
계층적 색인 구조를 갖는 다중 가우시안 기반의 배경 모델을 이용한 실시간 인간 행동 인식 연구

Real-time Human Activity Recognition
Using Multiple Of Gaussian based Background Model with Hierarchical Index Structure

최진, Jin Choi, 한태우, Taewoo Han, 조용일, Yong-il Cho, 양현승, Hyun S. Yang
한국과학기술원 전산학과 인공지능 미디어 연구실

요약 본 논문은 실내의 로비나 복도에 설치된 방법 카메라로부터 얻어진 일련의 영상으로부터 ‘걷기’, ‘뛰기’, ‘앉기’, ‘일어서기’, ‘넘어짐’의 비교적 짧은 시간에 일어나는 인간 행동들을 실시간으로 인식하는 시스템의 구현에 관해 다룬다. 먼저 입력으로 받은 영상을 계층적 색인 구조를 갖는 다중 가우시안 기반의 배경 모델을 이용하여 윤곽을 추출하고 객체를 인식하여 시간차에 의한 가중치로 누적하여 시간 템플릿을 만든다. 만들어진 시간 템플릿으로부터 특징을 추출하여 신경망 모델에 적용하여 5 가지 인간행동을 구분한다. 구현된 시스템으로 인간행동 인식 실험을 수행하였는데, 실험 참가자들의 행동 방식이 약간씩 달랐음에도 불구하고 높은 인식률을 보여주었다.

핵심어: Human Activity Recognition, Visual Surveillance, Adaptive Background Subtraction, Motion History Image

1. 서론

공공 안전에 대한 중요성이 점점 부각되고 있는 상황에서 영상 감시 시스템(Visual Surveillance System)이 널리 사용되고 있다. 하지만 기존의 시스템은 방법 카메라로부터 들어오는 영상을 사람이 직접 모니터링을 해서 비정상 행동을 하는 사람을 발견하는데, 이것은 굉장히 지루한 일이며 실수로 그 사람을 발견하지 못할 수 있다. 또한 기존 시스템은 주로 사건이 발생한 후에 저장된 영상을 검토하는 용도로만 사용되어 왔다. 자동으로 일련의 영상을 분석해서 사전에 비정상 행동을 하는 사람을 구분할 수 있다면 보다 효율적인 영상 감시 시스템을 구축 할 수 있을 것이다. 이러한 이유로 View-based Human Activity Recognition 연구가 활발히 진행 중이다.

View-based Human Activity Recognition 이란 일련의 영상에서 사람을 연속적으로 추적하고 이를 바탕으로 무엇을 하고 있는지를 인식하는 것을 말한다. [1] 이 연구는 영상 감시뿐만 아니라 지능공간(Smart Space) 내에서 사용자의 행동을 인지하고 의도를 파악해서 유용한 서비스를 제공하는데 쓰일 수 있으며, Human Computer Interaction 분야에도 적용할 수 있는 등 응용 범위가 넓어 연구의 활용도가 높다.

본 논문에서는 실내의 로비나 복도에 설치된 방법카메라

로부터 얻어진 일련의 영상으로부터 사람의 간단한 행동을 실시간으로 인식하는 시스템의 구현에 대해서 기술한다.

2. 관련연구

View-based human activity recognition은 전처리, 데이터 모델링, 행동인식의 3단계로 이루어진다. 전처리 과정에서는 배경제거(background subtraction)[2][3], 피부색 감지(skin detection)[4], 그림자 제거(shadow removal)[5] 등의 방법을 이용하여 인간의 움직임(movement), 윤곽(silhouette), 외양(appearance) 등의 정보를 추출한다.

데이터 모델링 과정에서는 인간의 움직임 정보 등을 이용하여 인식에 사용하기 위한 데이터를 모델링 한다. 이러한 방법으로는 크게 structure information -based approach와 global feature-based approach가 있다[6]. Structure information-based approach는 3D 인간 몸체 모델(human body model)을 이용하여 2D 영상으로부터 인간 몸체 모델의 각 부위를 감지하여 추적하고 관점으로 연결된 구조적인 형상으로 표현한다. 이런 접근방법에는 각 몸체 부위를 먼저 찾고 이들을 조합하는 Bottom-Up 방식[7][8][9]과 3D 관점 모델을 영상에 접합(fitting) 시키는 방식[10][11]이 있다.

Global feature-based approach는 각각의 2D 영상으로부터 추출한 형태, 윤곽, 텍스처, 모션 등의 정보를 이용하여 몸체 영역을 하나의 덩어리로 간주하여 전체적인 행동 특징을 모델링하고 기술한다. 이러한 방법들로는 MHI(Motion History Image)[12][13][14], Spatio-Temporal Volume Descriptor[15][16], Actions Sketch[17] 등이 있다. 관절 구조로 연결된 몸체 부위들을 2D 영상에서 감지하고 추적하는 것은 매우 어려운 문제이며, 전체적인 시각적 특징들을 사용하여 인간의 움직임을 추적하고 인식하는 것이 계산비용적인 측면에서 훨씬 더 효율적이다.

인간 행동인식 과정에서는 순차적인 영상들에 나타난 인간의 자세를 추적해 나가면서 쌓인 모델링 데이터를 이용하여 인간의 행동을 분류해 나간다. FSM (Finite State Machine)[18][19], Bayesian Network[20], HMM (Hidden Markov Model)[21][22] 등과 같이 인간 행동들을 각각의 state로 표현하고 이들간의 transition에 따라 인간의 행동을 구분하는 방법들과 미리 저장된 패턴들과의 차이에 주어진 영상들의 인간행동을 구분하는 template matching 방법들이 있다. Template matching 방법은 계산비용이 적게 들지만 움직임의 기간(duration)에 민감하다. State space 기반 방법은 각각의 정적인 자세(posture)를 하나의 state로 표현하고 이 state들은 특정한 확률로 연결되어 있다. 각 움직임의 기간은 크게 결과에 영향을 미치지 않는다. 그러나, State space 기반 방법은 적절한 state의 수와 특징 벡터의 차원을 고르는 것이 결과에 많은 영향을 미친다.

본 논문에서는 학습된 데이터와 차이가 있는 데이터가 들어오더라도 인간행동을 인식할 수 있도록 신경망 모델을 이용하여 인식기를 구현하였다.

3. 인간 행동 인식 시스템

3.1 인간 행동 인식 시스템 개요

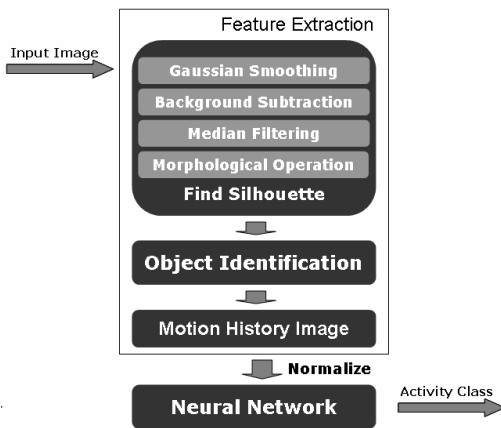


그림 1 인간 행동 인식 시스템의 구조

본 연구의 목적은 ‘걷기’, ‘뛰기’, ‘앉기’, ‘일어서기’, ‘넘어짐’ 등의 비교적 짧은 시간 (2~3초) 안에 일어나는 인간 행동들을 실시간으로 효율적으로 인식하는 시스템을 만드는 것이다. 그림 1은 이것을 토대로 제안된 인간 행동 인식 시스템의 전체 구조를 보여준다.

3.2 인간 형태 검출

카메라로부터 들어온 일련의 영상이 주어지면 먼저 노이즈를 제거하기 위해 Gaussian Smoothing을 적용한다. 우리가 제안하는 방법은 정확한 픽셀 위치 보다는 전체적인 사람의 윤곽을 찾는 것이 중요하기 때문에 Gaussian Smoothing을 적용하는 것이 타당하다. 그리고 영상 내에서 움직이는 물체의 윤곽 (전경 영상)을 배경과 분리해 낸다. 이때 계층적 색인 구조를 갖는 MOG (Multiple Of Gaussians)기반의 배경 모델을 적용하여 실행 속도가 빠르고 서서히 변하는 빛의 변화에는 적응하도록 구현하였다. 이어서 Median Filtering 과 Morphological Operation을 얻어진 전경 영상에 차례로 적용하여 노이즈로 유추되는 작은 영역들은 제거하고 사람이라고 판단되는 크기가 비교적 큰 영역들의 묶음인 Connected Component를 만든다. 우리는 문제를 간단히 하기 위해 위의 방법으로 추출된 Connected Component는 사람의 윤곽이라고 가정하였다. 아래의 그림 2는 위의 방법을 통해 얻어진 예를 보여 준다.



그림 2 입력영상(좌측)과 윤곽을 추출한 영상(우측)

이 단계에서 우리가 얻고자 하는 것은 움직이는 물체의 윤곽이다. 물론 픽셀 별로 정확하게 배경을 모델링 하는 것은, Background Subtraction시에 정확한 전경영역을 얻는데 도움이 된다. 하지만 이것은 이미지 사이즈가 커질수록, 전처리 단계로서 부적합한, 많은 오버헤드를 발생한다. 이런 오버헤드를 줄이기 위한 연구들에 의하면 클러스터링을 통해 뽑은 표본 값을 이용하거나, 랜덤 인덱스를 통해 픽셀 별로 갱신 주기를 조정하여도 Subtraction된 마스크 영상이 큰 차이가 없음을 보여준다.

이미지상에서 픽셀들의 시간-공간 응집성을 이용하면 인접한 픽셀들은 유사한 정보를 포함하고 있을 확률이 높다. 따라서 결과의 차이가 크지 않다면, 이 특징을 전처리 과정의 시간을 줄이는 것이 타당하다. 본 논문에서는 이 특징을 이용해 쿼드트리와 유사한 계층적 색인 구조를 사용하였다.

계층적 색인 구조는 Bottom-Up방식으로 만든 쿼드트리 비슷한 구조이다.. 우선 픽셀들을 패치라 불리는, 사용자가 감지하고자 하는 최소 사이즈의 영역으로 묶는다. 그리고 NW, NE, SW, SE 네 방향의 인접한 영역들을 묶어 새 레이어를 쌓아나간다. 최종적으로 패치들로 이뤄진 레벨 $L = n$ (L_{max}) 부터, 사용자가 지정한 검색 시작 레벨인 $L = 0$ (L_{init})까지 레이어가 쌓이게 된다.

초기 배경 모델이 생성된 후, 연속된 배경 이미지 시퀀스에 대해 쿼드트리 방식의 검색이 시작된다. L_{init} 의 각 노드들에 대해, 노드에 속한 픽셀들 중 랜덤 픽셀 하나를 선택하고, 전경인지 배경인지를 판단한다. 만약 노드가 전경이라 판단되면, L_{max} 레이어에 도달할 때까지 다음 레벨의 자식 노드들에 대해 같은 과정을 반복한다. 만약 L_{init} 상의 노드가

배경이라 판단된다면 다음 노드에 대해 위의 과정을 반복한다. 본 실험에서는 320×240 크기의 영상에 대해서는 2×2 크기의 패치로 2레벨로 레이어를 쌓아 사용하였다.

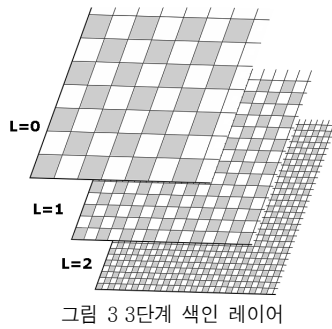


그림 3은 사이즈의 패치에 대해 3단계의 색인 레이어를 보여준다. 레벨 0의 레이어의 각 노드에 대해 랜덤 샘플을 이용해 위의 알고리즘을 적용해 레벨 2의 노드까지 분할해 나가고, 레벨 2의 노드가 전경이라 판단되면 최종적으로 전경이라 판단한다.

3.3 객체 인식

인간 윤곽이 추출되면 이를 하나의 객체로 가정 할 수 있다. 영상 내에서 얻어지는 객체가 지속적으로 동일한 객체로 인식되기 위해서는 객체마다 고유한 ID가 부여되어야 하는데, 추가적인 Tag의 도움 없이 주어진 영상만으로 구분 가능하도록 색상 정보를 주요 특징으로 사용한다. 구체적으로, 객체를 구성하는 색상 정보를 추출하기 위해 RGB 색상 공간에 있는 입력 영상을 HSV 색상 공간으로 변환 후, Hue Histogram을 구한다. 이렇게 얻어진 Hue Histogram은 객체 색상에 따른 분포를 갖기 때문에 객체 구분에 사용 가능하다. 하지만 객체의 색상이 유사하여 구분이 힘든 경우가 있기 때문에 추가적으로 객체의 크기와 화면 상의 위치를 추가적으로 고려한다.

하나의 객체는 영상 내에서 특정 출입 영역에서 나타나거나 사라진다고 가정하였다. 또한 그룹으로 나타나거나 사라지는 것은 고려하지 않았다. 이러면 영상 내에 존재하는 객체의 수를 파악하기 용이해진다. 영상 내의 현재 객체 수는 가림(Occlusion) 현상이나 부정확한 인간 윤곽 추출이 생겼을 때 추적에 사용되는 유용한 정보이다.

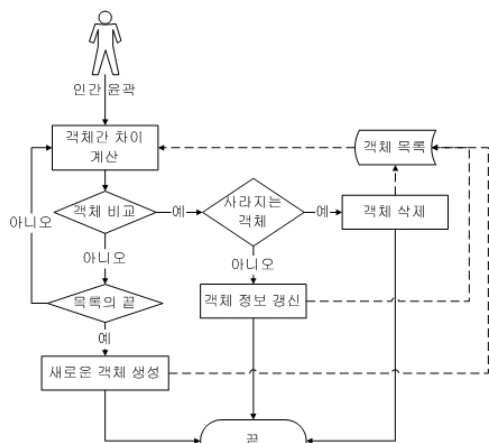


그림 4 객체 인식 순서도

그림 4는 객체 인식이 일어나는 과정을 보여주는 순서도이다. 먼저 인간 윤곽이 검출되면 기존의 객체 목록을 탐색하여 현재 정보와의 차이를 각각 계산한다. 객체 간의 차이는 세가지 정보, 즉 색상, 위치, 크기를 각각 비교하여 얻을 수 있다. 만약 차이가 특정 값 이하이면 객체가 인식된 것으로 볼 수 있다. 이때 이 객체가 영상 내에서 사라지고 있으면 객체 목록에서 삭제하여 주고 그렇지 않은 경우에는 이전 객체 정보를 현재 값으로 갱신하여 준다. 만약 목록을 모두 탐색하였는데도, 차이가 작은 객체가 나타나지 않은 경우는 새로운 객체가 나타난 것을 의미한다. 따라서 새로운 객체를 생성하고 객체 목록에 추가한다.

3.3 행동 모델링

인간 행동 인식을 하기 위해서는, 포즈 인식처럼 한 영상만을 고려하기 보다는 일련의 영상들을 고려해주어야 한다. 이를 위해 우리는 일련의 영상들이 주어졌을 때 이들의 시간에 따른 움직임 정보를 하나의 2D 템플릿으로 표현해주는 방법인 Motion History Image (MHI)를 이용하였다.[23] 일반적으로 MHI는 계산량이 적고 짧은 시간에 일어나는 동작을 표현하는데 적합하다고 알려져 있다.

t 프레임일 때 MHI는 아래 수식 1로 구할 수 있다. 여기서 δ 는 미리 정의된 누적에 사용될 이전 프레임 수를 말하고, $I'(x, y)$ 는 현재 영상을 나타낸다. 그리고 Ψ 는 (x, y) 좌표 상에 있는 윤곽의 유무를 나타낸다.

$$MHI_{\delta}^t(x, y) = \begin{cases} t/\delta & \text{if } \Psi(I'(x, y)) \neq 0 \\ MHI_{\delta}^{t-1}(x, y) & \text{otherwise} \end{cases} \quad (1)$$

실제 예를 들어 설명하면, 넘어지는 행동에 대해서 위의 수식 1을 적용하여 MHI를 구하면 아래 그림 5와 같다. 여기서 앞의 네 영상은 넘어지는 행동 중에 추출된 윤곽영상이고 마지막 영상이 이때의 MHI를 나타낸다.

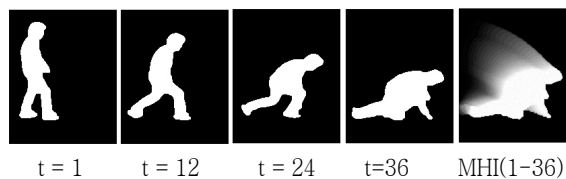


그림 5 넘어지는 행동에서 36 frames을 누적한 MHI

3.4 특징 추출

이렇게 얻어지는 MHI에서 다시 Connected Component를 찾고 Tracking Blob으로 정한다. 찾아진 Tracking Blob을 16 X 16 2D 영상으로 Normalize 해서 이를 Feature로 사용한다. 추가적으로 입력 영상 내에서 Tracking Blob의 상대적 크기 (상대적 높이와 폭의 두 가지 정보)를 feature에 추가하는데, 이 정보는 걷기와 뛰기 같이 MHI가 전반적으로 유사한 형태를 갖지만 Tracking Blob의 크기가 차이가 나는

행동들을 구분하는데 중요한 역할을 한다. 결과적으로 총 258개의 원소를 갖는 feature vector를 생성한다.

3.5 인간 행동 인식

한 사람이 같은 행동을 반복적으로 수행하더라도 똑 같은 방식으로 행동하지 않는다. 따라서 약간의 차이가 나는 행동들도 같은 클래스로 인식하기 위해서 신경망 모델을 적용하였다. 앞의 과정을 통해 얻어진 feature vector는 신경망의 입력으로 사용된다. 걷기, 뛰기, 앉기, 일어서기, 넘어짐의 총 5개의 클래스를 정의하고 미리 신경망을 학습시켜 인간 행동을 인식한다. 아래 그림 6은 신경망 학습에 적용된 클래스들의 대표적인 MHI의 예를 보여준다. 왼쪽부터 차례대로 걷기, 뛰기, 앉기, 서기, 넘어짐을 나타내는 MHI이다.

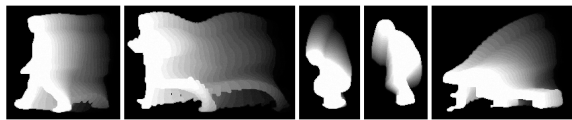


그림 6 학습에 쓰인 각 클래스들의 대표적인 MHI

추가적으로 인식의 신뢰도를 높이기 위해 신경망을 통해 얻어진 인간행동 클래스들의 시간대별 분포를 통계적으로 분석하여 최종 결과를 얻는다.

4. 실험 및 결과

본 논문에서는 3명의 피실험자를 대상으로 ‘걷기’, ‘뛰기’, ‘앉기’, ‘일어서기’, ‘넘어짐’의 행동 데이터를 각각 12개씩 얻어서 신경망 학습에 사용하였으며, 학습에 참여한 3명의 피실험자와 학습에 참여하지 않았던 피실험자 4명을 대상으로 총 350개의 행동 샘플을 자동으로 추출하여 인식 실험을 수행하였다. 피실험자 별로 각 행동들을 10개씩 반복하게 하였으며, 정확하게 인식한 행동의 개수를 표 1에 정리하였다.

표 1 인간행동 인식 실험 결과

	걷기	뛰기	앉기	일어서기	넘어짐	인식률
T1	10	10	10	10	10	100.0%
T2	10	10	10	7	10	94.0%
T3	8	10	8	5	10	82.0%
P1	10	9	10	10	10	98.0%
P2	7	10	10	8	10	90.0%
P3	4	10	10	10	10	88.0%
P4	9	10	10	3	10	84.0%
인식률	82.9%	98.6%	97.1%	75.7%	100.0%	90.9%

표 1에서 T1, T2, T3는 학습에 참여했던 피실험자를 나타내며, P1, P2, P3, P4는 학습에 참여하지 않았던 피실험자를 나타낸다. 학습에 참여했던 피실험자들을 대상으로 한 실험에서는 평균 92.0%의 인식 성능을 나타냈으며, 학습에 참여하지 않았던 피실험자들을 대상으로 한 실험에서는 평균 90%의 인식 성능을 나타내었다. 행동 별로는 비교적 특징이 큰 ‘넘어짐’, ‘뛰기’, ‘앉기’가 높은 인식률을 나타냈으며, ‘일어서기’ 행동이 가장 낮은 인식 성능을 보였다. 이는 피실험자의 상의 색과 배경색이 비슷하여 윤곽 추출이 잘못된 것이 주된 원인이었다. 피실험자 별로 행동의 패턴이 달라 정형화된 행동 모델을 구하기가 쉽지 않은 실험이었으나 현재 7명을 대상으로 한 실험에서 전체적으로 90.9%의 높은 인식률을 나타냈으며 윤곽 추출이 정확하게 이루어진다면 인식률은 더욱 향상될 것으로 생각된다.

5. 결론 및 향후 연구

본 논문에서는 지능화된 감시시스템에 사용할 수 있는 실시간 인간 행동 인식 시스템을 설계하고 구현하였다. 카메라를 통하여 들어온 일련의 영상을 입력으로 받아 계층적 색인 구조를 갖는 다중 가우시안 기반의 배경 모델을 적용하여 사람의 윤곽을 추출하고 객체를 인식하여 시간차에 의한 가중치로 누적하여 시간 템플릿을 만들었다. 이 시간 템플릿으로부터 중요한 특징들을 추출하여 신경망 모델의 입력으로 사용하여 ‘걷기’, ‘뛰기’, ‘앉기’, ‘일어서기’, ‘넘어짐’의 인간행동을 구분할 수 있는 인식 실험을 수행하였다. 사람의 윤곽 추출이 정확히 이루어지지 않는 상황에서 실험 참가자들의 행동 패턴이 조금씩 달랐음에도 불구하고 높은 인식 성능을 나타내었다.

향후 연구로는 인간 행동 인식 시스템의 성능향상을 위하여 전처리 과정에서 사람의 윤곽을 정확히 추출할 수 있는 연구가 필요하다. 배경색과 입고 있는 의상 색이 비슷하거나 그림자가 심하게 있는 경우에는 윤곽 추출이 정확히 이루어지지 않았다. 또한 현재는 정해진 한 방향에서 인간의 행동을 인식하고 있지만, 방향이 바뀌는 경우 인식이 잘 안 되기 때문에 여러 방향에서의 영상을 학습시켜 방향에 상관없이 인간 행동을 인식할 수 있는 시스템 개발이 필요하다.

Acknowledgement

본 연구는 21세기 프론티어 연구개발사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스컴퓨팅및네트워크원천기반기술개발사업과 첨단정보기술 연구센터를 통하여 과학재단의 지원에 의한 것임

참고문헌

[1] J. K. Aggarwal and Q. Cai, "Human Motion Analysis: A Review", Computer Vision and Image Understanding: CVIU, vol. 73, no. 3, pp. 428-440, 1999.

[2] KaewTraKulPong P, Bowden R, "An Improved Adaptive Background Mixture Model for Real-

- time Tracking with Shadow Detection", 2nd European Workshop on Advanced Video Based Surveillance Systems, Sep. 2001.
- [3] Ismail Haritaoglu, David Harwood, Larry S. Davis, "W4: Real-Time Surveillance of People and Their Activities", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, Issue 8, pp. 809-830, Aug. 2000.
- [4] M. Jones and J. Rehg., "Statistical color models with application to skin detection" International Journal of Computer Vision, Volume 46, No. 1, pp. 81-96, Jan. 2002.
- [5] A. Prati, I. Mikic, R. Cucchiara, and M. M. Trivedi, "Analysis and Detection of Shadows in Video Streams: A Comparative Evaluation", IEEE Computer Vision and Pattern Recognition Conference, Hawaii, Dec. 2001.
- [6] K. Aggarwal and Q. Cai, "Human motion analysis: a review", Computer Vision and Image Understanding, Volume 73, No. 3, pp. 428-440, 1999.
- [7] S. Ioffe and D.A. Forsyth, "Human tracking with mixtures of trees", IEEE Int'l Conf. on Computer Vision, pp. 690-695, July, 2001.
- [8] D. Ramanan and D.A. Forsyth, "Finding and tracking people from the bottom up", Computer Vision and Pattern Recognition (CVPR), Madison, Wisconsin, June 2003.
- [9] Ramanan, D., Forsyth, DA, Zisserman, A. "Strike a Pose: Tracking People by Finding Stylized Poses", Computer Vision and Pattern Recognition (CVPR), San Diego, CA, June 2005.
- [10] R. Urtasun, D. Fleet and P. Fua, "Monocular 3-D Tracking of the Golf Swing", Int'l Conf. on Computer Vision and Pattern Recognition, San Diego, CA, Vol. 1, pp. 932-939, June 2005.
- [11] Qi Zhao, Jinman Kang, Hai Tao, Wei Hua, "Part Based Human Tracking In A Multiple Cues Fusion Framework", 18th International Conference on Pattern Recognition, pp. 450-455, Aug. 2006.
- [12] Bobick, A. and J. Davis, "Action Recognition Using Temporal Templates", Motion-Based Recognition, pp. 125-146, 1997.
- [13] J. Davis, "Sequential Reliable-Inference for Rapid Detection of Human Actions", IEEE Workshop on Detection and Recognition of Events in Video, Washington DC, July 2, 2004.
- [14] J. Davis and A. Tyagi, "Minimal-Latency Human Action Recognition using Reliable-Inference", Image and Vision Computing, Vol. 24, pp. 455-472, 2006.
- [15] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance", the Ninth IEEE International Conference on Computer Vision, pp. 726-733, Washington, DC, USA, 2003
- [16] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes", IEEE International Conference on Computer Vision (ICCV), 2005.
- [17] A. Yilmaz and M. Shah "Actions Sketch: A Novel Action Representation", IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2005.
- [18] Aaron F. Bobick and Andrew D. Wilson, "A state-based approach to the representation and recognition of gesture", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 12, Dec. 1997.
- [19] P. Hong, M. Turk, and T. S. Huang, "Gesture modeling and recognition using finite state machines", Proc. the Fourth IEEE International Conference and Gesture Recognition, Grenoble, France, 2000.
- [20] A. Madabhushi and J.K. Aggarwal, "A Bayesian approach to human activity recognition", Proc. Second IEEE Workshop on Visual Surveillance, pp. 25-32, June 1999.
- [21] S. Fine, Y. Singer, and N. Tishby, "The hierarchical hidden Markov model: Analysis and applications", Machine Learning, vol. 32, pp.41-62, 1998.
- [22] N. Oliver, E. Horvitz, and A. Garg, "Layered representations for human activity recognition," Proc. IEEE International Conference on Multimodal Interfaces, pp. 3-8, Oct. 2002.
- [23] James W. Davis, "Sequential Reliable-Inference for Rapid Detection of Human Actions", IEEE workshop on Detection and Recognition of Events in Video, Washington DC, July 2, 2004.