

# 다해상도 동영상에서의 다중 객체 추적

강 성훈, 이 성환

고려대학교 컴퓨터학과/인공시각연구센터

E-mail:{shkang, swlee}@image.korea.ac.kr

## Multiple Object Tracking in Space-variant Image Sequences

Seonghoon Kang and Seong-Whan Lee

Dept. of Computer Science and Engineering/Center for Artificial Vision Research, Korea University

### 요 약

본 논문에서는 다해상도 영상에서 움직이는 다중 객체의 추적 방법을 다룬다. 일반적으로 객체 추적 알고리즘은 움직임 탐지, 예측, 정합, 갱신의 처리 단계로 구성되어 있다. 특히 다중 객체 추적일 경우, 정합 과정은 매우 중요하다. 일반적인 시각 시스템에서는 대상 객체가 강체(rigid object)라고 가정하면 이러한 정합 과정은 비교적 쉽게 구현될 수 있다. 그러나 다해상도 영상에서는 한 위치에서 다른 위치로 움직일 때 그 영역의 형태 및 크기가 변형되기 때문에 정합이 쉽게 이루어지지 않는다. 따라서 본 논문에서는 이러한 문제를 해결할 수 있는 다해상도 영상에서의 정합 방법을 제안한다.

### 1. 서론

능동 시각 시스템은 다음과 같은 세 가지의 요건을 만족해야 한다. 첫째, 관심 영역을 고해상도로 표현할 수 있어야 한다. 둘째, 넓은 시야를 제공해야 한다. 셋째, 빠른 응답 시간을 가져야 한다[1]. 그러나 일반적인 영상 표현 방법은 이러한 요건들을 동시에 만족하지 못하기 때문에, 이러한 요건들을 모두 만족시킬 수 있는 생물 시각 시스템의 특성에 기반한 방법의 연구가 많이 이루어져 왔다. 그 중 대표적인 것이 다해상도 시각(space-variant vision)이다.

이러한 다해상도 시각은 최근 들어 객체 추적, 버전스(vergence) 제어 등과 같은 능동 시각 분야에 많이 응용되고 있다. 이러한 다해상도 시각을 이용한 능동 추적 시스템은 카메라의 시선 위치를 오직 하나의 대상 객체에만 고정시키고 지속적으로 추적할 수 있다. 이때, 카메라의 시선 위치를 다른 객체로 이동하고자 할 경우 배경 상에서 움직이는 다른 객체들의 움직임을 파악하는 것이 매우 중요하게 된다. 즉 다중 객체 추적이 필요하게 된다. 이러한 다중 객체 추적의 경우에는 움직이는 다중 객체 탐지 및 그들 간의 정합 문제가 발생하게 된다.

다해상도 시각의 경우 다해상도 영상 좌표계에서 나타나는 움직임 벡터 들은 직교 영상 좌표계에서 나타나는 것들과는 그 크기와 방향들이 다르기 때문에, 이러한 벡터들로부터 다중 객체의 움직이는 영역을 분할하는 것은 매우 어렵다. 뿐만 아니라 다해상도 좌표계에서는 한 위치에서 다른 위치로 객체가 움직일 때 그 영역의 크기나 형태가 변형되기 때문에 정합 또한 매우 어려운 문제가 된다.

본 논문에서는 이러한 문제들을 효과적으로 해결할 수 있는 알고리즘을 제안한다.

### 2. 관련 연구

대표적인 다해상도 시각 모형으로는 Schwartz의  $\log(z)$  모형[2]을 들 수 있다. 이 모형과 함께 여러 변형된 모형들이 능동 시각과 같은 많은 분야에 적용되어 왔다. Panerai

등[3]은 log-polar 영상에서의 버전스 제어와 추적 방법을 개발 하였으며, Lim 등[4]은 다해상도 능동 시각에서의 추적 방법을 제안하였다. 최근에는 Juric[5]가 언클 탐지 및 추적을 위한 새로운 log-polar 매핑 방법을 제안하였다.

뿐만 아니라, 다중 객체 추적에 대해서도 많은 연구들이 이루어져왔다. 최근에 Bremond 등[6]은 여러 비강체(non-rigid object)들의 움직임을 추적하는 방법을 제시했으며, Haritaoglu 등[7]은 움직이는 사람들을 실시간으로 탐지하고 추적할 수 있는 시스템을 개발하였다.

### 3. 다해상도 시각

#### 3.1 다해상도 영상 표현 모형

이 장에서는 본 논문에서 사용된 다해상도 모형에 대해서 다룬다. 다해상도 모형은 인간 망막에 위치한 신경절(ganglion) 세포의 수용장(receptive field) 크기가 중심부(fovea)에서는 작으며 주변부(periphery)로 갈수록 커지는 특성을 이용한다. 이러한 특성으로 인하여, 그림 1에서 보이는 바와 같이 수용장의 크기가 중심부에서는 작고 균일하며, 주변부로 갈수록 커지는 모형을 얻을 수 있다.

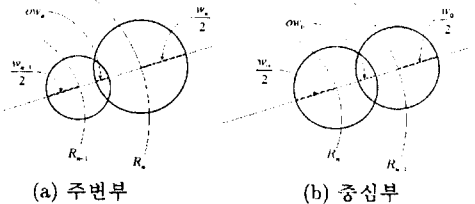


그림 1. 중심부와 주변부에서의 수용장 특성

다음 식 (1)과 (2)는 주변부에서 이심률(eccentricity)이  $n$ 일 때 수용장의 크기와 각 링에서의 수용장 개수들의 의미한다.

$$R_n = \frac{2(1-o)w_0 + (2+w_0k)R_{n-1}}{2 - (1-2o)w_0k} \quad (1)$$

$$K = \text{round}\left(\frac{2\pi R_0}{(1-o)(w_0 + w_0kR_0)}\right) \quad (2)$$

또한, 식 (3)과 (4)는 중심부에서 이심률이  $n$ 일 때의 수용장 크기와  $n$ 번째 링에서의 수용장 개수를 나타낸다.

$$R_n = R_{n-1} - \frac{R_0}{\text{round}(R_0/(1-o)/w_0)} \quad (3)$$

$$K_n = \text{round}\left(\frac{2\pi R_n}{(1-o)w_0}\right) \quad (4)$$

식 (1) ~ (4)에서,  $w_n$ ,  $w_0$ ,  $R_0$ 는 각각  $n$ 번째 링의 수용장 크기, 중심부의 수용장 크기, 중심부의 크기이다.  $k$ 는 비례상수이며,  $o$ 는 수용장의 겹치는 정도를 나타내는 상수이다.  $o$ 가 1이면 수용장은 완전히 겹쳐지며, 0이면 겹쳐지지 않는 상태가 된다. 그리고  $\text{round}(\cdot)$  함수는 반올림을 의미한다.

### 3.2 다해상도 영상에서의 변형 특성

영상 내의 객체 영역이 움직이게 되면 비록 그 움직임이 직교 영상 좌표계에서는 단순한 이동이라 할지라도 다해상도 영상 좌표계에서는 크기 및 형태의 변형이 일어나게 된다. 따라서 이러한 변형의 특성을 분석할 필요가 있다.

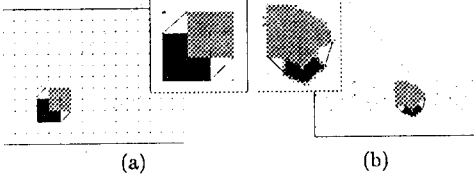


그림 2. (a) 직교 영상 좌표계에서의 움직임, (b) 다해상도 좌표계에서의 변형된 움직임

그림 3을 통하여 직교 영상 좌표계상의 움직인 위치에 대응되는 다해상도 영상 좌표계상의 변형된 위치를 쉽게 알 수 있다. 다해상도 영상 좌표계에서 주어진 위치의 반경  $R$ 과 각도  $\theta$ 를 매핑 함수 (1)과 (2)를 통하여 다음 식과 같이 얻을 수 있다.

$$R = SM_R(\xi) = \sum_{n=1}^{\xi} ab^{n-1} + R_0 b^{\xi} = \frac{(a + bR_0 - R_0)b^{\xi} - a}{b - 1} \quad (5)$$

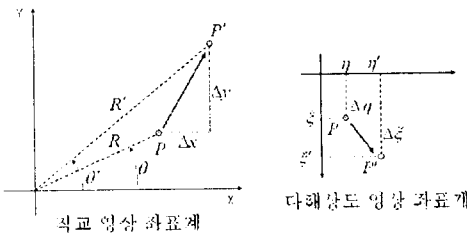


그림 3. 직교 영상 좌표계상의 위치에 대응되는 다해상도 영상 좌표계상의 위치

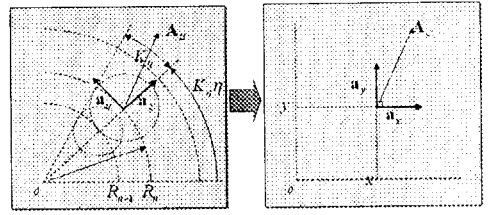


그림 4. 다해상도 영상 좌표계의 벡터에서 직교 영상 좌표계의 벡터로 변환

$$\theta = SM_{\theta}(\eta) = \frac{2\pi}{K}\eta \quad (6)$$

여기서  $a = \frac{2(1-o)w_0}{2-(1-2o)w_0k}$ ,  $b = \frac{2+w_0k}{2-(1-2o)w_0k}$  이다. 따라서,  $\Delta x$ ,  $\Delta y$ 만큼 움직인 직교 영상 좌표계상의 위치 ( $R'$ ,  $\theta'$ )는 다음과 같이 쉽게 구할 수 있다.

$$R' = \sqrt{(x + \Delta x)^2 + (y + \Delta y)^2} \quad (7)$$

$$\theta' = \arctan\left(\frac{y + \Delta y}{x + \Delta x}\right) \quad (8)$$

여기서  $x = R \cos \theta$ ,  $y = R \sin \theta$  이다. 위의 식을 통해 다해상도 영상 좌표계상의 변형된 위치 ( $\xi'$ ,  $\eta'$ )를 다음과 같이 얻을 수 있다.

$$\xi' = SM_R^{-1}(R') = \log_b\left(\frac{(b-1)R' + a}{(a + bR_0 - R_0)}\right) \quad (9)$$

$$\eta' = SM_{\theta}^{-1}(\theta') = \frac{K}{2\pi}\theta' \quad (10)$$

마지막으로, 다해상도 영상 좌표계상의 변량 ( $\Delta\xi$ ,  $\Delta\eta$ )가 다음 식과 같이 얻어진다.

$$\Delta\xi = SM_R^{-1}(R') - \xi \quad (11)$$

$$\Delta\eta = SM_{\theta}^{-1}(\theta') - \eta \quad (12)$$

위에서 보인 식을 통하여, 직교 영상 좌표계 상의 이동 변량 ( $\Delta x$ ,  $\Delta y$ )가 주어졌을 때, 다해상도 영상 좌표계상의 위치 ( $\xi$ ,  $\eta$ )에서 이동된 변량 ( $\Delta\xi$ ,  $\Delta\eta$ )를 구할 수가 있다.

## 4. 움직이는 객체 탐지

### 4.1 다해상도 영상에서의 움직임 추정

움직임 추정을 위해 Horn and Schunk의 광류(optical flow) 방법[8]을 사용하였다. 그리고 효율적인 움직임 영역 분할과 평균 속도 벡터 계산을 위하여, 그림 4와 같이 다해상도 영상 좌표계의 광류 벡터들을 직교 영상 좌표계로 변환한다.

다해상도 영상 좌표계의 벡터  $A_S = A_{\xi}a_{\xi} + A_{\eta}a_{\eta}$ 는 극 좌표계의 벡터  $A_S = A_{\xi}(R_n - R_{n-1})a_{\rho} + A_{\eta}K_{\eta}a_{\phi}$ 로 표현할 수 있다. 그리고 이 벡터는 직교 좌표계의 벡터  $A_C = A_x a_x + A_y a_y$ 로 쉽게 변환할 수 있다. 여기서  $A_x$ 와  $A_y$ 는 다음과 같이 결정된다.

$$\begin{aligned} A_x &= A_S \cdot a_x \\ &= A_{\xi}(R_n - R_{n-1})a_{\rho} \cdot a_x + A_{\eta}K_{\eta}a_{\phi} \cdot a_x \quad (13) \\ &= A_{\xi}(R_n - R_{n-1})\cos(K_{\eta}\eta) - A_{\eta}K_{\eta}\sin(K_{\eta}\eta) \end{aligned}$$

$$\begin{aligned}
 a_y &= A_s \cdot a_y \\
 &= A_\xi(R_n - R_{n-1})a_p \cdot a_y + A_\eta K_\eta a_\phi \cdot a_y \quad (14) \\
 &= A_r(R_n - R_{n-1}) \sin(K_\eta \eta) + A_\eta K_\eta \cos(K_\eta \eta)
 \end{aligned}$$

4.2 움직이는 객체 영역 분할

움직이는 객체의 영역을 분할하기 위해, 영역 기반의 분할 방법을 사용한다. 그림 5와 같이 이전 광류 맵을 구성한 뒤, 필터링과 연결 요소 분석을 통하여 분할된 영역에 대한 맵을 얻어낸다.



그림 5. 영역기반의 움직임 영역 분할

5. 다중 객체 추적

5.1 동적 대상 객체 모델

다중 객체 추적을 위해 객체 영역의 밝기 정보를 이용하여 동적인 대상 객체 모델(dynamic target model)을 구성한다. 이러한 대상 객체 모델은 추적 과정이 진행되는 동안 지속적으로 갱신되게 된다. 모델은 다음 식을 통하여 정의된다.

$$\Psi^i(\xi, \eta) = wI(\xi, \eta) + (1 - w)D_{\Psi^{i-1}, \Delta P}(\xi, \eta) \quad (15)$$

여기서  $\Psi^i(\xi, \eta)$ 는 새로 갱신되는 대상 객체 모델이며,  $I(\xi, \eta)$ 는 현재 입력 영상에서 탐지된 객체 영역을 말한다.  $D_{\Psi^{i-1}, \Delta P}(\xi, \eta)$ 는 이전의 대상 객체 모델이 현재의 위치로 이동해옴에 따라 변형된 모델을 의미한다. 그리고  $w(0 < w < 1)$ 은 가중치이다.

5.2 움직이는 대상 객체의 정합

현재 프레임에서 탐지된 객체와 대상 객체 모델과의 정합을 위하여, 다음과 같이 정의되는 상관관계를 이용한다.

$$C_{ij} = \frac{\sum_{\xi, \eta \in \Psi} (I(\xi + i, \eta + j) - D_{\Psi, \Delta P}(\xi, \eta))^2}{N^2} \quad (16)$$

여기서  $N$ 은 대상 객체 모델의 화소수를 말하며,  $W$ 는 탐색창의 크기이다.  $C_{ij}$ 가 특정 임계치 이하가 되었을 때, 탐지된 객체가 현재 대상 객체와 동일하다고 결정한다. 이때, 대상 객체 모델은 탐지된 객체로 갱신되게 된다.

6. 실험결과 및 분석

실험 시스템은 펜티엄 III 500MHz PC와 Matrox Meteor II 영상취득기 및 4자유도를 가지는 헤드아어 시스템을 사용하였으며, 실험 대상으로 실내환경에서 움직이는 사람들을 선정하였다.

그림 6은 두사람이 서로 반대 방향으로 엇갈려 지나갈 때 추적 모습을 보여준다. 그림에서 볼 수 있듯이 움직이는 사람들은 놓치지 않고 잘 추적하고 있음을 알 수 있다. 그러나 중첩(occlusion)의 경우는 고려되지 않았기 때문에 두 사람이 근접했을 때는 하나의 객체로 인식하게 된다. 보다 나은 추적성능을 위해 중첩의 문제도 고려되어야 할 것이다.



그림 6. 추적 결과: (a) 추적 결과 영상, (b) 추출된 객체 영역

7. 결론 및 향후 연구

본 논문에서는 다해상도 영상을 이용한 다중 객체 추적 시스템을 저가의 PC환경에서 구현해 보았다. 다해상도 영상 좌표계에서 직접 광류를 사용함으로써 보다 효율적인 움직임 탐지를 할 수 있었으며, 다해상도 영상 좌표계에서 나타나는 이동상의 변형 특성을 고려함으로써 직접적인 정합을 효과적이고 쉽게 할 수 있었다. 그러나 보다 나은 성능을 위해, 중첩 문제 등이 고려되어야 할 것이다.

참고 문헌

- [1] M. Bolduc and M. D. Levin, "A Real-time foveated Sensor with Overlapping Receptive Fields," *Real-Time Imaging*, Vol. 3, No. 3, 1997, pp. 195-212
- [2] E. L. Schwartz, "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception," *Biological Cybernetics*, Vol. 25, 1977, pp 181-194
- [3] F. Panerai, C. Capurro and G. Sandini, "Space Variant Vision for an Active Camera Mount," *Technical Report TR1/95, LIRA-Lab-DIST University of Genova*, 1995
- [4] F. L. Lim, G. A. W. West and S. Venkatesh, "Tracking in a Space Variant Active Vision System," *Proc. of 13th International Conference on Pattern Recognition*, Vienna, Austria, 1996, pp. 745-749
- [5] F. Jurie, "A new log-polar mapping for space variant imaging: Application to face detection and tracking," *Pattern Recognition*, Vol. 32, 1999, pp 865-875
- [6] F. Bremond and M. Thonnat, "Tracking Multiple Non-rigid Objects in Video Sequences," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 8, No. 5, 1998, pp. 585-591
- [7] I. Haritaoglu, D. Harwood and L. S. Davis, "Hydra: Multiple People Detection and Tracking Using Silhouettes," *Proc. of 2nd IEEE Workshop on Visual Surveillance*, Fort Collins, Colorado, 1999, pp. 6-13
- [8] A. M. Tekalp, *Digital Video Processing*, Prentice Hall, 1995