

역방향 어파인 변환을 이용한 축구 경기장면 매핑

윤 호섭*, 김 광용, 소 정, 민 병우, 양 영규
컴퓨터소프트웨어연구소 전자통신연구원
yoonhs@etri.re.kr

Soccer Game Scene Mapping Using Inverse Affine Transform

Ho-Sub Yoon, Kwang-Yong Kim, Jung Soh, Byung-Woo Min and Youngkyu Yang
Image Processing Division, Computer & Software Technology Lab., ETRI

Abstract

본 연구에서는 축구 경기 장면 분석시스템의 일부인 경기 장면에서 선수의 위치와 운동장 모델에서의 선수의 위치 간의 매핑에 관해 기술한다. 이때 두 좌표계 간의 사상(mapping)시 발생하는 hole에 관한 문제를 주로 사용되는 보간법이 아니라 역 방향 매핑을 이용하여 해결하고, 또한 운동장 모델의 전체 영역이 아니라 원 영상에서 얻어진 정보를 이용하여 최소한의 영역만을 변환함으로써 계산의 효율성을 높였다.

본 연구에서는 PC 상에서 초당 3 ~ 5 Frame로 입력된 280x640 해상도의 100여 개의 RGB 영상을 대상으로 실험하였다. 축구 경기장면을 분석한 결과 대부분의 경우 두 좌표계간의 매핑이 올바르게 이루어 졌다.

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 시작되었다.

먼저 MIT Media Lab에서 "Computer Watching Football"이라는 프로젝트를 수행하였는데 이 과제는 미식 축구 경기 비디오로부터 자동으로 선수를 추적하고 선수의 이동 궤적을 경기장 모델에 표현 해 주는 시스템을 개발하였다[1]. Princeton 대학의 Yow는 축구 영상에 모자이크 기법을 적용하여 하나의 큰 파노라마 영상을 구성하는 방법을 제안하였다[2]. 또한 일본의 Hasegawa 대학의 Taki는 하나의 카메라를 이용하는 대신 여러 대의 고정된 카메라로부터 얻은 축구 경기 장면으로부터 선수들과 팀의 경기력을 분석하기 위한 시스템을 구현하였다[3].

독일에서는 VITRA(Visual Translator)라는 장기적 프로젝트를 통해 연구를 수행하고 있는데 이 연구에서 SECCER, REPLAI 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 변환하고 한다. 이러한 시스템의 적용은 팀의 경기력을 분석하여 코치에게 도움을 주는 시스템을 개발하거나, 전체 경기 장면 중 원하는 중요 경기장면만을 추출하는 Hi-light 추출 시스템, 경기의 자동 해설 및 그래픽을 이용한 3D 3차원 Replay 시스템, 혹은 파노라마 뷰를 이용한 Virtual Replay 시스템 등의 개발이 가능하다.

본 연구에서는 이러한 시스템을 개발하기 위한 전단계로서 축구경기 장면으로부터 실제 운동장 좌표를 구하기 위한 영상 워핑 방법에 대해 기술한다.

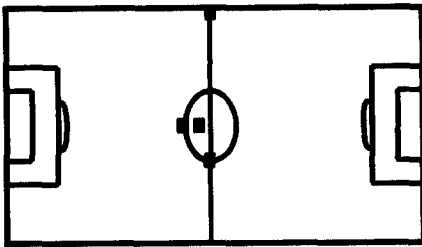
본 논문의 구성은 2장에서는 영상 Affine 변환에 대해 기술하고 3장에서는 워핑시 발생하는 hole을 제거하기

위한 역 변환 매핑 방법에 대해 기술하고 마지막으로 4 장에서는 실험 환경 및 결론에 대해 기술한다.

2. 어파인(Affine) 변환

어파인 변환[5]은 두 영상 좌표계간에 대응되는 좌표 점을 이용하여 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 운동장의 어떤 위치에 있는지를 알라내기 위해 이 변환을 사용하였다.

어파인 변환을 사용하기 위해서는 먼저 대응되는 두 영상 좌표계 및 특성에 따라 3~4 개의 대응점이 결정되어야 한다. 이를 위해 그림 1에서 운동장 모델 좌표계 및 이에 대응되는 경기 장면 좌표계와 두 좌표계 간의 대응점들을 보여준다.



(a) 운동장 모델 및 대응점



(b)경기 장면 및 대응점

그림 1. 어파인 변환을 위한 두 좌표계

그림 1의 (b)에서 운동장 모델 4 점에 대응되는 경기 장면의 대응점을 구하기 위한 처리는 이전 논문[6]을 참고하면 자세히 알 수 있고 본 연구에서는 전처리 격인 영상처리 알고리즘에 대해서는 기술하지 않는다.

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

Model Coordinate Image Coordinate

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

위 공식에서 모델 좌표계 $M = (x_2, y_2, 1)$ 은 입력영상 $X = (x_1, x_2, 1)$ 의 좌표에 대응됨을 보여준다. 이때 a_{11} 에서 a_{33} 까지 변환 행렬 T 를 구하면 두 좌표계 간의 변환이 이루어진다. 임의로 $a_{33} = 1$ 로 정한 것은 두 좌표계간의 scale에 무관함을 의미한다. 이식을 정리하면

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

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

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

이 되고 변환 행렬 T 를 구하기 위해 임의의 4 점을 이용하여 행렬 형태로 정리하면

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

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

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

위의 공식에서 역 행렬을 구하는 방법으로 본 연구에서는 Cramer's Rule[7]을 이용하였다. 이와 같은 처리 결과 얻어진 변환행렬 T 를 이용하여 모델 좌표계로 변환한 결과를 그림 2에서 볼 수 있다.

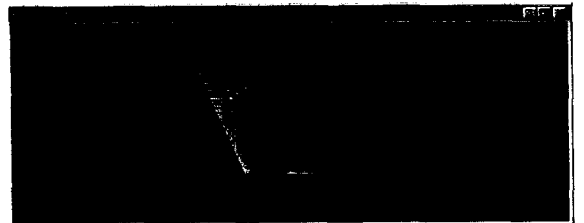
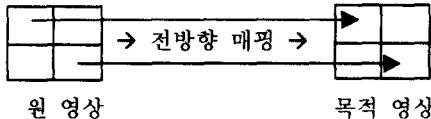


그림 2. 어파인 변환 결과

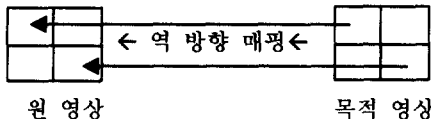
3. 역 변환 어파인 매핑

어떤 x 와 y 의 사상 함수에 있어 입력 화소가 출력 화소로 이동하는 것을 전방향 사상이라 불린다. 전방향 사상은 그림 2와 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 2의 결과는 아래와 같이 원 영상에서 목적 영상으로 좌표가 매핑되어 문제가 발생하였으나



역 방향 매핑은 그림 3에서의 결과와 같이 목적영상의 좌표점에 들어갈 화소를 원 영상에서 직접 찾으므로 hole 이 발생되지 않는다.



즉, 구해진 T 를 이용하여 $XT = M$ 의 공식으로 모델 영상의 좌표를 구하는 것이 아니라, 구해진 T 의 역 행렬을 다시 구하여 $X = MT^{-1}$ 의 공식을 이용해 모델 좌표계로부터 이에 대응되는 원 화소를 찾으므로 hole 이 발생되지 않는 것이다. 이때 경기 장면의 화소는 운동장 모델 영상의 전체 영역 중 일부의 영역을 차지하므로 전체 영역에 대하여 역 변환시키지 않고 원 영상의 최소, 최대점을 $XT = M$ 의 공식에 대입하여 모델 좌표의 최소, 최대점을 구한 후, 이 영역안에서만 변환을 수행한다.

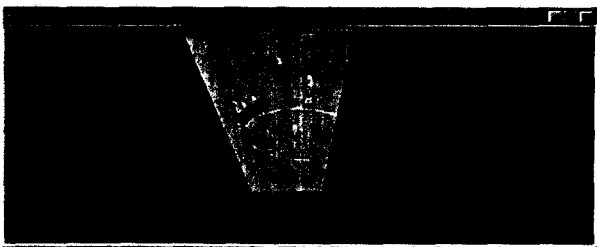


그림 3. 역 어파인 매핑 결과

4. 실험 환경 및 결론

본 연구에서는 Pentium II 400Mhz 의 PC 상에서 초당 3 ~ 5 Frame 로 입력된 240x640 해상도의 RGB 영상을 입력으로 사용하였다. 영상 캡처 보드로는 Metrox 사의 MeteorII 보드를 이용하였고 개발 언어로는 Visual C++ 5.0 을 사용하였다.

본 연구에서는 축구 경기 장면 분석시스템의 일부인 경기 장면에서 선수의 위치와 운동장 모델에서의 선수의 위치 간의 매핑에 관해 기술하였다. 이때 두 좌표계 간의 사상시 발생하는 hole 에 관한 문제를 주로 사용되는 보간법이 아니라 역 방향 매핑을 이용하여 해결하였고, 또한 운동장 모델의 전체 영역이 아니라 원 영상에서 얻어진 정보를 이용하여 최소한의 영역만을 처리함으로써 계산의 효율성을 높였다. 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 축구 경기를 좀더 재미 있게 즐길 수 있도록 함에 있다.

참고 문헌

- [1] A. F. Bobrick, "Video Annotation: Computer Watching Video," Proc. 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/vitra>
- [5] 최 형일, 이근수, 이양원, 영상처리 이론과 실제, 홍릉 과학출판사, pp.249-253. 1997
- [6] H. S. Yoon, and etc "Mosaicking and Panorama View Generation from Soccer Game Image Sequences", ITC-CSCC'99, Vol. I, pp. 378 - 381, Japan.
- [7] Erwin Kreyszig, Advanced engineering mathematics, EILEY Press, 5th Edition, pp.336-341.