동적 베이지안 네트워크를 이용한 다중 카메라기반 축구 비디오 요약

Summarization of Soccer Video based on Multiple Cameras Using Dynamic Bayesian Network

민준기, Jun-Ki Min *, 박한샘, Han-Saem Park**, 조성배, Sung-Bae Cho ***

요약 스포츠 경기의 비디오 중계는 생동감 있고 흥미로운 장면들을 시청자에게 제공해주기 위하여 여러 대의 카메라를 사용한다. 하지만 기존의 방송 시스템은 시청자에게 하나의 비디오로 편집된 장면만을 제공하기 때문에 시청자의 관심도를 고려하여 특정 장면을 요약해주거나 검색해주는 등의 지능형 방송 서비스가 어렵다. 본 논문에서는 여러 대의 카메라로 촬영한 축구경기 비디오를 요약 및 검색해주는 시스템을 제안한다. 이는 비디오에 주석으로 태깅되어있는 저수준 정보를 기반으로 하는 동적 베이지안 네트워크를 이용하여 슛, 크로스, 반칙, 세트플레이 등과 같은 주요장면을 추출하고, 해당 주요장면타입에 따라 자동으로 뷰를 선택한다. 따라서 제안하는 시스템은 사용자에게 주요장면 요약이나 선호하는 뷰의 선택기능을 제공하며, 사용자의 선호도를 고려할 경우 개인화 방송 서비스를 제공해줄 수 있다.

Abstract Sports game broadcasting system uses multiple video cameras in order to offer exciting and dynamic scenes for the TV audiences. Since, however, the traditional broadcasting system edits the multiple views into a static video stream, it is difficult to provide the intelligent broadcasting service that summarizes or retrieves specific scenes or events based on the user preference. In this paper, we propose the summarization and retrieval system for the soccer videos based on multiple cameras. It extracts the highlights such as shot on goal, crossing, foul, and set piece using dynamic Bayesian network based on soccer players' primitive behaviors annotated on videos, and selects a proper view for each highlight according to its type. The proposed system, therefore, offers users the highlight summarization or preferred view selection, and can provide personalized broadcasting services by considering the user's preference.

핵심어: Multiple video cameras, intelligent broadcasting service, soccer video summarization, dynamic Bayesian network

본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음.

*주저자 : 연세대학교 컴퓨터과학과 박사과정; e-mail: loomlike@sclab.yonsei.ac.kr

**공동저자 : 연세대학교 컴퓨터과학과 박사과정; e-mail: sammy@sclab.yonsei.ac.kr

***교신저자 : 연세대학교 컴퓨터과학과 교수; e-mail: sbcho@cs.yonsei.ac.kr

1. 서론

최근 방송기기의 발달로 많은 수의 비디오카메라를 동시에 이용하여 촬영하는 다중 카메라 방송 시스템이 널리 보급되었다. 이는 시청자에게 생동감 있고 흥미로운 장면을 보여줄 수 있어 특히 스포츠 경기 중계에서 많이 활용되고 있다. 또한 TV-anytime 과 같은 동영상 주석 표준이 제정되어 방송의 개인화나 요약, 내용 검색 등과 같은 지능형 방송 서비스에 대한 관심이 높아지고 있다. Huang 등은 영상처리를 통해 색상, 화면에서의 움직임, 텍스쳐, 등의 특징을 추출하였으며, 이를 기반으로 하는 동적 베이지안 네트워크(Dynamic Bayesian network, DBN)를 설계하여 골, 경고 및 퇴장 등 다양한 축구 경기의 이벤트를 인식하였다[1]. Tjondronegoro 등은 특정 이벤트의 인식 없이 심판의 휘슬소리, 관중의 함성소리, 영상에 보여지는 텍스트 정보 등을 이용하여 바로 주요장면(Highlight)을 선택하였다[2].

앞에서 소개한 연구들은 기존 방송 시스템을 가정하였는데, 이는 PD(Program director) 등과 같은 전문가가 다중 카메라로부터 촬영된 영상들을 하나의 정적 비디오 스트림으로 편집하여 사용자에게 제공해주는 방식으로, 시청자의 관심도를 고려하여 특정 장면을 요약해주거나 검색해주는 등의 지능형 방송 서비스가 어렵다. 따라서 본 논문에서는 다중 카메라로 촬영한 모든 영상들을 동적으로 편집하여 보여주는 시스템을 제안한다.

2. 배경 연구

최근 멀티미디어 데이터가 크게 늘어나면서, 다량의 정보 가운데 사용자가 필요로 하는 정보를 요약하거나 검색하는 것을 돕는 연구가 많이 수행되고 있으며, 특히 축구를 비롯한 스포츠 동영상을 대상으로 하는 연구가 활발히 수행되고 있다. A. Ekin 등은 영상 처리를 통해 축구 경기에서 골 장면, 심판, 페널티 박스 등을 인식하고, 이를 기반으로 주요 장면을 요약, 분석하였으며[3], Z. Xiong 은 coupled HMM(hidden Markov model)을 이용하여 골프 동영상에서 하이라이트를 추출하고 간단한 요약을 제공한다[4]. M. Albanese 등은 priority curve 알고리즘(PriCA)을 이용하여 50 개의 축구 비디오를 요약한 뒤 200 명에 대해 사용성 평가를 수행하였다[5].

대부분의 요약 및 검색 연구는 다중 카메라가 아닌 하나의 카메라로 촬영한 비디오 또는 이미 하나로 편집된 방송 경기 영상을 대상으로 하며, 본 논문처럼 다중 카메라 축구 경기를 대상으로 한 연구는 없다. 스포츠 비디오가 아닌 홈 환경이나 사무실 환경과 같은 실내 환경을 대상으로 한 경우 다중 카메라 비디오 영상을 이용하여 데이터를 수집하고, 간단한 요약 및 검색 서비스를 제공한 연구가 있었으나, 하나의 이벤트를 위한 다양한 카메라 뷰의 장점을 활용한 연구가 아닌 넓은 범위를 커버하기 위해 다중 카메라를 활용한 연구로 다중 카메라의 활용 목적이 다르다[6]. 본 논문은 다중 카메라를 활용하여 수집한 축구 비디오를 대상으로 하며, 제안하는 시스템은 동일한 화면에 대한 다양한 카메라 뷰를 제공하여 그 중 상황에 맞는 또는 사용자가 원하는 뷰를 선택할 수 있는 장점을 갖는다.

3. 제안하는 시스템

본 논문에서 제안하는 방법은 크게 DBN 을 이용하여 주요장면을 추출하는 부분과 최적의 뷰를 선택하는 부분의 두 파트로 구성된다. 그림 1 은 제안하는 방법의 전체 흐름을 나타낸다.

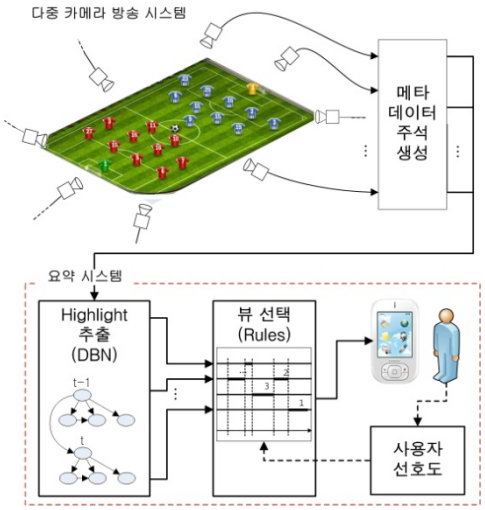

그림 1. 제안하는 시스템

본 논문에서 가정한 멀티 카메라 환경은 축구공을 찍는 6 대의 카메라 (경기장의 좌측골대에 1 대, 우측골대에 1 대, 상측과 하측에 각각 2 대씩)와 선수 각각을 찍는 22 대의 카메라(11 명의 선수×2 팀)로 총 28 대의 카메라를 사용한다. 표 1 은 각 카메라의 ID 와 시점, 줌(Zoom)거리를 나타낸다.

표 1. 멀티 카메라 시스템의 각 카메라 정보

카메라 ID	시점	거리
View00	Ball (Bottom)	Far
View01	Ball (Bottom)	Near
View02	Ball (Left)	Far
View03	Ball (Right)	Far
View04	Ball (Top)	Far
View05	Ball (Top)	Near
View06~16	Away team player 1~11	Near
View17~27	Home team player 1~11	Near

본 논문에서는 축구경기에서의 주요장면을 슛, 크로스, 득점, 경고, 공격진영에서의 패스, 공격진영에서의 드리블, 세트플레이, 역습의 8 가지로 정의하였다. 이를 자동으로 인식하기 위하여 먼저 이들의 기본 행동이 되는 저수준 행동(Primitive behavior)을 표 2 와 같이 정의하였다.

표 2. 축구 경기에서의 저수준 행동정보

저수준 행동	설명
Kick	공을 가지고 있는 선수가 공을 참
Receive	공을 가지고 있지 않은 선수가 공을 받음
Dribble	공을 가지고 있는 선수가 움직임
Intercept	공을 가지고 있지 않은 선수가 공을 가지고 있는 상대편 선수로부터 공을 뺏거나 찬 공을 막음 (공의 소유를 가져옴)
Block	공을 가지고 있지 않은 선수가 공을 가지고 있는 상대편 선수로부터 공을 뺏거나 찬 공을 막음 (공의 소유를 가져오지 못함)
Compete	공을 가지고 있지 않은 양팀의 선수 두명이 공을 다툼
Mix-up	공을 가지고 있지 않은 양팀의 선수 세명 이상이 공을 다툼
Goal	공이 골대에 들어감
Out	공이 아웃라인을 벗어남
Foul	경기장 내에서 경기가 중단되고 공이 멈춘 상태로 다시 시작함

본 논문에서는 이들 저수준 행동정보가 미리 비디오에 태깅되어있는 것을 가정하였으며, 따라서 이를 수동으로 태깅하기 위한 어노테이션 툴을 C++(Microsoft Visual Studio 2008, MFC)기반으로 그림 2 와 같이 제작하였다. 저수준 행동정보의 태깅은 표 2 의 정보 외에도 운동장에서의 선수의 위치를 9 개의 영역으로 구분하여 함께 태깅하였으며, 이는 해당 행동의 시작 프레임, 끝 프레임과 함께 저장된다.

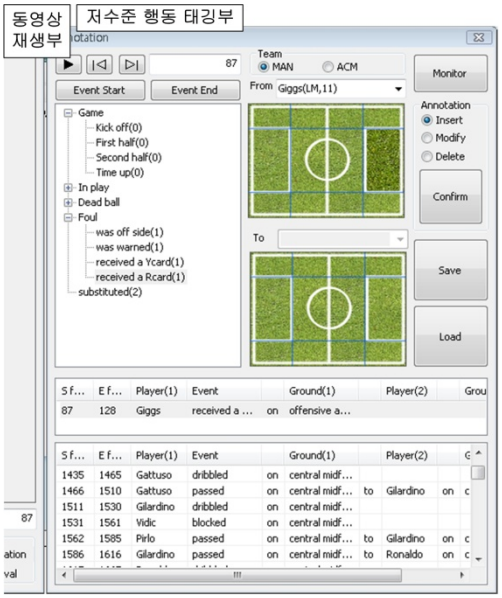

그림 2. 저수준 행동정보 태깅을 위한 어노테이션 툴

앞에서 태깅한 저수준 행동정보를 증거값으로 사용하여 8 가지 주요장면을 인식하는 DBN 을 설계하였다. 베이지안 네트워크는 확률기반 추론모델로, 불확실한 증거값을 처리할 수 있고, 모델의 구조의 설계나 확률값 설정에 전문가 지식을 반영하기 쉽다는 장점이 있으며, DBN 은 시간정보를 추가로 활용한 모델이다 [7]. 그림 3 은 제안하는 시스템을 위해 설계한 DBN 의 예를 보여준다. 각각의 모델 설계에 앞서 해당 주요장면의 요소(주요장면을 구성하는 저수준 행동, 위치정보 등)를 노드로 정의하고, 이들간의 인과관계, 시간관계를 고려하여 엣지를 연결한다.

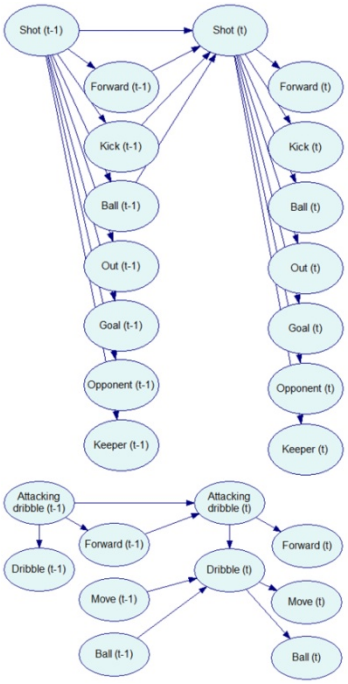

그림 3. 주요장면 인식을 위한 동적 베이지안네트워크 (슛과 공격진영에서의 드리블의 예).

주요장면을 인식한 후에는 식 (1)을 이용하여 장면 세그먼트별로 사용자의 관심도(Degree of interest, DOI)를 계산한다. 이때, i 는 각 세그먼트에 속하는 장면, 선수, 팀의 정보를 나타내며, w_i 는 해당 정보의 중요도이다. 장면 중요도는 도메인에 따라 미리 정의되어있는 값이며, 선수나 팀 중요도는 사용자가 본인의 선호도에 따라 입력한다. 요약 시 특정 뷰가 오래 선택되면 지루할 수 있기 때문에 식 (1)에 시간 t 에 대한 상수 c를 두어 뷰의 관심도를 조절한다.

$$DOI_S = \sum_{i \in S} w_i + tc \qquad (1)$$

4. 실험 및 평가

제안하는 시스템을 구현 및 평가하기 위하여 다중카메라 환경에서 수집한 데이터가 필요하지만, 실제 환경에서 수집된 축구경기 데이터는 획득하기 어렵기 때문에 본 논문에서는 PC 게임(EA Sports 사의 FIFA08)을 이용하여 가상의 축구경기 동영상 데이터 셋을 수집하였다. 데이터는 전반과 후반을 포함한 10 분 분량이며, 앞서 설명한 대로 28 대의 카메라를 가정하여 각각의 뷰로 재생한 화면을 캡쳐하였다. 그림 4 는 제안하는 시스템의 시연을 위한 플레이어를 나타내는 것으로, 주요장면의 요약, 검색, 선택이 가능하다. 그림 5 는 제안하는 시스템을 이용하여 뷰를 자동으로 선택한 결과를 나타내는 것으로, 드리블(그림 5 의 아래 왼쪽에서 두번째, View01)이나 슛(그림 5 의 아래 가장 오른쪽 2 개, View13, View03) 등 주요장면 타입에 따라 다양한 뷰가 선택된 것을 확인할 수 있다.

그림 4. 주요장면의 요약, 검색, 선택 기능을 제공하는 시스템 데모

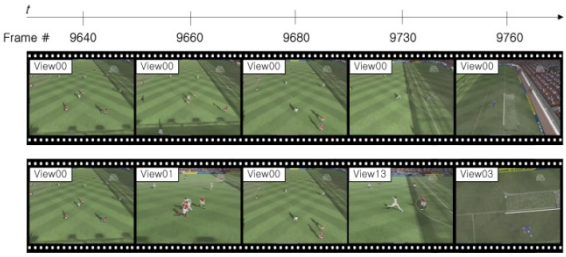

그림 5. 주요장면의 요약 시 뷰 고정(상)과 뷰 자동선택(하)의 예.

본 논문에서는 구현된 어플리케이션 시스템을 이용한 뷰 선택 결과와 선수 및 골 검색 결과를 보여준 뒤 시스템의 사용성 평가를 위해 설문 조사를 수행하였다. 설문조사는 사용성 평가를 위해 널리 사용되는 SUS (System Usability Scale) 10 문항이 사용되었다. 각 문항의 응답은 Likert 척도를 사용하여 강한 부정, 부정, 보통, 긍정, 강한 긍정을 의미하는 1 에서 5 까지 5 개의 답 중 하나를 선택하도록 하였으며 10 명의 사용자를 대상으로 하였다. SUS 는 이 평가결과를 바탕으로 0~100 사이의 점수로 환산할 수 있는 방법을 제공하는데, 사용성 평가 결과는 그림 6 과 같다. 환산 결과가 67.5 ~ 95 점 사이에 분포하며, 평균 80 점으로 시스템의 사용성이 높다고 해석할 수 있다.

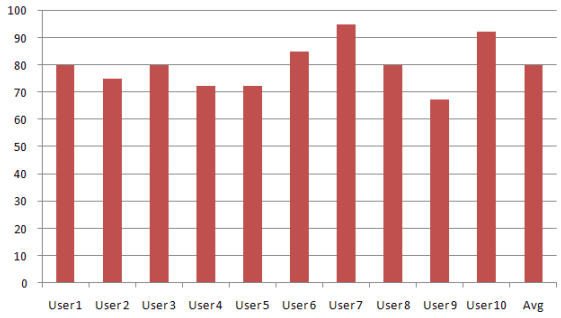

그림 6. 제안하는 시스템의 사용성 평가결과

5. 결론

본 논문에서 제안하는 시스템은 사용자의 선호도나 검색 질의에 맞게 비디오 영상을 동적으로 편집하여 제공해준다. 사용자가 스포츠 중계방송을 시청 할 때 특정 선수만을 보여주는 카메라 뷰를 선택하여 볼 수 있으며, 관심 있는 이벤트를 검색하거나 본인의 취향에 맞게 요약된 동영상을 시청할 수도 있다. 이를 이용하면 다양하면서 새롭게 구성되는 비디오 영상으로 지능형 개인화 방송을 제공하여 시청자의 흥미를 지속적으로 유지시킬 수 있다. 향후 연구로는 영상분석을 통한 저수준 행동정보의 자동 추출 및 제안하는 시스템의 모바일 버전 구현 등을 진행할 계획이다.

참고문헌

[1] C.-L. Huang, H.-C. Shih and C.-Y. Yao, "Semantic analysis of soccer video using dynamic Bayesian network," IEEE Transactions on Multimedia, vol. 8, no. 4, pp. 749-760, 2006.

[2] D. Tjondronegoro, Y. P. Chen and B. Pham, "Sports video summarization using highlights and play-breaks," ACM Workshop on Multimedia Informatioin Retrieval, pp. 201-208, 2003.

[3] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," IEEE Transactions on Image Processing, vol. 12, no. 7, pp. 796-807, 2003.

[4] Z. Xiong, "Audio-visual sports highlights extraction using coupled hidden Markov model," Pattern Analysis Application, vol. 8, pp. 62-71, 2005.

[5] M. Albanese, M. Fayzullin, A. Picariello, and V. S. Subrahmanian, "The priority curve algorithm for video summarization," Information Systems, vol. 31, pp. 679-695, 2006.

[6] G. C. Silva, T. Yamasaki, and K. Aizawa, "An interactive multimedia diary for the home," IEEE Computer, vol. 40, no. 5, pp. 52-59, 2007.

[7] K. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning," PhD thesis, University of California Berkley, 2002.