h</b>

선수명 DB 의 경우, 아래와 같이 지명도가 높은 20 명 가량의 남성/여성 선수들을 중심으로 구성하였다. 다음은 선수명 DB 에 포함된 선수명 목록이다.

남성 선수={"T. WOODS", "V. SINGH", "P. MICKELSON", "C. MONTGOMERIE", "J. OLAZABAL", "M. CAMPBELL", "S. GARSIA", "C. DIMARCO", "E. ELS", "J. FURYK", "A. CABRERA", "H. TANIHARA", "R. GOOSEN", "N. FALDO", "S. KATAYAMA", "G. OWEN"};

여성 선수={"A. SORENSTAM", "GRACE PARK", "K. STUPPLES", "J. MCGILL", "C. KOCH", "SHIHYUN AHN", "L. OCHOA", "CHRISTINA KIM", "JUMI KIM", "HEEWON HAN", "SERI PAK", "MICHELLE WIE", "BOBAE SONG", "C. MATTHEW", "MIHYUN KIM", "C. KERR", "L. DAVIES", "B. JONES", "K. GOLDEN"};

표 1 은 실험 동영상에 대한 성능 평가 결과이다. 여기서 길이는 동영상의 전체재생시간을 의미한다. 실험결과, 재현율은 약 84.3%, 정확율은 97.8 % 의 성능을 나타낸다 [7]. 또한 처리시간의 경우, 자막검증 및 선수명 인식에 걸린 시간이 1 시간 콘텐츠 당 33.7 sec 로 나타났다.

#### 4. 결론

본 논문에서는 동영상에 포함된 문자정보를 이용해 골프 동영상에서의 선수명을 인식하는 방법을 제안하였다. 선수명을 강건하게 인식하기 위해 SVM 에 의한 자막 검증 과정과 OCR 에 의한 자막 인식 결과를 이용하였다. 이렇게 인식된 선수명 정보는 골프 동영상에 대한 선수명 단위의 요약 및 검색 서비스를 가능케 한다. 뿐만 아니라, 자막에 포함되어 있는 다른 경기 정보들을 인식할 경우에 보다 양질의 경기정보를 제공해 줄 수 있다. 예를 들어, 홀 정보, 장면의 종류에 대한 정보, 스코어에 대한 정보 등이 그 예일 것이다.

#### 참고문헌

[1] M. Xu, N.C. Maddage, C. Xu, M. Kankanhalli, Q. Tian, "Creating Audio Keywords for Event Detection in Soccer Video," IEEE ICME 2003, Vol. 2, pp. 281-284, July 2003.

[2] D. Zhang, S. F. Chang, "Event Detection in Baseball Video Using Superimposed Caption Recognition," ACM Multimedia 2002, pp. 315 - 318, Dec. 2002.

[3] B. Li, I. Sezan, "Semantic Sports Video Analysis: Approaches and New Applications", IEEE ICIP 2003, Vol. 1, pp. 17-20, Sept. 2003.

[4] A. Ekin, A.M. Tekalp, "Shot Type Classification by Dominant Color for Sports Video Segmentation and Summarization," IEEE ICASSP 2003, Vol. 3, pp. 173-176, 2003.

[5] H. Pan, P. Beek, M.I. Sezan, "Detection of Slow-Motion Replay Segments in Sports Video for Highlights Generation," IEEE ICASSP 2001, Vol. 3, pp. 1649-1652, 2001.

[6] D. Zhang, S. F. Chang, "General and Domain-specific Techniques for Detecting and Recognizing Superimposed Text in Video," IEEE ICME 2002, Vol.1, pp. 593-596, 2002,

[7] 정철곤, 김의진, 이광국, 김희율, "스코어 정보를 이용한 농구 비디오의 자동요약," 한국통신학회 논문지, 제 32 권, 제 9 호, 881-887, 2007.

[8] Cheolkon Jung, Qifeng Liu, Joongkyu Kim, "Accurate Text Localization based on SVM Output Score," Submitted to Image and Vision Computing, 2007.

[9] N. Otsu, "A Threshold Selection Method from Gray-scale Histogram," IEEE Trans. on Systems Man Cybernet., Vol. 9, pp. 62-66, 1979.