비선형적 매니폴드를 이용한 임의 얼굴에 대한 얼굴 추적 및 인식

Face Tracking and Recognition on the arbitrary person using Nonliner Manifolds

주명호, Myung-Ho Ju*, 강행봉, Hang-Bong Kang**

요약 사람의 얼굴은 강체(rigid object)가 아니기 때문에 얼굴을 추적하거나 인식하기는 쉽지 않다. 또한 시스템에 미리 학습되어 있지 않은 임의의 얼굴의 경우 지속적으로 얼굴의 변화를 추적하고 인식하기는 어렵다. 본 논문에서는 시스템에 저장되어 있는 얼굴들에 대해 비선형적 매니폴드 모델을 구축하고 각 모델을 선형적으로 결합함으로써 비디오 기반의 영상으로부터 시스템이 알지 못하는 임의의 얼굴에 대해 추적하고 인식하는 방법을 제안한다. 입력된 임의의 얼굴은 얼굴 포즈나 표정 혹은 주위 환경 등에 따라 시스템에 저장되어 있는 서로 다른 얼굴들과 서로 다른 유사성을 갖는다. 따라서 입력 얼굴과 시스템에 저장되어 있는 얼굴들과의 확률적인 접근을 통해 유사성을 추정할 수 있고 추정된 유사성을 이용하여 입력 얼굴에 대한 새로운 비선형적 매니폴드 모델을 구축한다. 또한 추정된 모델은 매 프레임마다 입력 얼굴에 따라 실시간으로 갱신된다. 본 논문에서 제안하는 방법은 실험 결과를 통하여 효율적으로 임의의 얼굴에 대해 추적하고 인식할 수 있음을 보인다.

Abstract Face tracking and recognition are difficult problems because the face is a non-rigid object. If the system tries to track or recognize the unknown face continuously, it can be more hard problems. In this paper, we propose the method to track and to recognize the face of the unknown person on video sequences using linear combination of nonlinear manifold models that is constructed in the system. The arbitrary input face has different similarities with different persons in system according to its shape or pose. So we can approximate the new nonlinear manifold model for the input face by estimating the similarities with other faces statistically. The approximated model is updated at each frame for the input face. Our experimental results show that the proposed method is efficient to track and recognize for the arbitrary person.

핵심어: Nonlinear Manifolds, Face Recognition, Face Tracking, Arbitrary face

본 연구는 문화관광부 및 한국문화콘텐츠진흥원의 지역문화산업연구센터(CRC)지원사업의 연구결과로 수행되었음

*주저자 : 가톨릭대학교 컴퓨터공학과 대학원생 e-mail: hangel5@catholic.ac.kr
**공동저자 : 가톨릭대학교 컴퓨터공학과 교수 e-mail: hbkang@catholic.ac.kr

1. 서론

인증 시스템이나 비디오 감시 시스템, 로봇 컨트롤 등의 많은 분야에서 정지영상보다 동영상에서의 인식을 필요로 한다. 이에 최근 동영상에서 얼굴을 검출하고 추적, 인식하는 기술이 보다 많이 연구되고 있다. 이 중 얼굴에 대한 효율적인 추적 및 인식 방법은 얼굴 모델을 사전에 학습하고 이를 이용하여 입력된 얼굴을 추적 및 인식하는 방법이다. 하지만 이러한 방법은 추적하고 인식하려는 얼굴을 사전에 학습해야만 한다는 문제점이 발생한다. 즉, 시스템을 사용하기 위해서는 먼저 학습을 해야만 한다는 제약이 따른다.

본 논문에서는 시스템에 여러 사람의 얼굴 모델을 미리 학습하고 이 모델을 선형적으로 결합함으로써 임의의 얼굴에 대한 모델을 추정한다. 그리고 추정된 모델을 통해 임의의 얼굴에 대해 추적 및 인식하는 방법을 제안한다. 얼굴은 강체(rigid object)가 아니기 때문에 얼굴 이미지들로 생성되는 이미지 공간은 선형적으로 표현하기 어렵다. 따라서 얼굴 인식에 사용되는 얼굴 모델 역시 비선형적으로 표현되어야 한다. 본 논문에서는 여러 개의 선형적 모델들을 결합함으로써 각 사람의 얼굴 모델을 비선형적인 얼굴 모델로 근사화하고 각 모델들과 입력된 임의의 얼굴과의 확률적 유사도를 추정하여 입력된 얼굴에 대한 비선형적 모델을 추정한다. 각 사람에 대해서는 몇 개의 얼굴 포즈를 구분하여 각각의 선형적인 모델을 형성한다.

임의의 얼굴에 대한 추적 및 인식에 대한 문제는 주로 온라인 학습(Online Learning) 문제로 최근 들어서 연구되고 있다. 특히 Lee는 [6]에서 포즈마다 독립적인 가우시안 분포를 추정하고 각 분포마다 학습 데이터에서의 이동확률을 계산하여 베이지안 정리를 통해 얼굴을 인식하다. 하지만 포즈에 대한 분포간의 이동확률을 학습 데이터에 의존함으로써 하나의 포즈에 속한 모든 얼굴이 다른 포즈들에 대해 동일한 이동확률을 갖는 문제점이 있다. 또한 [7]에서 온라인 학습을 위해 입력 얼굴과 다른 포즈간의 평균 얼굴과의 차이를 최소 제곱근 문제로써 근사화 하여 계산함으로써 각 포즈 모델을 갱신하였지만 이는 입력 얼굴과 각 포즈의 평균 얼굴과의 차이를 계산하게 되어 올바르게 현재 얼굴의 포즈를 근사화하기 어렵다. 따라서 본 논문에서는 이러한 Lee의 문제에 대해 새로운 비선형적 모델을 제시하고 이를 확률적으로 갱신하여 임의의 얼굴에 대해 올바르게 추적하고 인식하는 방법을 제안한다.

입력 이미지에서 임의의 얼굴을 추적하거나 인식하기 위해서 먼저 학습되는 모든 사람에 대한 비선형적인 매니폴드 모델을 생성한다. 이를 위해 먼저 학습되는 각 사람의 얼굴을 몇 개의 정해진 포즈로 분류한다. 그 후 학습영상으로부터 추출한 각 포즈에 해당되는 얼굴 이미지들을 이용하여 주성분 분석법(Principal Component Analysis, PCA)을 통해 각 얼굴 포즈를 독립적인 가우시안 분포로 근사화한다. 각 얼굴 포즈의 독립된 가우시안분포를 가우시안 혼합 모델(Gaussian Mixture Model)로 재구성하고 입력 얼굴이미지로부터 각 사람마다의 확률분포함수(pdf)를 추정함으로써 입력된 동영상으로부터 얼굴 모델을 재구성하여 얼굴을 추적하거나 인식할 수 있다. 그림 1은 한 사람으로부터 구성된 비선형적 매니폴드의 구조를 보여준다.

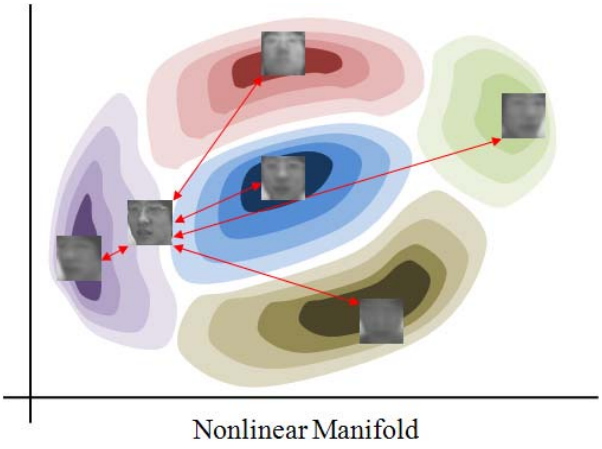

그림 1 한 사람으로부터 생성된 비선형적 매니폴드의 구조

본 논문의 구성은 다음과 같다. 2절에서는 한 사람으로부터 생성되는 비선형적 매니폴드에 대해 가우시안 혼합 모델과 모델의 학습 방법, 그리고 모델로부터 추정할 수 있는 확률분포함수에 대해 설명하고 3절에서 본 논문에서 제안하는 임의의 입력 얼굴에 대한 비선형적 매니폴드의 추정 방법을 소개한다. 4절에서 실험결과를 통해 본 논문에서 제시하는 방법이 임의의 얼굴에 대해 효율적으로 추적하고 인식할 수 있음을 보인다. 마지막으로 5절에서 추후 연구 사항을 논의하고 결론을 맺는다.

2. Nonlinear Manifold Model

얼굴이미지는 얼굴의 포즈에 따라서 매우 다른 이미지로 표현될 수 있다. 그림 2는 학습영상으로부터 추출된 얼굴이미지들로 이러한 포즈에 따른 얼굴 이미지의 차이를 보여준다. 얼굴 포즈에 따른 이미지의 변화를 처리하기 위해 각 포즈에 대해 선형적인 모델을 생성하고 이러한 선형 모델을 가우시안 혼합 모델을 이용하여 표현함으로써 비선형적인 얼굴 모델을 추정할 수 있다.

K개의 포즈로부터 근사화된 가우시안 혼합 모델의 확률분포 함수는 식(1)과 같이 정의할 수 있다.

$$P_G(x) = \sum_{i=1}^{K} \alpha_i Gauss(x|\mu_i, \Sigma_i) \qquad (1)$$

여기서 μ_i와 Σ_i는 각각 i번째 가우시안 성분의 평균벡터

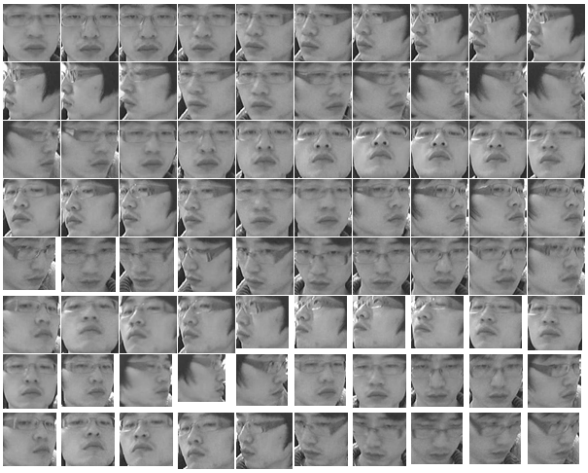

그림 2. 학습을 위한 서로 다른 포즈의 이미지 영상

와 공분산 행렬이고 α_i는 i번째 가우시안 성분의 가중치이다. 본 논문은 각 포즈의 가우시안 성분의 평균 벡터 μ_i와 공분산 행렬 Σ_i를 학습영상에서 i번째 얼굴 포즈에 해당되는 학습영상으로부터 근사화하고 각 가우시안 성분의 $Gauss(x|\mu_i, \Sigma_i)$를 가우시안 우도로써 추정한다. N차원의 입력 이미지에 대해 가우시안 우도(likelihood)는 식(2)와 같다.

$$P(x|\Omega_i) = \frac{\exp\left[-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\right]}{(2\pi)^{\frac{N}{2}}|\Sigma_i|^{\frac{1}{2}}} \quad (2)$$

[7]과 [9]에 따르면 i번째 가우시안 성분 Ω_i는 PCA를 이용하여 구성된 아핀부분공간(Affine subspace) $\widehat{\Omega}_i$로써 근사화할 수 있다. i번째 얼굴 포즈에 속하는 학습영상이 주어졌을 때, 집합 $\{\mu_i, \Sigma_i, \Phi, \Lambda\}$을 구할 수 있다. μ_i는 평균 벡터이고 Σ_i는 공분산 행렬, Φ는 Σ_i의 가장 큰 고유값의 순서대로 M개의 고유벡터를 포함하는 행렬이며 Λ는 $\Lambda_{jj} = \lambda_j$로 대각 성분을 고유값으로 가지는 대각행렬이다. 이미지 I는 $\widehat{\Omega}_i$에 선형적으로 투영되어 $y = [y_1, y_2, \ldots, y_M]^T = \Phi^T(I-\mu_i)$을 얻을 수 있다. 식(2)의 우도는 두 개의 가우시안 분포의 곱으로 표현할 수 있다.

$$P(x|\widehat{\Omega}_i) = \left[\frac{\exp\left(-\frac{1}{2}\sum_{j=1}^{M}\frac{y_j^2}{\lambda_j}\right)}{(2\pi)^{\frac{M}{2}}\prod_{j=1}^{M}\lambda_j^{\frac{1}{2}}}\right]\left[\frac{\exp\left(-\frac{\epsilon^2(x)}{2\rho}\right)}{(2\pi\rho)^{\frac{N-M}{2}}}\right] \quad (3)$$

여기서 M은 부분공간 $\widehat{\Omega}_i$의 차원 수이고 $\epsilon^2(x)$는 재구성 에러(residual reconstruction error)로 식(4)와 같이 정의된다.

$$\epsilon^2(x) = \sum_{j=M+1}^{N} y_j^2 = \| x-\mu_i \|^2 - \sum_{j=1}^{M} y_j^2 \quad (4)$$

식(3)에서 파라미터 ρ는 $\frac{1}{N-M}\sum_{j=M+1}^{N}\lambda_j$을 사용하거나 혹은 간단하게 $\frac{1}{2}\lambda_{M+1}$을 사용한다[7]. 본 논문에서는 후자를 선택하여 실험하였다.

식(3)을 사용하여 식(1)을 식(5)와 같이 다시 쓸 수 있다.

$$P_G(x) = \sum_{i=1}^{K}\alpha_i P(x|\widehat{\Omega}_i) \quad (5)$$

하지만 아직 가중치 파라미터 α_i를 구해야 하는 문제가 남아 있다. α_i는 각 포즈마다의 확률$P(x|\widehat{\Omega}_i)$에 대한 가중치 파라미터로 현재 입력되는 얼굴이미지의 포즈와 가장 유사한 포즈에서 가중치가 가장 높고 가장 다른 포즈에서 낮은 가중치를 갖는다. 본 논문에서는 입력되는 이미지에 따라 EM알고리즘[9]을 이용하여 매 프레임마다 가중치를 추정하여 갱신한다. EM알고리즘은 다음과 같은 두 가지 단계를 반복적으로 수행한다.

▶ E-step:

$$h_i^t(x) = \frac{\alpha_i^t P(x|\widehat{\Omega}_i)}{\sum_{j=1}^{K}\alpha_j^t P(x|\widehat{\Omega}_j)} \quad (6)$$

▶ M-step:

$$\alpha_i^{t+1} = \frac{h_i^t(x)}{\sum_{j=1}^{K} h_j^t(x)} \quad (7)$$

입력되는 비디오 영상의 각 프레임마다 식(2)의 확률분포함수를 이용하여 EM알고리즘을 통해 식(5)의 가중치 파라미터 α_i는 매 프레임마다 업데이트된다. 가중치 파라미터의 초기값은 최초 입력 얼굴로부터 식(7)의 M-step을 이용하여 초기화할 수 있다.

3. 임의의 사람에 대한 모델 추정

시스템은 미리 여러 명의 사용자에 대해 2절과 같은 방법을 통해 사전에 학습한다(본 논문의 실험에서는 10명의 사용자를 사전에 학습하였다.). 각 사람의 모델은 매 프레임마다 입력 얼굴에 따라 EM알고리즘을 통해 식(5)의 가중치 파라미터를 갱신한다. 하지만 시스템에 학습되지 않은 전혀 다른 사람의 얼굴이 입력되었을 경우 식(3)의 입력 얼굴에 대한 확률은 얼굴이 아닌 이미지와 같은 수준으로 현저히 떨어지

기 때문에 임의의 얼굴을 올바로 추적하고 인식하기 어렵다.

본 논문에서는 시스템에 학습되어 있는 모델을 선형적으로 결합하여 임의의 얼굴에 대한 모델을 추정한다. 시스템에 L명의 사람이 학습되어 있다면 입력된 임의의 얼굴은 시스템에 존재하는 각 모델과의 유사도에 따라 다음 식(8)과 같이 가우시안 혼합 모델 확률 분포를 추정할 수 있다.

$$P(x) = \sum_{i=1}^{L} \beta_i P_{G_i}(x) \tag{8}$$

여기서 $P_{G_i}(x)$는 식(5)로 추정된 입력 x에 대한 i번째 학습된 사람의 가우시안 혼합 모델 확률 분포이다. 그리고 β_i는 i번째 사람에 대한 가중치 파라미터로 입력 얼굴과 유사한 사람일수록 가중치가 크고 다른 얼굴일수록 가중치가 낮다. 만약 시스템에 등록되어 있는 사람이 입력 얼굴일 경우 해당 사람의 가중치가 매우 높게 추정될 것이다.

i번째 가중치 파라미터 β_i는 2절의 가중치 파라미터 α_i와 마찬가지로 EM알고리즘을 이용하여 매 프레임마다 가중치를 추정하고 갱신한다.

▶ E-step:

$$g_i^t(x) = \frac{\beta_i^t P_{G_i}(x)}{\sum_{j=1}^{L} \beta_j^t P_{G_j}(x)} \tag{9}$$

▶ M-step:

$$\beta_i^{t+1} = \frac{g_i^t(x)}{\sum_{j=1}^{L} g_j^t(x)} \tag{10}$$

가중치 파라미터 β_i의 초기값은 2절의 가중치 파라미터 α_i와 같이 최초 입력 얼굴로부터 식(10)의 M-step을 이용하여 초기화할 수 있다. 그림 3은 시스템에 10명의 사람이 학습되어 있을 경우, 입력 얼굴에 대한 각 얼굴의 분포에 대한 가중치의 예를 보여준다. 시스템은 가중치 파라미터를 통해 현재 입력된 얼굴이미지와 학습된 비선형적 매니폴드 모델들과의 관계를 갱신한다. 시스템은 매 프레임마다 얼굴 영역으로 추정되는 위치에 많은 표본을 생성하고 표본 영역 중 식(8)의 가우시안 혼합 모델 확률을 가장 높게 갖는 영역을 얼굴 이미지로 추적하고 또한 이 확률을 그대로 얼굴 인식으로 사용할 수 있다.

4. 실험 및 결과

본 논문의 실험은 조명 변화 등의 환경적 요소를 배제한다. 하지만 얼굴 포즈의 변화나 자연스러운 얼굴 표정의 변화, 그리고 자연스러운 얼굴의 움직임을 포함하여 실험한다.

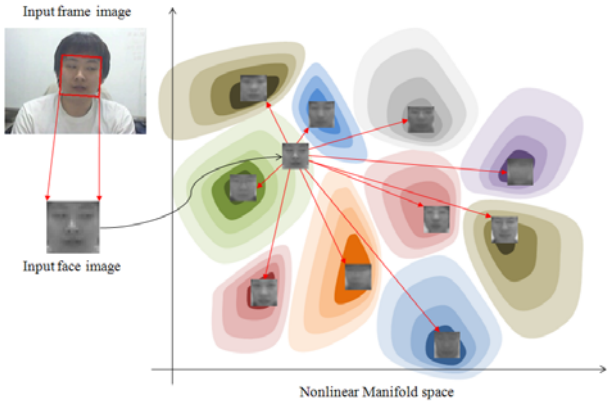

그림 3. 입력얼굴에 대한 비선형적 매니폴드 공간에서의 거리 가중치에 따른 추정

4.1 실험 환경 및 전처리

실험을 위해 15명의 서로 다른 사람들로부터 약 40초가량의 영상을 얻는다. 각 영상은 15fps로 약 615장의 이미지를 포함한다. 환경적인 요소를 배제하기 위해 모든 영상은 흰색의 배경과 동일한 조명상태를 유지하였다. 전체 영상 중 10명의 동영상은 되도록 많은 얼굴 포즈를 포함하여 학습 데이터로 이용된다. 그리고 나머지 동영상은 자유로운 얼굴 포즈와 얼굴 표정을 가지며 얼굴 추적 및 인식에 대한 실험을 위해 사용된다.

10명의 학습 영상은 각 프레임마다 얼굴 영역을 추출하고 추출된 얼굴 이미지를 20×20의 크기로 정규화 한다. 각 사람마다 2절에서와 같은 방법으로 비선형적 매니폴드 모델을 생성하기 위해 그림 4와 같이 7개 포즈의 주요한 얼굴 이미지를 시드(seed)로 하여 K-means알고리즘을 통해 포즈를 분류한다. 분류된 7개의 포즈로 각 사람마다 비선형적 매니폴드 모델을 형성하였다.

그림 4. 10명의 학습 얼굴이미지: 각 사람마다 7개의 포즈로 이루어져있다. 열은 10명의 사람을 나타내고 행은 서로 다른 7개의 포즈를 나타낸다.

그림 5 학습에 포함되지 않은 영상에 대한 얼굴 추적 결과

모든 실험은 P4-2.6Ghz의 CPU와 2.5Gb의 메모리를 가진 컴퓨터에서 수행되었다.

4.2 실험 결과

실험을 위해 입력된 동영상의 첫 프레임으로부터 Adaboot알고리즘[10]을 이용한 얼굴 검출방법을 이용하여 최초의 얼굴 위치를 초기화 한다. 이때 검출된 얼굴 이미지를 통해 3절에서 설명한 바와 같이 각 파라미터를 초기화 한다. 각 프레임에서 얼굴 영역을 추적하기 위해 이전 프레임의 위치를 중심으로 가우시안 분포로 80개의 표본 이미지를 추출한다. 그리고 추출된 이미지 중 식(8)의 확률을 가장 높게 갖는 이미지를 얼굴 이미지로써 추적하고 인식한다.

그림 5는 학습에 포함되지 않은 사람에 대한 얼굴 추적 결과를 40프레임 단위로 보여준다. 그리고 그림 6은 50프레임 단위로 추적 결과와 20×20의 크기로 정규화 된 추적 얼굴 이미지를 보여준다. 그림에서 보는 것과 같이 본 논문에서 제안한 방법을 이용하여 일반적으로 대부분의 입력 영상에서 얼굴에서 올바로 얼굴을 추적할 수 있었다. 하지만 학습 얼굴과 차이가 매우 큰 얼굴이나 학습에 많이 포함되어 있지 않은 얼굴 포즈(대각선 방향의 포즈 이동)에 대해서는 얼굴을 올바로 추적할 수 없었다.

만약 학습한 얼굴이 입력으로 들어왔을 경우 3절의 파라미터 β_i는 해당 얼굴에 대해 1에 가까운 값을 가지게 되고 식 (8)의 확률을 통해 입력 얼굴을 인식할 수 있었다.

5. 결론

본 논문에서는 얼굴을 학습하기 위해 비선형적 매니폴드 모델을 구성하였다. 그리고 학습된 모델들을 선형적으로 결합함으로써 임의의 얼굴에 대해 추적하거나 인식하는 방법을 제안하였다. 비디오 기반의 영상을 통해 해당 파라미터를 지속적으로 갱신함으로써 임의의 얼굴이 입력되었을 경우 효율적으로 얼굴을 추적할 수 있었다. 하지만 입력 얼굴이 학습한 얼굴과 크게 다르거나 포함되어 있지 않은 얼굴 포즈를 갖는 경우에는 올바르게 얼굴을 추적할 수 없었다. 또한 조명의 변화나 복잡한 배경 등과 같은 주위 환경에 따른 변화에서 얼굴을 올바르게 인식하기 어려웠다. 시스템은 얼굴을 추적하기 위해 시스템에 학습되어 있는 모든 사람과의 확률을 계산해야 하기 때문에 처리 시간이 오래 걸리는 단점을 보였다.

따라서 추후에는 보다 다양한 환경에서도 올바르게 시스템이 동작하고 확률적으로 거리가 매우 적은 사람의 확률은 배제시키는 등의 방법을 이용하여 시스템이 보다 빠르게 처리할 수 있는 방법을 연구하려 한다.

참고문헌

[1] G. Shakhnarovich and B. Moghaddam, "Face Recognition in Subspace", Handbook of Face Recognition, Eds, Stan Z, Li and Anil K, Jain, Springer-Verlag, 2004.

[2] A. Abate, M. Nappi, D. Riccio, G. Sabatino, "2D and 3D face recognition: A survey", Pattern

그림 6. 학습에 포함되지 않은 영상에 대한 얼굴 추적 결과 및 추적 얼굴

Recognition Letters, vol.28, pp.1885-1906, October 2007.

[3] A. Jepson, D. Fleet, T. El-Maraghi, "Robust online appearance models for visual tracking", Pattern Analysis and machine Intelligence, pp.1296-1311, 2003.

[4] Y. Wu, T-S. Huang, "A co-inference approach to robust visual tracking", In Proceeding of International Conference on Computer Vision, vol.2, pp.26-33, 2001.

[5] T. Sim, S. Zhang. "Exploring face space", In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop(CVPRW'04), vol.5, 2004.

[6] K-C. Lee, J. Ho, M-H. Y, D. Kreigman, "Visual tracking and recognition using probabilistic appearance manifolds", Computer Vision and Image Understanding, 99(3):303-331, 2005.

[7] K-C. Lee, D. Kriegman, "Online Learning of Probabilistic Appearance Manifolds for Video-Based Recognition and Tracking", CVPR' 05, vol.1, 2005.

[8] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, T. Darrell, "Face recognition with image sets using manifold density divergence", In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1:581-588, June 2005.

[9] B. Moghaddam, A. Pentland, "Probabilistic Visual Learning for Object Representation", Pattern Analysis and Machine Intelligence, PAMI-19 (7), pp. 696-710, July 1997

[10] P. Viola and M. Jones. Robust real time object detection. In IEEE ICCV Workshop on Statistical and Computational Theries of Vision, July 2001.