

연속적인 비디오 프레임에서의 히스토그램을 이용한 객체 인식 및 추적

종신회원 박 호 식*, 배 철 수**

Object Recognition and Tracking using Histogram Through Successive Frames

Ho-sik Park*, Cheol-soo Bae** *Lifelong Members*

요 약

히스토그램에 의한 객체 유형 인식 방법은 최근 들어 많은 연구가 이루어지고 있다. 그러나 대부분의 히스토그램 기반의 객체 추적이 칼라 모델을 사용하여 견실성을 개선하였지만 아직 충분히 견실하다고 할 수 없다. 이러한 단점을 보완하기 위하여 본 논문에서는 연속적인 프레임에서 히스토그램을 이용하여 객체를 표현하고 추적하는 방법을 제시하고자 한다. 자동차를 대상으로 실험한 결과 80m 거리 이내에서 신뢰성 있는 방법임을 확인하였다.

Key Words : object recognition, tracking, svm, mean-shift, color histogram

ABSTRACT

Recently, the research which concerns the object class recognition has been done. Although an object tracking based on most of histograms employs a colored model to improve robustness, the system is not reliable enough yet. In this paper, we presents a method to express and track an object by using the histograms which are composed with visual features through successive frames. The experimental results shows that this method is reliable to track a car within 80m distance from camera.

1. 서 론

실시간 영상에서 객체 인식 및 추적은 수년간 컴퓨터 비전 및 여러 실용적 응용 분야에서 관심을 가지는 주제 중 하나이다. 실제로 실시간 영상내의 객체 추적은 빠른 처리와 많은 연산을 요구하고 고가의 장비가 필요하기 때문에 많은 어려움이 따른다. 연산량을 줄이고자 객체를 모델링 하였다. 예를 들어, 특징적인 행동에 의한 모델^[1]이나 데이터 학습에 의한 모델^[2]을 대상으로 하는 경우가 있다. 그리고 윤곽선을 기반으로 한 모델링은^[3]은 추적된 객체의 외형 윤곽선을 이용하여 객체를 모델링한다.

그리고 최근에는 외관 모델의 형태 형성을 위하여 히스토그램을 이용하여 모델링을 하고 있다^[4].

히스토그램을 이용한 모델의 분명한 장점은 모델의 단순성 및 다양한 응용성이라 할 수 있다. 이러한 특징의 관찰 모델이 보이는 다른 장점은 추적된 객체의 히스토그램과 가장 유사한 히스토그램을 갖는 영상 영역을 찾는 효과적인 국부 탐색구조를 보이는 것이라 할 수 있다^[4].

컴퓨터비전에 의한 인식에는 객체 자신을 인식하는 객체 인식 및 객체의 유형을 인식하는 객체 유형 인식으로 구분할 수 있다. 객체 및 객체 유형 인식은 두 가지의 다른 문제를 갖는다. 객체 인식의 경우 다

* 오산대학 디지털전자과 (hspark@osan.ac.kr),

** 관동대학교 전자통신공학과 (baecs@kd.ac.kr)

논문번호 : KICS2008-12-569, 접수일자 : 2008년 12월 29일, 최종논문접수일자 : 2009년 3월 6일

른 객체들 또는 배경으로부터 개별 객체를 분리하는 모델이 필요하고 컬러 히스토그램을 주로 사용한다. 그러나 객체 유형 인식의 경우 객체의 유형의 영상 범위를 분할하는 유형 모델이 필요하게 된다.

인식을 위한 가장 바람직한 방법은 시각적 특징 구성^[5]이라 할 수 있다. 각각의 유형을 위한 학습 영상으로부터 추출된 국부적 서술자를 선택하고 클러스터링 알고리즘을 위한 데이터로 사용한다. 이 클러스터는 시각적 특징을 나타낸다. 하나의 새로운 영상은 이러한 클러스터를 통하여 히스토그램으로 표현될 수 있으며 각각의 영상이 갖는 서술자는 일부 거리 척도를 사용하여 해당 클러스터에 할당된다. 마지막으로, 영상으로부터 생성된 히스토그램은 영상의 내용을 추론하기 위한 영상 서술자로 사용된다.

객체 인식은 흔히 객체 추적 알고리즘의 일부가 되며 연속적인 비디오 프레임으로부터 특정 단일 객체를 인식하는데 적용된다. 나아가 객체의 추적에 있어서의 일부 객체 유형 인식 알고리즘은 어떤 유형의 객체가 추적되었는지를 결정하는데 적용되기도 한다.

본 논문에서는 추적과 객체 유형 인식의 구조에 대하여 제안한다. 객체 유형 인식을 위하여 시각적 특징인 히스토그램^[5]을 사용하여 객체 유형 인식에 의해 객체를 추적 하고자 하였다. 최근 대부분의 히스토그램 기반의 객체 추적 응용은 타겟 모델의 표시를 위하여 컬러를 사용한다. 컬러 히스토그램은 특정 객체의 외관을 표시하는데 그리 현실한 특성을 보이지 않으며 객체 유형 인식에는 특히 좋은 방법이 아니다. 그러므로 본 논문에서는 객체 추적에 있어서 히스토그램은 시각적 특징을 사용하여 구성하였다. 본 논문에서 보이는 구조는 객체 추적과 객체 유형 결정에 시각적 특징을 사용하여 추적을 수행한다.

II. 객체 유형 모델

객체 추적의 중요한 요소는 타겟 표현과 타겟 지정의 두 가지로 구분될 수 있다. 본 논문에서는 사용된 구조 내에서 객체의 분류를 시도하기 때문에 이러한 타겟 표현에 제약을 가지게 된다. 이러한 표현은 내부 유형 변이의 적용에 안정적이어야 하며 여러 객체 유형들의 구분에 충분히 차이를 보일 수 있어야 한다.

본 논문에서는 외관 기반 모델인 bag of keypoints 방법^[5]을 사용하였으며, 객체의 표현 및 유형 모델을 나타내기 위한 texton 기반 영상 서술자를 이용하였다.

2.1 타겟 표시

3D texton은 T. Leung과 J. Malik [7]에 의해서 처음 소개되었다. 이러한 3D texton에 의한 영상 특징은 본래 텍스처의 분류를 위하여 고안되었지만 이후 특정적 결과에 있어서 객체 분류를 위하여 적용되고 있다.

Texton의 기본 개념은 다음과 같다. 하나의 텍스처 또는 객체의 부류는 영상 필터들의 집합에 대한 반응으로서 규정된다. 기본 반응 벡터들 그리고 이때 반응에서 필터 공간의 유사한 반응에 존재하는 가장 높은 빈도 또는 최소한의 존재를 texton 또는 시각적 특징이라 부른다.

이의 적용 목적은 texton 이라 불리는 특정 필터에 대한 원형 반응이라 규정되는 마이크로 구조의 특징을 구축하기 위함이다. 상이한 텍스처/객체 부류들은 필터 반응의 연결 분산으로 나타내어진다.

2.2 필터들의 집합

본 논문에서 사용한 필터들의 집합은 11개의 개별 필터로 구성되지만 L, a, b 컬러 모델의 서로 다른 컬러 채널에 대한 필터의 중첩에 기인하여 총 17 개의 필터 반응을 나타낸다. 이들 필터는 그림 1에 나타낸 바와 같다. 필터 뱅크는 3 개의 Gaussians ($\sigma = 1, 2, 4$), 4 개의 Laplacian of Gaussians ($\sigma = 1, 2, 4, 8$), 그리고 4 차의 Gaussians 미분 방정식 ($\sigma_x = 2, 4$ and $\sigma_y = 2, 4$)으로 구성된다. 이때의 3 개의 Gaussians은 L, a, b의 세 개의 컬러 채널 모두에 적용된다. 그리고 4 개의 Laplacian of Gaussians 및 4 차의 Gaussian 미분 방정식은 L 채널에만 적용된다.

2.3 Texton 구성

Texton은 필터 반응의 원형으로 정의되기 때문에 본 논문에서는 각각의 서로 다른 픽셀에서 모든 필터의 반응이 다르게 나타나리라고 기대하지 않는다. 각각의 필터의 반응은 국부적으로 추출되고 한 개의 픽셀 값에만 의존하지 않는다. 그러므로 다른 필터에 대한 근사로서 추출된 필터의 반응은 크게 다르게 나타나지 않는다. 그러나 영상을 통하여 나타

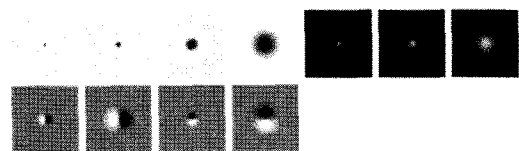


Fig. 1. Filter Bank
그림 1. 필터들의 집합

나는 필터의 반응은 매우 다르게 나타난다. 결국 이러한 결과는 여러 개의 매우 상이한 필터 반응이 존재하는 것을 암시하며 이러한 반응에는 다른 노이즈 형태의 변이가 있음을 의미한다.

명백한 반응 벡터들은 의도한 시각적 특징 또는 *texton*을 말한다. 모든 반응 벡터들은 여러 원형 반응 벡터들의 노이즈 형태의 변이라는 가정으로부터 본 논문에서는 필터 반응 공간에 있어서의 클러스터링을 통하여 이러한 원형들을 찾을 수 있었다. 결국, 시각적 특징은 얻어진 클러스터들의 중심을 의미한다.

데이터 집합에 나타난 각각의 학습 영상은 앞에서 언급한 필터들의 집합으로 둘러싸인 것이며 이는 필터 반응의 집합을 구하기 위한 것이다. 이러한 필터들은 시각적 기준 구축을 위한 트레이닝 데이터로부터의 모든 트레이닝 영상에 반응을 보인다. 전체 트레이닝 집합의 필터 반응은 *k*-평균 클러스터링을 사용하는 클러스터이다. 얻어진 클러스터 중심들은 본 논문에서 사용한 기준 *texton* 및 시각적 특징을 나타낸다.

본 논문에서는 *k*-평균 클러스터링을 위하여 자체적인 알고리즘을 사용하지 않고 *k*-평균의 급속 구현 기법을 적용하였다.

III. SVM을 이용한 객체 추적

특징의 구성이 이루어지면 다음 단계로서 영상 서술자를 이끌어낸다. 영상 서술자는 분류와 지정이 이루어지는 영상의 설명이나 특성을 나타낸다. 각각의 영상에 대한 이러한 영상 서술자는 서로 다르게 나타나지만 이는 영상의 부류가 다름을 기본으로 한다. 이러한 영상 서술자는 동일한 영상 부류에 있어서 유사한 영상 패턴을 보이며 상이한 부류의 영상에 대하여는 다른 패턴을 나타낸다.

시각적 특징 구축되면 영상 서술자의 규정은 간단히 이루어질 수 있다. 영상 서술자는 이전의 단계에서 만들어진 시각적 특징의 사용에 의하여 만들어진다. 본 논문에서는 각각의 픽셀에 대한 필터의 반응에 의한 적절한 필터 벡터를 사용한 영상을 사용한다. 이러한 단계에서 필터의 반응은 시각적 특징에 매핑되어 영상 서술자를 규정할 수 있게 하는 것을 가능하게 한다. 필터 반응을 시각적 특징에 매핑하는 것은 각각의 반응 벡터에 대한 대응되는 시각적 특징을 찾는 것을 의미한다. 이는 각각의 반응 벡터들 사이의 거리와 기준에서 시각적 특징을 구하여 이루어진다. 시각적 특징에 대한 각각의 필터

반응은 매우 긴밀하다. 마지막으로, 영상 서술자를 통합하였는데 이는 이러한 영상 서술자들을 시각적 특징 빈도의 벡터로 간주하여, 영상 내의 각각의 시각적 특징 발생 횟수를 계산하여 이루어졌다. 통상적으로 이러한 빈도 벡터는 서술자 히스토그램으로 불린다.

본 논문에서는 추적에 서술자 히스토그램을 적용하였기 때문에 영상에 대한 서술자 히스토그램의 규정은 계산에 있어서 효율성을 보일 수밖에 없다. 이러한 과정에 다른 방법을 적용하였다면 필터 반응에 관계하는 시각적 특징을 찾는데 있어서 벡터들 사이에 존재하는 수 많은 거리를 계산해야만 했다. 인접 중심 사이의 거리를 계산하는 이러한 직접적인 방법은 모든 거리를 계산해야 하며 가장 작은 값을 선택해야 한다. 그러나 이러한 방법과 달리 *kd-tree* 데이터 구조는 보다 효율적인 접근이 가능하게 하여 계산의 복잡성을 크게 줄여준다.

3.1 객체 유형 모델링

객체 유형들은 이들이 갖는 서술자 히스토그램의 분산에 의하여 모델링 될 수 있다. 이를 위하여 *Support Vector Machines (SVMs)*^[6]을 사용하였다. 그러나 여기에는 별도의 분류 과정은 적용되지 않았다. 학습 영상들은 단위 유형에 대하여 동일한 수의 양과 음의 보기가 적용되었다. 양의 보기는 하나의 특정 유형의 보기를 포함하는 영상들의 서술자 히스토그램이며 음의 보기는 다른 객체 유형들의 랜덤 사례의 컴파일로 이루어진다.

3.2 객체 추적

히스토그램 추적의 첫 번째 단계는 타겟 모델에 있다. 이 모델은 입력 영상에서 찾으려 하는 객체의 특성을 규정하는 것이며 특징 공간에 있어서의 확률 밀도 함수에 의하여 정의된다. 앞에서 언급한 바와 같이 본 논문에서는 타겟 및 적용 모델을 규정하기 위한 시각적 특징을 기본으로 하는 히스토그램을 사용한다.

추적은 평균-이동 방법의 확장에 의하여 이루어진다. 타원의 위치, 크기, 정렬은 히스토그램 기반 객체 추정을 위한 EM 과 같은 알고리즘을 사용하여 대상 영역을 추정한다^[4].

IV. 실험 결과 및 고찰

제안된 *texton* 기반 서술자 히스토그램에 의한

추적 알고리즘에 의한 추적 및 인식 성능을 입증하고자 일련의 실험을 진행하였다. 실험에 사용한 장비는 Honeywell 사의 HSDN-251NS DOME 형 카메라와 CryptoTelecom의 S-152 비디오서버를 통하여 초당 30프레임의 CIF(352×240) 규격의 연속적인 영상을 획득하였다.

4.1. 상세 데이터 집합

객체 유형들은 4800 개의 주석이 달린 영상을 갖고 모델링 하였다. 본 논문에서 사용한 유형은 자동차를 사용하였다. 또한 모든 영상은 시산이나 조명과 같은 조건으로부터 다른 변이를 나타낸다. 학습 영상들은 객체 둘레에 경계선을 그려 넣어 객체와 배경을 수작업으로 구분하였다.

4.2. 분류의 정확성

분류의 정확성을 피하기 위하여 본 논문에서는 10-fold 교차 검증을 적용하였다. 각각의 반복 단계에 있어서 데이터 집합을 단일 트레이닝 집합으로 지정하였으며 여기에 90%의 데이터를 포함시키고 시험 집합에 있어서 나머지 10%의 데이터를 포함시켰다.

시각적 기준은 전체 데이터 집합에 대한 반응 벡터를 사용하여 교차 검증이 이루어지기 전에 구성하였다. 여기에서 k-평균 클러스터링이 데이터의 최적의 클러스터링을 찾는 것을 보장 할 수 없기 때문에 클러스터링은 여러 번 반복적으로 적용되었다. 데이터의 가장 바람직한 클러스터링을 위하여 가장 낮은 내부 유형 변이를 나타내는 중심이 시각적 기준의 구축에 적용되었다.

표 1은 카메라와 객체와의 거리에 따른 SVM 분류에 의한 인식 결과를 나타내었다. 카메라와 객체와의 거리가 멀수록 객체의 화소수가 작아짐에 따라 인식율에 영향을 미치게 된다. 제안된 알고리즘에서는 약 객체가 약 1,200화소 이상이 되는 80m

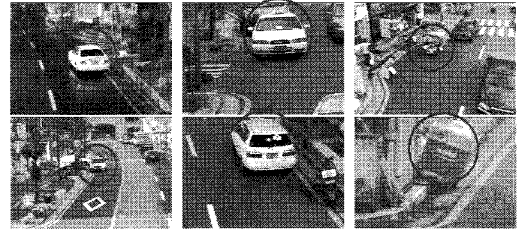


Fig. 2. Examples of classified instances. 그림 2. 분류 결과에 따른 예

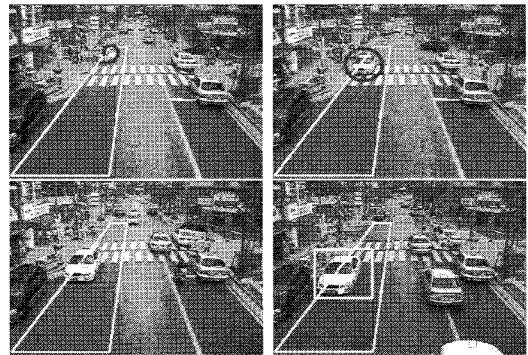


Fig. 3. Examples of Tracking 그림 3. 추적 예

이내의 거리에서는 약 97%의 인식율을 나타내지만 80m 이상이 되는 거리에서는 급격히 인식율이 떨어짐을 알 수 있었다.

그림 2는 분류 결과에 따른 몇 가지 영상을 보여 준다. 각각의 열에 2 개의 영상을 나타낸다. 처음 2 개의 영상은 정확히 자동차로 분류한 경우를 나타내고 있고 나머지 한 개의 영상은 자동차가 아닌 것을 자동차로 인식한 경우를 나타내고 있다.

4.3. 추적 결과

본 논문에서는 객체 추적에 의한 여러 실험을 수행하였다. 모든 실험들은 texton 기반 서술자 히스토그램을 적용하였다. 자동차는 성공적으로 추적되었으며 분류를 위하여 서술자 히스토그램이 정확히 규정 되었다.

V. 결론

본 논문에서 추적과 객체 유형 인식의 방법에 대하여 제안하였다. 객체 유형 인식을 위하여 시각적 특징 히스토그램으로 객체를 표현하였고, 이러한 객체 유형 인식에 의해 추적된 객체의 실효성을 나타내고자 하였다.

Table 1. Result for SVM classification 표 1. SVM 분류 결과

classify distance	cars	other	recognition rate (%)
0 ~ 30 m	624	12	98.1%
30 ~ 50 m	528	14	97.4%
50 ~ 80 m	568	22	96.3%
80 ~ 100 m	320	153	67.7%

제안된 알고리즘의 실효성을 입증하기 위해 인식과 추적 실험을 진행하였다. 인식의 경우 카메라와 객체와의 거리에 따른 SVM 분류에 의해 인식하였다. 그 결과 객체가 약 1,200화소 이상이 되는 80m 이내의 거리에서는 약 97%의 인식율을 나타내지만 80m 이상이 되는 거리에서는 급격히 인식율이 떨어짐을 알 수 있었다.

그리고 객체 추적에 있어 texon 기반 서술자 히스토그램을 적용하여 자동차가 실시간으로 성공적으로 추적되어 분류를 위한 서술자 히스토그램이 정확히 규정되었음을 나타내었다.

참 고 문 헌

[1] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82-98, 1999.

[2] M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking articulated objects using a view based representation. *International Journal Computer Vision*, 26(1):63-84, 1998.

[3] M. Isard and A. Blake. Condensation conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5-28, 1998.

[3] M. Swain and D. Ballard. Color indexing. *Intl. J. of Computer Vision*, 7(1):11-32, 1991.

[4] Z. Zivkovic and B. Krose. An em-like algorithm for color-histogram-based object tracking. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 2004.

[5] Chris Dance, Jutta Willamowski, Lixin Fan, Cedric Bray, and Gabriela Csurka. Visual categorization with bags of keypoints. In *ECCV International Workshop on Statistical Learning*

in *Computer Vision*, 2004.

[6] Corinna Cortes and Vladimir Vapnik. Support vector networks. *Machine Learning*, 20(3): 273-297, 1995.

[7] Thomas Leung and Jitendra Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. J. Comput. Vision*, 43(1):29-44, 2001.

[8] Derk Crezee Athanasios K. Noulas Zoran Zivkovic Ben. J.A. Krose. "Visual Object Tracking and Recognition in a Visual Vocabulary Framework", *ASCI* 2007.

박 호 식 (Ho-Sik Park)

중신회원



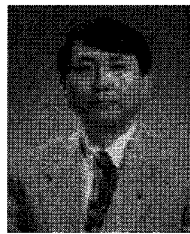
1994년 2월 연세대학교 의용전
자공학과 졸업 (공학사)
2001년 2월 관동대학교 대학원 전
지통신공학과 졸업 (공학석사)
2005년 2월 관동대학교 대학원 전
지통신공학과 졸업 (공학박사)
2008년 3월~현재 오산대학 디지털

탈전자과 교수

<관심분야> 영상처리, 신호처리시스템, 영상압축

배 철 수 (Cheol-Soo Bae)

중신회원



1979년 2월 명지대학교 전자공
학과 졸업(공학사)
1981년 2월 명지대학교 대학원
전자공학과졸업 (공학석사)
1988년 8월 명지대학교 대학원
전자공학과졸업 (공학박사)
1999년 3월~2001년 5월 관동

대학교공과대학 학장

2001년 6월~2003년 8월 관동대학교 평생교육원장
1981년~현재 관동대학교 전자정보통신공학부 교수
<관심분야> 영상처리, 신호처리시스템, 영상압축