DOI: 10.3745/KIPSTD.2009.16B.5.385

환경에 적응적인 얼굴 추적 및 인식 방법

주 명 호[†] · 강 행 봉[††]

요 약

사람의 얼굴은 강체(Rigid object)가 아니기 때문에 얼굴을 추적하거나 인식하는 일은 쉽지 않다. 특히 얼굴의 포즈나 주변 조명의 변화에 따른 입력 영상의 차이는 얼굴 인식을 어렵게 하는 주된 원인이다. 본 논문에서는 비디오 영상으로부터 얼굴을 추적하고 인식할 때 발생하는 이 두 가지의 문제를 해결하기 위한 프레임웍과 전처리 방법을 제안한다. 얼굴 포즈의 변화에도 효과적으로 얼굴을 추적 및 인식하기 위해 먼저 학습 영상으로부터 주성분 분석법(Principal Component Analysis)을 이용하여 각 얼굴 포즈마다 하나의 독립된 가우시안 분포를 추정하고 이를 이용하여 각 사람마다 가우시안 혼합 모델(Gaussian Mixture Model)을 구성한다. 본 논문에서는 서로 다른 조명 상태를 가진 얼굴 영상을 처리하기 위해 먼저 입력된 얼굴 영상을 SSR(Single Scale Retinex) 모델을 이용하여 반사율(Reflectance)과 조도(Illuminance)로 분해한다. 반사율은 사전 정의된 범위 안에서 히스토그램 평활화를 수행함으로써 재조정되고 조도는 조명의 변화를 포함하고 있지 않은 영상들으로부터 학습된 매니폴드 모델로 다시 근사된다. 이 두 특징을 결합함으로써 실내 환경이나 실외 환경에서 촬영된 영상에서 효율적으로 얼굴을 추적 및 인식한다. 비디오 기반의 영상으로부터 보다 효율적으로 얼굴을 추적하기 위해 본 논문에서는 구성된 모델의 가중치를 각 프레임마다 이전 프레임의 추적 결과에 의해 EM 알고리즘을 이용하여 갱신함으로써 비디오 영상내의 연속적으로 변화하는 얼굴 포즈를 추정하였다. 본 논문에서 제안된 방법은 실내에서의 다양한 조명환경과 실외의 여러 장소에서 획득한 실험 영상을 이용하여 기존에 연구되어 온 다른 방법에 비해 우수한 성능을 보였다.

키워드 : 얼굴 인식, 얼굴 추적, 매니폴드, 레티넥스

A New Face Tracking and Recognition Method Adapted to the Environment

Myung-Ho Ju[†] · Hang-Bong Kang[††]

ABSTRACT

Face tracking and recognition are difficult problems because the face is a non-rigid object. The main reasons for the failure to track and recognize the faces are the changes of a face pose and environmental illumination. To solve these problems, we propose a nonlinear manifold framework for the face pose and the face illumination normalization processing. Specifically, to track and recognize a face on the video that has various pose variations, we approximate a face pose density to single Gaussian density by PCA(Principle Component Analysis) using images sampled from training video sequences and then construct the GMM(Gaussian Mixture Model) for each person. To solve the illumination problem for the face tracking and recognition, we decompose the face images into the reflectance and the illuminance using the SSR(Single Scale Retinex) model. To obtain the normalized reflectance, the reflectance is rescaled by histogram equalization on the defined range. We newly approximate the illuminance by the trained manifold since the illuminance has almost variations by illumination. By combining these two features into our manifold framework, we derived the efficient face tracking and recognition results on indoor and outdoor video. To improve the video based tracking results, we update the weights of each face pose density at each frame by the tracking result at the previous frame using EM algorithm. Our experimental results show that our method is more efficient than other methods.

Keywords : Face recognition, Face tracking, Manifold, Retinex

1. 서 론

최근 몇 년 동안 정지 영상뿐만 아니라 비디오 영상에 대한 얼굴 검출이나 얼굴 추적, 얼굴 인식 등에 대한 많은 연구가 진행 되어 왔다. 특히 비디오 영상 기반의 시스템은 정지 영상 기반의 시스템에 비해 생체 인증이나 비디오 감

※ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2009년도 문화콘텐츠산업기술지원사업의 지원 및 2008년도 가톨릭대학교 교비연구비 지원으로 이루어졌음.
† 준 회 원 : 가톨릭대학교 컴퓨터공학과 박사과정
†† 종신회원 : 가톨릭대학교 디지털미디어학부 교수
논문접수 : 2009년 4월 28일
수 정 일 : 1차 2009년 7월 20일, 2차 2009년 8월 3일
심사완료 : 2009년 8월 3일

시 시스템, 인간과 로봇간의 상호 작용 등의 보다 많은 분야에서 응용될 수 있다. 하지만 비디오 영상에서 얼굴을 추적하거나 인식하는 일은 얼굴의 포즈 변화나 얼굴 표정의 변화, 주변 조명이나 장소, 시간 등 환경에 대한 영향을 크게 받는 문제점을 가진다. 특히 얼굴 조명의 변화와 포즈의 변화는 얼굴을 추적하거나 인식할 때 발생하는 주된 문제로 꼽을 수 있다. 이러한 조명과 얼굴 포즈의 변화를 많이 포함한 영상에서 얼굴을 추적하고 인식하는 일은 매우 중요한 문제이며 이를 해결하기 위한 다양한 연구가 진행되어 왔다.

카메라로부터 입력된 얼굴 영상은 주위 조명이나 얼굴의 형태, 조명에 따른 그림자 등에 의해 입력되는 얼굴의 픽셀 값이 크게 달라진다. 이러한 조명 변화에 강건한 얼굴 인식을 하기 위한 방법은 조명 변화에도 픽셀 값이 크게 변하지 않는 근적외선 카메라(Near infrared camera)를 이용하는 방법과 일반적으로 사용되는 카메라(Visual camera)를 이용하지만 정규화 과정을 통해 영상을 정규화 한 후 얼굴 인식을 하는 방법으로 나눌 수 있다.

근적외선을 이용하여 획득한 얼굴 영상은 주위 조명 변화에 민감하지 않으며 보다 정확한 얼굴 인식을 가능하게 한다[1]. 그러나 근적외선 기반 얼굴 인식은 근적외선 얼굴 영상간의 매칭을 기반으로 하기 때문에 항상 사용자는 근적외선으로 입력된 얼굴 영상으로 시스템에 등록되어 있어야만 한다. 얼굴 인식이 응용되는 많은 분야에서 비주얼 영상을 이용한 얼굴 등록을 요구하기 때문에 이러한 적외선 영상을 등록하는 일은 쉽지 않다[2].

얼굴의 조명 변화를 정규화 하기 위한 연구들은 크게 두 가지로 분류된다. 첫 번째는 이미지 전체를 정규화 하는 방법으로 히스토그램 평활화 (HE)[3], Shape-from-shading[4], Quotient image relighting 방법 (QI)[5] 등이 있다. 그러나 이러한 방법은 전체 이미지를 정규화 하는 과정에서 얼굴의 특징을 왜곡하는 문제점을 갖기 때문에 효과적으로 얼굴을 인식하기 어렵다. 두 번째는 얼굴 인식에 있어 조명 상태에 영향을 받지 않는다고 생각되는 얼굴의 특징을 이용하여 얼굴 인식을 수행하는 방법으로 Single Scale Retinex (SSR) 모델[6, 7]이나 Logarithmic total variation (LTV) 모델[8], Self quotient image (SQI)[9] 등이 있다. 이러한 방법들은 얼굴 영상으로부터 조도(Illuminance)와 반사율(Reflectance)을 추정하고 조명 변화에 영향을 받지 않는 반사율을 이용한다. 반사율은 물체의 알베도(Albedo)와 표면 법선(Surface normal)에 의해서만 변화하기 때문에 조명의 영향을 받지 않지만 이미지의 2차원 정보만으로 정확한 반사율을 추정하는 것은 쉽지 않다. 또한 조도에 포함되는 얼굴 정보를 함께 고려해야 보다 정확한 얼굴 인식을 할 수 있다.

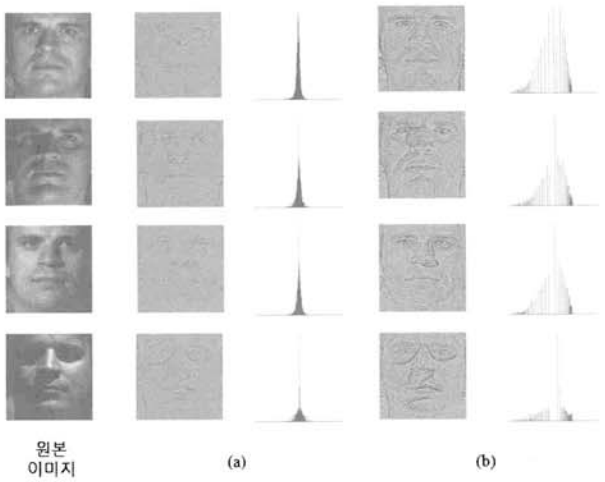
동영상의 경우 매 프레임마다 얼굴 영상을 처리해야 하기 때문에 보다 빠른 계산 속도가 요구된다. SSR 모델은 반복 과정 없이 정의된 가우시안 필터의 분산 값을 이용하여 단순한 계산만으로 반사율 영상을 추정할 수 있다. 그러므로 SSR 모델을 사용할 경우 매우 빠르면서도 LTV 모델과 유사한 결과를 얻을 수 있는 장점을 갖는다. 본 논문에서는 SSR 모델을 이용하여 빠르게 얼굴의 반사율을 추정하고 정규화 과정을 통해 보다 정확한 반사율을 획득한다.

비디오 영상에서 다양한 포즈의 얼굴을 추적하고 인식하기 위해 Lee[10]는 사람의 얼굴을 몇 개의 구분되는 포즈로 나누고 각 포즈를 선형적인 방법으로 근사하였다. 그리고 각 포즈를 학습 영상으로부터 직접 계산된 변환 확률을 이용하여 연결하였다. 또한 [11]에서 입력받은 얼굴 영상을 이용하여 모델을 갱신함으로써 보다 효율적인 인식을 도모하였다. 그러나 포즈 간의 변환 확률을 학습 영상 내의 포즈 변환 횟수로부터 직접적으로 계산하였기 때문에 학습 영상이 충분한 수의 얼굴 포즈를 포함하지 않거나 입력 영상에서의 얼굴 포즈의 변화가 학습 영상과 다른 경우 올바르게 포즈를 인식하지 못한다. 비디오 기반의 얼굴 포즈 인식에서 학습 영상에 의존적이지 않은 포즈간의 변환 확률을 계산하기 위해서는 실제 입력되는 프레임간의 얼굴 포즈 변화에 따라 연속적으로 확률이 갱신되어야 한다.

Sim[12]은 각 사람의 얼굴 영상 전체를 PCA를 통하여 저차원의 데이터로 투영한 이후 가우시안 혼합 모델로 근사하고 이 모델을 이용하여 확률분포함수를 구함으로써 얼굴 인식을 시도하였다. 이때 학습 영상에 다양한 포즈나 조명 상태의 얼굴을 포함시킴으로써 포즈 및 조명에 대해 좋은 인식률을 보였다. 그러나 이러한 방법은 이미지 전체를 저차원으로 투영하기 때문에 [10]의 방법에 비해 많은 정보가 손실되며 만약 입력되는 영상의 얼굴 포즈나 조명 변화가 실험 영상에 포함되어 있지 않은 경우 좋은 성능을 보이기 어렵다. 보다 효율적인 인식을 위해서는 입력되는 얼굴 영상과 학습 영상과의 차이를 최소화하기 위한 전처리 과정이 필요하다.

본 논문에서는 한 사람의 얼굴을 몇 개의 구분되는 포즈로 나누고 [10]과 같이 각 포즈를 PCA를 이용하여 선형적인 방법으로 근사하였다. 그러나 [10]의 방법과는 다르게 근사된 포즈 분포들의 가우시안 혼합 모델로 전체 얼굴의 매니폴드를 표현함으로써 다양한 얼굴 포즈에 대해 효과적으로 인식하는 방법을 제안한다. 비디오 기반의 입력 영상에서 프레임간의 얼굴 포즈 변화를 효율적으로 인식하기 위해 본 논문에서는 매 프레임마다 입력되는 얼굴 영상에 따라 EM알고리즘을 이용하여 각 분포의 가중치를 갱신하고, 갱신된 가중치를 다음 프레임에서 얼굴 포즈간의 변환 확률로 이용한다. 제안된 방법은 학습 영상에 의존적이지 않으면서 비디오 기반의 연속적인 입력 얼굴에 대해 효율적으로 얼굴 포즈 인식을 수행한다.

입력된 얼굴 영상에서 조명을 정규화하기 위해 본 논문에서는 SSR 모델을 이용하여 입력 영상을 반사율과 조도로 분해한다. 반사율은 주로 얼굴의 구조적인 정보를 포함한다. 그러나 얼굴 영상이 다양한 조명을 포함할 경우 반사율은 보다 약한 구조적 정보를 갖는다. 본 논문에서는 이러한 반사율을 사전 정의된 범위 안에서 히스토그램 평활화를 수행함으로써 보다 정확한 반사율을 획득한다. 입력 영상으로부터 분해된 조도는 조명 변화에 의한 대부분의 왜곡을 포함



(그림 1) 향상된 반사율 영상: (a) SSR을 이용한 반사율 영상과 해당 히스토그램, (b) 범위적 히스토그램 평활화된 반사율 영상과 해당 히스토그램

한다. 왜곡을 포함하지 않는 조도를 추정하기 위해 조명 변화를 포함하고 있지 않은 영상들로부터 학습 된 매니폴드 모델로부터 새로이 입력 영상의 조도를 추정하고 이를 반사율과 결합함으로써 실내 환경이나 실외 환경으로부터 획득한 영상에서 효율적으로 얼굴을 추적 및 인식한다.

본 논문의 구성은 다음과 같다. 2절에서 얼굴 조명의 정규화를 위해 SSR 모델을 이용한 반사율 추정과 정규화 방법을 소개하고 조명 왜곡을 포함하지 않은 조도의 추정 방법을 설명한다. 두 특징을 결합하여 조명 정규화 된 얼굴 영상을 재구성한다. 3절에서는 가우시안 혼합 모델과 모델의 학습 방법, 그리고 모델로부터 추정할 수 있는 확률분포 함수를 소개하고 4절에서 이를 이용하여 얼굴을 추적하고 인식하는 방법을 설명한다. 5절에서 실험결과를 통해 본 논문에서 제시하는 방법이 기존의 연구보다 우수한 성능을 가짐을 보이고 6절에서 결론을 맺는다.

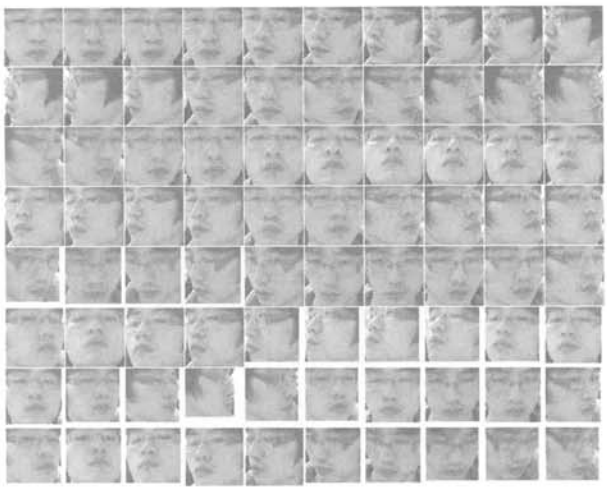
2. 얼굴 조명의 정규화

Lambertian 모델을 가정하면 이미지 $I(x,y)$는 반사율(Reflectance)과 조도(Illuminance)의 곱으로 식(1)과 같이 표현할 수 있다.

$$I(x,y) = R(x,y)L(x,y) \tag{1}$$

여기서 $R(x,y)$과 $L(x,y)$는 점(x,y)에서의 반사율과 조도를 나타낸다. 반사율 $R(x,y)$은 이미지의 알베도(Albedo)와 표면 법선(Surface normal)에 의해서만 변화하기 때문에 조명 변화에 영향을 받지 않는다. 그러나 이미지는 2차원 정보이기 때문에 정확한 반사율을 추정하는 것은 쉽지 않다[13]. 식(1)은 다음과 같이 다시 전개할 수 있다.

여기서 $\hat{R}(x,y)$는 얼굴의 구조적 정보만을 포함하며



(그림 2) 학습을 위한 서로 다른 포즈의 이미지 영상

$\hat{L}(x,y)$는 얼굴의 조명과 음영 정보를 포함한다.

2.1 반사율 추정

Retinex는 고급 레벨의 동적 범위 압축과 색의 불변성을 갖는 이미지 향상 기법이다. Retinex 알고리즘은 이미지 $I(x,y)$의 비율로써 반사율을 추정하고 저주파 버전으로써 조도를 추정한다. SSR 모델[6][7]는 Land[14]의 가장 최근 버전으로 이미지에서 한 점(x,y)의 SSR은 다음 식(3)과 같이 정의된다.

$$R(x,y) = \log I(x,y) - \log[F(x,y) \otimes I(x,y)] \tag{3}$$

여기서 $R_i(x,y)$는 Retinex 출력 결과이고 $I(x,y)$는 입력 영상의 픽셀 값이다. 그리고 $\otimes$는 컨볼루션(Convolusion) 연산을 나타내며 $F(x,y)$는 가우시안 필터 함수이다.

SSR 모델을 이용하여 추정된 반사율은 (그림 1(a))와 같이 좁은 형태의 히스토그램 분포를 가지며 입력 영상의 조명에 따라 얼굴 구조가 부분적으로 약하게 표현되는 문제점을 갖는다((그림 1(a))에서는 조명에 따라 턱 선의 형태가 약하게 나타나는 예를 보여준다). 본 논문에서는 이러한 반사율을 사전 정의된 범위 안에서 히스토그램 평활화하여 정규화 함으로써 보다 정확한 반사율 $\hat{R}$을 추정한다.

영상의 밝기 레벨이 범위 [0, 1]로 정규화 된 값이라고 가정할 때, 특정 범위 $[\theta_{\min}, \theta_{\max}]$에서의 밝기 레벨 $\hat{s}$는 다음과 같이 얻어진다.

$$\hat{s} = T(r) = \int_{\theta_{\min}}^{r} P(x)dx \tag{3}$$

여기서 $P(x)$는 주어진 영상의 밝기 레벨에 대한 확률밀도 함수(PDF)를 나타낸다. 일반적으로 히스토그램 평활화는 이미지의 히스토그램 전체에 대해 균등화를 수행하는 반면 식(3)의 범위적 히스토그램 평활화는 히스토그램의 특정 범위로 균등화를 수행한다. 따라서 식(3)의 $\hat{s}$는 다음 식(4)

와 같이 영상의 밝기 레벨 범위를 변화시킨다.

$$\begin{aligned} &\hat{s}:[0,1]\rightarrow[\theta_{\min},\theta_{\max}] \\ &where\ 0\le\theta_{\min}<\theta_{\max}\le 1 \end{aligned} \tag{4}$$

(그림 1(b))는 식(3)을 이용하여 반사율 영상에 범위적 히스토그램 평활화를 적용한 예를 보여준다.

2.2 조도 특징 추정

식 (2)와 같이 얼굴 영상은 반사율과 조도의 곱으로 계산될 수 있다. 조도는 얼굴 영상의 조명과 음영을 포함한다. 그러므로 정규화 된 반사율 $\hat{R}$과 조명 정규화 된 조도 $\acute{L}$을 결합하여 식 (5)와 같이 조명 정규화 된 얼굴 영상 $\acute{I}$을 추정할 수 있다.

$$\acute{I}(x,y)=\hat{R}(x,y)\acute{L}(x,y) \tag{5}$$

식 (5)와 마찬가지로 식(6)과 같이 조명 정규화 된 얼굴 영상으로부터 조명 정규화 된 조도를 계산할 수 있다.

$$\acute{L}(x,y)=\frac{\acute{I}(x,y)}{\hat{R}(x,y)} \tag{6}$$

임의의 조명을 가지고 있는 입력 얼굴 영상 I의 조명 정규화 된 조도 $\acute{L}$을 추정하기 위해 학습 영상으로부터 추정된 조도를 이용한다. 학습 영상은 조명 정규화 된 영상이기 때문에 추정된 조도는 조명의 변화를 포함하지 않는 조명 정규화 된 조도이다.

i번째 학습 영상 I_i^{tra}은 식(3), (4)와 식(6)을 이용하여 반사율 $\hat{R}_i^{tra}$과 조도 $\acute{L}_i^{tra}$의 결합으로 분해할 수 있다. 반사율은 조명의 영향을 받지 않기 때문에 임의의 조명을 갖는 입력 영상 I의 조명 정규화 된 조도 $\acute{L}$은 추정된 반사율 $\hat{R}$과 i번째 학습 영상의 반사율 $\hat{R}_i^{tra}$의 비율로써 식 (7)과 같이 추정할 수 있다.

$$\gamma_i=\frac{\hat{R}(x,y)}{\hat{R}_i^{tra}(x,y)}=\frac{\acute{L}(x,y)}{\acute{L}_i^{tra}(x,y)} \tag{7}$$

본 논문의 매니폴드 프레임웍에서는 K개의 포즈 분포에 대해 식(7)을 이용하여 K개의 조도를 추정하고 식(8)과 같이 각 분포의 가중치 합으로 최종적인 조명 정규화 된 조도를 추정한다.

$$\acute{L}(x,y)=\sum_{i=1}^{K}\alpha_i\gamma_i\acute{\bar{L}}_i^{tra}(x,y) \tag{8}$$

여기서 $\acute{\bar{L}}_i^{tra}$은 i번째 포즈 분포의 평균 영상으로부터 추정된 조도이고 γ_i는 식(7)을 이용하여 계산된 i번째 포즈에 대한 반사율간의 비율이다. 그리고 α_i는 i번째 포즈 분포에 대한 가중치이다.

2.3 조명 정규화 된 얼굴 영상 추정

임의의 조명을 갖는 얼굴 영상 I는 식(2)와 같이 정규화 된 반사율 $\hat{R}$과 조도 $\hat{L}$의 결합으로 표현된다. 본 논문에서는 I로부터 SSR 모델을 이용하여 추정된 반사율에 사전 정의된 범위 안에서 히스토그램 평활화하여 정규화 함으로써 반사율 $\hat{R}$을 추정하였다. 그리고 식(8)과 같이 매니폴드 프레임웍으로부터 조명 정규화 된 조도 $\acute{L}$을 추정하였다. 조명 정규화 된 얼굴 영상 $\acute{I}$는 식(5)와 같이 정규화 된 반사율 $\hat{R}$과 조명 정규화 된 조도 $\acute{L}$의 결합으로 추정할 수 있다.

3. 포즈 기반 매니폴드 모델

얼굴 영상은 얼굴의 포즈에 따라서 매우 다른 이미지로 표현될 수 있다. (그림 2)는 학습 영상으로부터 추출된 얼굴 영상들로 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 근사된 가우시안 혼합 모델의 확률 분포 함수는 식(9)와 같이 정의된다.

$$P(\mathbf{x})=\sum_{i=1}^{K}\alpha_i Gauss(\mathbf{x}|\mu_i,\Sigma_i) \tag{9}$$

여기서 μ_i와 Σ_i는 각각 i번째 가우시안 성분의 평균벡터와 공분산 행렬을, α_i는 i번째 가우시안 성분의 가중치를 나타낸다. i번째 포즈 가우시안 분포의 평균 벡터 μ_i와 공분산 행렬 Σ_i는 학습 영상에서 i번째 얼굴 포즈에 해당되는 학습 영상으로부터 근사화하며 가우시안 확률 분포 $Gauss(\mathbf{x}|\mu_i,\Sigma_i)$는 가우시안 Likelihood로써 추정한다. N차원의 입력 이미지에 대해 가우시안 Likelihood는 식(10)과 같다.

$$P(\mathbf{x}|\Omega_i)=\frac{\exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)^{\mathrm{T}}\Sigma_i^{-1}(\mathbf{x}-\mu_i)\right]}{(2\pi)^{\frac{N}{2}}|\Sigma_i|^{\frac{1}{2}}} \tag{10}$$

[10]과 [15]에 따르면 i번째 가우시안 분포 Ω_i는 PCA를 이용하여 구성된 어파인 부분 공간(Affine subspac-e) $\widehat{\Omega}_i$로써 근사할 수 있다. $\widehat{\Omega}_i$는 i번째 얼굴 포즈에 속하는 학습 영상이 주어졌을 때, 집합 $\{\mu_i,\Sigma_i,\Phi,\Lambda\}$을 계산함으로써 추정된다. 여기서 μ_i는 데이터의 평균벡터이고 Σ_i는 공분산 행렬, Φ는 Σ_i의 고유값의 큰 고유벡터를 순서대로 M개 ($M\ll N$) 포함하여 데이터를 고유공간으로 투영하기 위한

행렬이며 Λ는 대각 성분 $\Lambda_{jj}=\lambda_j$로 Φ의 고유벡터에 해당하는 고유값을 가지는 대각행렬이다.

입력된 얼굴 영상 I는 근사된 고유공간 $\widehat{\Omega}_i$에 선형적으로 투영되어 $y=[y_1,...,y_M]^T=\Phi^T(I-\mu_i)$의 고유 얼굴을 얻는다. 식(5)의 Likelihood는 $\widehat{\Omega}_i$에 의해 근사된 고유 얼굴을 이용하여 식(10)과 같이 두 개의 가우시안 분포의 곱으로 표현될 수 있다[15].

$$P(\mathbf{x}|\widehat{\Omega}_i)=\left[\frac{\exp\left(-\frac{1}{2}\sum_{j=1}^{M}\frac{y_j^2}{\lambda_j}\right)}{(2\pi)^{\frac{M}{2}}\prod_{j=1}^{M}\lambda_j^{\frac{1}{2}}}\right]\left[\frac{\exp\left(-\frac{\epsilon^2(x)}{2\rho}\right)}{(2\pi\rho)^{\frac{N-M}{2}}}\right] \tag{11}$$

여기서 M은 부분 공간 $\widehat{\Omega}_i$의 차원 수이고 $\epsilon^2(x)$는 재구성 에러(Residual reconstruction error)로 식(12)와 같이 정의된다[15].

$$\epsilon^2(\mathbf{x})=\sum_{j=M+1}^{N}y_j^2=\|\mathbf{x}-\mu_i\|^2-\sum_{j=1}^{M}y_j^2 \tag{12}$$

식(11)에서 파라미터 ρ는 $\frac{1}{N-M}\sum_{j=M+1}^{N}\lambda_j$를 사용하거나 혹은 간단하게 $\frac{1}{2}\lambda_{M+1}$을 사용한다. 본 논문에서는 후자를 선택하여 실험하였다.

식(11)을 이용하면 식(9)를 다음 식(13)과 같이 바꾸어 쓸 수 있다.

$$P(\mathbf{x})=\sum_{i=1}^{K}\alpha_i P(\mathbf{x}|\widehat{\Omega}_i) \tag{13}$$

α_i는 i번째 포즈의 확률 분포 $P(\mathbf{x}|\widehat{\Omega}_i)$에 대한 가중치 파라미터로 입력된 영상의 얼굴 포즈와 유사한 포즈 분포에서 높은 가중치를 갖는다.

비디오 기반의 입력 영상은 프레임간의 연속성을 갖는다. 그러므로 프레임간의 연속성을 고려할 때 보다 효율적으로 얼굴을 추적 및 인식할 수 있다. 본 논문에서는 First order Markov 가정으로 현재 프레임에서 추적 및 인식한 결과를 이용하여 다음 프레임의 각 포즈분포의 가중치를 추정한다. t시점에서 추정된 각 확률분포 및 가중치 파라미터를 이용하여 EM알고리즘[15]을 통해 $t+1$시점에서의 가중치 파라미터를 추정한다. EM알고리즘은 다음과 같은 두 가지 단계를 반복적으로 수행한다.

▶ Initial:

$$\alpha_i^0=\frac{1}{K}$$

▶ E-step:

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

▶ M-step:

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

시스템은 E-step에서 t시점의 새로운 입력 $\mathbf{x}$의 기댓값으로 사전 확률 $h_i^t(\mathbf{x})$을 계산한다. K개의 사전확률이 계산되면 M-step에서는 입력에 대한 결합 Likelihood(J-oint Likelihood)를 최대화한다. EM알고리즘은 점증적으로 Likelihood를 수렴하기 때문에 학습 데이터의 전체 L-ikelihood에서 지역적 최대값(Local Maximum)을 찾을 수 있다. 따라서 입력 얼굴에 대해 각 포즈 분포의 가중치를 올바르게 갱신할 수 있다.

4. 비디오 기반 얼굴 추적 및 인식

비디오 기반 얼굴 추적을 위해 먼저 현재 프레임 F_t에서 이전 프레임의 얼굴 추적 결과 $\mathbf{x}_{t-1}^*$을 중심으로 갖는 가우시안 분포로써 후보 얼굴 $\mathbf{x}_t$를 샘플링한다. 샘플링 된 각 후보 얼굴마다 식(13)의 확률분포함수가 계산되고 시스템은 식(14)와 같이 가장 큰 확률분포함수를 갖는 후보 얼굴을 현재 프레임에서의 얼굴 위치로 추적한다.

$$\mathbf{x}_t^*=\underset{\mathbf{x}_t}{\arg max}\,P(\mathbf{x}_t) \tag{14}$$

여기서 $P(\mathbf{x}_t)$는 식(13)의 확률분포함수이다.

시스템은 식(14)를 이용해 추적된 얼굴 영역 x_t^*을 입력으로 현재 사용자를 인식한다. 시스템에 S명의 사람이 등록되어 있을 때, 시스템은 각 사람마다 독립적으로 식(9)와 같은 매니폴드를 사전 학습한다. 입력 영상 x_t^*는 S명의 사람마다 확률분포를 계산하고 시스템은 식(14)와 유사하게 가장 높은 확률분포를 갖는 사람 s^*를 현재 입력된 사람으로 인식한다.

$$s^*=\underset{s}{\arg max}\,P_s(\mathbf{x}_t^*) \tag{15}$$

본 논문에서 제안하는 프레임웍의 전체적인 알고리즘은 <표 1>과 같다.

〈표 1〉 제안된 추적 및 인식 알고리즘 요약

Tracking and Recognition Algorithm:

Input Parameters: (W, S)
$W=\{x, y, w, h, \theta\}$: The set of 5 parameters for sampling windows, represented by a rectangular box in the image centered at (x, y) and of size (w, h) with orientation θ.
S: The number of windows sampled for each frame.
Output: (x^*, k^*)
x^*: The image of the tracked face.
k^*: Current identity of the tracked face.
Model Parameters: (K, N, x^*)
K: The number of the pose Gaussian densities
N: The number of persons for identity.
$x^* = (x, y, w, h, \theta)$: the location of the face in the image.

Initialization:
The tracker is initialized either manually or by a face detector (e.g., Adaboost algorithm) in the first frame. Let x_0^* be the initial face window from the first frame. Using x_0^*, the initial identity k^* is determined. The weight of the front pose is set 1 and others are set 0.
Begin
1. Sample Windows: Draw S samples of windows $\{W_1, \ldots, W_i, \ldots, W_s\}$ in current image frame at various locations of different orientations and sizes according to a 5-dimensional Gaussian distribution centered at x_{t-1}^*.
2. Tracking: Rectify each window W_t to a 20-by-20 image and rasterize it to form a vector I_t^t. Compute the pdf of equation (5) and choose x_t^* that gives the maximum pdf as the tracking output and update the weight for each Gaussian density using EM algrorithm.
3. Recognition: Compute the pdf for each person. And identity is computed using equation (15). Loop back to step 1 until the last frame.
End

5. 실험 결과

본 논문에서 제안하는 방법의 성능을 평가하기 위해 20명의 사람으로부터 임의의 조명 상태를 지속적으로 변화시킨 실내 영상과 옥상이나 도로, 공원과 같은 실외 환경에서 촬영한 실험 영상을 이용하여 얼굴을 추적 및 인식하고 Lee[10]의 방법과 성능을 비교하였다. 또한 다양한 조명의 변화를 포함하는 YalefaceB[16] 데이터를 이용하여 본 논문에서 제안한 전처리 방법을 적용했을 경우와 기존의 이미지 보정 기법을 적용했을 경우의 인식률을 비교하였다.

5.1 전처리

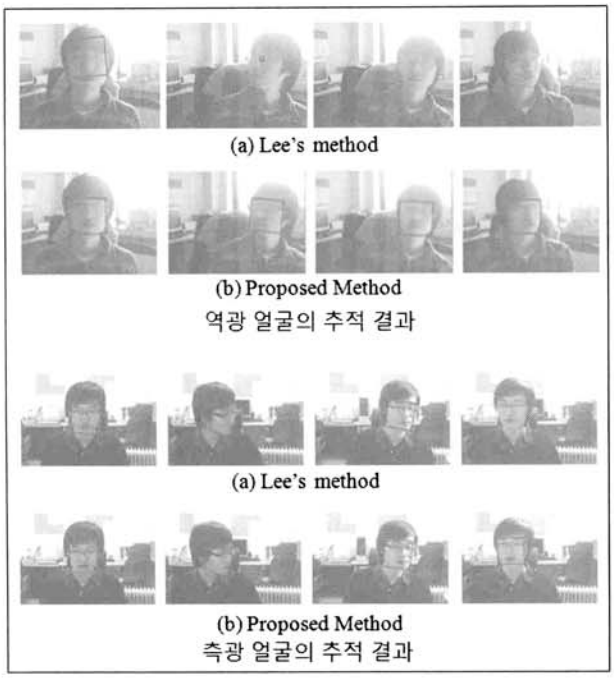
YalefaceB[16]는 조명변화에 대한 얼굴 인식 성능 평가에 널리 사용되는 실험 영상 중 하나이다. YalefaceB는 10명의 사람을 각 사람마다 8개의 얼굴 포즈로 나누고 포즈마다 서로 다른 65개의 조명 영상을 갖는 이미지로 총 5,200장(10명×8포즈×65조명=5,200)의 조명변화를 갖는 얼굴 영상을 제공한다. 본 논문에서는 YalefaceB로부터 정면 포즈 영상 650장을 이용하여 비교적 조명 변화가 적은 정면 조명의 영상 10장을 학습 영상으로 사용하고 나머지 640장의 영상을 이용하여 실험하였다.

조명 변화에 대한 얼굴 추적 및 인식 성능을 평가하기 위한 대표적인 비디오 영상이 아직 존재하지 않는다. 따라서 본 논문에서는 20명의 사용자로부터 다양한 조명 상태를 포함하는 실내 영상과 옥상이나 도로, 공원 등과 같은 실외 환경에서 획득한 영상을 이용하여 비디오 기반 얼굴 추적 및 인식 실험에 사용하였다. 사용자는 6명의 여성과 14명의 남성으로 구성되며 각 사용자마다 비교적 고른 조명 상태에서의 영상 1개를 획득하여 학습 영상으로 사용하였다. 그리고 다양한 조명 변화와 얼굴 포즈 변화를 포함하는 영상을 사용자마다 2~4개 획득하여 실험 영상으로 사용하였다. 획득한 모든 영상은 15fps이며 약 30초의 길이를 갖는다. 실험에 사용된 실내 영상은 창문을 통한 태양빛을 이용하여 사용자의 얼굴에 측광이나 역광으로 환경 변화를 주거나 책상 위의 스탠드, 거울 등을 이용하여 얼굴의 조명 상태를 지속적으로 변화시켰다. 실외 영상은 사용자의 얼굴이 직접적으로 태양빛을 받기 때문에 학습 영상과는 전혀 다른 조명 상태를 가지며 도로나 공원에서 획득한 영상은 움직이는 카메라를 이용하여 나무나 건물 등으로 인한 음영에 따른 조명 변화를 포함한다. 또한 모든 영상은 사용자 개개인에 따라 자연스러운 임의의 얼굴 포즈 변화를 포함한다.

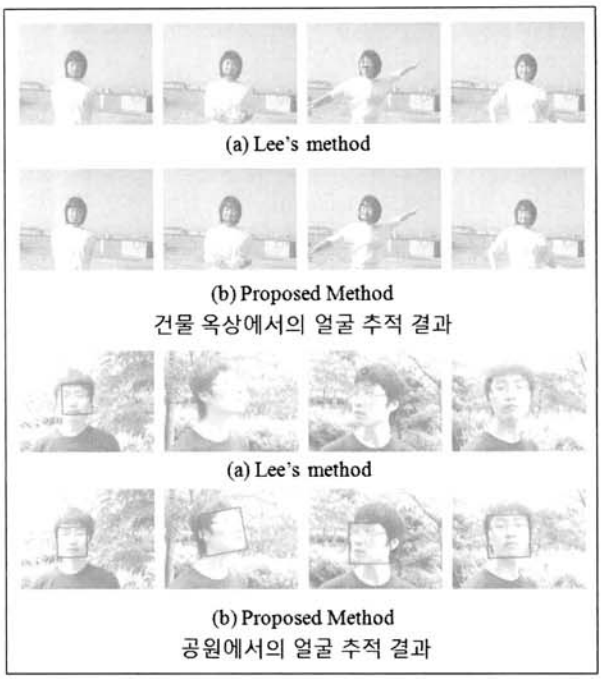
시스템의 학습을 위해 먼저 학습 영상으로부터 수동적으로 사용자의 얼굴을 추출하였다. 추출된 얼굴은 20×20 크기의 영상으로 크기를 조정되고 K-means 알고리즘을 이용하여 7개의 포즈로 분류하였다. 사용자의 매니폴드 모델을 구성하기 위해 분류된 포즈 영상은 PCA를 이용하여 각 포즈 분포를 근사한다. 포즈 분포의 차원 M은 클수록 시스템의 성능을 향상시키지만 속도를 저하시킨다. 본 논문에서는 적절한 포즈 분포의 차원 M을 정하기 위해 포즈 분로를 이루는 고유값을 누적하여 90%가 되는 M을 선택하였다.

5.2 얼굴 추적 결과

(그림 3)과 (그림 4)는 본 논문에서 제안하는 방법을 이용한 얼굴 추적 결과의 주요 프레임을 보여준다. (그림 3)은 실내에서 촬영된 조명 변화 영상의 예이며 (그림 4)는 실외



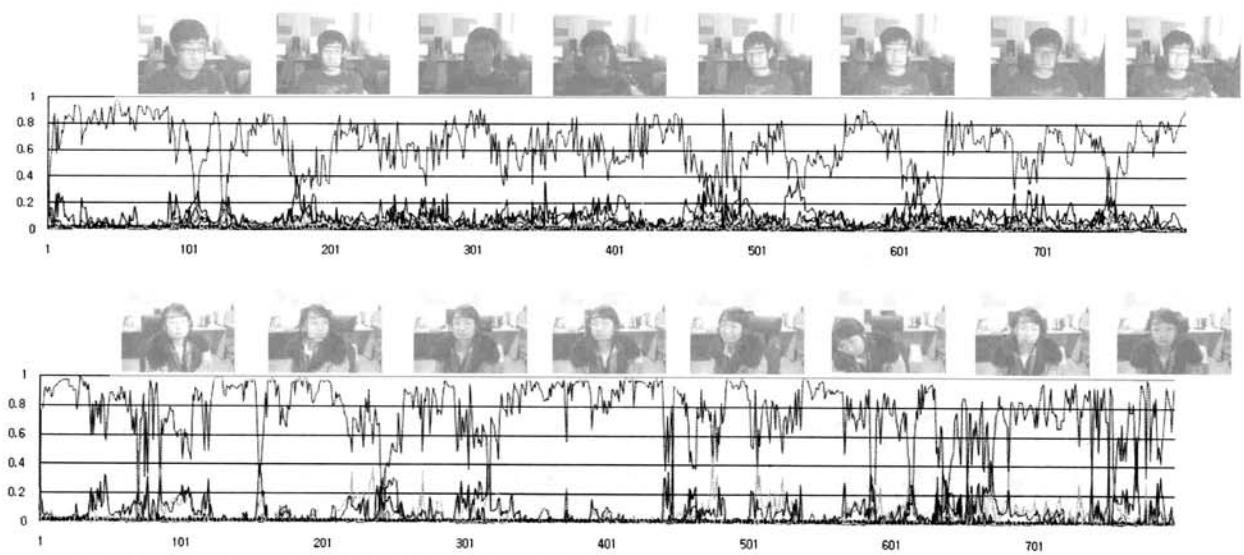
(그림 3) 실내 영상에서의 얼굴 추적 결과: (a) Lee의 방법을 이용한 얼굴 추적 결과, (b) 제안된 방법을 이용한 얼굴 추적 결과.



(그림 4) 실외 영상에서의 얼굴 추적 결과: (a) Lee의 방법을 이용한 얼굴 추적 결과, (b) 제안된 방법을 이용한 얼굴 추적 결과.

에서 촬영된 영상의 예다. (그림 3)과 (그림 4)의 (a)는 Lee[10]의 방법을 이용하여 얼굴을 추적한 결과로 조명의 상태가 크게 변하기 때문에 조명이 변화하기 시작할 때부터 얼굴을 올바르게 추적하지 못했다. 반면 (b)는 본 논문에서 제시한 방법을 이용한 얼굴 추적 결과로 환경이나 장소 등에 따라 얼굴의 조명 상태가 크게 변하지만 모든 영상에서 올바르게 얼굴을 추적하였다. Sim[12]의 방법은 모든 변화 가능한 조명 상태의 얼굴 영상을 모두 학습하지만 본 논문에서 제시한 방법은 사용자로부터 비교적 고른 조명상태를 가진 1개의 학습 영상만으로 학습한 후 모든 실험 영상에서 사용자의 얼굴을 추적한다.

실내 영상은 태양 빛과 임의적인 조명의 변화에 따라 추출되는 얼굴 영상이 부분적으로 밝거나 어둡게 왜곡된 영상이 획득된다. 또한 야간에 실내등을 이용하여 전체 조명을 변화시킴으로써 프레임간의 얼굴 밝기 차이가 큰 영상을 실험 영상으로 사용하였다. 실외 영상의 경우 태양 빛에 의한 직접적인 얼굴 밝기 변화뿐만 아니라 음영로 인한 왜곡을 포함한다. 본 논문에서는 음영에 의한 얼굴 왜곡을 (그림 5)



(그림 5) 지속적인 조명 변화에 따른 각 사용자의 확률분포함수: 붉은 선은 실제 입력된 사용자의 확률을 나타낸다.

와 같이 얼굴 자체의 굴곡으로 인해 발생되는 왜곡(Self Occlusion)과 건물이나 나무 등과 같은 주변 환경에 의해 발생되는 왜곡(External Occlusion)으로 정의하고 실험 영상에는 두 가지 얼굴 왜곡을 모두 포함하여 실험하였다.

얼굴을 추적하기 위해 시스템은 매 프레임마다 100개의 얼굴 샘플을 생성하고 Manifold를 이용하여 가장 적합한 얼굴을 추정한다. 모든 얼굴 샘플은 20×20의 크기로 조정되기 때문에 매 프레임에서의 연산 시간은 동일하다. 본 논문에서 제안된 방법은 매 프레임에서 약 350ms 의 시간이 소요되었다.

5.3 얼굴 인식 결과

비디오 기반의 실험 영상에서 얼굴 인식 성능을 평가하기 위해 본 논문에서는 각 사람의 확률분포함수를 매 프레임마다 비교하였다. (그림 6)은 매 프레임마다의 사용자 간의 확률분포함수의 변화를 보여준다. 그래프에서 붉은 선은 실제 입력된 올바른 사용자의 확률을 나타내고 그 외의 선은 시스템에 등록되어 있는 나머지 사람의 확률을 나타낸다. (그림 3)과 (그림 4)와 같이 본 논문에서 제안한 방법은 지속적인 조명의 변화에서도 항상 올바르게 사용자를 인식하였다. 본 논문의 방법에서 사용자의 인식률이 저하되는 원인은 조명의 변화보다는 샘플링 결과에 따른 얼굴 피팅(Fitting)에 보다 많은 영향을 받았다. <표 2>는 실험에 사용된 20명의 사용자에 대한 개인별 인식률을 보여주며 20명의 전체 사용자에 대해 평균 93.9%의 인식률을 보였다.

<표 3>은 히스토그램 평활화(Histogram Equalization), 감마 보정(Gamma Correction), 로그 이미지(Log Image), SSR 모델 등의 대표적인 조명 보상 기법과 본 논문에서 제안한 전처리 방법 간의 얼굴 인식률을 비교한 결과이다. 얼굴 인식은 YalefaceB 데이터에서 정면 조명 영상 10장으로 PCA를 통해 간단한 Eigenface 분류기를 학습하고 나머지 640장에 대한 얼굴 인식 결과를 측정하였다. PCA의 축의 개수가 증가함에 따라 기존의 방법은 히스토그램 평활화가 73.6%로 가장 높은 인식률을 보인 반면 본 논문에서 제안한 방법은 91.1%로 기존의 다른 전처리 방법에 비해 크게 인식률을 향상시킬 수 있었다.

(a) 얼굴 굴곡 등에 의한 음영 왜곡 (Self Occlusion)

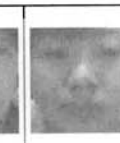
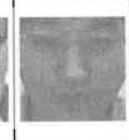
(b) 주변 환경에 의한 얼굴 왜곡 (External Occlusion)

(그림 6) 실내 영상과 실외 영상에서의 음영에 의한 얼굴 왜곡의 예

〈표 3〉 yalefaceB를 이용한 얼굴 인식 결과

Method	The number of PCA components		
	7	8	9
Histogram Equalization	60.5	73.1	73.6
Gamma Correction	52.3	54.8	57.9
Log Image	55.2	57.7	59.8
The SSR model	26.4	27.8	30.3
Proposed Method	63.3	82.9	91.1

〈표 2〉 20명에 대한 각 사용자 인식률

사용자										
인식률	98.0	99.2	97.6	99.0	98.9	90.4	93.2	98.4	97.0	97.4
사용자										
인식률	95.7	80.7	89.0	86.7	99.0	97.0	96.7	92.6	88.0	82.1

6. 결 론

본 논문에서는 각 얼굴 포즈를 선형적으로 근사하여 이를 가우시안 혼합모델로 구성하고 매 프레임마다 가중치를 갱신함으로써 비디오 영상에서 효율적으로 얼굴을 추적하고 인식하는 방법을 제안하였다. 그리고 임의의 조명 상태를 가진 입력 얼굴 영상으로부터 SSR 방법과 사전 정의된 범위에서의 평활화 방법을 이용하여 보다 향상된 반사율을 획득하고 학습된 매니폴드로부터 새로운 조도를 추정하여 두 특징을 결합함으로써 조명 정규화 된 얼굴 영상을 획득하였다.

본 논문에서 제안한 방법은 조명 변화에 의해 얼굴의 밝기가 전체적 또는 부분적으로 크게 변하는 실내 영상과 얼굴의 굴곡으로 인해 발생되는 그림자에 의한 왜곡(Self Occlusion), 주변 환경으로 인해 발생되는 그림자에 의한 왜곡(External Occlusion)을 모두 포함하는 실외 영상을 이용하여 기존의 다른 연구에 비해 효과적으로 얼굴 영역을 추적하고 인식할 수 있음을 보였다.

그러나 실외환경에서 얼굴을 추적하고 인식할 때, 본 논문에서 제안한 방법은 강한 태양빛에 의하여 얼굴의 그림자가 매우 선명할 경우 때때로 얼굴 추적에 실패하였다. 진한 그림자는 시스템이 추정한 반사율에 크게 영향을 미치기 때문에 본 논문의 방법으로 얼굴 조명을 정규화 할 경우, 그림자로 인한 에지 영역을 완벽하게 정규화 하기 어렵다. 그러므로 이러한 그림자를 제거

하여 보다 향상된 얼굴 추적 및 인식을 시도하는 것이 앞으로 나아가야 할 방향이라 하겠다.

참 고 문 헌

[1] S. Z. Li, R. Chu, S. Liao, and L. Zhang. "Illumination invariant face recognition using near-infrared images". IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4):627-639, 2007.

[2] Z. Lie and S. Z. Li. "Coupled Spectral Regression for Matching Heterogeneous Faces". IEEE Conference on CVPR, 1123-1128, 2009.

[3] R. Basri and D. Jacobs. "Photometric Stereo with Genera, Unknown Lighting". IEEE Conference on CVPR, 374-381, 2001.

[4] W. Zhao and R. Chellappa. "Symmetric shape from shading using self-ratio image". International Journal of Computer Vision. 45(1):55-75, 2001.

[5] A. Shashua and T. Riklin-Raviv. "The Quotient Image: Class-Based Re-rendering and Recognition with Varying Illuminations". TPAMI, 2001.

[6] R. Kimmel, M. Elad, D. Shaked, R. Keshet and I. Sobel. "A Variational Framework for Retinex". International Journal of Computer Vision, Vol.52, No.1, pp.7-23, 2003.

[7] D. J. Jobson, Z. Rahman and G. A. Woodell. "A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes". IEEE Transactions on Image Processing, 1997.

[8] T. Chen, X. S. Zhou, D. Comaniciu and T. S. Huang. "Total Variation Models for Variable Lighting Face Recognition". TPAMI, 28(9):1519-1524, 2006.

[9] H. T. Wang, S. Z. Li and Y. S. Wang. "Face Recognition under Varying Lighting Conditions using Self Quotient Image". International Conference on FGR, 2004.

[10] K-C. Lee, J. Ho, M-H. Y, D and Kreigman. "Visual tracking and recognition using probabil-istic appearance manifolds". Computer Vision and Image Understanding, 2005.

[11] K-C. Lee and D. Kriegman. "Online Learning of Probabilistic Appearance Manifold for Video-based Recognition and Tracking". CVPR, 2005.

[12] T. Sim and S. Zhang. "Exploring face space". CVPRW'04, 2004.

[13] X. Xie, W. S. Zheng, J. Lai and P. C. Yuen. "Face Illumination Normalization on Large Small Scale Features". CVPR, 2008.

[14] E. Land. "The Retinex Theory of Color Vision". Scientific American, 1977.

[15] B. Moghaddam and A. Pentland. "Probabilistic Visual Learning for Object Representation". Pattern Analysis and Machine Intelligence, 1997.

[16] A. S. Georghiades, P. N. Belhumeur and D. J. Kriegman. "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose". IEEE Trans. Pattern Anal. Mach. Intelligence, Vol.23, No.6, pp.643-660, 2001.

주 명 호

e-mail : hangel5@catholic.ac.kr

2005년 가톨릭대학교 컴퓨터공학과(학사)

2007년 가톨릭대학교 컴퓨터공학과(석사)

2007년~현 재 가톨릭대학교 컴퓨터공학과 박사과정

관심분야 : 영상처리, 인공지능, 컴퓨터비전

강 행 봉

e-mail : hbkang@catholic.ac.kr

1980년 한양대학교 전자공학과(학사)

1986년 한양대학교 전자공학과(석사)

1989년 Ohio State Univ. 컴퓨터공학(석사)

1993년 Rensselaer Polytechnic Institute 컴퓨터 공학(박사)

1993년~1997년 삼성종합기술원 수석연구원

1997년~현 재 가톨릭대학교 디지털미디어학부 교수

2005년 UC Santa Barbara, Visiting Professor

관심분야 : 컴퓨터비젼, HCI, 컴퓨터그래픽스, 인공지능