

사전 지식을 이용한 축구 경기장면 분석 및 좌표 변환

윤 호섭(尹 皓燮), 소 정(蘇 楨), 민 병우(閔 丙雨), 양 영규(梁 英奎)

한국전자통신연구원 컴소연 영상처리부

전화 : (042) 860-5233 / 팩스 : (042) 860-4844

Soccer Scene Analysis and Coordinate Transformation using a *priori* Knowledge

Ho-Sub Yoon, Jung Soh, Byung-Woo Min and Youngkyu Yang

Image Processing Div. / Computer Software Technology Lab. in ETRI

E-mail : yoonhs@etri.re.kr

Abstract

This paper presents a method for soccer scene analysis and coordinate transformation from scene to ground model using a *priori* knowledge. First, the ground and spectator regions are separated, and various objects are extracted from the separated ground region. Second, an affine model is used for mapping the object positions on the soccer image into the position on the ground model. Problems regarding holes arising from mapping processing are solved using inverse mapping instead of a usual interpolation method. Experiments are performed on a PC using about 100 RGB images acquired at 240*640 resolution and 3~5 frames per second.

1. 서론

스포츠는 우리 일상 생활의 한 부분으로서 아주 밀접한 관계를 가지고 발전해 왔다. 특히 축구 경기는 2002년 월드컵을 한일 양국이 주최하게 됨으로써 일반 대중의 큰 관심을 끌고 있으며, 이에 따라 경기를 시청하는 시청자에게 더 많은 정보를 제공하기 위해 방송국에서는 다양한 첨단 방송 기술을 개발하고 있다. 이와 더불어 축구 경기를 컴퓨터를 이용하여 자동으로 분석하고자 하는 시스템의 개발이 다음과 같이 시작되었다.

먼저 MIT Media Lab[1]에서는 Computer Watching Football이라는 프로젝트를, Princeton 대학의 Yow[2]

는 축구 영상에 모자이크 기법을 적용하여 하나의 큰 파노라마 영상을 구성하는 방법을 제안하였다. 또한 일본의 Hasegawa 대학의 Taki[3]는 하나의 카메라를 이용하는 대신 여러대의 고정된 카메라로부터 얻은 축구 경기 장면으로부터 선수들과 팀의 경기력을 분석하기 위한 시스템을, 독일에서는 VITRA (Visual Translator)라는 장기적 프로젝트를 통해 연구를 수행하고 있는데 이 연구에서 SOCCER, REPLAY 같은 축구 경기 분석 시스템을 개발하였다[4].

이들 시스템은 기본적으로 영상 분석 기법을 운동 경기에 적용하여 선수, 경기장, 공 등의 경기 장면에서 나타나는 모든 개체들을 검출하고 이들의 이동 궤적 등의 정보를 추출하고 분석하여 의미 정보로 변환한다. 이러한 시스템의 적용은 팀의 경기력을 분석하여 코치에게 도움을 주는 시스템을 개발하거나, 전체 경기 장면 중 원하는 중요 경기장면만을 추출하는 Hilight 추출 시스템, 경기의 자동 해설 및 그래픽을 이용한 3D 3차원 Repaly 시스템, 혹은 파노라마 뷰를 이용한 Virtual Replay 시스템 등의 개발이 가능하다.

본 연구에서는 이와 같이 축구 경기를 자동으로 분석하기 위한 기반 연구로서 경기장내에 나타나는 각종 개체들을 추출하여 분석하고, 이 정보를 이용해 경기 장면에서의 위치를 축구 경기장 모델의 위치로 변환하는 방법을 기술한다. 즉, 첫 번째 단계에서 경기 장면을 분석하여 경기장 및 관중석을 분리하고, 경기장 안에서 각종 선수 및 선들을 추출한다. 두 번째 단계는 추출된 정보를 이용하여 운동장 모델의 4점과 매핑하여 운동장 모델의 좌표로 변환한다.

2. 경기 장면 분석

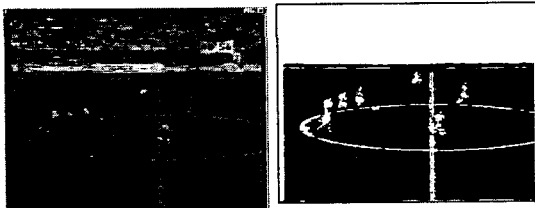
2.1 경기장 및 관중석 분리

축구 경기 장면에서 경기장과 관중석을 분리 할 수 있는 가장 중요한 요소는 컬러 정보이다. 즉, 경기장을 분리하기 위해선 주로 컬러 정보를 이용하게 되는데 이는 일반적으로 경기장의 색상은 Green 컬러로 구성되어 있다는 사실에 기반을 둔다. 이는 분석하고자 하는 경기 장면이 월드컵과 같이 잘 정돈된 그라운드일 경우 확실한 값으로 사용될 수 있다. 이에 따라 RGB로 구성된 각 화소 중 Green값이 Red나 Blue보다 강하게 나타난 화소는 경기장이 될 확률이 높다. 또한 TV 화면에서 경기장 영역이 관중석 영역보다 더 많이 포함된다는 정보를 사용하여 전체 영역의 RGB 히스토그램의 Peak값을 이용하면 비교적 정확히 경기장 영역만을 추출할 수 있다.

즉, 전체 입력영상에서 다음과 같은 공식을 사용하여 경기장 및 관중석을 분리하였다.

$$O(x, y) = \begin{cases} 1: & \begin{cases} |I_R(x, y) - R_{peak}| < Rt \\ |I_G(x, y) - G_{peak}| < Gt \\ |I_B(x, y) - R_{peak}| < Bt \\ I_G(x, y) > I_R(x, y) \\ I_G(x, y) > I_B(x, y) \\ \text{and } G(x, y) < Gt \end{cases} \\ 0: & \text{otherwise} \end{cases}$$

위 식에서 $I_R(x, y)$ 는 입력 영상 I라 할 때 x, y위치의 Red값을 의미하며, R_t 는 경기장에 속하는 Red의 임계치 값을 의미한다. 또한 $G(x, y)$ 는 R, G, B로부터 얻어진 명암 정보를 의미하며 선을 검출하기 위해 사용된다. 이와 같은 처리에 의해 입력 영상 그림 1 (a)로부터 처리된 그림 1 (b)의 영상을 볼 수 있다.



(a) 입력영상 (b) 이진화된 영상
그림 1. 경기장 영역 추출

일반적인 경우에 경기장 장면에서는 보통 윗 부분이 관중석이 위치하게 된다. 그러나 풀 에어리어에서는 이러한 조건을 항상 성립되지 않으므로 경기장 영역만을 추출하는 알고리즘이 필요하다.

경기장 영역을 추출하는 알고리즘은 그림 1로 (b)부터 가장 큰 영역이 경기장 영역이라는 사전 지식을 바탕으로 레이블링 알고리즘[5]을 수행하여 고립 영역을 찾고, 고립 영역 중 가장 큰 영역을 경기장 영역으로 결정한다. 레이블링 알고리즘을 사용하면 경기장 영역 뿐 아니라 선수 후보 영역 및 선 후보 영역에 관한 정보 및 잡음 제거를 위한 정보를 부가적으로 얻을 수 있다.

2.2 선 및 선수 분리

일반적으로 경기 장면에서 선수의 유니폼 색상은 TV방송시 구별이 잘 가도록 경기장의 Green색상과 차별되는 색상을 갖게 된다. 다만 선수의 유니폼이 흰색일 경우 경기장의 선과 동일한 색상을 가지므로 선수와 선의 Occlusion 발생 시 이를 처리하기 어렵다. 본 입력 영상에서도 프랑스 선수들의 유니폼이 흰색이므로 이러한 문제가 발생되었다. 이러한 문제는 컬러 정보만을 이용하여 해결하기 어렵고 위치 정보 등을 이용한 분할 정보를 사용하여야 한다.

그림 1 (b)의 영상으로부터 선수 영역만을 추출하는 방법은 비교적 간단히 수행된다. 즉 앞서 수행한 레이블링 알고리즘에서 얻어진 결과를 바탕으로 경기장 영역 안에 존재하나 경기장 색상을 갖지 않는 부분이 선수 후보 영역이 된다. 이 후보 영역에는 주로 선수 영역 및 각종 선 영역이 포함되게 되는데 이 후보 영역 중에서 선수 영역을 판정하면 된다.

경기장 영역에서 경기장과 다른 색상을 갖는 영역은 선과 선수 영역밖에 없다. 그러므로 두 영역 중에서의 한 영역이 결정되면 나머지 영역은 자동적으로 다른 영역으로 판정된다. 그러나 선과 선수 영역의 컬러 정보 특성이 다르므로 두 영역을 찾을 수 있는 각각의 판정 기준을 설정하여 각자 확실한 영역을 먼저 추출하고 이에 속하지 않는 영역은 좀더 세밀한 처리를 거치는 것이 적절하다.

먼저 후보 영역 중 선수 영역을 판정하는 기준은 다음과 같다.

- 1) 후보 영역의 면적이 일정 임계치 영역 안에 속하는지 판정
- 2) 후보 영역의 Compactness(영역의 복잡도: 면적 대 둘레)가 일정 임계치 값 이상인지 판정
- 3) 후보 영역에 존재하는 화소의 R, G, B값의 차이가 큰지 판정

다음 후보 영역중 선 영역을 판정하는 기준은 다음과 같다.

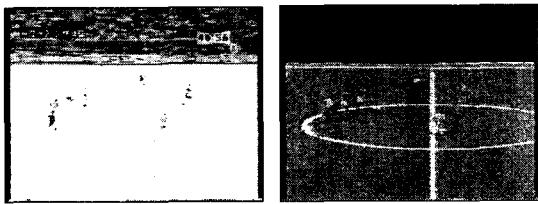
- 1) 후보 영역의 색상 R, G, B가 유사한 값을 갖는지 판정
- 2) 후보 영역의 Hough[6] 변환 값이 큰 직선 선분율

갖는지 판정

3) 후보 수평 수직 길이의 비가 1: 3 이상의 값인가를 판정

4) 후보 영역의 면적 대 배경비가 1 : 1(직선) 근방이거나, 1: 30(Center Circle) 이상인가를 판정

이와 같은 사전지식을 이용한 제약조건을 사용하면 대부분의 후보 영역에서 선과 선수의 분리가 가능하다. 그림 2 (a)에서는 추출된 선수 영상을 컬러로 그림 2 (b)에서는 추출된 선 영역을 흰색으로 보여준다.



(a) 추출된 선수 영역 (b) 추출된 선 영역
그림 2. 추출된 영역

그림 2에서 추출된 선 영역에는 앞서 언급한 바와 같이 프랑스 선수들의 하의 유니폼과 브라질 선수들의 스타킹 영역도 포함되게 된다. 이 경우 다음과 같은 두 가지 조건을 검사하여 선수 영역인지 선 영역인지 판정하게 된다.

- 1) 주변에 같은 색상을 갖는 영역이 존재하는지 조사
- 2) 바로 위의 영역에 일정 크기의 선수 영역이 존재하는지 조사

위의 두 조건은 선수인지 선 영역인지 불확실한 영역을 주변에 위치한 영역을 이용하여 판정하기 위한 제한 조건이다. 즉, 첫 번째 조건은 선 영역은 연속적으로 나타난다는 사전지식을 이용해 혹시 선 영역이 아닌가 판정하는 조건이고, 두 번째 조건은 위의 영상에서 프랑스 선수의 유니폼이 파랑 색 상의에 흰색 바지이므로, 결과적으로 파랑 색 영역이 위에 존재하는 후보 영역은 선수 영역이 될 가능성이 높음을 알 수 있다.

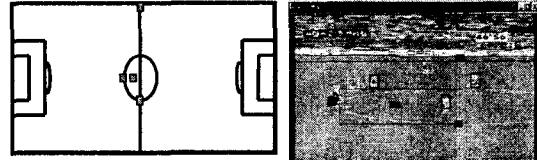
3. 운동장 모델 좌표 변환

3.1 어파인(Affine) 변환

어파인 변환[5]은 두 영상 좌표계간에 대응되는 좌표점을 이용하며 영상을 기하학적으로 변형하는 기법이다. 이 기법은 스케일, 회전, 이동 그리고 이들의 결합을 포함하며, 선을 따라 등 간격이 유지되는 특징이 있다. 본 연구에서는 경기 장면에 나타난 선수가 운동

장의 어떤 위치에 있는지를 측정하기 위해 이 변환을 사용하였다.

어파인 변환을 사용하기 위해서는 먼저 대응되는 두 영상 좌표계 및 특성에 따라 3~4개의 대응점이 결정되어야 한다. 이를 위해 그림 3에서 운동장 모델 좌표계 및 이에 대응되는 경기 장면 좌표계와 두 좌표계간의 대응점들을 보여준다.



(a)운동장 모델 및 대응점 (b)경기 장면 및 대응점
그림 3. 어파인 변환을 위한 두 좌표계

이제 주어진 4개의 대응점을 이용하여 입력영상과 모델 사이의 변환은 다음의 공식으로 구해진다.

$$\begin{bmatrix} x_{2i} \\ y_{2i} \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix} \begin{bmatrix} x_{1i} \\ y_{1i} \\ 1 \end{bmatrix}$$

위 공식에서 모델 좌표계 M = (x₂, y₂, 1)은 입력영상 X = (x₁, x₂, 1)의 좌표에 대응됨을 보여준다. 이때 a₁₁에서 a₃₃ 까지 변환 행렬 T를 구하면 두 좌표계 간의 변환이 이루어진다. 이때 a₃₃ = 1로 정한 것은 두 좌표계간의 scale에 무관함을 의미한다. 이식을 정리하면

$$x_{1i}a_{11} + y_{1i}a_{12} + a_{13} - x_{1i}x_{2i}a_{31} - y_{1i}x_{2i}a_{32} = x_{2i},$$

$$x_{1i}a_{21} + y_{1i}a_{22} + a_{23} - x_{1i}y_{2i}a_{31} - y_{1i}x_{2i}a_{32} = y_{2i},$$

이 된다. 이 두 식을 a_{ij}에 대해 다시 정리하면

$$x_{2i} = \frac{a_{11}x_{1i} + a_{12}y_{1i} + a_{13}}{a_{31}x_{1i} + a_{32}y_{1i} + 1}, \quad y_{2i} = \frac{a_{21}x_{1i} + a_{22}y_{1i} + a_{23}}{a_{31}x_{1i} + a_{32}y_{1i} + 1}$$

이 되고 변환 행렬 T를 구하기 위해 임의의 4점을 이용하여 행렬 형태로 정리하면 아래와 같다.

$$\begin{bmatrix} x_{11} & y_{11} & 1 & 0 & 0 & 0 & -x_{11}x_{21} & -y_{11}x_{21} \\ 0 & 0 & 0 & x_{11} & y_{11} & 1 & -x_{11}x_{21} & -y_{11}y_{21} \\ x_{12} & y_{12} & 1 & 0 & 0 & 0 & -x_{12}x_{22} & -y_{12}x_{22} \\ 0 & 0 & 0 & x_{12} & y_{12} & 1 & -x_{12}x_{22} & -y_{12}y_{22} \\ x_{13} & y_{13} & 1 & 0 & 0 & 0 & -x_{13}x_{23} & -y_{13}x_{23} \\ 0 & 0 & 0 & x_{13} & y_{13} & 1 & -x_{13}x_{23} & -y_{13}y_{23} \\ x_{14} & y_{14} & 1 & 0 & 0 & 0 & -x_{14}x_{24} & -y_{14}x_{24} \\ 0 & 0 & 0 & x_{14} & y_{14} & 1 & -x_{14}x_{24} & -y_{14}y_{24} \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \\ a_{31} \\ a_{32} \end{bmatrix} = \begin{bmatrix} x_{21} \\ y_{21} \\ x_{22} \\ y_{22} \\ x_{23} \\ y_{23} \\ x_{24} \\ y_{24} \end{bmatrix}$$

따라서 벡터 T는 임의의 4점으로 구성된 X 좌표계의 역 행렬로 구해진다.

$$XT = M, T = X^{-1}M$$

위의 공식에서 역 행렬을 구하는 방법으로 본 연구에서는 Crammers Rule[6]을 이용하였다. 이와 같은 처리 결과 얻어진 변환행렬 T를 이용하여 모델 좌표계로 변환한 결과를 그림 4에서 볼 수 있다.

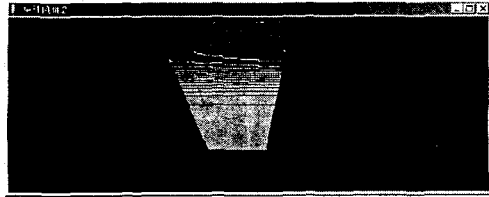


그림 4. 어파인 변환 결과

3.2 역 변환 어파인 매핑

어떤 x와 y의 사상 함수에 있어 입력 화소가 출력 화소로 이동하는 것을 전 방향 사상이라 불린다. 전 방향 사상에는 그림 4와 같이 홀(hole)과 오버랩(overlap) 발생의 두 가지 문제점이 있다. 홀은 정의되지 않은 화소들이며 목적 화소에 대응되는 원 화소가 없어서 발생된다. 이와 같은 문제점을 해결하기 위해 nearest neighbor, bilinear, higher order, cubic convolution, B-spline[5]등의 보간법(interpolation)이 주로 사용되어져 왔으나, 이러한 보간법을 사용하지 않고 역 변환 매핑을 이용하면 쉽게 hole 문제를 해결할 수 있다.

즉, 그림 4의 결과는 아래와 같이 원 영상에서 목적 영상으로 좌표가 매핑되어 문제가 발생하였으나 역 방향 매핑은 그림 5에서의 결과와 같이 목적영상의 좌표점에 들어갈 화소를 원 영상에서 직접 찾으므로 hole이 발생되지 않는다.

즉, 구해진 T를 이용하여 $XT = M$ 의 $X = MT^{-1}$ 의 $XT = M$ 의 공식에 대입하여 모델 좌표의 최소, 최대점을 구한 후, 이 영역 안에서만 변환을 수행한다. 그림 5에서 역 어파인 매핑의 수행 결과를 볼 수 있다.

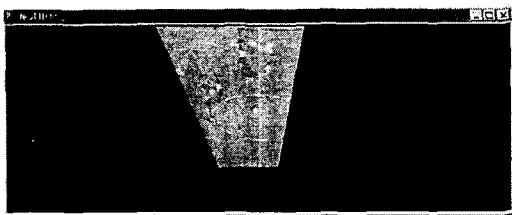


그림 5. 역 어파인 매핑 결과

4. 실험 환경 및 결론

본 연구에서는 Pentium II 400Mhz의PC상에서 초당 3 ~ 5 Frame로 입력된 240x640 해상도의 RGB영상을 입력으로 사용하였다. 영상 캡처 보드로는 Metrox사의 MeteorII 보드를 이용하였고 개발 언어로는 Visual C++ 6.0을 사용하였다.

본 연구에서는 축구 경기 장면 분석시스템의 일부로서 경기 장면에서 선수의 위치와 운동장 모델에서의 선수의 위치간의 매핑에 관해 기술하였다. 이를 위해 먼저 경기장과 관중석을 분리하고, 분리된 경기장내에서 각종 개체들을 추출하는 방법을 기술하였다. 또한 경기 장면에서의 위치와 운동장 모델과의 사상을 위해 어파인 모델을 사용하였다. 이때 두 좌표계간의 사상시 발생하는 hole에 관한 문제를 주로 사용되는 보간법이 아니라 역 방향 매핑을 이용하여 해결하였다. 현재 이 시스템은 계속 개발 중에 있으며 연속적으로 입력된 영상을 대상으로 운동장 좌표계의 영상을 모자이크하여 파노라마 뷰의 합성에 관한 연구가 수행중이다 또한 운동장에서 추출된 선수들의 위치를 이용하여 3D 실감 합성 그래픽 Replay 기능 연구를 추가로 다른 팀에서 수행되고 있다. 이러한 연구의 최종 목표는 영상 처리를 이용해 경기 장면을 분석하여 보다 많은 사람이 각자의 목적에 맞게 축구경기를 좀더 재미있게 즐길 수 있도록 함에 있다.

참고 문헌

- [1]A. F. Bobrick, "Video Annotation: Computer Watching Video", Preceeding of Second Asian Conference on Computer Vision, Singapore, Vol. I, pp.19-23, 1995.
- [2]D. Yow, B. L. Yeo, M. Yeung and B. Liu, "Analysis and presentation of Soccer Highlight from Digital Video", Proc. Of second Conference on Computer Vision, Singapore, Vol. II, pp.499-503, 1995.
- [3]H.S. Sawhney, S. Ayer and M. Gorkani, "Model-based 2D and 3D dominate Motion Estimation for Mosaicking and Video Representation", Proc. of International Conference on Image processing, Cambridge, Vol. III, PP.815-818, 1995
- [4]VITRA: <http://www.dfki.uni-sb.de/vista>
- [5]최형일, 이근수, 이양원, 영상처리 이론과 실제, 홍릉과학출판사, pp.249-253, 1997
- [6]Erwin Kreyszig, *Advanced engineering mathematics*, EILEY Press, 5th Edition, pp.336-341.