한국통신학회논문지 '13-01 Vol.38A No.01
http://dx.doi.org/10.7840/kics.2013.38A.1.1

가중치 기반 Bag-of-Feature와 앙상블 결정 트리를 이용한 정지 영상에서의 인간 행동 인식

홍 준 혁◆, 고 병 철°, 남 재 열*

Human Action Recognition in Still Image Using Weighted Bag-of-Features and Ensemble Decision Trees

June-hyeok Hong◆, Byoung-chul Ko°, Jae-yeal Nam*

요 약

본 논문에서는 CS-LBP (Center-Symmetric Local Binary Pattern) 특징과 공간 피라미드를 이용한 BoF (Bag of Features)를 생성하고 이를 랜덤 포레스트(Random Forest) 분류기에 적용하여 인간의 행동을 인식하는 알고리즘을 제안한다. BoF를 생성하기 위해 영상을 균일한 패치로 나누고, 각 패치 마다 CS-LBP 특징을 추출한다. 행동 분류 성능을 향상시키기 위해 패치들마다 추출한 특징벡터들에 대해 K-mean 클러스터링을 적용하여 코드 북을 생성한다. 본 논문에서는 영상의 지역적인 특성을 고려하기 위해 공간 피라미드 방법을 적용하고 각 공간 레벨에서 추출된 BoF에 대해 가중치를 적용하여 최종적으로 하나의 특징 벡터로 결합한다. 행동 분류를 위해 결정 트리의 앙상블로 이루어진 랜덤 포레스트는 학습 단계에서 각 행동 클래스를 위한 분류 모델을 만든다. 가중 BoF가 적용된 랜덤 포레스트는 다양한 인간 행동 영상을 포함하고 있는 Standford Actions 40 데이터를 성공적으로 분류하였다. 또한 기존 방법에 비해 분류 성능이 유사하거나 우수하며, 한 장의 영상에 대해 빠른 인식속도를 보였다

Key Words : Bag-of-feature, action recognition, random forest, CS-LBP, spatial pyramid, code book

ABSTRACT

This paper propose a human action recognition method that uses bag-of-features (BoF) based on CS-LBP (center-symmetric local binary pattern) and a spatial pyramid in addition to the random forest classifier. To construct the BoF, an image divided into dense regular grids and extract from each patch. A code word which is a visual vocabulary, is formed by k-means clustering of a random subset of patches. For enhanced action discrimination, local BoF histogram from three subdivided levels of a spatial pyramid is estimated, and a weighted BoF histogram is generated by concatenating the local histograms. For action classification, a random forest, which is an ensemble of decision trees, is built to model the distribution of each action class. The random forest combined with the weighted BoF histogram is successfully applied to Standford Action 40 including various human action images, and its classification performance is better than that of other methods. Furthermore, the proposed method allows action recognition to be performed in near real-time.

※ 본 연구는 교육과학기술부와 한국연구재단의 지역혁신인력양성사업으로 수행된 연구결과임.
◆ 주저자 : 계명대학교 컴퓨터공학과 멀티미디어통신 연구실, plastiqvox@kmu.ac.kr, 준회원
° 교신저자 : 계명대학교 컴퓨터공학과 컴퓨터비전&패턴인식 연구실, niceko@kmu.ac.kr, 정회원
* 계명대학교 컴퓨터공학과 멀티미디어통신 연구실, jynam@kmu.ac.kr, 정회원
논문번호 : KICS2012-11-527, 접수일자 : 2012년 11월 13일, 최종논문접수일자 : 2012년 12월 24일

Ⅰ. 서 론

멀티미디어 데이터의 증가와 함께 데이터를 의미 있는 카테고리로 분류하기 위한 자동 인식 및 분류기술이 새로운 연구 분야로 떠오르고 있다. 특히 인간 행동 인식은 비디오 감시, 색인, 검색, 인간-컴퓨터 상호작용 등에서 폭넓게 응용되고 있는 컴퓨터 비전에서의 중요한 이슈들 중 하나이다.

인간 행동인식은 주로 동영상에서 모션정보를 사용하여 인식하는 방법을 사용한다. 모션정보는 행동 인식을 위한 시간적 객체 이동 정보와 공간적 방향 정보를 제공하기 때문에 인간의 행동 패턴과 특성을 분석하여 행동 인식에 유용한 정보를 제공한다. 하지만, 동영상내에서 '서있기','독서','사진 찍기'와 같이 모션이 없는 인간 행동의 경우 시간적 정보가 제한적임으로 인식이 불가능하거나 잘못된 인식결과를 초래할 수 있다[1].

정지영상을 이용한 인간의 행동 인식은 영상 분류, 주석 생성, 영상검색 [2]등에서 다양하게 적용될 수 있다. 하지만 동영상과 달리 정지영상에서의 인간 행동 인식은 상대적으로 많이 연구가 이루어지지 않고 있다.

정지영상을 이용한 행동인식에 관한 최근의 연구들 [1-7]에서는 정지영상의 한계를 극복하고 정지영상에서 인간의 행동을 인식 하기 위해 두 가지 방향의 연구가 진행되어 오고 있다.

첫 번째 방법[4][5]은 신체 부위를 식별하고 인체 구조의 전체적인 포즈(pose)에 대한 사전 지식을 구축 하여 행동인식을 위한 단서로 사용하는 방법이다. 그러나 이러한 방법들은 가려짐 정도가 심하거나 카메라 각도의 큰 변화가 있는 영상에서는 행동을 감지가 하기가 어렵다는 단점이 있다[1,3].

Ikizler [4]등은 인간의 행동을 분석하고, 사각형의 원형 막대그래프를 사용한 공간적인 정보와 방향정보를 가진 빈(bin)의 축적을 통해 각 행동들에 대한 특징을 추출하였다. 그런 다음 차별적인 특징을 얻기 위해 선형 판별 분석(Linear Discriminant Analysis, LDA)을 사용하고, 인간 행동 분류를 위해 이진 SVMs (Support Vector Machines)를 사용하였다.

Thurau 와 Hlavac [5]는 포즈에 기초하여 인간 행동을 인식하였다. 포즈를 나타내기 위해, HoG (Histogram of Oriented Gradient)를 기반으로 포즈 기술자를 확장하는 비음수 행렬 분해(nonnegative matrix factorization)를 사용하였다. 행동 클래스들은 포즈 원형(pose primitives)의 HoG 히스토그램으로 나타내고, 행동 인식은 간단한 HoG 히스토그램 비교를 사용하였다.

행동 인식을 위한 두 번째 방법은 포즈렛(poselet) [2][3][6]을 사용하는 방법이다. 포즈렛은 인체를 여러 부위로 세분화 할 경우, 학습 집합에서 각 세분화된

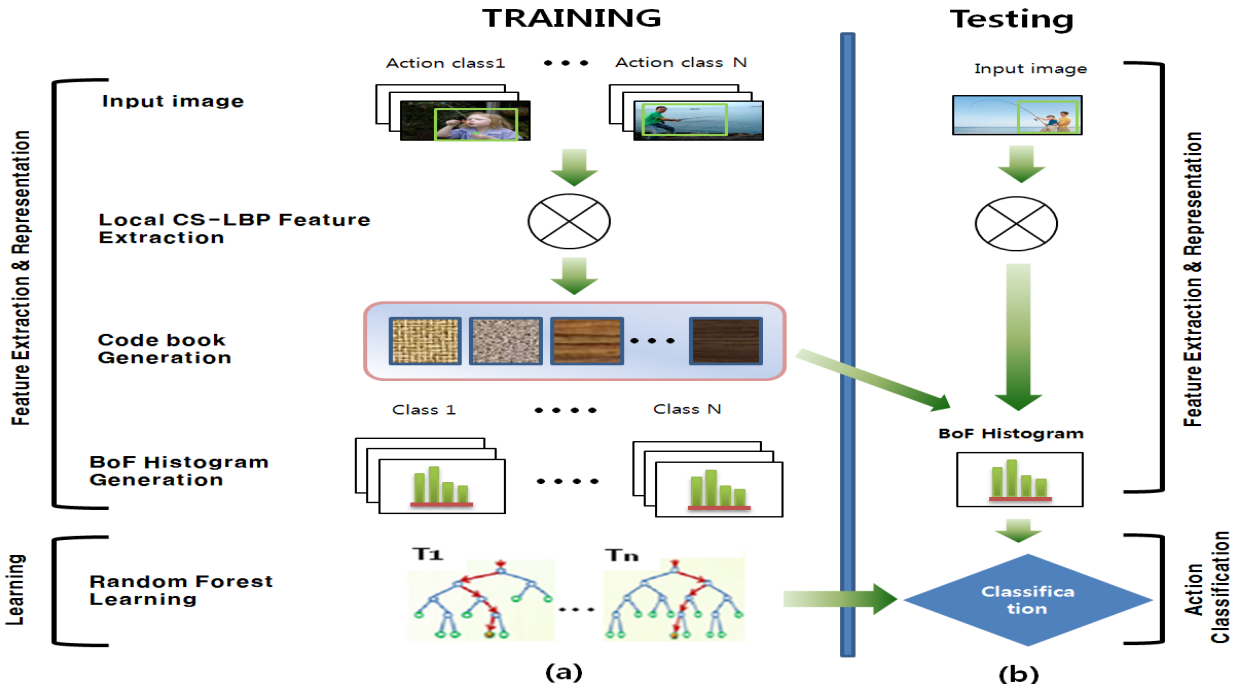


그림 1. 인간행동 인식을 위한 전체 시스템 개요도: (a) 행동 클래스에 대한 코드북 생성 및 BoF 히스토그램 생성, 이를 이용한 랜덤 포레스트 학습 과정 (b) 새로운 입력 영상에 대한 BoF 히스토그램 생성과 학습된 랜덤 포레스트를 이용한 행동 분류 과정
Fig. 1. Overall system diagram: (a) generation of code book and BoF histogram, random forest training using BoF histogram (b) action classification on input image using BoF histogram and trained random forest

인체 부위에 속하는 유사한 값을 갖는 영역들의 집합이다. 따라서 하나의 신체 부위 (예, 머리 영역)는 '정면', '측면', '뒤면'등을 포함하는 다양한 영상의 포즈렛으로 구성될 수 있다. Yao [2]등은 신체 부위들을 객체와 포즈렛으로 구성하여, '타기', '앉기'와 같은 행동과 관련된 속성을 정의하였다. 신체 부위와 속성에 기반한 영상 특징의 표현을 위해 희소 기저(sparse bases)를 생성하고 이를 기반으로 학습 영상의 특징을 학습시킨다. 이후 알고리즘에서 각 행동 영상에 대해 동일한 특징을 추출한 후에 희소 계수(sparse coefficients)를 적용시켜 각 행동들에 대한 가중치 값을 결정하는 방법으로 행동을 인식하도록 하였다.

Maji [3]등은 다 해상도방식에서 방향과 포즈의 고유 모호성을 표현하는 포즈렛 활성 벡터(poselet activation vector)를 사용하였다. 이 알고리즘은 정적인 영상에서 인간의 3D포즈뿐만 아니라 인간 행동도 인식할 수 있는 방법을 제안하고 있다.

Yang [6]등도 마찬가지로 행동 인식을 위해 유용한 중간 정보로써 영상에 포함된 인간의 포즈를 다루었다. 포즈만을 사용한 다른 알고리즘과 달리, 이 알고리즘은 포즈와 행동을 함께 고려한 통합된 방식으로 학습하고 같은 방법으로 행동을 인식하는 알고리즘을 제안하였다.

위에서 언급한 세 가지 방법과는 대조적으로 Delaitre [1]등은 BoF(bag-of-features)와 파트 기반의 latent SVM을 이용한 행동 인식 기법을 제안하였다. 이 방법에서는 통계적인 방법과 파트 기반의 표현 방법을 결합하고, 인간 중심의 특징 기술에 배경 정보도 행동인식에 중요하다는 가정 하에 배경을 추가하는 방법으로 행동 인식 성능을 높이는 방법을 사용하였다.

본 논문에서는 특징 추출의 성능을 높이고 행동 인식 시간을 단축시키기 위해 이전연구인 [1][7]에 서 제안되었던 공간 피라미드 BoF (spatial pyramid BoF)의 구성 방법을 수정 하였다. 또한, 본 논문에서는 인식 성능을 향상시키고 인식 속도를 높이기 위해, SVM분류기 대신 랜덤 포레스트 분류기(Random Forest)를 사용하는 방법을 제안하고 있다.

본 논문에서 제안한 알고리즘의 블럭도는 그림 1과 같다. 첫 번째로 그림 1-(a)의 학습 과정에서 보듯이 영상을 일정한 패치 단위로 나누고, 각 패치들로 부터 공간 피라미드를 고려한 CS-LBP(Center-Symmetric Local Binary Patterns)[7]를 추출한다. 다음으로, 각 패치들에 대한 코드 북(code book)을 생성하기 위해, 학습 집합에서 랜덤하게 추출된 패치들의 특징을 기반으로 K-mean 클러스터링을 수행한다. 코드 북이 완성되면 각 행동 클래스의 패치들로 부터 동일한 CS-LBP 특징들을 추출하고 코드 북에 매핑하여 K-차원의 BoF히스토그램을 생성하게 된다.

두 번째로, 결정트리의 앙상블로 구성된 랜덤 포레스트는 그림 1-(a)의 학습단계에서 각 행동 클래스들의 BoF특징 분포를 모델링하는 약한 결정트리를 모아서 앙상블 형태로 구성된다.

마지막으로 테스트를 위해서는 그림1-(b)의 과정과 같이 테스트 영상의 패치들로부터 CS-LBP특징을 추출하고 코드 북을 이용하여 BoF 특징 히스토그램을 추출한 뒤, 학습된 랜덤 포레스트에 입력하여 인간의 행동을 인식하도록 설계하였다.

본 논문의 구성은 다음과 같다. 2장에서는 공간 피라미드를 사용한 BoF 생성 알고리즘을 설명한다. 3장에서는 랜덤 포레스트와 BoF 히스토그램을 사용한 본 논문의 행동 분류 방법을 소개한다. 4장에서는 본 논문에서 제안한 행동 분류의 실험 결과를 제시하고, 5장에서는 결론과 향후 연구방향에 대해 서술한다.

Ⅱ. 가중치 BoF (bag-of-features) 생성

인체 행동 인식을 위해서, 정확한 특징 추출은 필수적이다. 본 논문에서는 영상의 전역적 특징과 지역적 특징을 동시에 고려하기 위해 Lazebnik [7]등이 제안한 공간 피라미드로 부터 CS-LBP 특징을 추출하고 이를 기반으로 코드 북을 생성하여 각 행동 데이터 집합으로 부터 가중된 BoF (weighted BoF)을 생성하는 알고리즘을 제안한다.

2.1. 코드 북 (Code Book) 생성

BoF는 각 영상을 순서 없는 지역 특징 집합으로 표현하기 위해 고안되었다. 각 영상에서 추출된 지역 특징들은 K-mean 군집화 방법에 의해 군집화 되고 각 군집의 중심벡터는 코드 워드(code word)로 정의되어 K개의 코드 워드로 구성된 코드 북이 생성된다. 코드 북이 생성되면 한 영상의 지역 특징들은 코드 북에 매핑 되고, 매핑 유사성에 따라 가장 거리가 가까운 코드 워드의 히스토그램 빈에 누적 시키는 방법으로 BoF를 생성하게 된다[9].

본 논문에서는 기존의[7][9] 방법들과 마찬 가지로 영상을 표현하기 위해 BoF를 사용한다. 우선, 학습영상들로 부터 사람을 포함하는 영역을 잘라 내고 잘라낸 각 영역에서 8픽셀씩 이동하며 16 픽셀 ×16 픽셀 크기의 패치(patch)를 생성한다.

패치를 추출하기 위해 SIFT (scale invariant feature transform) 가 주로 사용되었지만 Fei-Fei [2]등의 실험에서 균일한 패치의 추출이 영상 분류에 있어서 오히려 SIFT를 이용한 패치 추출보다 성능이 우수하다는 실험결과에 따라 본 논문에서는 균일한 패치 추출 방법을 적용하였다. 추출된 패치에서 특징 차원을 줄이기 위해 본 논문에서는 객체 분류에 주로 사용되고 있는 SIFT 대신에 CS-LBP를 추출한다. LBP(Local Binary Patterns)는 그레이레벨의 변화와 조명 변화에 대한 견고한 특징을 갖고 간단한 연산으로 얻을 수 있는 있는데, 중심 픽셀과 특정 범위내의 이웃 픽셀간의 그레이 스케일 값의 차이로 256()의 패턴을 형성한다. LBP는 큰 차원의 패턴 히스토그램을 생성함에도 불구하고 최근 영상분류 및 인식에서 많이 사용되는 질감 연산자이다. CS-LBP [8][10]는 계산 비용을 줄이기 위해 기존 LBP의 이웃 픽셀 간의 비교 방법을 수정한 것으로, LBP와 마찬가지로 단조로운 그레이 스케일 변화와 조명변화에 강한 특성을 유지한다.

CS-LBP는 중심 픽셀과 각 이웃 픽셀을 비교하는 것이 아니라 이웃 픽셀들에서 대칭되는 위치에 있는 픽셀간의 차이를 구하기 때문에 연산의 수가 절반으로 줄어들며, 16()의 다른 이진 패턴을 생성한다.

각 패치 영역에서 CS-LBP를 추출 후, L2 정규화가 적용되고, 각 패치들의 정규화된 CS-LBP 특징과 k-mean 클러스터링을 통해 코드 북을 생성한다. 본 논문에서는 실험을 통해 코드 북의 크기가 200일 때 가장 좋은 성능을 보여 주었음으로 K=200으로 설정하였다.

2.2. K차원의 BoF 생성

일반적으로 BoF히스토그램의 차원은 코드 북의 개수와 일치하고, 입력된 패치들의 시각적 특징이 코드 북내의 특정 코드 워드와 일치할 경우 해당 코드 워드에 해당하는 히스토그램 빈을 1로 누적시키는 방법에 의해 BoF히스토그램을 생성한다. 하지만 이 방식은 유사한 코드 워드가 여러 개 존재할 경우에도 하나의 히스토그램 빈에만 누적해야 함으로 정확도가 떨어지는 문제점이 있다. 이러한 문제점을 해결하기 위해서, Jiang [9]등은 soft-weighting방식에 의한 BoF히스토그램 생성방법을 제안하였다. 이 방법에서는 영상으로부터 추출된 패치들의 특징과 한 개의 코드 워드간에 유사도를 측정하여 상위 N개의 패치들을 선택하고 코드 워드와의 유사도합을 계산하여 이 값을 해당 코드 워드의 히스토그램 빈에 누적하는 방법을 사용하였다. 하지만, 코드 워드 별로 유사한 패치를 선택하는 방법임으로 경우에 따라서는 코드 북의 어떠한 코드 워드에도 선택되지 못하는 패치가 발생 할 수 있음으로 영상의 특성을 정확하게 반영하지 못하는 문제점이 있다.

따라서, 본 논문에서는 soft-weighting방법과는 반대로 입력 패치와 가장 유사한 N개의 코드 워드를 선택하고 각 코드 워드와의 거리의 따라 각 코드에 해당하는 BoF 히스토그램 빈에 서로 다른 가중치를 할당하여, 기존의 soft-weighting방식에서 발생할 수 있는 문제점을 해결 하였다.

본 논문에서 제안하는 개선된 가중치 BoF 히스토그램 생성 방법은 다음과 같다.

(1) 하나의 영상 I에서 패치 P를 추출한다.

(2) K차원의 BoF 히스토그램 의 모든 요소들을 0으로 초기화 시킨다. 각 요소 는 각 패치와의 유사성 측정에 의해 선택된 코드 워드로 수식 (3)에 의해 가중치 값을 갖는다.

(3) 입력 패치에 대해 가장 유사한 상위 N개의 코드 워드들을 찾고, 패치 P에서의 특징과 코드 워드(V) 간의 거리 D를 아래 수식에 의해 계산 한다.

$$D_{i=1\ldots N} = \| P - V_i \| \tag{1}$$

(4) For i<=N

(4-1) 코드 워드 t_k에 대해 패치 P와 i번째 코드 워드 사이의 가중치 거리 W(D_i)를 누적시킨다.

$$t_k = t_k + W(D_i) \tag{2}$$

(4-2) 가중치 함수 $W(\bullet)$는 거리 D에 따라 수식 (3)이용하여 가변적으로 계산된다.

$$W(D_i) = \frac{1}{\exp(T \bullet D_i)},\ T > 0 \tag{3}$$

$i = i + 1$

End For

위의 알고리즘에서, T 는 가중치 함수 를 최대화 또는 최소화 시킬 수 있는 파라메터로 T가 1일 경우 는 지수형태로 에 반영되고 T가 0일 경우 거리에 상관없이 1의 값을 갖게 된다. 지수 함수를 이용한 가중치 계산 방식은 로컬 특징의 변화에 민감하게 반응 하고 성능 향상에도 효과적이다[9]. 본 논문에서는 T값을 0.5로 설정하였다.

2.3. 공간 피라미드를 이용한 BoF의 확장

본 논문에서는 입력된 사람의 크기 변화와 부분적 가림의 경우에도 올바른 행동인식이 가능하도록 입력 영상의 지역적 특성을 고려한 공간 피라미드 (spatial pyramid) [7]기법을 사용하여 BoF를 확장하였다.

그림 1과 같이 우선 입력 영상을 레벨 0 (1 × 1 = 1 block), 레벨 1 (2 × 2 = 4 blocks), 레벨 2 (4 × 4 = 16 blocks)의 피라미드 구조로 분할한다. 각 레벨에서 패치를 생성하고 CS-LBP 특징을 추출하여 코드북과의 연산을 통해 BoF를 생성한다. 코드 북의 크기 K가 200인 경우, 레벨 0에서의 BoF 히스토그램 특징 차원 수는 200 차원 (1 × 1 × 200)이 된다. 마찬가지로 레벨 1에서는 800차원 (2 × 2 × 200), 레벨 2에서는 3200차원 (4 × 4 × 400)이 생성된다. 마지막으로 이렇게 생성된 각 레벨의 BoF 히스토그램을 하나로 결합하여 4200차원 (3200+800+200)의 BoF가 생성된다.

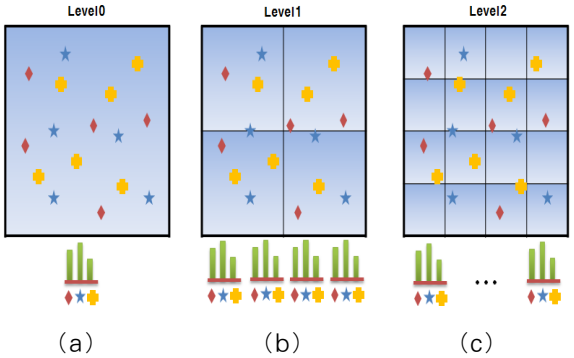


그림 2. K가 3일 경우, 3 레벨의 공간 피라미드의 구성 예제 (a) 0 레벨의 BoF 히스토그램 (3차원) (b) 1 레벨의 BoF 히스토그램 (12차원) (c) 2 레벨의 BoF 히스토그램 (48차원)
Fig. 2. Example of 3-level spatial pyramid when K is 3 (a) 0-level BoF histogram (3 dimension) (b) 1-level BoF histogram (12 dimension) (c) 2-level BoF histogram (48 dimension)

Ⅲ. Random Forest를 이용한 인간 행동 분류

인간 행동 인식을 위해 기존의 많은 논문 [1][2][4][6][7][13]들에서는 MSVM(multi-class SVM)을 행동 분류를 위한 알고리즘으로 사용하고 있다. MSVM 분류기는 높은 성능과 정확도를 가지기 때문에 일반적인 패턴 분류를 위해서 주로 사용되는 방법이지만, 계산의 복잡성 때문에 높은 차원을 가진 특징이나, 방대한 양의 클래스를 분류 할 때는 적합하지 않다. 따라서 인간 행동들을 분류하기 위해 본 논문에서는 특징 벡터 차원의 크기에 덜 민감하여 분류 속도가 빠르고 정확한 분류 결과를 나타내는 것으로 알려진 랜덤 포레스트 (Random Forest) 분류기를 사용한다.

3.1. Random forest 분류기

Breiman [11]에 의해 제안된 랜덤 포레스트는 다수의 결정 이진 (binary) 트리를 앙상블 형태로 결합한 것으로, 각 이진트리에서는 랜덤한 방법으로 트리들을 성장 시킨다. 랜덤 포레스트는 결정 트리들을 기본으로 하고 있기 때문에, 빠른 학습속도와 많은 양의 데이터 처리 능력을 가지고 있다[12].

그림 3과 같이 랜덤 포레스트의 각 트리의 구조는 이진이며, 하향식(top-down) 형태를 갖는다. 학습 데이터로 부터 각 행동들을 위한 가중된 BoF 히스토그램을 생성한 후, 랜덤 포레스트의 각 결정 트리는 아래와 같은 알고리즘에 의해 구축 되도록 설계 하였다.

입력 :
D_{target} : 트리의 최대 확장 깊이
S_n : 모든 액션 샘플을 포함하는 학습 데이터
초기값 : $i=0, j=0, k=0, F_i=1.0$

(1) n개의 부스트랩 샘플을 S_n에 할당하고 각 샘플들에 대해 BoF를 생성
(2) S_n으로부터 m개의 서브(sub) 부스트랩 샘플을 선택

(2-1) Loop : $D_k < D_{target}$ 또는 종료 조건 만족
$k=k+1$
1) m개의 부스트랩을 사용하여 초기의 트리를 확장
2) 각 내부 노드는 무작위로 샘플로부터 p개의 특징을 선택하고, 이들을 이용하여 최적의 분할 함수 (split function)를 결정.
3) 다른 p번째 특징을 사용하여 분할함수 $f(v_p)$는 수식(4)에 의해 반복적으로 m개의 샘플 데이터를 left(I_l) 과 right(I_r)의 서브셋으로 분할

$$I_l = p \in I_n | f(v_p) < t, \quad I_r = I_n \setminus I_l \tag{4}$$

4) 임계값 t는 분할 함수 $f(v_p)$에 의해 범위 $t \in (\min_p f(v_p), \max_p f(v_p))$에서 랜덤하게 선택. Loop ends

반복적인 학습은 다음의 두 가지 조건에 의해 종료할 수 있다. 즉 information gain 이 0이거나 종단 노드가 최대 트리 깊이에 도달할 때 종료 된다.

3.2. 인간 행동 분류

학습 데이터를 이용하여 랜덤 포레스트 분류기가

학습된 이후에 그림 3과 같이 테스트 영상을 입력받아 동일한 차원의 BoF를 추출하여 랜덤 포레스트에 적용하고 행동을 분류한다.

입력된 영상에 대해 각 트리의 종단 노드의 분류 결과에 따라 트리는 클래스별 확률 값을 갖게 되고 최종적으로 수식 (5)를 이용하여 T개 트리의 사후 확률을 결합하고 정규화 하여 최종 클래스를 결정한다.

수식(5)에서, T는 트리의 개수이고, $p(c_i|l_t)$는 l번째 트리에서 클래스 C_i에 대한 확률 값을 의미한다. 각 트리의 확률 값에 대한 합인 $p(c_i|L)$가 최댓값을 가질 때, 입력 영상의 최종 클래스로써 c_i를 선택한다.

그림 3의 경우 테스트 영상은 최대 사후 확률 값을 갖는 2번 클래스(Jumping)로 분류되게 된다.

Ⅳ. 실험 및 결과 분석

본 논문에서는 일상생활에서 볼 수 있는 인간의 행동들, 즉, 달리기(running), 박수치기(applauding), 풍선 불기(blowing bubble), 나무 자르기(cutting trees), 우산 들고 있기(holding an umbrella)와 같은 40가지의 다양한 행동이 포함된 Stanford 40 Action [2] 데이터를 실험에 사용하였다. Stanford Action 40은 각 행동 클래스 당 180~300개의 영상을 포함하며, 총 9532개의 영상 데이터를 제공한다. 각 행동 클래스의 영상들은 다양한 인간의 포즈, 외형, 복잡한 배경 등을 포함하고 있다. 본 논문에서는 학습을 위해 임의로 각 행동 클래스 당 100개의 영상을 선택하고 나머지 영상들을 테스트에 사용하였다. 또한 각 학습 영상에서 배경을 제외한 사람 영역을 수작업으로 잘라내어 학습에 사용하였다.

제안한 방법의 성능을 검증하기 위해, 평균 정밀도(average precision)를 사용하였다. 그림 4는 가장 높은 정밀도를 갖는 10개의 행동 클래스들의 혼동 행렬(confusion matrix) 테이블을 보여준다. 40개의 행동 중 달리기(running)는 65%로 가장 높은 정밀도를 보였다. 반면 프리즈비 던지기(throwing frisby) 는 42%의 정밀도를 보여 주었다. 프리즈비 던지기에서 가장 낮은 정밀도를 보여준 이유는 테스트 영상에 복잡한 배경이 많이 포함되어 있고 프리즈비를 던지는 사람들의 포즈가 다양하기 때문이다.

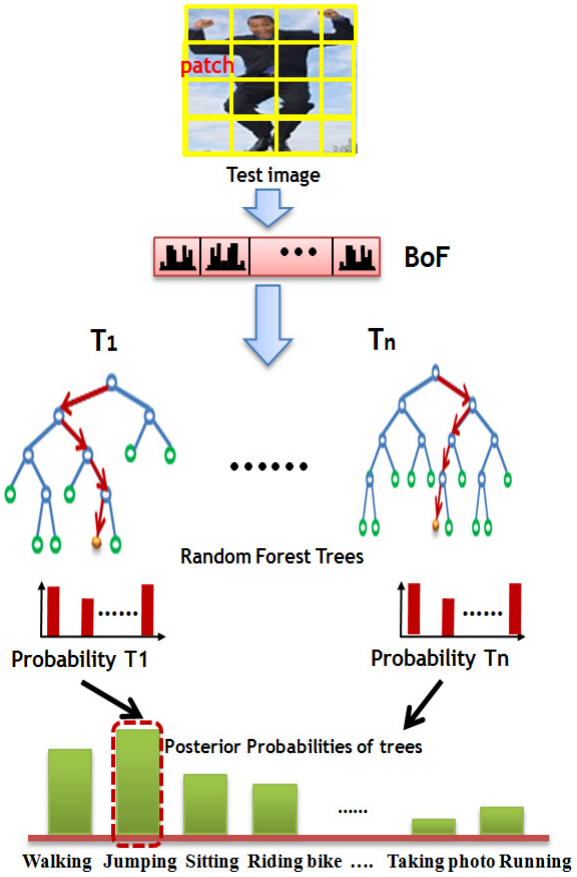


그림 3. 랜덤 포레스트에 의한 입력 영상 분류 과정
Fig. 3. Input image classification using random forest

본 논문에서는 기존 알고리즘과의 성능 비교를 위해 LLC [14]방법과, 본 논문에서 생성한 가중치 기반의 BoF와 가장 일반적으로 사용되는 패턴 분류기인 MSVM를 결합한 알고리즘과(BoF+MSVM)의 행동 분류 성능을 측정하였다.

표1은 40개의 행동에 대한 각 알고리즘의 분류 성능을 보여주고 있다. 표1에서 볼 수 있듯이, 논문에서 제안한 알고리즘은 40개 클래스에 대해 평균 정밀도가 약 34.8%로 LLC의 분류 성능보다 0.3%우수한 성능을 보이고 있으며, MSVM에 비해서는 1.3%가 향상된 성능을 보여주고 있다.

계산 복잡도는 테스트 데이터에 대한 평균 분류시간을 측정하여 평가하였다. LLC와 제안한 방법을 동일한 시스템 환경에서 실험한 결과 영상당 평균 처리속도는 0.3초와 0.1초로 측정되었다. MSVM의 경우 3초가 소요 되었는데, 이것은 MSVM의 연상량 특징벡터의 차원에 매우 민감하다는 것을 보여준다. 시간 측정 실험 결과로부터, 제안한 방법은 LLC와 비슷한 분류를 가지면서, 빠른 처리속도로 방대한 양의 데이터를 처리 할 수 있는 능력을 가지고 있다는 것을 알 수

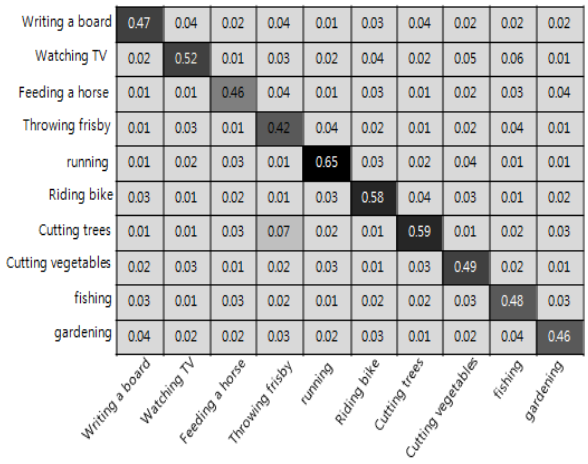


그림 4.높은 정밀도를 보이는 상위 10개의 행동 클래스에 대한 혼동 테이블
Fig. 4. Confusion matrix on top 10 action classes having high precision.

있고, MSVM은 성능도 두 알고리즘에 비해 떨어지며 특히 처리시간이 차원에 비례하여 증가함으로 실시간 처리에는 적합하지 않음을 알 수 있다.

그림 5는 본 논문에서 제안하는 방법을 Stanford Action 40 데이터에 적용한 결과영상의 예를 보여주고 있다. 그림 5에서 보는 것과 같이 제안하는 방법은 각각 다른 크기의 사람과 복잡한 배경에서도 비교적 정확하게 인간의 행동을 분류하는 것을 볼 수 있다.

표 1. 40개의 행동 클래스를 위한 LLC, BoF+MSVM 방법과 제안 알고리즘 간의 비교
Table 1. Performance comparison between LLC, BoF+MSVM, and proposed method on 40 action classes

Action classes	Accuracy per class (%)		
	LLC	MSVM	제안방법
applauding	21	28	27
blowing bubble	33	29	25
brushing teeth	37	42	41
cleaning the floor	60	30	36
climbing	62	25	29
cooking	18	25	26
cutting trees	38	65	57
cutting vegetables	37	51	53
drinking	18	29	24
feeding a horse	31	39	39
fishing	41	33	50
fixing a bike	11	47	44
fixing a car	38	23	23
gardening	41	45	47
holding on umbrella	41	23	25
jumping	61	35	42
looking through a microscope	22	37	39
looking through a telescope	11	29	28
phoning	17	21	28
playing guitar	50	25	25
playing violin	35	29	27
pouring liquid	10	25	25
pushing car	10	27	25
reading	21	27	21
riding bike	70	48	57
rowing a boat	75	35	38
running	55	65	61
shooting on arrow	56	33	26
smoking	20	25	25
taking photos	10	21	23
texting message	7	23	23
throwing frisby	56	30	38
using a computer	17	29	33
walking the dog	57	27	21
washing dishes	17	25	26
watching TV	31	55	55
waving hands	10	21	21
writing on a board	38	53	51
writing on a book	35	27	30
평균 정밀도	34.5	33.5	34.8

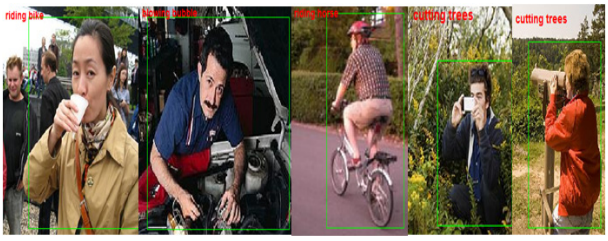
그림 5. 제안하는 방법에 의해 올바르게 분류된 인간의 행동의 예
Fig. 5. Examples of correctly classified human action

그림 6은 복잡한 배경과 카메라 시점, 불분명한 사람의 포즈로 인해 오 분류 된 예를 보여주고 있다.

V. 결 론

본 논문은 공간 피라미드와 가중치 방법에 의해 BoF를 생성하고 랜덤 포레스트 분류기를 이용하여 인간의 행동을 인식하는 알고리즘을 소개하였다. 입력 영상을 균일한 패치로 나누고, 각 패치로부터 CS-LBP를 추출해서 코드 북을 생성함으로써 코드 워드를 위한 특징벡터의 차원을 감소 시켰다. 또한, BoF 생성을 위해 모든 코드 워드에 동일한 가중치를 부여하지 않고 입력 패치와의 유사성 측정을 통해 각기 다른 가중치를 부여하는 방법을 제안하였다.

분류를 위해 결정트리의 앙상블로 이루어진 랜덤 포레스트는 빠른 학습 및 테스트 속도와 분류 결과에 있



(a) (b) (c) (d) (e)

그림 6. 복잡한 배경 및 불분명한 포즈로 인해 오 분류된 인간 행동의 예. (a) throwing frisby (b) gardening (c) cutting tree (d) watching TV (e) cutting vegetable

Fig. 6. Example of false classification caused by clutter background and ambiguous pose

어서도 MSVM에 비해 우수한 결과를 보여 주었다.

향후 연구에서는 현재 35%의 인식율을 45%이상 향상시키기 위해 인간의 신체부위를 포즈렛(poselet)으로 분할하고 이를 기반으로 행동을 인식하는 알고리즘을 현재의 시스템에 접목하여 포즈의 변화에 강인한 인식 알고리즘을 개발하고자 한다.

Reference

[1] V. Delaitre, I. Laptev, and J. Sivic. "Recognizing human action in still images: a study of bag-of-features and partial-based representations," in *Proc. British Machine Vision Conf.*, pp. 1-11, Wales, UK, Sep. 2010.

[2] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas, and L. Fei-Fei. "Human action recognition by learning bases of action attributes and parts," in *Proc. Int. Conf. on Computer Vision*, pp. 1331-1338, Barcelona, Spain, Nov. 2011

[3] S. Maji, L. Bourdev, and J. Malik. "Action recognition from a distributed representation of pose and appearance," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 3177-3184, Providence, RI, Jun. 2011.

[4] N. Ikizler, R.G. Cinbis, S. Pehlivan, and P. Duygulu. "Recognizing actions from still images," in *Proc. Int. Conf. of Pattern Recognition*, pp. 1-4, Tampa, Florida, Dec. 2008

[5] C. Thurau and V. Hlavac. "Pose primitive based human action recognition in videos or still images," in *Proc. IEEE Int. Conf. on Pattern Recognition*, pp. 1-8, Tampa, Florida, Dec. 2008

[6] W. Yang, Y. Wang, and G. Mori. "Recognizing human actions from still images with latent poses," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 2030-2037, San Francisco, USA, Jun. 2010.

[7] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 2169-2178, NY, USA, Jun. 2006.

[8] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recogn.*, vol. 42, no. 3, pp. 425-436, Mar. 2009.

[9] Y. G. Jiang C. W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in *Proc. ACM Int. Conf. on Image and Video Retrieval*, pp. 494-501, Amsterdam, Netherlands, Jul. 2007.

[10] B. C. Ko, J. Y. Kwak, and J. Y. Nam, "Object tracking using particle filters in moving camera," *J. KICS,* vol. 37A, no. 5, pp. 35-40, May 2012.

[11] L. Breiman. "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, Oct. 2001

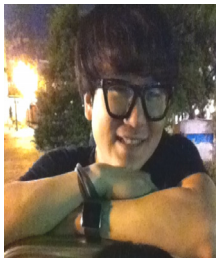
[12] B. C. Ko, S. H. Kim, and J. Y. Nam, "X-ray image classification using random forests with local wavelet-based CS-local binary patterns," *J. Digit. Imaging*, vol. 24, no. 16, pp. 1141-1151, Oct. 2011

[13] L. Bourdev and J. Malik, "Poselets: bady part detectors trained using 3d human pose annotations," in *Proc. European Conf. on Computer Vision*, pp. 3178-3179, Kyoto, Japan, Sep. 2009

[14] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear

coding for image classification," in *Proc. IEEE Int. Conf. on Pattern Recognition*, pp. 3360-3367, Istanbul, Turkey, Aug. 2010.

홍 준 혁 (June-hyeok Hong)



2012년 2월 계명대학교 컴퓨터공학과 졸업
2012년 3월~현재 계명대학교 컴퓨터공학과 석사과정
<관심분야> 컴퓨터비전, 패턴인식

고 병 철 (Byoung-chul Ko)

1998년 2월 경기대학교 컴퓨터공학과 졸업
2000년 2월 연세대학교 컴퓨터공학과 석사
2000년 4월 연세대학교 컴퓨터공학과 박사
2005년 8월~현재 계명대학교 컴퓨터공학과 부교수
<관심분야> 컴퓨터비전, 패턴인식

남 재 열 (Jae-yeol Nam)

1983년 2월 경북대학교 전자공학과 졸업
1985년 2월 경북대학교 전자공학과 석사
1991년 4월 UTA 전자 공학과 박사
1995년 8월~현재 계명대학교 컴퓨터공학과 교수
<관심분야> 컴퓨터비전, 패턴인식