볼쯔만 머신을 이용한 인간 모션 학습 및 생성*

이충연[1○], 김지섭[2], 김은솔[2], Karinne Ramírez Amaro[3], Michael Beetz[3], 장병탁[1,2]
{cylee, jkim, eskim} @ bi.snu.ac.kr, ramirezk@in.tum.de, beetz@cs.tum.edu, btzhang@bi.snu.ac.kr
[1]서울대학교 뇌과학협동과정, [2]서울대학교 컴퓨터공학부, [3]뮌헨공과대학교 컴퓨터공학부

Learning and Generation of Human Motion using Boltzmann Machines

Chung-Yeon Lee[1○], Ji-seob Kim[2], Eun-Sol Kim[2], Karinne Ramírez Amaro[3], Michael Beetz[3], Byoung-Tak Zhang[1,2]
[1]Interdisciplinary Program in Neuroscience, Seoul National University, [2]School of Computer Science and Engineering, Seoul National University, [2]Fakultät für Informatik, Technische Universität München

요 약

카메라 영상 기반 모션 캡처 시스템을 이용하여 수집한 주방 공간에서의 인간 행동 데이터로부터 기본 동작들의 모션 데이터를 추출 및 전처리한 후, non-linear generative model인 cRBM을 이용하여 각 모션 데이터를 학습하였다. 초기 모션 일부를 seed로 사용하여 자동으로 생성한 새로운 모션들은 학습에서 사용된 모션들과 동일한 작업을 수행하는 결과를 나타냈다. 본 실험 결과는 일상 환경에서 인간으로부터 획득한 모션 데이터를 이용하여 생성된 가상의 모션 데이터를 통해 로봇이나 아바타의 움직임을 보다 유연하게 제어할 수 있으며, 또한 인간이 유아기 때부터 주위 사람들의 행동을 따라하며 자연스럽게 자신의 행동 방법을 배우는 방식과 유사한 행동 학습 메커니즘이 로봇에게도 적용시킬 수 있음을 보인다.

1. 서 론

일상 속에서 인간의 움직임은 인체 구조, 이동 환경, 물리 법칙, 개인의 심리나 개성 등 다양한 요소에 의해 결정되며, 이처럼 복합적인 요소들이 고려된 사실적인 인간 모션을 컴퓨터에서 학습하거나 생성하는 것은 단순한 작업이 아니다.

모션 생성을 위해서는 먼저 위치, 속도, 방향 등 측정 대상의 움직임에 대한 정보를 획득해야 하는데, 이것을 모션 캡처(motion capture)라고 한다. 모션 캡처 시스템의 종류로는 인체의 각 부위에 회전식 측정 장치를 부착하여 관절의 움직임을 획득하는 기계식 방법, 자기장 계측 센서를 이용하는 자기식 방법, 그리고 적외선 광학식 카메라를 이용하는 광학식 방법 등이 있다. 하지만 이러한 모션 캡처 시스템의 경우 기계 장비나 마커와 같은 측정 장치를 신체에 붙여야 하며, 고가의 모션 캡처 장비가 설치된 특정 실험실에서만 측정이 가능하다는 점 등의 문제점이 있기 때문에 일상에서의 모션 데이터를 획득하는데 많은 제약이 따른다. 이에 본 논문에서는 마커를 부착하지 않고 영상 기반 모션 캡처 시스템을 이용하여 주방 공간에서의 인간 행동을 기록한 TUM Kitchen Data Set [1]을 사용함으로써 보다 자연스러운 일상에서의 인간 모션을 학습하고 생성할 수 있도록 하였다.

모션 캡처 시스템을 통해 측정된 인간의 사실적인 움직임 데이터는 교육, 보안, 영화, 게임 산업 등 다양한 분야에서 활용 가능하지만, 새로운 동작이 필요할 때마다 매번 모션 캡처를 수행해야 하는 문제로 인하여, 키프레임 애니메이션 또는 역운동학(inverse kinematics), 역동역학(inverse dynamics)과 같은 시뮬레이션 방법과 달리 기존 데이터를 재사용하기 어렵다는 단점이 있다.

이에 반해 데이터 기반 모션 생성[2]은 모션 데이터베이스로부터 대상 개체가 움직이는 원리를 유추하여 실제 사람과 같은 자연스러운 움직임을 재현해낼 수 있는 기법이다. 또한 데이터 기반 접근 방법은 학습 데이터로부터 모델을 유추하는 기계학습의 목적과도 부합되기 때문에[3, 4], 현재 기존의 다양한 기계학습 알고리즘들을 인간 모션 학습과 생성 문제에 적용하려는 시도들이 이루어지고 있다.

개인의 개성이나 스타일이 반영된 모션을 생성하는 방법으로, Amaya[5]는 frequency domain에서의 speed와 spatial amplitude를 이용하여 추출한 감정적 모션의 특징을 반영하여 일반 모션으로부터 새로운 감정적 모션을 생성하는 방법을 제안했으며, Hsu[6]는 서로 다른 모션의 시공간적인 dense correspondences를 자동으로 계산하는 기법과 선형시불변(linear time invariant system) 모델을 이용하여 내용을 유지하면서 다른 스타일을 갖는

* 이 논문은 교육과학기술부의 재원으로 한국연구재단의 지원을 받아 수행된 연구(0421-20110032, 지능형 추천 서비스를 위한 인지기반 기계학습 및 추론기술, Videome), (2010-0018950, 뇌정보처리 기반 사용자 의도 변화 모델링 및 예측 기술 개발), (2010-0018950, 로봇팔의 유연한 모션 생성을 위한 기계학습 연구)이며, 교육과학기술부의 BK21-IT사업에 의해 일부 지원되었음.

모션을 생성하는 방법을 제시했다. 또한 Brand[7]는 다양한 댄스 동작에 대한 모션 데이터를 Hidden Markov Model (HMM)을 이용하여 학습함으로써 새로운 댄스 동작을 생성하는 방법을 제시했다.

본 논문에서는 Conditional Restricted Boltzmann Machine (cRBM)[8]을 이용하여 주방 환경에서의 자연스러운 동작을 학습하고, 동일한 작업 수행시의 모션에 대한 새로운 모션 데이터를 생성 하였다.

2. 연구 내용 및 방법

2.1. 모션 데이터 획득 및 전처리

본 논문에서 사용한 TUM Kitchen Data Set은 그림 1과 같이 실제 주방 환경을 갖춘 실험실에서 다수의 피험자가 선반과 서랍장에서 식기를 꺼내어 식탁에 올려두는 등의 행동을 트래킹한 것으로, 마커를 부착하지 않은 피험자의 동작을 비디오 카메라 4대(25Hz)를 이용하여 촬영한 영상 데이터로부터 피험자들의 모션을 획득하였다.

(a) 서랍을 여는 동작 (b)선반을 여는 동작

그림 1. 모션 데이터 획득 환경

데이터 포맷은 BioVision사에서 제안한 BVH 형식으로 skeleton hierarchy information과 각 joint들의 angle degree들로 구성된다. 실험에 사용한 skeleton 모델은 그림 2와 같으며, 이때 중심 관절인 BEC의 경우 position과 orientation에 대해 6 DOF (degree of freedom)의 정보를 가지고, 다른 관절들의 경우 모두 3 DOF의 orientation 정보를 갖는 구조이다.

본 실험에서는 28개의 관절로 이루어진 기존 skeleton 모델에 5개의 dummy joints를 머리와 팔, 다리의 각 말단 부위에 더하여 33개의 관절을 사용하였으며, 각 관절의 중심 관절에 대한 오프셋을 각 프레임에서의 위치값으로 사용하여, 총 198차원에 해당하는 정보가 사람의 동작을 표현하는 데이터로 사용하였다. 단, 기존 데이터의 경우 피험자가 조리대와 테이블을 수차례 이동하면서 정해진 다양한 동작을 임의로 수행한 것을 모두 포함하고 있기 때문에, 하나의 모션을 학습하기에 적합하지 않다. 이에 본 논문에서는 그림 1의 (a)와 (b)와 같이 서랍을 여는 동작(모션 A)과 선반을 여는 동작에 대한 모션 데이터(모션 B)를 따로 추출하였다. 한 가지 동작을 학습하기 위해 서로 다른 피험자로부터 추출된 3개의 모션을 사용하였으며, 평균적으로 모션 A는 270프레임(10.8초), 모션 B는 110프레임(4.4초)의 모션 시퀀스로 구성되었다.

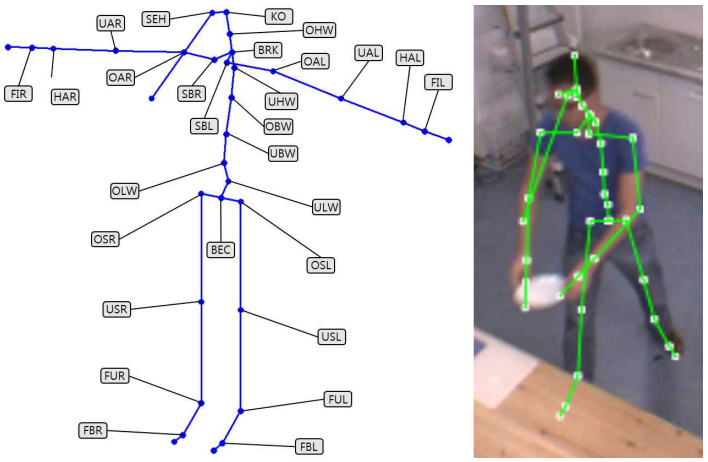

그림 2. 모션 캡쳐에 사용된 skeleton 모델 구조

2.2. 모션학습 및 생성 알고리즘

모션캡쳐 데이터에 기반하여 새로운 모션을 생성해내기 위해 본 논문에서는 cRBM을 이용하여 세 피험자로부터 획득한 각 모션을 학습시켰다. cRBM은 visible data로 이진값만을 사용할 수 있었던 기존의 RBM을 수정하여 실수값을 사용할 수 있기 때문에 각 관절의 회전각이 실수로 이루어진 모션캡쳐 데이터를 학습시키는 것이 가능하다. 또한 cRBM은 temporal information을 처리하기 위해 visible variables의 이전 값들을 조건적으로 사용하여 추가적인 directed connection으로부터 가중치를 갱신한다. 이는 기존의 RBM이 layer내 connection을 허용하지 않는 것에 반해, cRBM이 그림 3과 같이 두 개의 directed connection을 추가로 허용하기 때문에 가능하다. 먼저 visible layer에서 visible layer로 연결되는 것이 있고(그림 3-b), visible layer에서 hidden layer로 연결되는 것이 있다(그림 3-c). 이러한 조건적인 directed connection은 모션 데이터의 학습이 진행되는 동안 현재 프레임에서의 입력값과 함께 이전 프레임의 입력값을 이용하여 latent variable을 갱신함에 따라 dynamic한 bias의 변화를 반영할 수 있기 때문에 temporal information을 처리하는 것이 가능하게 한다.

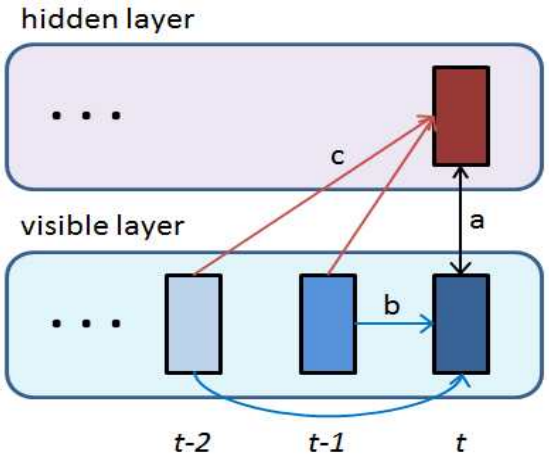

그림 3. cRBM 구조도

3. 실험 결과

최종적으로 학습된 cRBM에 첫 번째 학습 데이터의 초기 6프레임을 seed 모션으로 주고 새로운 모션을 생성하였다. 그림 4의 모션 시퀀스는 서랍을 열고 식기를 꺼내는 동작(모션 A), 그림 6의 모션 시퀀스는 선반의 문을 열고 그 안에 있는 식기를 꺼내는 동작(모션 B)이며, 학습에 사용한 3개 모션(a, b, c)과 이를 이용하여 생성한 모션(d)의 주요 동작을 출력한 결과이다.

모션 A는 그림 4(a,b,c)와 같이 먼저 앞으로 이동하면서 왼발을 이루는 관절들이 다른 관절들에 비해 많이 움직이고, 이후 서랍을 여는 왼팔, 식기를 꺼내는 오른팔, 서랍을 닫는 왼팔, 그리고 테이블로 가기 위해 돌아서는 몸통 순으로 관절들이 주로 움직이는 모션을 학습한 것으로, 새로 생성된 모션도 그림 4(d)와 같이 학습 결과를 반영하여 생성됨을 확인할 수 있다. 모션 B에서도 마찬가지로 그림 6과 같이 선반이 위치한 곳으로 이동한 후, 왼손으로 선반을 여닫고, 오른손으로 컵을 꺼낸 후, 다시 테이블이 위치한 곳으로 이동하여 테이블에 컵을 올려두는 일련의 동작을 학습하였으며, 생성된 모션이 각 동작의 패턴을 반영하여 생성되었다.

그림 5는 모션 A에서 동작 시간에 따른 각 신체부위의 움직임을 정량적으로 확인하기 위해 각 관절을 몸통, 목, 왼팔, 오른팔, 왼발, 오른발을 구성하는 6개의 그룹으로 나눈 후 각 프레임에서의 회전각 정도를 분석한 결과이다. 학습 데이터의 모션을 이루는 각 부위별 관절의 정도가 우세한 프레임 구간에서, 생성된 모션의 변화도 우세함을 확인할 수 있으며, 특히 seed 모션으로 사용한 첫 번째 학습 데이터의 값과 유사한 패턴을 보인다.

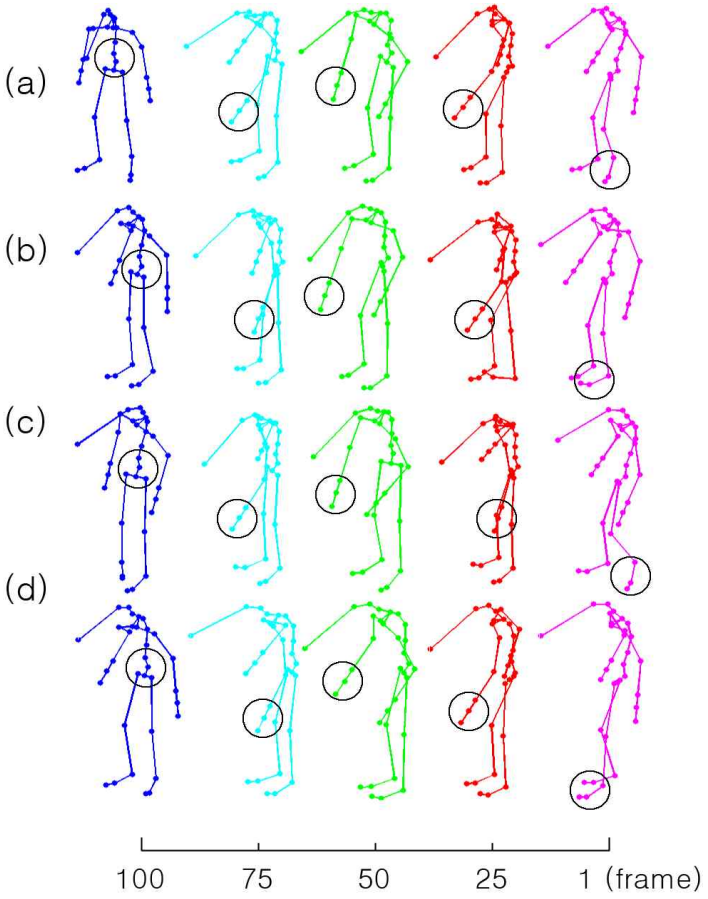

그림 4. 모션 시퀀스 A의 프레임별 주요 관절 움직임 (a-c: 학습 모션, d: 생성 모션)

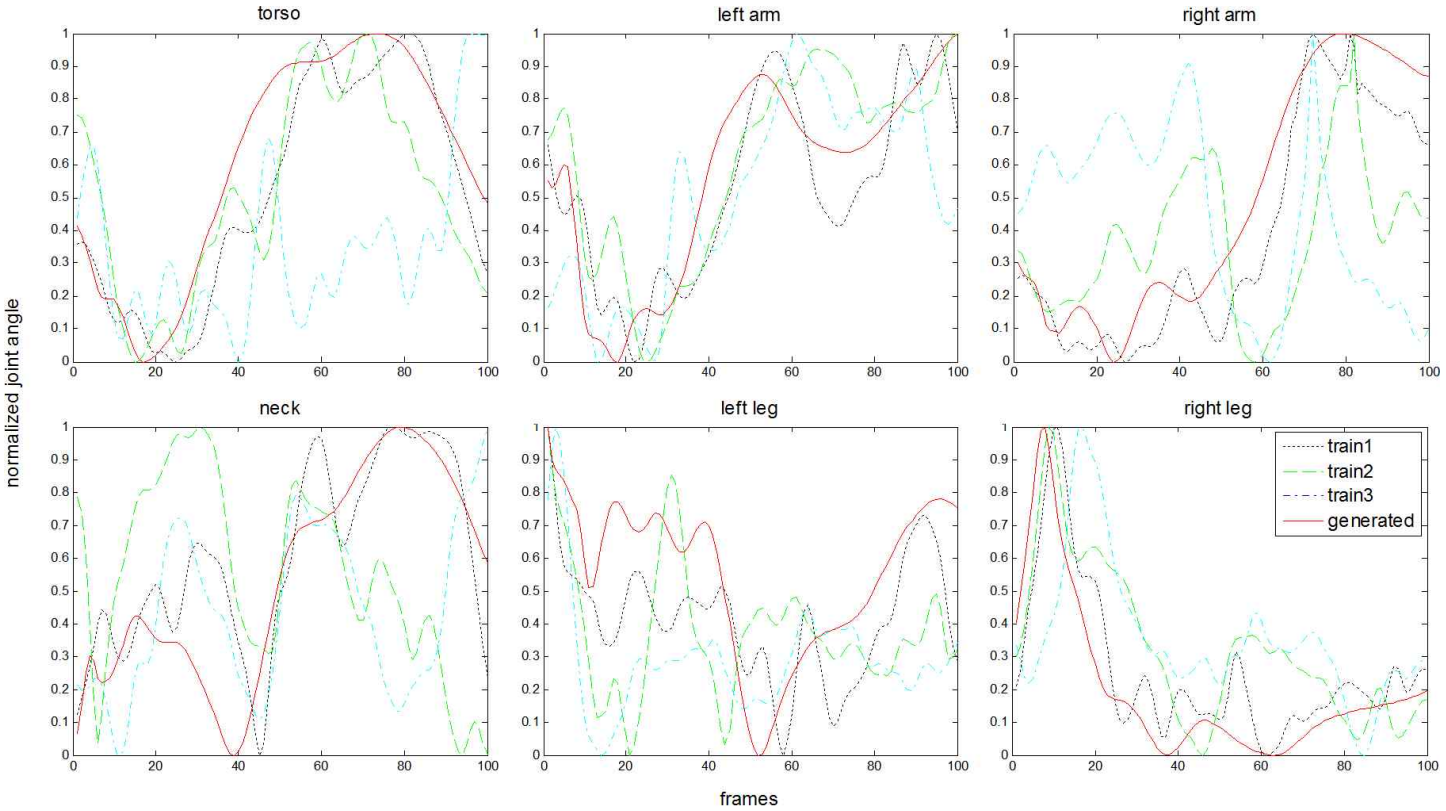

그림 5. 모션 A의 프레임에 따른 신체부위별 관절 회전각 변화 (점선: 학습 모션, 실선: 생성 모션)

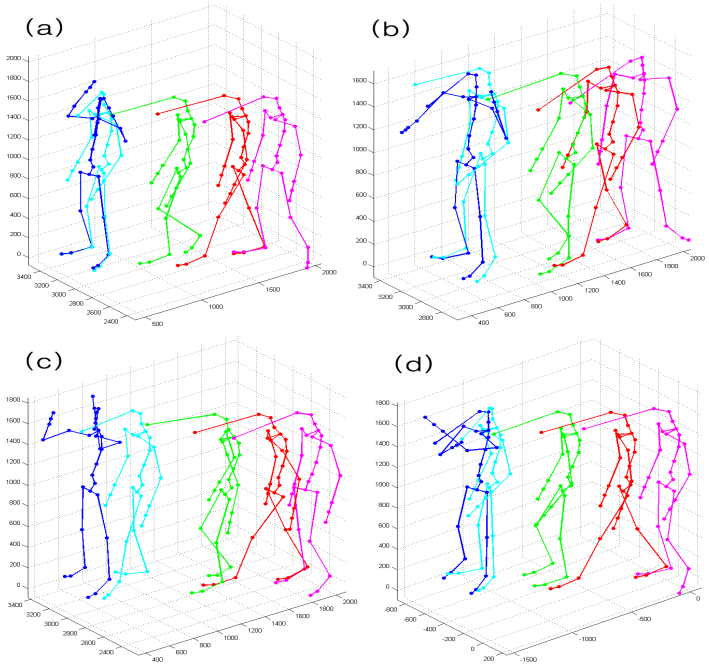

그림 6. 모션 시퀀스 B의 프레임별 주요 모션
(a-c: 학습 모션, d: 생성 모션)

하지만 생성된 모션 데이터가 학습 데이터의 세부적인 동작까지는 표현하지 못하고 모든 관절에서 완만한 동작 변화를 보이고 있기 때문에, 보다 복잡한 동작을 학습하기 위해서는 이 부분에 개선이 필요할 것이다.

4. 결론 및 향후연구

본 논문에서는 카메라 영상 기반의 모션 캡처 시스템을 이용하여 수집한 주방 공간에서의 인간 행동 데이터로부터 선반과 서랍을 열어 식기를 꺼내는 기본 동작들에 대한 모션 데이터를 추출하고, non-linear generative model인 cRBM을 이용하여 각 모션 데이터를 학습하였다. 모션 생성 실험에서는 학습에서 사용된 원본 모션들과 동일한 작업을 수행하는 새로운 모션을 6프레임의 초기 모션 일부를 seed로 사용하여 자동으로 생성하였다.

cRBM을 이용하여 걷기 및 뛰기 모션을 생성하였던 기존 연구[8]의 경우 팔과 다리가 단순한 패턴으로 반복 운동하는 것을 학습한 것이다. 본 연구에서는 여기서 나아가 주방 환경에서와 같이 보다 복잡한 인간 행동을 학습 및 생성함으로써 이러한 연구 내용이 실제 생활에 적용될 수 있음을 제시하였으며, 생성된 모션 데이터를 신체 부위별로 분석하여 학습 데이터와의 유사성을 보였다.

본 실험 결과는 로봇이나 아바타가 인간의 모션 데이터를 활용하여 점차적으로 인간 수준의 유연한 움직임을 학습해나갈 수 있음을 나타낸다. 즉, 인간이 유아기 때부터 주위 사람들의 행동을 따라하며 자연스럽게 자신의 행동 방법을 배우는 방식과 유사한 행동 학습 메커니즘이 로봇에게도 적용될 수 있을 것이다[4].

향후 연구에서는 이러한 비디오 데이터 기반의 모션 생성 기법을 모션 인식에 활용할 계획이다. 인간은 감각기관을 통해 습득한 정보들을 해석할 때 이전에 경험한 데이터를 prior knowledge로 사용하여 그 효율성 및 정확도를 높이는 것으로 알려져 있다[9, 10]. 컴퓨터를 이용한 모션 인식의 경우에도 이미 학습한 데이터를 prior knowledge로 사용한다면, 동일한 효과를 얻을 수 있을 것이다. 즉, 사용자가 특정 동작을 나타낼 때, 초기의 모션 일부를 이용하여 새로운 모션을 생성해내고 이를 다음 부분의 모션과 비교하는 작업을 계속해나가면서 보다 정확하고 효율적인 모션 인식을 수행할 수 있을 것이다. 인식 결과는 다시 학습 데이터로 사용되어 기존의 모션 인식 엔진을 업데이트함으로써 점차적으로 보다 정확한 모션 인식을 수행할 수 있도록 할 것이다.

참고문헌

[1] M. Tenorth, J. Bandouch, M. Beetz, “The TUM Kitchen Data Set of everyday manipulation activities for motion tracking and action recognition,” IEEE 12th International Conference on Computer Vision Workshops, pp. 1089-1096, 2009.

[2] 이제희, “데이터 기반 애니메이션과 기계학습,” *정보과학회지*, 제25권, 제3호, pp. 52-56, 2007.

[3] 장병탁, “차세대 기계학습 기술,” *정보과학회지*, 제25권, 제3호, pp. 3-144, 2007.

[4] 장병탁, “SNU Videome Project: 인간수준의 비디오 학습 기술,” *정보과학회지*, 제29권, 제2호, pp. 3-127, 2011.

[5] K. Amaya, A. Bruderlin, T. Calvert, “Emotion from motion,” In Proceedings of the conference on Graphics interface, pp. 222-229, 1996.

[6] E. Hsu, K. Pulli, J. Popovi, “Style translation for human motion,” *ACM Transactions on Graphics*, Vol. 24. No. 3, pp. 1082-1089, 2005.

[7] M. Brand, A. Hertzmann, “Style machines,” In Proceedings of SIGGRAPH, pp. 183-192, 2000.

[8] G. W. Taylor, G. E. Hinton and S. Roweis, “Modeling human motion using binary latent variables,” In Advances in Neural Information Processing Systems (NIPS), Vol. 19, 2006.

[9] A. Seydell, D. C. Knill, J. Trommershäuser, “Priors and learning in cue integration,” In J. Trommershäuser, M. S. Landy, & K. P. Körding (Eds.), Sensory Cue Integration. New York, NY, Oxford University Press, 2011.

[10] P. Geradin, Z. Kourtzi, P. Mamassian, “Prior knowledge of illumination for 3D perception in the human brain,” *PNAS*, Vol. 107, No. 37, 2010.