

모바일 환경 신뢰도 평가 학습에 의한 다중 객체 추적

한우리 · 김영섭[†] · 이용환^{*}

[†]단국대학교 전자공학과, ^{*}극동대학교 스마트 모바일학과

Multi-Object Tracking based on Reliability Assessment of Learning in Mobile Environment

Woo ri Han, Young-Seop Kim[†] and Yong-Hwan Lee^{*}

[†]Dankook University of Electronic Engineering

^{*}Far East University, Department of Smart Mobile

ABSTRACT

This paper proposes an object tracking system according to reliability assessment of learning in mobile environments. The proposed system is based on markerless tracking, and there are four modules which are recognition, tracking, detecting and learning module. Recognition module detects and identifies an object to be matched on current frame correspond to the database using LSH through SURF, and then this module generates a standard object information that has the best reliability of learning. The standard object information is used for evaluating and learning the object that is successful tracking in tracking module. Detecting module finds out the object based on having the best possible knowledge available among the learned objects information, when the system fails to track. The experimental results show that the proposed system is able to recognize and track the reliable objects with reliability assessment of learning for the use of mobile platform.

Key Words : Multi-Object Tracking , LSH (Locality Sensitive Hashing), SURF (Speeded-Up Robust Features), Optical Flow

1. 서 론

객체 인식과 추적은 사진 혹은 동영상에서 특정 객체를 판별하여 인식하고, 인식된 객체의 움직임을 추적, 관측하여 그 궤도 및 위치를 정하는 것을 의미한다. 컴퓨터 비전을 기반으로 하는 추적 방법은 다양한 분야로 사용되고 있으며 그 방법 또한 다양하다. 현재 모바일 기기의 성능이 발전함에 따라 컴퓨터 비전기반의 추적기술을 모바일에 적용하여 다양한 콘텐츠와 시장을 형성 하고 있지만 모바일 환경에서의 추적기술은 마커기반의 추적기술이 많은 비중을 차지 하고 있으며, 마커를 이용함으로써 발생하는 제약 때문에 다양한 분야에서 활용이 어렵다. 비 마커 방법의 객체 추적 방법은 일반적으로 정확도를 높이기 위해선 계산 비용이 증가되고, 이는 정합의 효율성을 감소시킨다. 반면, 인식 속도를

향상시키기 위해 정확도에 대한 비중을 낮추면 잘못된 객체 인식 오류를 발생 시켜 오정합 결과를 초래하기 때문에 객체들의 특징점을 추출하여 객체를 정확하고 빠르게 인식하기 위한 방법들이 제안되었다[1].

이는 크게 점(0D feature), 선(1D feature), 영역(2D feature)을 이용하는 방법들로 분류 할 수 있다. 먼저 점을 추출하는 방법에는 Harris Corner detector [2], Hessian detector [3], Kitchen Rosenfeld [4], SUSAN (Smallest Univalued Segment Assimilating Nucleus) [5] 등이 있다. 점과 선을 이용한 방법들은 계산 비용이 적어 빠르게 수행 될 수 있으나 영상내 객체의 크기와 회전등이 변화된 영상 내에서는 정확도가 떨어지는 단점이 있기 때문에 영역 내에서 특징들을 추출하고 영역 기술자를 생성하는 SIFT [6]와 SURF [7]등이 개발 되었다. 이러한 알고리즘들은 정확도가 높아 영상 스티칭(image stitching)을 위한 특징 탐지에서도 사용된다.

가우시안 DoG (Differential of Gaussian)영상을 이

[†]E-mail : wangcho@dankook.ac.kr

용하여 특징점을 추출하는 SIFT는 연산속도가 매우 느리다는 단점이 있어 실시간 시스템에서는 적용되기 힘들다. SIFT의 단점을 줄이고자 개발된 SURF는 Box filter를 이용한 고속 헤이시안 탐지기(fast hessian detector)를 통해 특징점의 탐지과정과 기술자 생성 과정을 단순화 하여 수행 속도를 향상 시켰다[8].

본 연구에서는 객체 추적을 위해 모바일 환경에서의 비 마커 기반 객체 인식과 추적, 신뢰도를 평가한 뒤 학습을 통한 객체 검출을 이용한 다중 객체 추적 방법을 제안한다. 먼저 다양한 객체를 학습시켜 얻은 database를 기반으로 추적하고자 하는 객체를 LSH (Locality Sensitive Hashing)을 이용하여 빠르게 탐색 및 인식 한 뒤, 인식된 객체 정보를 기반으로 추적한다. 추적을 하면서 추적에 성공한 객체 정보의 신뢰도를 평가함과 동시에 학습 시켜 학습된 객체 정보의 신뢰도를 현재 영상과 비교 평가하여 객체 추적에 실패할 경우 신뢰 할 수 있는 객체를 다시 검출하고, 검출된 객체정보를 기반으로 다시 추적함으로써 모바일 환경에서 보다 정확하고 신뢰 할 수 있는 객체 추적 시스템을 제안한다.

본 논문의 2장에서는 추적할 객체에 대한 SURF를 이용하여 탐색 및 인식 하는 방법에 대하여 설명하고, 3장에서는 추적과 신뢰도를 평가 및 학습하는 방법을, 4장에서는 추적에 실패할 경우 신뢰도 평가에 의한 학습된 이미지 정보를 기반으로 탐색하여 재 추적하는 방법을 설명한다. 마지막으로 5장에서는 결론 및 향후 과제에 대해 논의 한다.

2. SURF와 LSH (Locality Sensitive Hashing)

현재 영상에 대해 SURF알고리즘을 기반으로 한 128 차원의 기술자와 특징점을 추출한 뒤, 그 정보를 기반으로 미리 학습된 Database를 LSH 통해 많은 양의 데이터를 고속으로 탐색하여 객체를 인식 한다.

2.1. SURF

SURF 알고리즘은 특징점을 추출하는 feature extraction 부분과 특징점 주변 영상에 대한 정보를 서술하는 feature description 부분으로 나눌 수 있다.

Feature extraction 부분에서 적분 영상을 생성하는 이유는 SURF 알고리즘에서 박스 타입의 필터와 컨벌루션 연산을 빠르게 수행하기 위함이며, 고속 헤이시안 행렬(Fast-Hessian Matrix)에서는 x, y, xy 축으로 근사화된 2차 미분 LoG (Laplacian of Gaussian) 박스필터

와 적분 영상간의 컨벌루션 연산을 통해 헤이시안 행렬을 구한 후 행렬의 디터미넨트 값을 계산한다. 그리고 Non-maxima Suppression을 통해 헤이시안 행렬식(Determinant) 이 임계값보다 크고 인접 스케일과 비교하여 디터미넨트가 가장 클 경우 특징점으로 결정된다. Feature Description 부분 에서는 추출된 특징점의 주 방향(dominant direction)을 결정하고 그 주변 영상의 서술자를 생성한다[9-10].

2.2. Locality Sensitive Hashing (LSH)

LSH는 특정 벡터간 유사성이 보존되도록 설계된 해시 함수를 통해 최소거리의 객체가 검색 대상 특징 벡터와 동일한 해시 함수 값을 가질 확률이 높도록 개선한 해싱 방법이다.

먼저 d차원의 특징벡터 $p = (x_1, \dots, x_d)$ 를 식 (1)과 같은 변환함수 $v(p)$ 을 이용하여 d차원의 해밍 공간으로 변환한다.

$$p' = v(p) = \text{Unary}_c(x_1) \cdots \text{Unary}_c(x_d) \quad (1)$$

여기서 d'은 $C \cdot d$ 로 계산된다. C는 x_i 가 가질 수 있는 최대 값이다. Unary_c 를 x개의 1과 (C-x)개의 0으로 이루어진 길이 C인 비트열을 나타낸다.

해밍거리를 이용하는 LSH는 주어진 데이터를 해밍 공간에 매핑하여 실제 데이터를 이진 표현으로 나타낸다. 이 때, 실제 데이터 p는 d개의 값을 갖는 벡터이며, 각 차원의 값은 모든 데이터 값 중 가장 큰 값 C의 길이의 2진수로 변환된다 p의 이진표현 $v(p)$ 는 이러한 C 길이의 2진수 d개를 결합한 길이 dC의 2진수로 표현된다. 각 차원의 값은 해당 값 만큼의 연속된 1과 C에서 해당 값을 뺀 만큼의 연속된 0으로 이루어진다. 해밍 거리를 이용하는 해시 함수 집합은 이렇게 변환된 p의 이진표현으로부터 미리 임의로 정해진 k개의 비트를 선택하는 방식으로 정의된다.

해밍 거리를 이용하는 LSH는 해밍거리 해시 함수를 L개 정의하고 각각의 해시 함수에 상응하는 해시 테이블을 L개 생성한다. 전체 데이터에 대해 L개의 해시 테이블을 이용한 인덱스를 생성한 이후 쿼리 데이터 아이템 q가 주어지면 q를 L개의 해시 함수를 통해 해시하여 나온 해시 값에 해당하는 L개의 해시 버킷들에 들어있는 아이템들을 후보 아이템들로 간주하고, 이러한 아이템들의 수는 전체 아이템 수에 비해 매우 적기 때문에 k-최근접 검색 방법에서 굉장히 빠른 속도를 가진다. p-Stable 함수를 이용하는 LSH는 해시 함수를 구성하기 위한 k개의 요소를 각각 다음 식(2) 통해 계산한다.

$$h(a, b)(v) = ((a * v + b)/r) \tag{2}$$

이때, 주어진 데이터의 벡터표현 v 를 미리 임의로 정해진 같은 길이의 벡터 a 와 역시 임의로 정해진 상수 b 를 이용하여 하나의 값으로 바꾸고, 이를 r 개의 구간으로 나누어 해석한다.

이렇게 여러 종류의 객체 정보를 각각 인덱스와 그에 따른 SURF를 통한 특징점, 128차원의 기술자를 학습시켜 Database에 저장한 뒤, 현재 영상에 대한 정보와 Database를 통한 정보를 LSH를 사용하여 빠르게 검색을 한다. LSH를 통한 객체 검출 및 인식 결과에서 추출된 객체의 정보는 가장 신뢰 할 수 있는 Level 1의 정보로 학습되고 이 정보는 신뢰도를 평가 할 때 기준 정보가 되어 추적에 성공하여 학습될 객체의 정보를 비교 및 평가되는데 사용된다[11].

3. 추적 및 추적 객체 신뢰도 평가 학습

객체 검출 및 인식 과정에서 추출한 객체의 정보 특징점, ROI, 카메라의 위치 정보, 히스토그램 정보 등을 기반으로 광류를 이용한 객체 추적을 한다. 추적에 성공하게 되면 학습을 통하여 추적에 성공한 객체를 평가 학습하게 되는데, 평가는 현재 성공한 객체에 대해 k-means를 통해 히스토그램을 군집화하고 각 특징점이 위치한 히스토그램 라벨과 신뢰도 Level0의 히스토그램 라벨과의 비교를 통해 매칭된 히스토그램의 라벨이 다양하고 고르게 분포 되어 있을수록 높은 점수의 신뢰도를 가지며, Level0의 신뢰도를 갖는 객체에 대하여 매칭된 특징점의 분포가 넓을수록 높은 점수의 신뢰도를 갖는다. 또한 PCA를 통한 매칭 영역의 고유 벡터가 유사 하고, 추적에 성공한 영상이 가장 최신일수록 높은 점수의 신뢰도를 얻게 된다.

3.1. 객체 추적

객체 추적은 옵티컬 플로우의 LK 알고리즘을 통해 수행되고 3가지 가정에 기초하고 있다.

첫 번째 밝기 항상성(Brightness constancy)은 추적하고 있는 물체의 픽셀의 밝기 값은 프레임이 변하더라도 변하지 않는다는 가정이다. 이러한 가정에 의해 다음의 식 (3)이 유도된다.

$$f(x, t) = I(x(t), t) = I(x(x + dx), t + dt) \tag{3}$$

여기서, $\partial(x)/\partial t = 0$

두 번째, 시간 지속성(Temporal persistence)은 영상

에서 물체의 움직임에 비하여 시간의 변화가 더 빠르게 진행되며, 이는 연속된 프레임 사이에서 물체의 이동량이 많지 않음을 의미한다. 이러한 가정을 2차원으로 확장하여 y축 방향으로의 속도 성분을 v 라 하고, x축 방향으로의 속도 성분을 u 라고 표시하면 다음 식 (4)가 성립된다.

$$I_x u + I_y v + I_t = 0 \tag{4}$$

여기서 원하는 y축 방향으로의 속도 성분 v 와 x축 방향으로의 속도 성분 u 를 구하려 할 때 식을 이용해서는 유일한 해를 구할 수 없다. 대신 법선 방향의 움직임 성분을 Fig. 1과 같이 구할 수 있다.

LK의 마지막 가정인 공간 일관성(spatial coherence)이 사용된다. 공간 일관성은 공간적으로 서로 인접하는 점들은 동일한 물체에 속할 가능성이 높고, 동일한 움직임을 갖는다는 가정이다. 식 ()을 이용하여 5×5 크기의 윈도우를 사용하면 다음과 같은 25개의 방정식 (5)를 유도 할 수 있다.

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ \vdots & \vdots \\ I_x(p_{25}) & I_y(p_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -I_t(p_1) \\ \vdots \\ -I_t(p_{25}) \end{bmatrix} \tag{5}$$

위 식을 풀기 위해 최소 자승법을 사용하며, $\min \|Ad-b\|^2$ 에 해당하는 값을 다음 식 (6)과 같이 구할 수 있다.

$$(A^T A)d = A^T b \text{ 다음과 같이,}$$

$$d = \begin{bmatrix} u \\ v \end{bmatrix} = (A^T A)^{-1} A^T b \tag{6}$$

$(A^T A)$ 의 역행렬이 존재하면 위 수식을 풀 수 있고

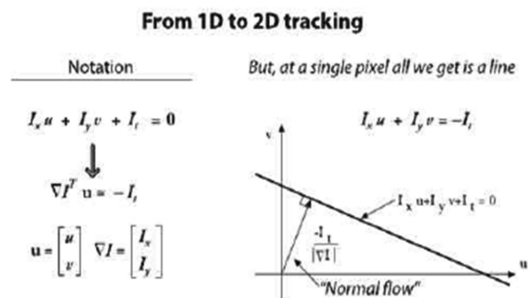


Fig. 1. Normal Flow.

원하는 y축 방향으로의 속도 성분 v 와 x축 방향으로의 속도 성분 u 를 구할 수 있다[12].

3.2. 신뢰도 평가 및 학습

신뢰도에 대한 평가 및 객체 학습은 객체 검출, 인식 과정에서 생성된 가장 높은 신뢰도를 갖는 Level0의 객체 정보를 기준으로 객체 추적에 성공한 객체를 평가하여 신뢰할 수 있는 객체인지 평가하여 학습을 하고, 신뢰도 수준을 결정 한다. Fig. 2는 추적에 성공하여 신뢰도를 평가한 뒤, 학습된 객체를 보여준다.

기준이 되는 Level 0의 객체 정보를 분석하여 픽셀을 Fig. 3와 같이 군집화 $K_i(k_1, k_2, k_3, \dots, k_i)$ 한다.

또한 Level 0의 객체 정보에 포함된 특징점의 좌표 $X_k(x_1, x_2, x_3, \dots, x_k)$ 를 군집화된 픽셀에 매칭하여 해당 특징점이 어느 픽셀의 집합에 속해 있는지 라벨링 Label을 수행하여 식 (7)과 같이 표시한다.

$$\begin{aligned} \text{Label}_0(k_1 \supset X_n(x_1, x_2, \dots, x_n)) \\ \vdots \\ \text{Label}_1(k_j \supset X_j(x_1, x_2, \dots, x_j)) \end{aligned} \quad (9)$$

Level0의 기준 객체와 추적에 성공한 객체 사이의 매칭된 특징점이 포함된 라벨의 종류가 많을수록 넓고 다양한 종류의 픽셀을 포함한 특징점이 매칭되었을 확률이 높기 때문에 높은 점수를 받고, 매칭된 특징점의 수가 많을수록 높은 점수를 받는다.



Fig. 2. Trained objects.

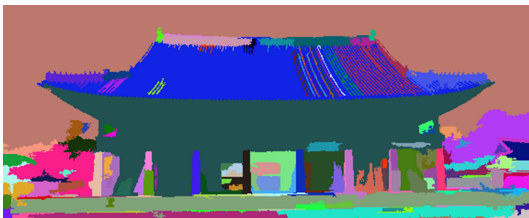


Fig. 3. Clustered pixel.

또한 매칭된 특징점을 기반으로 매칭 영역을 설정한 뒤, PCA(principal component analysis)를 수행하여 k 개의 고유벡터 $P_k = (C_1 \dots C_k)$ 를 추출하여 Level 0와 근사한 계수인 근사 계수 H 를 식 (8)과 같이 조사한다.

$$0 \leq H(P_{\text{level}0_k} = (C_1 \dots C_k), P_k = (C_1 \dots C_k)) \leq 1 \quad (8)$$

추적에 성공한 객체가 최신의 영상영상부터 추적에 성공한 객체 일수록 높은 점수를 받고, 신뢰도를 식 (9)와 같이 평가 및 결정한다.

$$\left(\begin{aligned} \text{Level}_{n-1} &= \frac{\text{NLabel}}{\text{total Label}} \times \frac{\text{frame}}{\text{total frame}} \times \frac{\text{matching} X_n}{X_k} \\ 0 &\leq H((P_{\text{level}1_k} = (C_1 \dots C_k), P_k = (C_1 \dots C_k)) \leq 1) \end{aligned} \right) \quad (9)$$

식 (9)와 같이 추적에 성공한 객체에 대한 신뢰도를 평가하여 Level0의 신뢰도를 갖는 객체를 기준으로 Level5까지 각각의 객체를 평가하여 총 6가지의 신뢰도 갖는 객체들로 구성한다. 이렇게 구성된 객체들은 추적에 실패 하였을 때나, 객체를 검출 또는 인식 할 때 사용된다.

4. 객체 탐색

객체가 추적에 실패 했을 때, 학습된 객체에 대한 정보를 이용하여 현재 영상과 학습된 객체간신뢰도를 재 평가하여 학습된 객체 정보를 정렬하고, 가장 신뢰 할 수 있는 학습된 객체 정보를 이용하여 현재 검출된 객체 정보를 업데이트 한다. 또한 정렬된 학습 객체 정보는 Level0, Level1, Level2의 신뢰도에 대한 객체 정보를 제외하고 초기화 한 뒤, 다시 객체 추적 모듈로 돌아가 업데이트된 객체의 정보를 기반으로 추적을 시행한다.

객체 탐색에 수행된 특징점과 기술자는 ORB를 사용하여 빠른 탐색 결과와 객체 및 객체를 포함한 영역(ROI)을 재 설정할 때 6자유도를 갖는 homography를 수행, 회전과 스케일 변화에 강건한 탐색을 수행 하였다.

5. 결 론

모바일 환경에서 Database를 기반으로 빠르고 다양한 객체를 추출 하고 인식한 뒤, 인식된 객체를 가장 신뢰도가 높은 Level0로 학습하고, 이를 기준으로 5가지의 신뢰도를 갖는 객체를 추적과 동시에 학습하였다. 또한 객체 추적에 실패 하였을 경우, 신뢰 할 수 있는 학습

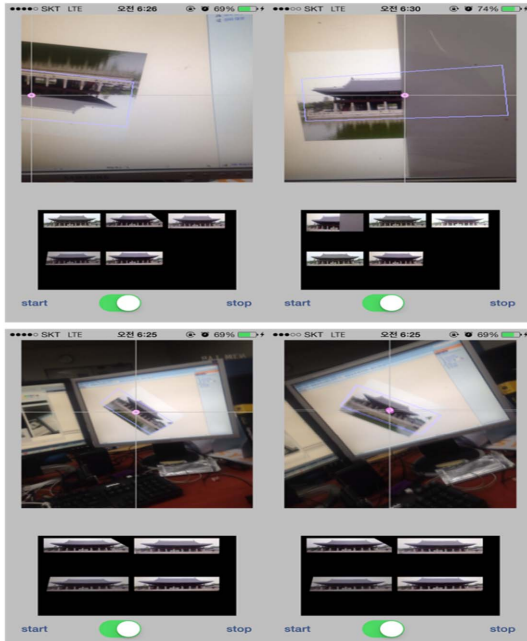


Fig. 4. Tracking in mobile environment.

된 객체 정보를 이용하여 객체를 재 탐색 하였기 때문에 재 탐색된 객체 또한 신뢰 할 수 있는 객체의 정보 라고 할 수 있으며, 특징점과 6 자유도를 갖는 Homography 정보를 이용하여 회전과 스케일 변화에 강건한 객체를 추적 할 수 있었다. Fig. 4는 모바일 환경에서 결과 영상이다. 실험은 IOS OS의 iPhone5 환경에서 수행하였으며 평균 15 fps의 결과를 얻었다.

감사의 글

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2013R1A2A2A 03068794).

참고문헌

1. Lee, S.-G., "Survey on mixed-reality R&D," *Journal of the Korea Computer Graphics Society*, 13(2), pp.

1-15, (2007).
 2. Harris, C. and Stephens, M., "A combined corner and edge detector," *Proc. Alvey Vision Conf.*, pp. 147-151, (1998).
 3. Lindeberg, T., "Feature detection with automatic scale selection," *International Journal of Computer Vision*, 30(3), pp.79-116, (1998).
 4. Kitchen, L. and Rosenfeld, A., "Gray level corner detection," *Pattern Recognition Letters*, 1(2), pp. 95-102, (1982).
 5. Smith, S. M. and Brady, J. M., "SUSAN-A new approach to low level image processing," *International Journal of Computer Vision*, 23(1), pp. 45-78, (1997).
 6. Lowe, D., "Distinctive image feature from scale-invariant keypoints," *International Journal of Computer Vision*, 60(2), pp. 91-110, (2004).
 7. Bay, H., Tuytelaars, T. and Cool, L. V., "SURF: speeded up robust features," *ECCV*, 3951, pp. 404-417, (2006).
 8. Lee, H. J., Lee, S.-G., "Improvement method of tracking speed for color object using Kalman filter and SURF," *Journal of Korea Multimedia Society*, 15(3), pp. 336-344, (2012).
 9. Do, Y.-S. and Jeong, Y.-J., "Hardware design of SURF-based feature extraction and description for object tracking," *Journal of The Institute of Electronics Engineers of Korea*, 50(5), (2013).
 10. Lee, Y.-H., Park, J.-H. and Kim, Y.-S., "Comparative analysis of the performance of SIFT and SURF," *Journal of the Semiconductor & Display Technology*, 12(3), pp. 59-64, (2013).
 11. Ji-Won Hong, Byung-Moon, Sang-Wook Kim, "A study on LSH parameters for large multimedia databases," *The Korea Contents Association*, 2015(5), (2015).
 12. Choi, J. and Cho, Y., "Moving object recognition and tracking algorithm using parallel processing of SURF and optical flow," *Proceedings of KIIS Fall Conference*, 21(2), (2011).

접수일: 2015년 9월 4일, 심사일: 2015년 9월 14일,
 게재확정일: 2015년 9월 22일