

인체 골격 정보를 이용한 Multiclass SVM 기반의 자세 인식 분류 기법

강민주, 강제원
이화여자대학교, 이화여자대학교
joo2950@ewhain.net, jewonk@ewha.ac.kr

요 약

본 논문에서는 효율적인 자세인식을 위해 인체 골격 정보를 활용한 멀티클래스 SVM(Multiclass Support Vector Machine) 학습 기반의 자세 인식 분류 기법을 제안한다. RGB 카메라로 취득한 영상을 활용하거나 깊이 카메라로부터 취득한 골격 정보를 그대로 사용하는 기존 연구와 달리 제안 기법에서는 깊이 정보로부터 추출한 인체의 3 차원 골격 정보를 이용하여 고차원의 특징을 추출하고 그로부터 자세 인식 분류를 수행한다. 제안 기법의 특징 벡터는 깊이 정보에서 취득한 골격 정보의 관절간 각도의 조합으로 구성하여 인체의 골격 편차에 강인할 뿐 아니라 특징의 차원을 효과적으로 감소시킬 수 있다. 또한 분류기로는 멀티클래스 SVM 방식 중 one-vs-one 분류 방식을 이용하여 학습 및 판별을 수행함으로써 제안 기술의 성능을 평가한다. 실험을 통해 제안 기법은 다수의 자세에서 비교하는 다른 학습 기법보다 비교적 높은 자세인식률을 보인다.

1. 서론

최근 자세 인식 기술은 카메라로부터 입력된 영상을 기반으로 사람이 취한 자세의 의미를 인식하고 공유하는 기술로 컴퓨터 비전 분야에서 많은 주목을 받고 있다. 자세는 정지되어 있는 동작을 의미하며 자세 인식 분류 기술은 보다 정밀한 인식률로 사전에 정의한 다양한 동작과 자세를 인지하기 위해 많은 연구가 진행되고 있다. 현재 자세 인식 기술은 감시 시스템, 모니터링 시스템, 스마트 TV 와 같은 실감형 인터페이스에서 널리 사용되고 있는 실정이다.

기존의 자세인식은 영상의 화소를 이용해 사람의 자세를 분류하는 연구가 주를 이루었다. Ardizzone et al. 은 사람을 검출하기 위해 영상에서 slice 를 이동시키며 고유벡터를 구하여 자세를 분류하는 방식 [1] 을 제안하였고, Tahir et al. 은 배경 차이를 이용해 사람의 실루엣을 획득하여 자세를 분류하는 방법 [2] 을 제안하였다. 최근에는 기계학습을 통해 깊이 영상으로부터 인체의 골격 정보를 획득 할 수 있게 되면서 Microsoft 의 Kinect 와 같이 사람의 인체를 쉽게 인식하는 장비가 개발되어 인체의 골격정보를 활용한 자세인식 연구들이 진행되고 있다[3~5]. Patsadu et al. 은 Kinect 를 통해 획득한 인체 골격 정보 기반의 자세 분류 기법의 성능 분석을 위해 신경망, support vector machine, decision tree, naïve Bayes 학습 방식을 이용하였다. 신경망 학습 방식에서는 인체 골격정보 20 개에 대한 3 차원 좌표 값 그대로 특징점으로 사용하였으며 서있기, 앉아있기, 누워있기에 대한 단순한 자세를 구별 할 수 있음을 보였다[5]. 하지만 사람의 위치의 변화에 따라 좌표가 변하는 경우 정확한 결과를 내기 힘들며 다양한 골격을 가진 사람을 적용하였을 때 동일한 모션을 판별하기 힘든 단점이 있다. Kang et al. 은 사람의 골격 편차를 극복하기 위해 3 차원 골격정보로부터 고차원의 각도특징 벡터를 추출하여 신경망 학습을 이용한 자세 분류 기법을 제안하였다 [6].

본 논문에서는 자세의 인체 골격 정보를 활용한 멀티클래스 SVM(Multiclass Support Vector Machine) 기반의

자세 인식 분류 기법을 제안한다. 제안기법에서는 Kinect 센서를 이용해 영상의 깊이 정보로부터 3 차원 골격정보를 획득한 후 특정 관절의 각도를 계산하고 조합하여 고차원의 특징 벡터를 생성한다. 기존연구와 달리 영상 전체 혹은 관심 영역의 화소를 그대로 사용하지 않기 때문에 제안 기법은 특징벡터 차원을 효율적으로 감소 시키는 것이 가능하며 사람의 골격 편차에도 강인한 특성을 가진다. 추출한 각도 특징은 멀티클래스 SVM 학습을 진행하여 사람간 자세 모양 편차에 강인한 분류기를 생성함으로써 새로 입력된 영상에 나타난 자세가 어떤 자세인지 분류한다.

2. 제안기법

각도 특징은 먼저 인체의 관절의 3 차원 좌표 값 획득을 통해 추출된다 [6]. 사람이 취한 자세를 4 분할($i=1,2,3,4$) 하여 왼쪽 팔, 오른쪽 팔, 왼쪽 다리, 오른쪽 다리로부터 3 개씩($j=1,2,3$) 추출된다. 오른쪽 팔을 예로 들자면, 그림 1 에서 보듯이 어깨-팔꿈치, 팔꿈치-손목 사이의 각도 u^i 를 어깨의 좌표가 $P_1^i(x_1^i, y_1^i, z_1^i)$ 팔꿈치의 좌표가 $P_2^i(x_2^i, y_2^i, z_2^i)$ 손목의 좌표가 $P_3^i(x_3^i, y_3^i, z_3^i)$ 이라고 할 때, 다음과 같은 수식을 사용하여 구한다.

$$u^i = 180^\circ - \left\{ \arctan\left(\frac{y_2^i - y_1^i}{x_2^i - x_1^i}\right) - \arctan\left(\frac{y_3^i - y_2^i}{x_3^i - x_2^i}\right) \right\} \times \frac{180^\circ}{\pi} \dots (1)$$

두 번째 특징점은 팔 모션에서의 각도 ϕ 이며 골격 정보 중 어깨와 팔꿈치의 좌표를 사용한다. 어깨 $P_1^i(x_1^i, y_1^i, z_1^i)$ 에서 $z = z_1^i$ 평면과 어깨-팔꿈치 좌표인 P_2^i 직선이 이루는 각을 ϕ^i 라고 한다. ϕ^i 의 계산은 다음과 같다.

$$\phi^i = \arctan\left(\frac{z_2^i - z_1^i}{x_2^i - x_1^i}\right) \times \frac{180^\circ}{\pi} \dots (2)$$

세 번째 특징점은 θ 이며, 골격 정보 중 어깨와 팔꿈치의

좌표를 사용하여 어깨 $P_1^i(x_1^i, y_1^i, z_1^i)$ 에서 $y = y_1^i$ 평면과 어깨-팔꿈치 좌표인 $p_1^i - p_2^i$ 직선이 이루는 각을 θ^i 라고 한다. θ^i 의 계산은 다음과 같다.

$$\theta^i = \arctan\left(\frac{y_2^i - y_1^i}{x_2^i - x_1^i}\right) \times \frac{180^\circ}{\pi} \dots (3)$$

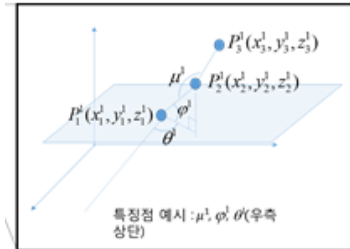


그림 1. 추출하는 각도 그림 [6]

4 개의 모든 분할에 대해 3 개의 각도를 동일한 방식으로 추출하여, 12 개의 각도 특징으로 구성된 집합 A 를 구성한다 ($A = \{u^1, \varphi^1, \theta^1, u^2, \varphi^2, \theta^2, u^3, \varphi^3, \theta^3, u^4, \varphi^4, \theta^4\}$). 그리고 각 자세에 따라 label 1~ label N 을 할당하여, 학습을 진행하고, 테스트 단계에서 새로운 영상에서 관찰된 자세의 label 을 잘 예측하는지 확인한다.

특징벡터를 학습하는 기법으로 멀티클래스 SVM 기법을 사용한다. SVM 은 기본적으로 이진 분류기로 그림 2 와 같이 두 개의 클래스의 데이터 분포를 구분하는 초평면을 찾는 문제이다. 최적 초평면은 평면에서 가장 가까운 거리의 두 클래스의 각 특징 벡터 (support vectors)간 거리를 최대로 갖게 하여 찾는다. SVM 은 기본적으로 이진 분류기이므로 멀티클래스를 분류하기 위해서 one-vs-one 접근법 혹은 one-vs-all 접근법을 사용한다. One-vs-one 접근법은 전체 분류하고자 하는 클래스의 개수가 m 개 라고 할 때 m 개의 클래스 중 서로 다른 2 개의 클래스들을 반복적으로 선택하여 support vector 간 거리 (d)가 가장 큰 초평면을 만든다. 최종적으로 $m(m-1)/2$ 개의 분류기를 생성하여 새로운 데이터에 대한 클래스를 예측한다. 새로운 데이터에서 추출된 특징점은 모든 이진 분류기에 입력되어 가장 많이 예측된 클래스 i 를 최종 클래스로 선정한다 [8]. 제안 기법에서는 one-vs-one 접근법을 one-vs-all 대비 성능 비교하여 실험적으로 선택하였다.

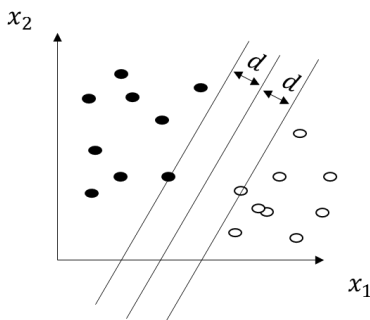


그림 2. 이진 분류 SVM 의 예 [7]

3. 실험 결과

제안 기법은 효율적인 자세 인식 분류 성능을 보이기 위해 기존 학습 기법들과 자세 인식 분류 성능을 비교하였다. 규정이 정해진 자세 6 가지를 인식하는 실험을 구성하였으며 [9], 그림 3 은 표 1 에서 사용한 자세 중 ‘ 이해 불가, ’ 옆으로 나란히’ 자세를 취했을 때 획득한 깊이 영상의 예시이다. 제안 기법에서 사용하는 골격 정보를 획득하기 위한 깊이 정보는 Microsoft 사의 Kinect 장비를 통해 획득 하였다. 획득한 깊이 영상의 해상도는 480x680 화소 이다. 제안 기법의 학습단계에 사용된 데이터의 개수는 240 개 이며, 그 중에서 10%는 과 적합을 방지 하기 위해 사용되었으며, 학습 후 분류기의 자체 테스트를 위해 10%의 데이터를 사용하였다. 제안 기법이 적용된 분류기를 테스트 하기 위해 120 개의 데이터 셋을 사용하였으며, 성능은 표 1 에 나타내었다.

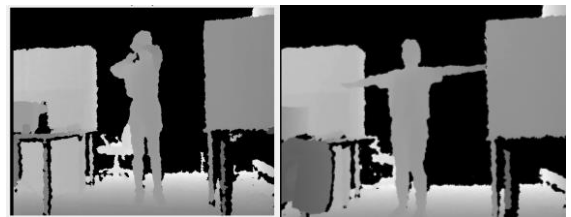


그림 3. 획득한 깊이 영상의 예 (좌: 이해 불가, 우: 옆으로 나란히)

기존 학습 기법과 성능 비교를 위해 멀티클래스 SVM 기법 중 one-vs-all 기법과 인공신경망 학습 기법을 함께 비교하였다. 다른 기법에도 위와 같은 실험 환경을 적용하였다. 표 1 에서 사용한 각 자세의 자세 인식률(%)은 아래의 계산 방식을 사용하였다.

$$\text{자세동작의 인식률}(\%) = \frac{\text{올바르게 인식한 자세}}{\text{해당 자세의 전체 영상 수}}$$

표 1 에서 사용한 동작은 군대의 수신호로 사용되고 있는 대기, 이해 불가, 정지등과 같은 정지된 자세이다[9]. 실험 결과, 제안 기법의 평균 성능은 94.3%로 멀티클래스 SVM one-vs-all 방식의 성능 93.3%와 인공신경망의 성능 87.9% 보다 높으며 제안 기법이 총 6 개의 자세 중 다수의 자세에서 인식 성능이 가장 높음을 보이고 있다. 평균 성능은 총 10 번의 실험으로 평균을 계산하여 산출하였다.

(%)	Ready to Move	As you were	Do not understand	Halt or stop	Wed ge	Line
Multiclass -SVM one vs one	94.1	97.4	95.5	94.6	88.4	97.3
Multiclass-SVM	89.1	97.0	91.1	94.3	94.3	96.0

one vs all						
Neural network [6]	59.5	86.5	96.5	97.0	91.0	97.0

표 1. 제안 기법과 기존 기법의 성능 비교

4. 결론

본 논문에서는 사람의 골격 정보를 추출하여 12 개의 각도를 특징으로 삼아 멀티클래스 SVM 학습 기법 기반의 자세 인식 기법을 제안하였다. 제안 기법에 따르면 자세를 취하고 있는 주 표현 부분을 4 분할 하여, 각 분할마다 3 개의 관절에 대한 공간좌표를 얻은 뒤 각도를 계산하고, 고차원의 특징 벡터를 사용해 one-vs-one 접근법을 적용한 멀티클래스 SVM 학습을 진행하였다. 제안 기법은 다른 학습 기법인 와 one-vs-all 접근법을 적용한 멀티클래스 학습 기법과 인공지능망에 대해 비교하였을 때 다수의 자세에서 가장 높은 자세 인식률을 보였다.

5. 참고문헌

[1] Ardizzone, Edoardo, Antonio Chella, and Roberto Pirrone. "Pose classification using support vector machines." Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on. Vol. 6. IEEE, 2000.

[2] Tahir, Nooritawati Md, et al. "PCA- Based Human Posture Classification." Jurnal Teknologi 46.1 (2012): 35-44.

[3] Hachaj, Tomasz, and Marek R. Ogiela. "Rule-based approach to recognizing human body poses and gestures in real time." Multimedia Systems 20.1 (2014): 81-99.

[4] Raptis, Michalis, Darko Kirovski, and Hugues Hoppe. "Real-time classification of dance gestures from skeleton animation." Proceedings of the 2011 ACM SIGGRAPH/Eurographics symposium on computer animation. ACM, 2011.

[5] Patsadu, Orasa, Chakarida Nukoolkit, and Bunthit Watanapa. "Human gesture recognition using Kinect camera." Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on. IEEE, 2012.

[6] Minjoo Kang, Sukyung Ryu, Nayoung Kim, Jieun

Lee, Jewon Kang. "Neural network based human posture classification using 3D skeleton structure." Conference of The Korean Society Of Broadcast Engineers, Aug., 2015, 322-325.

[7] Tong, Simon, and Daphne Koller. "Support vector machine active learning with applications to text classification." The Journal of Machine Learning Research2 (2002): 45-66.

[8] Hsu, Chih-Wei, and Chih-Jen Lin. "A comparison of methods for multiclass support vector machines." Neural Networks, IEEE Transactions on 13.2 (2002): 415-425.

[9] Lampton, Donald R., et al. Gesture Recognition System for Hand and Arm Signals. No. ARI-RN-2003-06. ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA, 2002.