Proceedings of the Korean Data And Information Science Society,
October 22-23, 2004, pp. 105-109

동작 인식 방법에서 주성분 분석법의 활용에 관한 연구

권용만[1] · 홍연웅[2]

요 약

동작(motion) 인식 방법 있어서 2차원 정보는 영상이라는 2차원 정보만을 이용하기 때문에 여러 가지 행동의 제약이 있으며 이것은 인식률을 저하시킬 뿐 아니라, 그 응용 면에서 자연스럽지 못하게 된다. 이러한 문제점을 보완하기 위하여 3차원 정보를 사용하는 시스템으로 발전하게 되었지만 영상 기반의 3차원 정보는 에러가 많이 포함되어 있을 뿐만 아니라 차원수가 높기 때문에 일정한 특징을 찾아내기 어렵다. 본 연구에서는 동작을 모델링하고 분석하기 위해 주성분 분석법을 사용하는 방법을 기술한다. 주성분 분석법은 낮은 차원의 영상 공간을 얻기 위해서 사용되는데, 이 방법을 사용함으로써 3차원 데이터가 가지는 에러의 영향을 줄일 수 있게 되고, 차원 축약의 효과를 얻을 수 있다.

주요용어: 동작 인식 방법, 주성분 분석법

1. 서론

동작 인식은 사람과 컴퓨터의 상호작용으로 이용되는 한 수단을 말하는데, 사람의 동작이 컴퓨터에 의해 인식된다. 동작을 컴퓨터의 입력으로 인식하는 것은 컴퓨터가 동작을 인식할 수 있고, 컴퓨터 디스플레이 상에 실시간으로 그것들을 표현해준다. 다른 형태의 동작 인식에 대한 연구가 현재 진행 중이다. 얼굴 추적, 안구의 움직임, 입술의 움직임을 읽는 것 등 역시 컴퓨터와의 상호작용을 제공하기 위한 방편으로서 고려되고 있다. 또한 사람의 몸 전체와, 몸 동작의 범위가 컴퓨터 효과를 만들어 내기 위해 사용되는 멀티미디어 실험들이 행해져왔다.

인간의 동작을 분석하고 이를 인식한다는 것은 인간이 복잡한 3차원 관절체임을 고려해 볼 때 매우 어려운 일이다. 이러한 문제를 해결하기 위해 신체 부위의 위치 혹은 움직임 정보를 알 수 있는 기계적인 센서를 이용하는 방법을 사용한다. 그러나 외부의 추가적인 장치를 사용하기 때문에 자연스러운 인간과 컴퓨터의 상호작용이라 할 수 없다. 또한 시스템으로의 제약이 많기 때문에 다양한 시스템으로의 응용이 어려워지게 된다. 이러한 이유에서 보다 자연스러운 인간-컴퓨터 인터페이스(interface)에 대한 연구가 필요해졌고, 추가적인 장치의 제약을 없애기 위하여 비디오 카메라를 사용하게 된다. 비디오 카메라를 이용한 동작 인식 기술은 컴퓨터 비젼 기술을 기반으로 카메라를 통해 입력되어진 영상을 분석함으로서 인간의 행동을 분석하는 기술이다.

[1]501-759, 광주광역시 동구 서석동 375번지, 조선대학교 컴퓨터통계학과 부교수.
E-mail : ymkwon@chosun.ac.kr
[2]750-711, 경북 영주시 풍기읍 교촌동, 동양대학교 전자상거래 · 정보산업학부 교수.

대표적인 예로 PFinder[1]를 들 수 있다. PFinder에서는 카메라를 통하여 입력된 영상을 분석하여, 대략적인 인간의 모델을 만들고 이 모델이 어떠한 형태로 움직이는가를 분석함으로써 인간의 움직임을 인식 할 수 있다고 기술하고 있고, 이 기술을 기반으로 하는 다양한 응용도 소개하고 있다. 그러나 이 기술은 영상이라는 2차원 정보만을 이용하기 때문에 여러 가지 행동의 제약이 있으며, 이러한 제약은 인식률을 저하시킬 뿐 아니라, 그 응용 면에서 자연스럽지 못하게 된다. 이러한 단점을 보완하기 위하여 3차원 정보를 사용하는 시스템으로 발전하게 되었다. 2차원의 형상 정보를 이용하는 대신에 동작의 의미를 많이 포함하고 있는 손과 발의 3차원 움직임 정보를 이용함으로써 움직임의 제약을 많이 줄일 수 있었다[2]. 또한 Murase 와 Nayar[4]는 Parametic Eigenspace에서 객체 영상집합들에 대한 픽셀 값들의 공간적 위치 값들이 주로 각 영상에서 어디에 분포하는가를 계산하여 확률 빈도가 높은 값들을 고유값에 비례하여 재구성하는 방법을 제안하였다. 여기서 Parametic Eigenspace란 N명의 사람이 M개의 다른 관점을 가지고 있을 때, $N \times M$개의 이미지 조합으로부터 얻어지는 모든 사람의 고유공간(eigenspace)에서 인식과 위치를 판단하는 방법이다. 3차원 객체의 비교를 위하여 하위의 조각들로 분해한 후 그 분해된 것 끼리 비교를 하여 유사성을 검사하는 이론을 제시하였다. 그러나, 3차원 정보의 사용으로 인한 오차 혹은 오인식의 부작용도 무시할 수 없다. 다시 말해서 움직임 정보를 얻기 위해 한 장의 영상을 사용할 때 보다 여러 장의 영상을 사용함으로써 오차가 발생할 수 있는 가능성이 더욱 커지게 된 것이다.

본 연구에서는 불안정한 3차원 정보에 의해 인식률이 저하되는 문제를 해결하기 위해서 주성분 분석법(principal component analysis)을 이용하였다. 주성분 분석법은 주로 다루기 힘든 고차원의 신호를 낮은 차원으로 줄여 다루기 쉽게 해주는 통계적 방법을 일컫는다. 이 방법은 80년대 정체기에 있던 외관 인식에 관한 연구를 90년대 초반에 들어서 다시 불을 붙였던 방법으로 외관 인식 분야에서 가장 보편적으로 쓰이는 방법이라고 할 수 있다. 이 방법은 외관만이 존재하는 낮은 차원의 영상공간을 얻기 위해서 사용되었는데 그렇게 하여 구한 외관만을 위한 공간을 고유공간이라 하며 그 공간을 구성하는 좌표계에 해당하는 벡터들은 학습영상들의 공분산행렬에 대해서 선형대수학에서의 고유값(eigenvalues), 고유벡터(eigenvectors) 문제를 풀어서 계산 되어진다. 등록되는 영상들은 외관만을 위한 고유공간에서의 새 좌표계로 변환되어 저장되며 나중에 인식할 때에는 새로 들어온 영상들을 역시 외관만의 고유공간상의 좌표계로 변환하여 그 둘 사이의 떨어진 거리를 측정함으로써 등록된 외관과의 일치 여부를 결정하게 된다.

2. 주성분 분석법을 이용한 동작 모델링

연구하고자 하는 시스템에서는 동작을 인식하기 위하여 <표 1>과 같이 6개의 동작으로 정의하기로 한다. <표 1>에서 정의한 동작들을 구별하기 위하여 수학적으로 모델링 하기 위하여 확실한 특징 벡터의 선정하기로 하자[2]. 본 시스템에서는 3차원 상에서 머리를 기준으로 하는 양손(x_{lh}, x_{rh}), 양발(x_{lf}, x_{rf})의 상대 위치와 머리 이동 속도(v_h)와 양발 이동 속도(v_{rf}, v_{lf})벡터를 특징으로 사용한다.

<표 1> 정의한 동작

동작	내용
왼쪽 펀치	왼손을 앞으로 뻗는 동작
오른쪽 펀치	오른손을 앞으로 뻗는 동작
왼쪽 킥	왼발을 앞으로 차는 동작
오른쪽 킥	오른발을 앞으로 차는 동작
뛰기	가볍게 뛰는 동작
앉기	다리를 구부려 앉는 동작

동작의 특징은 총 21차원의 입력 영상벡터로 생성되며, 이를 분석함으로써 동작을 구별해 낼 수 있다. 그러나 입력 데이터의 차원이 높으면 계산량이 많아서 실시간 영상을 인식하는데 어려움이 따르며, 또한 입력 데이터의 모든 차원이 모두다 일반적인 특징을 갖고 있지 않은 경우도 있기 때문에 고차원의 입력 데이터를 낮은 차원으로 줄여 이를 효과적으로 모델링 하는 방법을 사용하여야 한다. 따라서 본 연구에서는 주성분 분석법을 사용하여 고유공간에서 동작을 모델링 하기로 한다[3, 4].

입력 영상벡터를 다음과 같이 정의하면,

$$\underline{i}=[i_1, i_2, \cdots, i_{21}]^T=[\underline{r}_{lh}, \underline{r}_{rh}, \underline{r}_{lf}, \underline{r}_{rf}, \underline{v}_h, \underline{v}_{rf}, \underline{v}_{lf}]^T$$

여기서, $\underline{r}_{lh}=[r_{lh}^{(1)}, r_{lh}^{(2)}, r_{lh}^{(3)}]$이고 비슷하게 $\underline{v}_{lf}=[v_{lf}^{(1)}, v_{lf}^{(2)}, v_{lf}^{(3)}]$이다. 또한 여기서. 은 각 $r_{lh}^{(1)}, r_{lh}^{(2)}, r_{lh}^{(3)}$ 각 3차원상에서 머리를 기준으로 하는 왼손의 상대 위치이다. 인식 작업에서 미리 정하여진 영상 크기를 맞추거나 인식 방법이 크기(scale)를 변하지 않게 하게 하기 위하여 얻어진 모든 영상의 크기를 같게 하면 $\hat{\underline{i}}_j=\underline{i}_j/\|\underline{i}_j\|$이 된다.

총 영상수를 M이라 하면 완전한 영상 세트는 다음과 같다.

$$\{\hat{\underline{i}}_1, \hat{\underline{i}}_2, \hat{\underline{i}}_3, \cdots, \hat{\underline{i}}_M\}$$

입력 벡터의 평균영상 c 와 공분산 행렬(covariance matrix) Q는 다음과 같이 구해진다.

$$c=(1/M)\sum_{j=1}^{M}\hat{\underline{i}}_j,$$

$$P \triangleq [\hat{\underline{i}}_1-c, \hat{\underline{i}}_2-c, \hat{\underline{i}}_3-c, \cdots, \hat{\underline{i}}_M-c]^T,$$

$$Q \triangleq PP^T$$

P는 $21\times M$, 여기서 21은 각 영상당 픽셀의 수이며 M은 세트당 총 영상수이다. Q 는 21×21이고 아주 큰 행렬이다. 고유벡터 e_k와 공분산 행렬 Q 에 대응하는 고유값 λ_k

를 식 (1)에서 특이값 분해(singular value decomposition)를 이용하여 구할 수 있다.

$$\lambda_k e_k = Q e_k \tag{1}$$

공분산 행렬 Q 에 대하여 식 (1)을 만족하는 고유치 λ_k를 구해보면 <표 2>과 같다. <표 2>는 21차원 입력 영상벡터간의 공분산계수를 통해서 구한 각 축의 고유값과 누적점유율이다. 주성분공간의 차원을 결정하는 몇 가지 기준에 의하여 고유값이 0.7이상이면서 누적점유율이 80%이상이 되는 첫 5개의 주성분을 택하기로 한다. <표 2>에서도 알 수 있듯이 가능한 모든 입력 영상벡터에 대한 주성분 분석 결과 고유값의 크기가 큰 쪽의 5개의 벡터가 고유공간에서의 기여도가 높은 것으로 나타났다. 따라서 본 논문에서는 주성분 분석법을 이용하여 21차원 입력 영상벡터를 5차원의 고유공간으로 차원을 축약시킨 뒤 고유공간 내에서 동작을 모델링하고 분석한다.

<표 2> 고유공간에서 고유벡터의 기여도

성분	고유값	누적점유율(%)
1	179.556152	58.88
2	66.333160	80.58
3	22.883926	87.70
4	13.751337	91.97
5	6.012637	93.94
6	4.006974	.
7	3.146442	.
8	2.440615	.
9	1.912645	.
10	1.274066	.
.	.	.
.	.	.
.	.	.
21	.	.

미리 정의된 동작에 대하여 고유공간에서 모델링하고, 새로 입력되는 동작을 고유공간 내에서 가장 가까운 동작으로 결정함으로써 입력 동작을 분석할 수 있다. 모델 동작과 입력 동작의 유사도를 측정하기 위하여 각 모델 동작에 대한 고유공간에서의 평균값을 구하고, 이 평균값과의 5차원의 유클리드(Euclidean) 거리를 계산한다.

3. 결론

본 연구에서는 3차원 동작 정보를 이용한 동작 인식 시스템에 대하여 기술하였다. 인간의 동작을 수치적으로 표현하기 위해 5개의 신체 특징점을 정의했고, 이들 특징점의 3차원 정보를 계산하여 동작 정보를 추출하였다. 이러한 데이터를 모델링 하기

위해서 주성분 분석법을 사용하였다. 본 시스템은 고유공간에서 동작을 비교하기 위해 유클리드 거리 기반의 방법을 사용한다. 유클리드 거리 기반의 방법은 고유공간 내에서의 모델 동작의 분포형태를 고려하지 않기 때문에 모델 동작이 일반적이지 않을 경우 인식이 되지 않는 경우가 생길 수 있다. 이러한 문제는 모델 동작의 분포형태를 고려한 비교 방법으로 개선함으로서 해결할 수 있을 것이다. 또한 3차원 동작 데이터 또한 에러를 많이 포함하고 있음을 확인하였다. 이는 에러를 보정하는 프로세스를 추가함으로서 해결 할 수 있으며, 보다 안정적인 결과를 보일 것으로 예상된다.

참고문헌

[1] Chistopher Wren, Ail Azarbayejani, Trevor Draeerl, and Alex Pentland. Pfinder: Real-time tracking of the human body. In Photonics East, SPIE Proceedings Vol. 2615 Bellingham, WA, 1995. SPIE.

[2] Lee W. Campbell, David A. Becker, Ail Azarbayejani, Aaron F. Bobick, Alex Pentland. Invariant features for 3-D gesture recognition. In Second International Workshop on Face and Gesture Recognition, Killington VT Oct., 1996

[3] Richard Hartley and Andrew Zisserman, Multiple View Geometry in computer vision.

[4] Hiroshi Murase and Shree K, Nayar, "Visual Learning and Recogntion 3-D object from appearance", international journal of Computer Vision, Vol,14,1995.