

배경의 특징 추적을 이용한 물체의 이동 거리 추정 및 정확도 평가

이종현, 남시욱, 이재철, 김재희
연세대학교 전기·컴퓨터공학과
Tel : (02) 361-2869, FAX : (02) 362-0413

A Distance Estimation Method of Object's Motion by Tracking Field Features and A Quantitative Evaluation of The Estimation Accuracy

Jonghyon Yi, Siwook Nam, Jae Cheol Lee, Jaihie Kim
Dept. of Electrical & Computer Engin, Yonsei Univ.
e-mail://johnyi@seraph.yonsei.ac.kr

Abstract

This paper describes a distance estimation method of object's motion in soccer image sequence by tracking field features. And we quantitatively evaluate the estimation accuracy.

We suppose that the input image sequence is taken with a camera on static axis and includes only zooming and panning transformation between frames. Adaptive template matching is adopted for non-rigid object tracking. For background compensation, feature templates selected from reference frame image are matched in following frames and the matched feature point pairs are used in computing Affine motion parameters. A perspective displacement field model is used for estimating the real distance between two position on input image.

To quantitatively evaluate the accuracy of the estimation, we synthesized a 3 dimensional virtual stadium with graphic tools and experimented on the synthesized 2 dimensional image sequences. The experiment shows that the average of the error between the actual moving distance and the estimated distance is 1.84%.

I. 서론

최근 스포츠 영상 분석[1]-[5]에 대한 관심이 증가함에 따라, 스포츠 영상에 컴퓨터 비전 기술들이 광범위하게 적용되고 있다. 특히, 선수나 공과 같은 물체의 위치를 추적하고, 실제 경기장에서의 물체 위치를 구함으로써 물체의 이동거리나 속도를 추정하기 위한 연구가 활발히 진행되고 있다.

Intille[1]등은 사전지식(prior knowledge)을 이용하여 미식 축구에서 선수들의 자동 추적에 대한 연구를 수행하였고, Seo[2]등은 축구경기에서 선수와 공을 추적하고 실제 경기장에서의 위치를 구하는 방법에 대해 연구하였다. Kim[3]등은 축구영상으로부터 축구공의 3차원 위치를 구하기 위한 방법을 제안하였고, Reid[4]등은 시점(point of view)이 다른 두 영상시퀀스로부터 축구공의 정확한 위치를 구하는 방법에 대해 연구하였다. 또한 스포츠 영상 처리를 위한 상용 소프트웨어도 개발되어 영상으로부터 물체의 이동 거리 및 이동 속도를 추정하는 방법이 구현되기도 하였다[8].

본 논문에서는 축구 영상에서 투영 변위 필드 모델(perspective displacement field model)을 이용하여 물체의 이동 거리를 추정하고 그 정확도를 정량적으로 평가하였다. 이를 위해서 선수와 공을 추적하고 카메라에 의한 배경 차이를 보정한 후, 축구 영상과 축구 경기장 모델의 변환관계를 구하였다.

축구 영상에서 선수와 같이 형태가 계속 변하는 물체의 위치를 추적하기 위하여 템플릿을 매 영상마다 갱신하는 적응 템플릿 매칭(adaptive template matching)[1]을 사용하였다. 카메라의 움직임으로 인한 배경의 차이를 보상하기 위하여 경기장에 존재하는 특징으로부터 템플릿을 설정하여 추적함으로써 각 영상간의 변환을 계산하였다. 또한, 물체의 이동거리를 추정하기 위하여 축구 영상과 축구 경기장의 대응되는 특징점과 투영 변위 필드 모델[7]을 이용하여 상호 변환관계를 구하였다.

본 논문의 구성은 다음과 같다. II절에서는 물체의 이동 거리 추정을 위한 전체 과정에 대하여 설명하고, III절에서는 모양이 변하는 물체 추적에 대하여 설명한다. IV절에서는 카메라의 움직임에 의한 배경 차이 보상에 대하여 설명하고, V절에서는 투영 변위 필드 모델에 대하여 설명한다. VI절에서는 추정한 거리의 정확도를 정량적으로 평가하고, 마지막으로 VII절에서는 결론을 맺는다.

II. 물체의 이동 거리 추정 과정

축구 영상에 존재하는 물체의 실제 이동 거리를 추정은 영상 내에서 물체의 위치 추적, 카메라의 움직임으로 인한 배경 차이 보상, 영상과 축구 경기장 모델간의 변환관계 파악의 세 과정으로 구성된다. 영상 내에서 물체위치 추적은 영상내의 물체의 좌표를 계속 찾는 것이다. 카메라의 움직임으로 인한 배경 차이 보상은 카메라의 panning이나 zooming으로 인한 영상간의 변화를 고려하여 영상에서의 물체 좌표를 보상하는 것이다. 영상과 축구 경기장 모델간의 변환관계 파악은 축구 영상과 경기장 모델간의 대응점으로부터 두 영상간의 변환을 표시하는 8개의 파라미터를 추출하는 것이다. 물체의 이동 거리 추정과정은 그림 1과 같다.

III. 모양이 변하는 물체의 위치 추적

일반적으로 스포츠 중계에서는 움직이는 카메라로 물체를 촬영하기 때문에 촬영된 물체의 형태나 크기가 시간에 따라 변하므로 자동으로 물체를 추적하기가 어렵다. 물체 위치 추출에 많이 사용되는 방법중의 하나인 템플릿 매칭(template matching)은 우리가 찾고자 하는 물체의 형태에 해당하는 고정 템플릿을 영상의 각 부분과 비교하여 유사도가 높은 위치를 물체의 위치로 판정하는 방법이다. 그러나, 축구와 같은 스포츠 경기 중계화면에서는 확대/축소(zooming in and out)가 존재할 뿐만 아니라, 추적대상이 되는 사람이 움직이기 때문에 그 형태가 매우 다양하게 변화하므로 고정된 형태의 템플릿으로 유사도 비교에 의해 물체의 위치를 찾아내기가 어렵다. 이러한 문제점을 해결하기

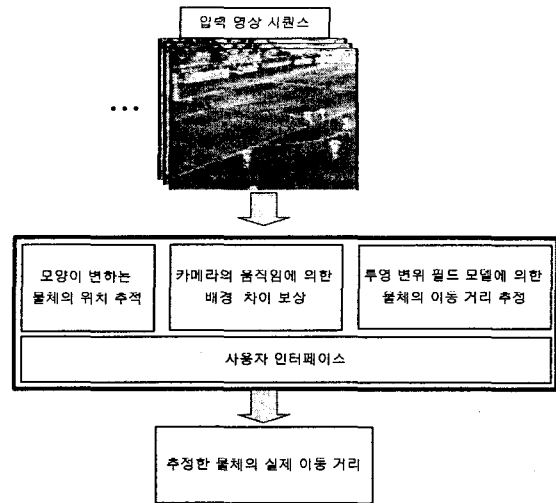


그림 1. 물체의 이동 거리 추정 과정

위하여 초기에 정한 고정 템플릿을 사용하는 것이 아니라 시간에 따라 적응적으로 템플릿을 변화시키는 방법이 적응 템플릿 매칭(adaptive template matching)이다. 즉, 적응 템플릿 매칭방법에서는 현재 영상에서 찾은 물체를 분할하여 다음 영상에서 물체의 위치를 찾기 위한 템플릿으로 설정한다. 영상 시퀀스에서 인접 영상간에는 큰 형태 변화가 없다고 가정할 수 있으므로 이 방법을 이용하여 형태가 변하는 물체의 위치를 추적할 수 있다.

그림 2와 같이 우선 초기 영상에서 사용자가 추적하고자 하는 물체 주위에 표시를 하면 그 영역 안에서 물체를 분할하여 템플릿을 설정하고, 이 템플릿을 이용하여 다음 영상에 대해 매칭을 하여 물체의 위치를 찾는다. 찾은 물체의 위치를 중심으로 다시 물체를 분할하여 템플릿을 설정한 후 그 다음 영상에 대해 매칭을 하여 물체의 위치를 찾는 과정을 계속 반복한다.

초기 영상에서 템플릿으로 설정할 물체를 분할하는 과정은 다음과 같다. 그림 2에서와 같은 영상의 초기 화면에서 사용자가 마우스나 전자펜으로 물체주위에 원을 그리면 원 위에 있는 화소 $P_1(R_1, G_1, B_1)$, $P_2(R_2, G_2, B_2)$, ..., $P_N(R_N, G_N, B_N)$ 에 대하여 평균 칼라 M 과 평균 거리 D 를 식 (1), (2)에 의해 구한다.

$$M = (m_R, m_G, m_B) = \left(\frac{1}{N} \sum_{i=1}^N R_i, \frac{1}{N} \sum_{i=1}^N G_i, \frac{1}{N} \sum_{i=1}^N B_i \right) \quad (1)$$

$$D = \frac{1}{N} \sum_{i=1}^N \|P_i - M\|, \quad \text{단, } \|\cdot\| \text{은 Euclidean Norm} \quad (2)$$

식 (1), (2) 를 이용하여 M 과 D 가 구해지면 사용자가 표시한 영역 내의 임의의 화소 P 에 대해 식 (3)을 이용하여 배경인지 물체인지를 판단한다.

$$P = \begin{cases} 1, & \|P-M\| > c \cdot D \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

여기서, c 는 상수이며 1은 물체 화소, 0은 배경화소를 의미한다.

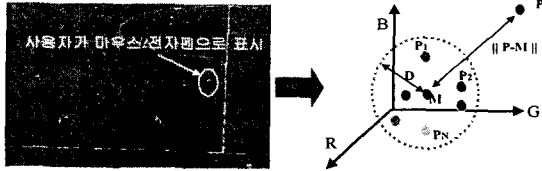


그림 2. 초기영상에서의 템플릿 설정

IV. 카메라 움직임에 의한 배경 차이 보상

본 논문에서는 기준 프레임 영상에서 특징 템플릿을 설정하고, 이후 프레임에서 템플릿을 매칭함으로써 기준 영상과 i 번째 영상 사이의 카메라의 움직임에 의한 배경 차이를 계산하였다. 배경의 차이를 계산하기 위한 움직임 모델은 식 (4)와 같은 Affine 움직임 모델을 사용하였다.

$$\begin{pmatrix} u_n \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_n \\ y_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (4)$$

Affine 움직임 모델의 미정 파라미터의 개수는 6 개로서, 해를 구하기 위해서는 3개의 특징점 쌍이 필요하다. 따라서, 본 연구에서는 경기장에서 추출할 수 있는 3개의 특징 템플릿을 설정하고, 추적하였다.

특징 템플릿을 구성하기 위해서는 경기장에 존재하는 모서리 근처의 좌표에 마우스/전자펜 입력에 의해 표시하면 그 주위에 탐색 영역이 설정되며, 그 영역에 그림 3과 같이 SUSAN Corner Operation[6]을 수행하게 된다. 출력값이 가장 크게 나온 좌표를 중심으로 일정한 크기의 템플릿을 설정하고 이후 프레임 영상에서 매칭하여 추적한다.

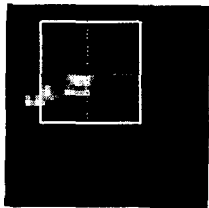


그림 3. SUSAN Corner 영상 및 템플릿 중심 결정

템플릿 매칭과정의 처리 시간을 단축하기 위하여 식 (5)와 같은 등가속도 운동 가정에 의한 예측을 수행하였다. 즉, 이전 세 프레임에서의 특징의 좌표를 이용하여 현재 프레임에서의 위치를 예측하고, 예측된 위치

를 중심으로한 탐색 영역에 대하여 템플릿 매칭을 수행하였다.

$$\hat{x}_T = 3x_{T-1} - 3x_{T-2} + x_{T-3} \quad (5)$$

기준 프레임과 이후 프레임 영상 사이의 세 쌍의 특징 대응점으로부터 기준 프레임에 대한 영상 시퀀스 전체의 배경 차이 정보를 Affine 움직임 파라미터로 표현하였다.

V. 투영 변위 필드 모델에 의한 거리 추정

시각 t , t' 에서의 물체의 3D 위치를 각각 $P=[X,Y,Z]^T$, $P'=[X',Y',Z']^T$ 라 하고 그 투영위치를 $p=[x,y]^T$, $p'=[x',y']^T$ 라 하면 perspective projection model에 의해 식(6)를 얻을 수 있다[7].

$$x' = \frac{ax+by+e}{gx+hy+1} \quad y' = \frac{-cx+dy+f}{gx+hy+1} \quad (6)$$

여기서 8개의 파라미터 (a, b, c, d, e, f, g, h)를 추정하기 위하여 4쌍의 대응되는 특징점 쌍이 필요하다. 본 논문에서는 선수가 슈팅을 하는 경우에 슈팅거리를 추정하기 위하여 그림 2에서의 같이 goal area 내의 10개의 특징점 중 4개를 사용자가 선택하여 변환관계를 구할 수 있도록 하였다.

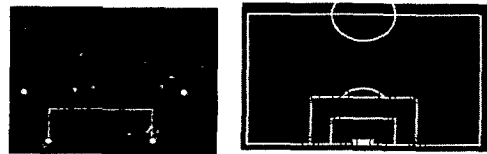


그림 4. 경기장모델과 축구영상에서 특징점의 대응

물체의 이동거리 s 는 다음 과정에 의해 구한다.

- ① 이동거리를 추정하려는 구간에 대해 물체의 위치를 추적한다.
- ② 이동거리를 추정하려는 구간의 첫 영상에서의 물체의 위치를 카메라의 움직임을 보상하여 구간의 마지막 영상에서의 물체 위치로 변환한다.
- ③ 그림 4의 경기장 모델의 4 점과 축구 영상의 4 점을 다음 관계식에 대입하여 a, b, c, d, e, f, g, h 파라미터를 계산한다.

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} = \frac{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}}{\begin{pmatrix} g & h \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} + 1}, \quad i=1, 2, 3, 4$$

- ④ ①,②에서 구한 물체의 위치를 ③에 의해 구한 파라미터에 의해 경기장 모델에서의 좌표로 변환하여 두 점 사이의 거리를 구한다.

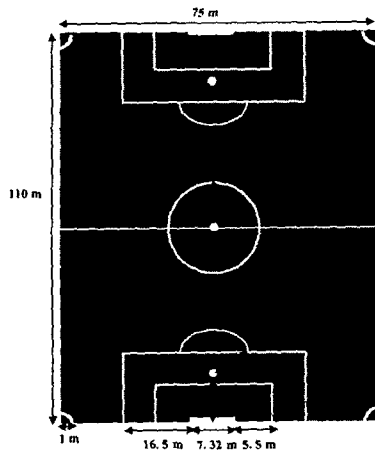


그림 5. 국제 경기용 축구장 규격

VI. 실험 결과

물체의 이동 거리 추정 정확도를 정량적으로 평가하기 위해 720×486의 해상도를 갖는 32 bit의 합성 축구 영상을 Adobe Photoshop 과 3D Studio Max 2.5를 이용하여 합성하였다. 합성된 축구 경기장은 국제 경기용 표준 규격이다[9]. 합성한 3D Stadium 과 2D 축구 영상의 예를 그림 6에 나타내었다.

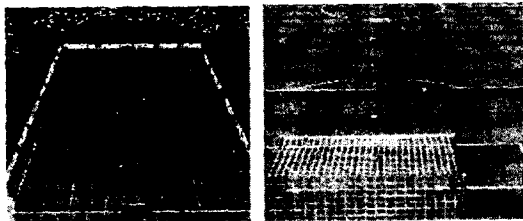


그림 6. 합성 영상의 예

합성한 축구 영상으로부터 알 수 있는 물체의 실제 이동 거리와 알고리즘에 의해 추정된 이동 거리의 오차는 식(7)에 의해 측정하였다.

$$e = \frac{|실제이동거리 - 추정이동거리|}{실제이동거리} \times 100(\%) \quad (7)$$

세 개의 합성 영상에 대하여 이동 거리 예측을 5회 반복한 결과의 평균과 실제 이동 거리를 표 1과 같이 비교하였다. 추정된 이동 거리의 오차는 평균 1.84%이며, 물체의 이동거리가 길수록 오차가 크게 나타나는 것을 확인 할 수 있다. 이 오차의 원인은 입력 영상과 운동장 모델 사이의 특징점 대응관계에서 발생하는 것으로서, 네 쌍의 특징점 좌표로부터의 거리에 비례하여 오차가 증가함을 알 수 있다.

표1. 합성영상별 추정 오차

| 번호 | 이동거리(m) | | e(%) |
|----|---------|------|------|
| | 실제 | 추정 | |
| 1 | 실제 | 46.2 | 2.16 |
| | 추정 | 45.2 | |
| 2 | 실제 | 13.8 | 1.44 |
| | 추정 | 13.6 | |
| 3 | 실제 | 32.1 | 1.93 |
| | 추정 | 32.7 | |

VII. 결론

본 논문에서는 축구 중계 영상 시퀀스 내에서 물체의 위치를 추적하고 카메라의 움직임으로 인해 발생하는 배경의 차이를 보정한 후, 투영 변위 필드 모델에 의하여 영상의 물체 위치를 경기장 모델의 위치로 변환한 후 물체의 실제 이동거리를 추정하였다.

물체 이동거리 추정의 정량적 평가를 위해 그래픽 툴로 합성한 축구 영상에 대해서 실험한 결과 평균 1.84%의 오차를 나타내었다. 영상과 경기장 모델의 변환관계 파악에 사용되는 대응되는 네 쌍의 특징점에서 멀리 떨어진 물체일수록 그 추정오차가 증가함을 알 수 있었다.

참고문헌

- [1] S.S. Intille, A.F. Bobick, "Tracking using a local closed-world assumption : Tracking in the football domain", *MIT Media Lab Perceptual Computing Group*, TR No.296, MIT, August, 1994
- [2] Y. Seo, S. Choi, H. Kim and K. S. Hong, "Where are the ball and players? Soccer game analysis with color-based tracking and image mosaic", *Int'l Conf. on Image Analysis & Processing*, Florence, Italy, 1997
- [3] T. Kim, Y. Seo, and K.-S. Hong, "Physics-based 3D position analysis of a soccer ball from monocular image sequences," *Int. Conf. on Computer Vision*, Bombay, India, pp. 721-726, Jan. 1998
- [4] Ian Reid and Andrew Zisserman, *Goal-directed Video Metrology*, *In Proc. European Conference on Computer Vision*, 1996
- [5] T. Bebie, H. Bieri, "SoccerMan-Reconstructing Soccer Games from Video Sequences, *International Conference on Image Processing*, pp. 898-902, 1998
- [6] S.M. Smith and J.M. Brady, "SUSAN - a new approach to low level image processing". *Int. Journal of Computer Vision*, 23(1):45--78, May 1997.
- [7] A. Murat Tekalp, *Digital Video Processing*. Prentice-Hall, pp.153-155, 1995
- [8] ORAD High Tech Systems
(<http://www.orad.co.il/sport/dreplay.htm>)
- [9] FIFA Home page (<http://www.fifa.com>)