

특집논문 (Special Paper)

방송공학회논문지 제19권 제2호, 2014년 3월 (JBE Vol. 19, No. 2, March 2014)

<http://dx.doi.org/10.5909/JBE.2014.19.2.166>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 비디오 행동 인식을 위하여 다중 판별 결과 융합을 통한 성능 개선에 관한 연구

김 세 민<sup>a)</sup>, 노 용 만<sup>b)†</sup>

### A Study for Improved Human Action Recognition using Multi-classifiers

Semin Kim<sup>a)</sup> and Yong Man Ro<sup>b)†</sup>

#### 요 약

최근 다양한 방송 및 영상 분야에서 사람의 행동을 인식하려는 연구들이 많이 이루어지고 있다. 영상은 다양한 형태를 가질 수 있기 때문에 제약된 환경에서 유용한 템플릿 방법들보다 특징점에 기반한 연구들이 실제 사용자 환경에서 더욱 관심을 받고 있다. 특징점 기반의 연구들은 영상에서 움직임이 발생하는 지점들을 찾아내어 이를 3차원 패치들로 생성한다. 이를 이용하여 영상의 움직임을 히스토그램에 기반한 descriptor(서술자)로 표현하고 학습기반의 판별기로 최종적으로 영상내에 존재하는 행동들을 인식하였다. 그러나 단일 판별기로는 다양한 행동을 인식하기에 어려움이 있다. 따라서 이러한 문제를 개선하기 위하여 최근에 다중 판별기를 활용한 연구들이 영상 판별 및 물체 검출 영역에서 사용되고 있다. 따라서 본 논문에서는 행동 인식을 위하여 support vector machine과 sparse representation을 이용한 decision-level fusion 방법을 제안하고자 한다. 제안된 논문의 방법은 영상에서 특징점 기반의 descriptor를 추출하고 이를 각각의 판별기를 통하여 판별 결과들을 획득한다. 이 후 학습단계에서 획득된 가중치를 활용하여 각 결과들을 융합하여 최종 결과를 도출하였다. 본 논문에 실험에서 제안된 방법은 기존의 융합 방법보다 높은 행동 인식 성능을 보여 주었다.

#### Abstract

Recently, human action recognition have been developed for various broadcasting and video process. Since a video can consist of various scenes, keypoint approaches have been more attracted than template based methods for real application. Keypoint approaches tried to find regions having motion in video, and made 3-dimensional patches. Then, descriptors using histograms were computed from the patches, and a classifier based on machine learning method was applied to detect actions in video. However, a single classifier was difficult to handle various human actions. In order to improve this problem, approaches using multi classifiers were used to detect and to recognize objects. Thus, we propose a new human action recognition using decision-level fusion with support vector machine and sparse representation. The proposed method extracted descriptors based on keypoint approach from a video, and acquired results from each classifier for human action recognition. Then, we applied weights which were acquired by training stage to fuse each results from two classifiers. The experiment results in this paper show better result than a previous fusion method.

Keyword : Human action recognition, Multi classifiers, Decision-level fusion, Support vector machine, Sparse representation.

a) 한국과학기술원 정보통신공학과 (Dept. Information and Communications Engineering, KAIST)

b) 한국과학기술원 전기및전자공학과(Dept. Electrical Engineering, KAIST)

† Corresponding Author : 노용만(Yong Man Ro)

E-mail: ymro@ee.kaist.ac.kr

Tel: +82-42-350-8094

※ 본 연구는 미래부가 지원한 2013년 정보통신-방송(ICT) 연구개발사업의 연구결과로 수행되었음.

※ 이 논문의 연구결과 중 일부는 “2013년 한국방송공학회 추계학술대회”에서 발표한 바 있음.

· Manuscript received January 13, 2014 Revised February 21, 2014 Accepted February 21, 2014

## I. 서론

현재 다양한 방송 콘텐츠에서 사람의 특정한 행동 또는 행위를 검출해내는 연구가 진행되고 있다. 또한 스마트 TV 등의 보급으로 시청자들의 움직임에 반영하려는 연구도 많이 이루어 졌다. 이러한 영상 콘텐츠에서 사람의 크기 및 위치가 다양하기 때문에 최근 특징 점에 기반한 행동 인식에 관한 연구가 많이 진행되고 있다.

특징점 기반의 연구는 특징점 추출, local descriptor 추출, bag-of-words 추출, 행동 판별 등의 총 4가지 단계로 이루어진다. 먼저 영상에서 움직임이 발생하는 영역 검출하여 3차원 패치들로 추출해낸다. 특징점을 추출해내는 다양한 방법들이 제안되고 있는데 대표적으로 cuboids<sup>[1]</sup>, Harris3D<sup>[2]</sup>, Hessian<sup>[3]</sup> 등의 방법들이 있다. 이후 local descriptor 들이 3차원 패치들로부터 추출되는데 HOG<sup>[4]</sup>, HOF<sup>[5]</sup>, HOG3D<sup>[6]</sup> 등의 방법들이 주로 사용된다. 추출된 local descriptor들은 ‘Bag-of-Words’<sup>[7]</sup> 라는 히스토그램으로 표현이 되고 판별 모델 등을 통하여 해당 영상이 어떠한 행동을 포함하고 있는지를 판별하였다.

현재 대부분의 행동인식 방법들은 다양한 local descriptor를 융합하는데 초점을 두며 하나의 판별 모델을 사용하였다. 그러나 최근 보행자 검출이나<sup>[8]</sup> 영상 분류<sup>[9]</sup>를 위하여 다수개의 판별 모델을 사용하여 영상 인식 성능을 향상시킨 연구가 많이 진행 되었다. 이러한 방법들의 특징은 다양한 특성의 판별 모델을 사용함으로써 서로 상호보완적 효과를 얻을 수 있기 때문에 판별 성능을 향상시킬 수가 있었다.

본 논문에서는 support vector machine (SVM)과 sparse representation (SR)을 등을 활용하여 다중 판별 모델 기반한 행동 인식 방법을 제안하고자 한다. SVM과 SR 현재 대부분의 행동 인식 연구에서 많이 사용되어지고 있다. 또한 Liu와 Li<sup>[9]</sup>에 따르면 이 두가지 판별 모델을 사용하는 것이 영상 판별에 효율적인 성능을 가져다 주었다고 한다. 그러나 Liu의 방법은 주어진 테스트 셋(set)에서 성능을 극대화시키는 것에 초점을 주었기 때문에 실제 응용과는 거리가 있었다.

따라서 본 논문에서는 위의 SVM과 SR을 통한 다양한

행동인식 결과를 효율적으로 융합하는 것에 목적을 둔다. 이를 위하여 학습단계에서 각 판별 모델의 성능을 행동의 종류마다 측정하고, 두 판별 모델의 정확성에 따라 각 행동에 대한 가중치(weight)들을 계산한다. 따라서 각 판별 모델들은 행동 마다 정확도가 비교가 가능하고 가중치의 차이만큼 정확도가 높은 판별모델의 결과가 중요시 될 수 있게 된다. 이후 실제 테스트 영상이 입력되면 각 판별모델을 통하여 일차적인 결과를 획득한다. 이 후 각 결과들은 정규화(normalization) 과정을 거치고 학습단계에서 획득한 가중치를 이용하여 두 결과를 융합(fusion)하게 된다. 마지막으로 융합된 결과 중 가장 높은 값을 보여주는 행동으로 테스트 영상을 판별하게 된다. 본 논문에서 제안한 방법을 사용할 경우 기존의 방법보다 약 2% 이상의 성능 향상을 얻을 수가 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 영상으로부터 특징을 추출해내는 과정을 서술하며 3장에서는 본 논문에서 제안한 융합 방법을 설명한다. 4장에서는 제안된 융합 방법을 통하여 획득한 행동 인식 결과를 보여주며, 마지막 장에는 본 논문의 결론을 보여준다.

## II. 행동 인식 특징 추출

일반적으로 행동 인식 특징은 keypoint 검출과 descriptor 추출과정으로 구분할 수 있다. 본 논문에서는 행동 인식을 위한 영상 descriptor를 추출하기 위하여 cuboid<sup>[1]</sup>와 histograms of oriented gradients (HOG)<sup>[4]</sup>를 활용하였다.

### 1. Keypoint 검출

Cuboid는 많은 기존 연구에서 다른 특징점 추출보다 행동 인식에서 우수한 성능을 보여 주었다<sup>[7]</sup>. 먼저 Cuboid는 영상 내에서 특징점을 추출하기 위하여 움직임이 많이 발생하는 지점을 찾는다. 영상의 움직임 정보를 구하는 함수 R은 아래 식과 같이 정의 된다.

$$R = (I * g * h_{ev})^2 + (I * g * h_{od})^2, \quad (1)$$

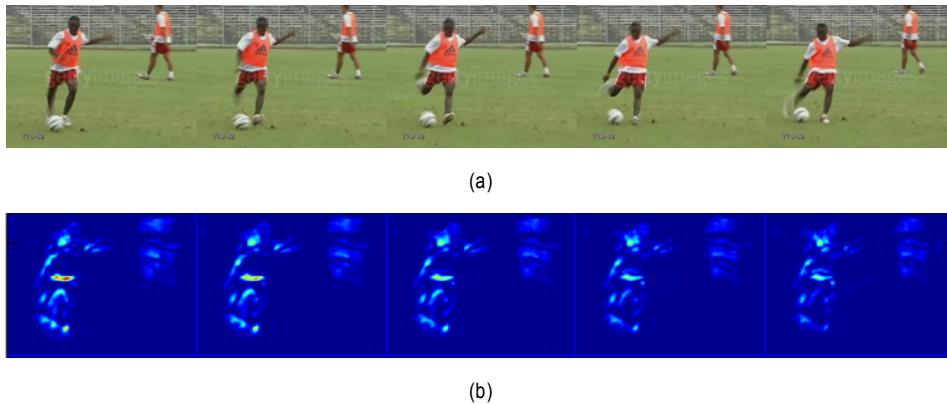


그림 1. Cuboid를 실제 영상에 적용한 예제 (원본 영상 (a)에서 식 (2)와 (3)을 활용하여 움직임이 많은 부분들이 (b)의 붉고 밝은 지점처럼 표현된다.)  
 Fig. 1. An example of cuboid for a video clip. Using Eq. (2) and (3), large motion areas are displayed with red and bright color

여기서  $I$ 는 입력된 영상이며,  $g$ 는 Gaussian smoothing 함수,  $g(x, y; \sigma)$  이며  $\sigma$ 는 필터의 scale이다. 그림 1은 실제 영상에서 함수  $R$ 을 적용한 결과이다. 이때에  $h_{ev}$ 와  $h_{od}$ 는 1D-Gabor 필터의 quadrature pair 함수들이며 아래와 같다.

$$h_{ev}(t; \tau, \omega) = -\cos(2\pi t\omega)e^{-t^2/\tau^2}, \quad (2)$$

$$h_{od}(t; \tau, \omega) = -\sin(2\pi t\omega)e^{-t^2/\tau^2}, \quad (3)$$

여기서  $\omega = 4/\tau$  이며,  $\tau$ 는 시간 축 scale이다.

그림 1은 식 (2)와 (3)을 이용하여 실제 킥(kick) 모션 영상에 적용시킨 예제이다. 그림 1(b)처럼 움직임이 많이 발생하는 부분에서 밝은 색으로 표현이 되는데 사람의 팔과 다리 등에서 발생된다. 다음 non-maximum suppression 등의 알고리즘을 이용하여 최종적으로 특징들을 추출할 keypoint들을 검출한다.

## 2. HOG 추출

검출된 keypoint에서 그림 2처럼 3D 패치들을 추출해낸다. 이때 추출된 3D 패치에서 Laptev<sup>[5]</sup>의 방법에 따라 HOG descriptor를 추출해낸다. 먼저 각 3D 패치들은  $3 \times 3 \times 2$ 의 셀로 나뉘지고 각 셀은  $X \times Y \times T$ 의 크기를 가진다. 각 셀  $I$ 에서 가로 방향과 세로 방향의 미분 값이 식

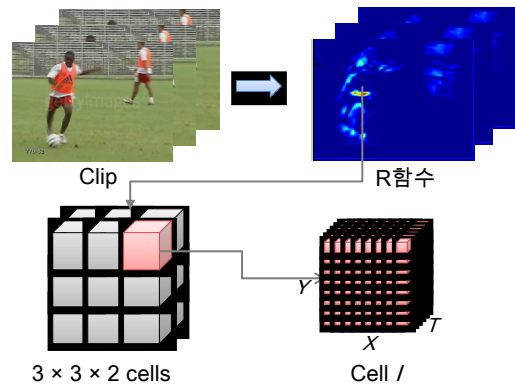


그림 2. R함수에서 하나의 keypoint를 선택한 후 3D 패치를 선택하는 예제  
 Fig. 2. An example of 3d-patch extraction from a keypoint by R function

(4)와 (5)와 같이 계산되어지고 미분 값들의 크기와 각도 등이 식 (6)과 (7)처럼 계산된다.

$$G_x = I(x+1, y, t) - I(x-1, y, t), \quad (4)$$

$$G_y = I(x, y+1, t) - I(x, y-1, t), \quad (5)$$

$$M_G = \sqrt{G_x^2 + G_y^2}, \quad (6)$$

$$\theta = \tan^{-1}\left(\frac{G_y}{G_x}\right) \in [-\pi, \pi]. \quad (7)$$

이렇게 계산된 미분의 크기와 각도들은 4개의 방향으로 양자화 되고 최종적으로 하나의 3D 패치는 72차원( $3 \times 3$

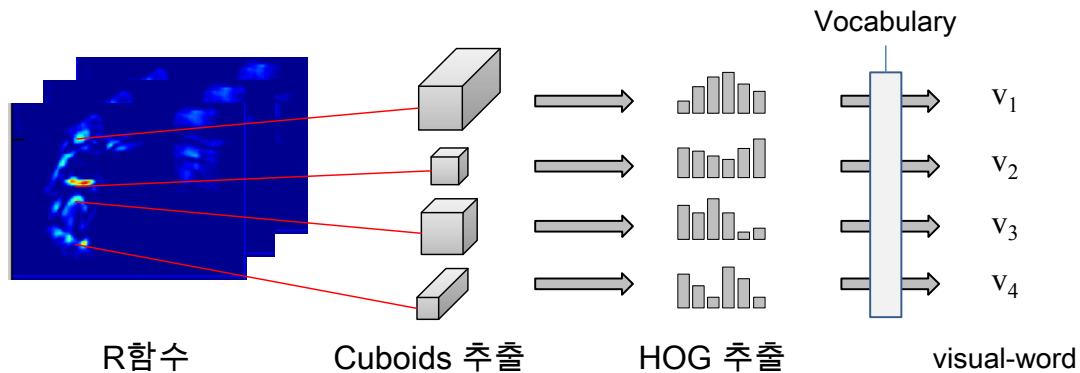


그림 3. Bag-of-words 예제 (영상에서 R합수로 cuboid들을 선택하고 HOG를 추출된다. 추출된 HOG들은 사전에 만들어진 vocabulary에 따라 visual-word로 할당된다.)

Fig. 3. An example of 'Bag-of-words'. HOG descriptors are extracted from cuboids in a clip, and assigned into visual-words of pre-built vocabulary.

$\times 2 \times 4$ ) 크기의 HOG descriptor로 표현된다.

### 3. Bag-of-visual words

영상에서 HOG들이 추출되면 bag-of-word<sup>[7]</sup>를 활용하여 하나의 히스토그램으로 표현한다. 왜냐하면 영상마다 다양한 수의 HOG들이 추출되어 직접적인 매칭으로 비교가 힘들기 때문에, 이를 동일한 크기를 가지는 히스토그램으로 정규화(normalization)를 해야 되기 때문이다. Bag-of-word는 HOG들을 사전에 정의된 vocabulary로부터 visual-word들로 할당한다. Vocabulary는 학습 영상으로부터 HOG들을 추출하고 클러스터링을 알고리즘을 이용하여 각 클러스터 센터들이 모여 있는 집합이다. 이러한 센터들은 일반적으로 visual-word라 지칭된다. 본 논문에서는 k-means 알고리즘을 활용하여 visual-word들을 생성하였다. 따라서 테스트 영상이 입력되면 HOG들을 추출하여 Wang<sup>[7]</sup>의 방법을 참고하여 visual-word들로 할당하고 bag-of-word 히스토그램으로 영상을 표현하였다. 그림 3은 cuboid로부터 visual-word들이 할당되는 개념도를 보여준다.

## III. 행동 인식 다중 판별 및 융합

본 장에서는 다중 판별기를 활용한 행동 인식 방법에 대

하여 제안한다. 1장에서 언급한 것처럼 SVM과 SR을 활용하여 행동 인식 결과를 융합(fusion)한다. SVM은 기본적으로 두 가지 클래스를 다루는 이진(binary)판별기이다. 이는 두 가지 클래스를 구성하고 있는 샘플들을 특징 공간(feature space)로 투영하여 각 클래스를 가장 잘 구분하는 hyper-plane을 찾는 것에 초점을 두고 있다. 따라서 두 개 이상의 클래스를 구분하기 위하여 모든 클래스 간에 이진 모델을 생성하고 이를 활용하여 다중 판별이 가능한 SVM이 제안되었다<sup>[10]</sup>. 본 논문에서는 histogram intersection kernel을 활용한 SVM이 사용되었다.

SR은 신호를 분해하여 이를 재구성할 때 필요한 가장 대표적인 계수들만 선택한다. 이를 통하여 신호의 크기를 줄일 수 있고 또한 신호가 가장 많이 분포된 영역을 알 수 있기 때문에, 이를 활용한 판별이 가능하다. 현재 다양한 SR 방법이 제안되고 있는데 본 논문에서는 Liu<sup>[9]</sup>의 방법에 따라 orthogonal matching pursuit (OMP)<sup>[11]</sup>를 사용하였다.

그림 4는 본 논문에서 제안하는 다중 판별 융합 순서도를 보여주고 있다. 영상에서 HOG를 추출하여 이를 bag-of-words 히스토그램으로 표현하고 SVM과 SR 판별기를 통하여 각각 일차적인 판별 결과를 추출한다. 그리고 각 결과들을  $\alpha$ 와  $\beta$ 로 표현되는데 각각 SVM과 SR의 결과 값들을 0에서 1사이의 확률 값으로 표현한 것이며 아래와 같이 정의된다.

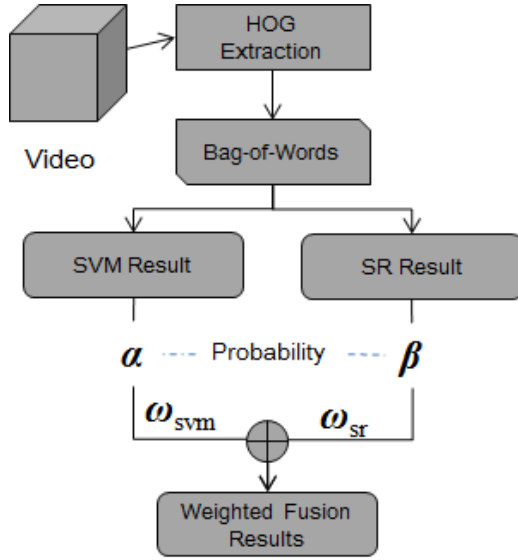


그림 4. 다중 판별 융합 개념도  
Fig. 4. The diagram of the proposed multi-classifiers fusion

$$\alpha = \alpha_1, \alpha_2, \dots, \alpha_J, \tag{8}$$

$$\beta = \beta_1, \beta_2, \dots, \beta_J, \tag{9}$$

여기서 J는 영상이 판별될 수 있는 행동의 가지수이다. 그리고 각 판별기에 대한 가중치는  $\omega_{svm}$ 와  $\omega_{sr}$ 은 학습 단계에서 각 판별 모델의 정확도로 계산된다. 이 가중치들은 학습 단계에서 계산되는데 본 논문에서는 N-fold cross-validation을 이용하여 계산하였다.

본 논문에서는 5-fold cross-validation을 사용하였는데 그림 5와 같다. 학습 데이터들은 5개의 그룹으로 나뉘지고 각 그룹은 M개의 샘플들을 가지고 있다. 이때 하나의 그룹을 선택하여 실험 데이터로 사용하고 나머지 그룹들은 SVM 또는 SR기반의 학습 모델 생성으로 이용한다. 따라서 실험 데이터로 선택된 그룹에 속한 샘플들은 나머지 그룹으로 생성된 판별기를 통하여 해당 샘플들이 각 행동으로 인식될 확률 값들을 SVM과 SR에 대하여 각각  $p = \{p_1, p_2, \dots, p_J\}$ 의 형식으로 표현된다. 이를 통하여 각 학습 모델에 대하여 precision(P)과 recall(R)을 구할 수 있고 다음과 같이 정의 할 수 있다.

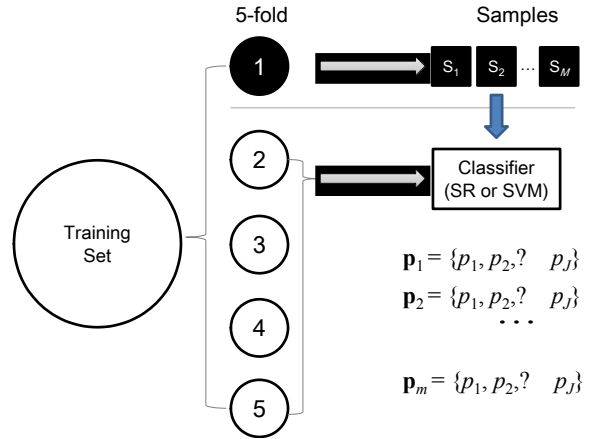


그림 5. 가중치를 획득하기 위하여 학습 데이터로부터 5-fold cross-validation을 적용하는 예제

Fig. 5. An example of weight computation from training data using 5-fold cross-validation

$$P_{svm} = p_{svm,1}, p_{svm,2}, \dots, p_{svm,J}, \tag{10}$$

$$P_{sr} = p_{sr,1}, p_{sr,2}, \dots, p_{sr,J}, \tag{11}$$

$$R_{svm} = r_{svm,1}, r_{svm,2}, \dots, r_{svm,J}, \tag{12}$$

$$R_{sr} = r_{sr,1}, r_{sr,2}, \dots, r_{sr,J}, \tag{13}$$

이렇게 구해진 P와 R을 이용하여 각 모델에 대한 F1-score<sup>[7]</sup>를 구하여 가중치를 다음과 같이 정의 할 수 있다.

$$\omega_{svm}(P_{svm}, R_{svm}) = \omega_{svm,1}, \omega_{svm,2}, \dots, \omega_{svm,J}, \tag{14}$$

$$\omega_{sr}(P_{sr}, R_{sr}) = \omega_{sr,1}, \omega_{sr,2}, \dots, \omega_{sr,J}, \tag{15}$$

$$\omega_i = \frac{2 \times p_i \times r_i}{p_i + r_i}. \tag{16}$$

따라서 일차 판별 결과인  $\alpha$ 와  $\beta$ 와 위에서 구해진 가중치  $\omega_{svm}$ 와  $\omega_{sr}$ 를 통하여 최종 판별 결과 x를 아래와 같은 공식으로 구할 수 있다. 즉, 최종 행동 인식 결과는 xi가 가장 높은 클래스로 판별되어 진다.

$$\mathbf{x} = x_1, x_2, \dots, x_f, \quad (17)$$

$$x_i = \frac{\omega_{\text{svm},i} \times \alpha_i}{\omega_{\text{svm},i} + \omega_{\text{sr},i}} + \frac{\omega_{\text{sr},i} \times \beta_i}{\omega_{\text{svm},i} + \omega_{\text{sr},i}}. \quad (18)$$

#### IV. 실험 결과

본 장에서는 제안된 다중 판별 행동 인식의 방법을 UCF-sports<sup>[12]</sup>를 활용하여 분석하였다. UCF-sports는 총 10가지의 행동들이 존재하는데 그림 6과 같다. 각 클래스별로 6 ~ 22개의 영상이 존재하고 총 150개의 영상으로 구성되어 있다. 본 실험에서는 Wang<sup>[7]</sup>의 실험 방법에 따라 실험을 진행 하였다.

먼저 SVM과 SR을 각각 단독적으로 사용하였을 경우에 성능을 confusion table로 계산하였다. 표 1과 2는 각각의 성능을 보여준다. 이때에 ID는 그림 2의 각 행동 들을 나타낸다. 두 표에서 보이는 것 처럼 전체적으로 비슷한 성능을 보여주고 있으나 3번 kick 행동이나 4번 lifting과 같이 한쪽 판별기가 우수한 성능을 보이고 있는 것을 확인 할 수 있다. 실제 UCF-sports 행동인식 DB에서 kick 행동들은 다양한 포즈를 취하고 있다. 반면 lifting 행동들은 유사한 포즈들로 이뤄져있다. 본 실험의 결과에서 SVM은 좀더 복잡한 환경에서 상대적으로 SR보다 높은 성능을 가져다주었고, 반면에 단순한 환경에서는 SR이 상대적으로 SVM보다 높은 성

표 1. SVM을 이용한 행동 인식 결과 (Confusion table)  
 Table 1. Recognition results using SVM.

ID	1	2	3	4	5	6	7	8	9	10
1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.72	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.17
3	0.00	0.20	0.60	0.00	0.05	0.10	0.05	0.00	0.00	0.00
4	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.17
5	0.08	0.00	0.17	0.00	0.58	0.17	0.00	0.00	0.00	0.00
6	0.08	0.15	0.23	0.00	0.08	0.46	0.00	0.00	0.00	0.00
7	0.00	0.08	0.00	0.00	0.00	0.08	0.67	0.00	0.00	0.17
8	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.95	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
10	0.00	0.18	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.73

표 2 SR을 이용한 행동 인식 결과 (Confusion table)  
 Table 2. Recognition results using SR.

ID	1	2	3	4	5	6	7	8	9	10
1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.72	0.06	0.00	0.00	0.00	0.00	0.00	0.06	0.17
3	0.00	0.20	0.30	0.05	0.05	0.30	0.05	0.00	0.00	0.05
4	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.08	0.00	0.08	0.00	0.58	0.25	0.00	0.00	0.00	0.00
6	0.08	0.23	0.15	0.00	0.15	0.38	0.00	0.00	0.00	0.00
7	0.00	0.08	0.00	0.00	0.00	0.00	0.75	0.00	0.00	0.17
8	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.95	0.00	0.00
9	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.92	0.00
10	0.00	0.14	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.82

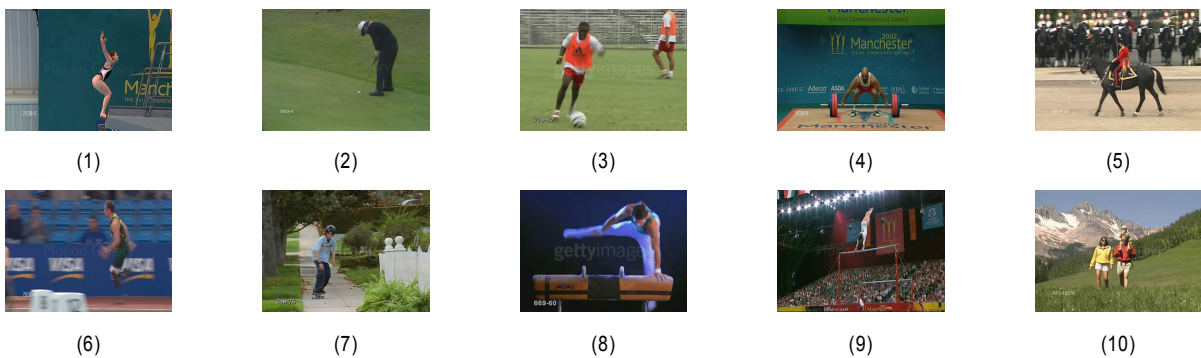


그림 6. UCF-sports의 행동 구성. (1) diving, (2) golf, (3) kick, (4) lifting, (5) riding-horse, (6) run, (7) skateboarding, (8) swing-bench, (9) swing-side, (10) walk  
 Fig. 6. Actions in UCF-sports database. (1) diving, (2) golf, (3) kick, (4) lifting, (5) riding-horse, (6) run, (7) skateboarding, (8) swing-bench, (9) swing-side, (10) walk

능을 보였다. 이는 두 개의 비디오의 bag-of-words를 비교할 때에 SVM은 공통된 visual-words들만 비교하여 공통된 점을 찾는 것에 초점을 둔다. 반면에 SR은 서로가 동일한 형태로 복원될 수 있는지를 좀 더 초점을 두기 때문에 포즈의 영향을 상대적으로 많이 받기 때문이다. 따라서 kick 행동을 판별할 때에 SVM 결과에 더 높은 가중치를 부여하고 lifting을 판별할 때에는 SR 결과에 더 높은 가중치를 부여한다면 최종적으로 향상된 판별 결과를 얻을 수 있다. 표 3은 본 논문에서 각 판별 모델을 단독으로 사용했을 때와 기존 융합 방법<sup>[9]</sup>, 본 논문에서 제안한 방법을 비교하여 최종적인 행동인식결과를 보여주고 있다. 표에서 보이는 것처럼 제안된 방법이 가장 우수한 성능을 보여주고 있는 것을 확인 할 수 있다. 따라서 포즈가 일정한 환경에서의 행동 인식은 SR에 좀 더 높은 가중치를, 반면에 포즈가 다양한 영상에서는 SVM에 좀 더 높은 가중치를 부여하면 전체적인 성능을 향상 시킬 수 있다.

표 3. 최종 행동 인식 결과 비교

Table 3. Comparison results of action recognition with the previous methods

Actions	SVM	SR	Liu [9]	Proposed
Diving	1.000	1.000	1.000	1.000
Golf	0.722	0.722	0.667	0.667
Kick	0.600	0.300	0.600	0.550
Lifting	0.833	1.000	0.833	1.000
Horse	0.583	0.583	0.667	0.667
Run	0.462	0.385	0.462	0.462
Skate	0.667	0.750	0.750	0.750
Bench	0.950	0.950	0.950	0.950
Side	1.000	0.923	1.000	1.000
Walk	0.727	0.818	0.727	0.818
Average	0.754	0.743	0.766	0.786

## V. 결론

본 논문에서는 SVM과 SR 판별기를 이용하여 영상내에 행동을 인식하는 방법을 제안하였다. 제안된 방법은 두 개의 판별 결과를 융합하기 위하여 학습모델 단계에서 각 판

별기에 대한 가중치를 계산하고, 이를 활용하여 기존의 융합 방법보다 우수한 성능을 보였다. 이는 각 행동별에 따른 판별기들의 가중치를 달리 함으로써 각 행동 인식의 성능을 최대화시킨 결과이다.

현재 행동 인식은 CCTV, 온라인 콘텐츠 검색, 가상 현실 오락 등 다양한 분야에서 이용되고 있다. 이러한 연구 분야들은 대부분 판별기를 활용한 행동 인식이 주가 이루어진, 따라서 제안한 방법은 위 연구 분야들에 대한 인식 성능 향상에 기여 할 것이다.

## 참 고 문 헌 (References)

- [1] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, Behavior recognition via sparse spatio-temporal features, in Proc. IEEE Int. Work. Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005, pp. 65-72.
- [2] I. Laptev, and T. Lindeberg, Space-time interest points, in Proc. IEEE Int. Conf. Computer Vision, 2003, pp. 432-439.
- [3] G. Willems, T. Tuytelaars, and L. Gool, An Efficient Dense and Scale-Invariant Spatio-Temporal Interest Point Detector, in Proc. Euro. Conf. Computer Vision, 2008, pp. 650-663.
- [4] N. Dalal, and B. Triggs, Histograms of oriented gradients for human detection, in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, 2005, pp. 886-893.
- [5] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, Learning realistic human actions from movies, in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, 2008, pp. 1-8.
- [6] A. Klaser, M. Marzalek, and C. Schmid, A spatio-temporal descriptor based on 3D-gradients, in Proc. British Machine Vision Conf., 2008, pp. 995-1004.
- [7] H. Wang, M.M. Ullah, A. Klaser, I. Laptev, and C. Schmid, Evaluation of local spatio-temporal features for action recognition, in Proc. British Machine Vision Conf., 2009.
- [8] O.L. Junior, D. Delgado, V. Goncalves, and U. Nunes, Trainable Classifier-Fusion Schemes: an Application to Pedestrian Detection, in Proc. IEEE conf. Intelligent Transportation Systems, 2009, pp. 432-437.
- [9] H. Liu, and S. Li, Decision fusion of sparse representation and support vector machine for SAR image target recognition, Neurocomputing Vol. 113, 2013, pp. 97-104.
- [10] C.C. Chang, and C.J. Lin, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [11] A.Y. Yang, S.S. Sastry, A. Ganesh, and YiMa, Fast  $\ell_1$ -minimization algorithms and an application in robust face recognition: A review, in Proc. IEEE Int. Conf. Image Processing, 2010, pp.1849-1852.
- [12] UCF Sports, [http://crcv.ucf.edu/data/UCF\\_Sports\\_Action.php](http://crcv.ucf.edu/data/UCF_Sports_Action.php)

---

저 자 소 개

---



김 세 민

- 2006년 : 안동대학교 멀티미디어공학과 공학사
- 2008년 : 경북대학교 컴퓨터공학과 공학사
- 2008년 : 한국산업기술재단 연구원
- 2008년 ~ 현재 : 한국과학기술원 정보통신공학과 박사과정
- 주관심분야 : 영상 패턴 추출, 이미지/비디오 분류 및 검색



노 용 만

- 1985년 : 연세대학교 전자공학과 공학사
- 1987년 : 한국과학기술원 전기 및 전자공학과 공학석사
- 1992년 : 한국과학기술원 전기 및 전자공학과 공학박사
- 1987년 ~ 1988년 : Columbia University 연구원
- 1996년 ~ 1997년 : University of California Berkeley 연구원
- 2006년 ~ 2007년 : University of Toronto 방문교수
- 1997년 ~ 현재 : 한국과학기술원 전기 및 전자공학과 정교수
- 주관심분야 : 영상 및 비디오 신호처리, 패턴인식, MPEG-7, Image and Video indexing, Spectral analysis of image