

## 합성 대표영상에 기반한 조명 변화 무관 얼굴 인식

문송항<sup>0</sup>, 이상웅, 이성환  
고려대학교 정보통신대학 컴퓨터학과

{shmoon<sup>0</sup>, sangwlee, swlee}@image.korea.ac.kr

### Illumination Invariant Face Recognition based on the Synthesized Exemplars

Song-Hyang Moon, Sang-Woong Lee, and Seong-Whan Lee  
Department of Computer Science and Engineering, Korea University

#### 요약

최근 생체 인식에 대한 관심이 고조됨에 따라, 다양한 환경 변화에 강한 얼굴 인식 방법들이 연구되고 있다. 특히, 조명 변화를 처리하기 위한 연구들이 세계적으로 발표되고 있다. 그러나 기존 방법들은 다수의 등록 영상이나 조명에 대한 사전 지식이 필요하다는 제약조건을 가지고 있다. 본 논문에서는 기존 방법의 한계점을 해결하기 위해 조명 대표영상의 선형 분석을 이용한 새로운 방법론을 제안하였다. 또한 제안 방법의 효율성을 입증하기 위하여 공인된 얼굴 데이터베이스를 이용하여 다양한 실험을 시도하였으며, 이를 통해 제안된 방법이 기존의 다른 방법에 비하여 안정적인 인식 성능을 보이는 것을 확인할 수 있었다.

#### 1. 서론

일반적으로 동일한 사람일 경우에도 조명, 포즈, 배경 등의 변화에 따라 다른 얼굴 영상들로 나타나기 때문에 얼굴 인식의 성능이 크게 저하된다. 특히 조명 변화에 따른 영상의 변화는 얼굴 인식 성능 저하의 가장 큰 문제점이라고 할 수 있다. 따라서 이러한 조명 변화에서도 강한 얼굴 인식에 대한 연구가 진행되어 왔다.

얼굴 인식 방법으로 가장 널리 알려진 Eigenface 방법에서는 조명 문제를 해결하기 위하여 최상위 3개의 주성분을 제외하고 인식을 수행한다. 이 방법은 조명 변화에 해당하는 특징도 제거되지만 얼굴 고유 특징도 제거되기 때문에 인식을 성능이 저하되는 단점을 가지고 있다[1].

최근의 조명 변화 문제를 처리하는 대표적인 얼굴 인식 방법에는 다른 조명을 가지는 다수의 영상들로부터 반사도(reflectance)를 추정하여 임의의 조명 영상을 합성하는 photometric stereo 방법[2][3]이 있다. 이 방법들은 합성된 여러 조명 영상들의 부분공간(subspace)으로 입력 영상을 투영시켜 조명 변화에 무관한 인식을 수행한다. 그러나 임의의 조명 영상을 합성하기 위해 3장의 등록 영상이 필요한 한계점이 존재한다.

또 다른 방법으로는 한 장의 등록 영상으로도 인식이 가능한 학습 기반 선형 분석이 있다. 이러한 접근방법에서는 입력 영상을 얼굴이 가지는 고유 특징과 조명에 의해 변하는 특징으로 분리한다. 학습 데이터를 구성하는 객체를 3장 이상의 조명으로 구성하고 이중선형 분석(bilinear analysis)을 통해 이 영상들로 입력 영상을 표현한다[4][5]. 하지만 객체가 가지는 조명 집합이 모두 동일해야 하는 제약 조건이 따르고 얼굴이 가지는 고유 특성과 조명 특성을 분리하기가 쉽지 않다는 한계점을 가지고 있다. Blanz와 Vetter는 3차원 모델 정보를 이용하여 학습 기반 선형 분석을 수행한다[6]. 3차원 정보를 사용하기 때문에 포즈 및 조명 변화에 강인하지만 3차원 얼굴 형태와 질감에 대해 반복적으로 조명 및 포즈 파라미터를 추정하는 과정을 필요로 하기 때문에 속도가 상당히 느리다는 단점이 따른다.

#### 2. 합성 대표영상의 선형 분석을 통한 얼굴 인식

앞에서 언급했듯이 조명 문제를 해결하는 많은 얼굴 인식 방법에서 제기되고 있는 한계들을 극복하려면 조명 정보를 모르는 학습 영상들과 한 장의 입력 영상에서도 인식이 가능하여야 한다. 따라서 1장에서 언급한 기존 방법들의 한계를 극복하고 조명 변화에 안정된 성능을 보이는 방법을 제안하고자 한다.

본 논문에서는, photometric stereo 방법을 토대로 학습 데이터의 임의의 조명 영상들로부터 대표영상(Exemplar)을 합성한 후 합성된 대표영상들의 선형 분석을 통해 입력 영상을 표현한다 이때의 선형 계수를 이용하여 인식을 수행하게 된다. 그림 1은 임의의 조명 영상이 들어왔을 때 인식되는 과정을 보여준다.

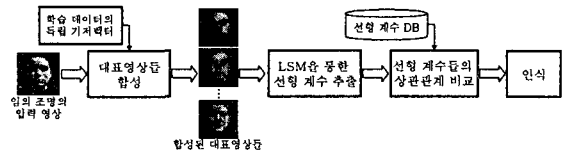


그림 1. 임의의 조명 영상의 인식 과정

##### 2.1 조명 대표영상 합성

사람의 얼굴이 Lambertian surface를 가지는 illumination cone이라고 가정하면, 임의의 조명하에 있는 얼굴 영상은 다른 조명 영상들의 선형조합으로 표현될 수 있다. 그때의 영상들을 photometric stereo 영상들이라 하고, 최소 3장 이상의 다른 조명하의 영상들로 구성되어 있다[2][3]. 임의의 조명을 가지는 영상은 다음과 같이 나타낼 수 있다.

$$I_{d \times 1} = \rho I_n = I_t \quad (1)$$

여기서  $t$ 는 albedo,  $\rho$ 와 surface normal,  $n$ 의 곱을 의미하며  $I$ 은 조명의 방향을 의미하고  $d$ 는 픽셀 수이다. 한 객체에 대해 다른

조명이 주어지면, 같은 위치의 각 픽셀은 조명의 영향을 받아 다른 밝기 값을 가지게 된다. 이 밝기의 차이는 조명 때문에 발생하는 특징이라고 볼 수 있다. 한 객체의 같은 포즈를 가지고 다른 조명을 가지는 3장의 영상 벡터를 표현하면 다음과 같다.

$$I_{d \times 3} = \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix}^T = \begin{bmatrix} l_1 t \\ l_2 t \\ l_3 t \end{bmatrix}^T = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix}^T t = Lt \quad (2)$$

만약  $L$ 의 조명 벡터,  $l_1, l_2, l_3$  등이 서로 독립적이라면 이들 벡터의 역을 이용하여  $t$ 를 구할 수 있다. 여기서  $t$ 는 조명 성분이 배제된 객체의 고유 특징이 된다. 즉, 이 객체 고유 특징에 특정 조명에 해당하는 조명 벡터가 곱해지면 입의 조명 영상을 합성할 수 있는 것이다.  $t$ 는 식 (3)에 의해 구해진다.

$$t = L^+ I \quad (3)$$

여기서  $L^+$ 은  $L$ 의 의사역행렬(pseudo inverse)이다. 식 (3)은 기본적인 photometric stereo 방법을 나타내고 있으며 세 조명 벡터,  $l_1, l_2, l_3$  등이 독립적이라는 것으로 직교하는 벡터들임을 의미한다. 직교하는 벡터들의 선형 조합은 임의의 조명 영상을 합성할 수 있다. 즉, 각 벡터와 곱해지는 선형 계수,  $a_i$ 에 따라 어떤 조명 하의 영상이 되는지가 결정된다. 우리는 입력 영상을 참조로 하여 얻어지는 계수를 이용하여 입력 영상 조명에 해당하는 조명 대표영상,  $e$ 를 합성한다. 입력 영상,  $h$ 를 참조로 LSM(Least Square Minimization)방법을 사용하여 각 벡터에 곱해질 계수 값을 계산하게 된다.

$$a^* = \arg \min_a \left\| h - \sum_{i=1}^3 a_i l_i \right\| \quad (4)$$

여기서 얻은  $a^*$ 를 사용하여 학습 영상들로부터 합성되는 대표영상은 다음과 같이 표현된다.

$$e = a^* I \quad (5)$$

그림 2는 입의 조명 영상들로부터 합성된 대표영상의 예들을 보여준다. 그림에서 볼 수 있듯이 입력영상을 참조 하기 때문에, 대표영상들은 입력 영상의 조명 방향에 근사한 조명 영상으로 합성된다.

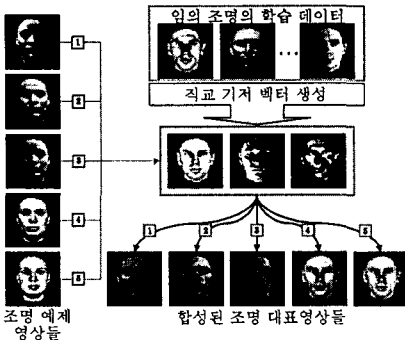


그림 2. 입력 영상의 조명에 근사한 대표영상 합성의 예

입력 영상이 들어오면 해당 조명의  $m$ 개의 대표영상들이 생성되고, 이 대표 영상들의 선형 조합으로 입력 영상을 나타낼 수 있게 된다.

### 2.2 조명 대표영상들의 선형 분석

학습 데이터  $m$ 명의 photometric stereo 영상들로부터 대표영상

들이 합성된다. 이  $m$ 개의 합성 대표영상들의 선형 조합으로 입력 영상을 표현하고 이때의 선형 계수를 인식에 사용한다. 합성 대표영상들로부터 입력 영상을 표현 할 수 있는 선형 계수를 구하는 식은 다음과 같다.

$$h = \sum_{k=1}^m \mu_k e_k = \mu E \quad (6)$$

$\mu$ 는 각 대표영상들로부터 가져오는 계수들,  $(\mu_1, \mu_2, \dots, \mu_m)$  이고 조명 대표영상들의 벡터,  $E$ 는  $(e_1, e_2, \dots, e_m)$  들로 나타낼 수 있다. 입력 영상을 학습 데이터로부터 구해지는 대표영상들로 표현하므로 한 장의 등록 영상으로도 인식이 가능하며, 여기서 구해지는 선형계수들은 입력 영상의 조명에 근사한 조명으로 합성된 대표영상들로부터 추출되므로 조명에 무관한 얼굴 특징이 될 수 있다. 합성된 대표 영상들의 선형 조합을 통해 입력 영상을 복원한 결과는 그림 3과 같고, 조명 특징뿐 아니라 객체의 고유 특징도 유사하게 복원된 것을 확인할 수 있다.



그림 3. 입력 영상들의 복원 결과 예

### 2.3 인식

합성 대표영상들로부터 추출된 선형 계수들의 상관관계를 통해 인식을 수행한다. 등록 영상을 나타내는 선형 계수  $\mu_g$ 와 인식 영상을 나타내는 선형 계수  $\mu_p$ 의 상관관계를 구하는 식은 다음과 같다.

$$corr(g, p) = \frac{cov(\mu_g, \mu_p)}{sd(\mu_g)sd(\mu_p)} \quad (7)$$

여기서  $sd(a)$ 는  $a$ 의 표준편차(standard deviation)를 의미하고,  $Cov(a, b)$ 는  $a$ 와  $b$ 의 공분산(covariance)을 의미한다.

선형 계수로 인식을 수행하는 원리는 그림 4과 같다. 같은 객체에 대해 조명이 다른 영상들이 들어오면 해당 조명의 대표영상들로부터 선형 계수가 추출될 수 있다. 이 선형 계수가 입력 영상을 나타내는 고유 특징으로 사용되어 DB에 등록되어 있는 다른 영상들의 선형 계수와 상관관계를 계산하여 가장 유사한 상관관계를 가지는 영상을 인식하게 된다.

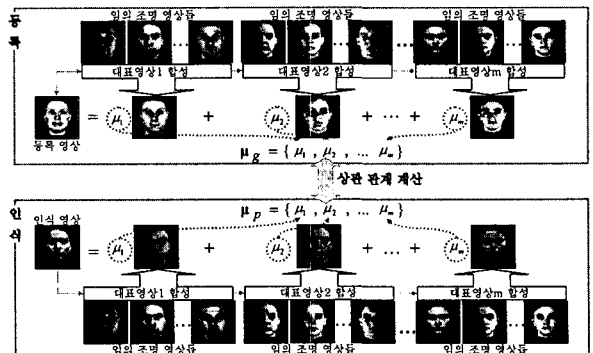


그림 4. 합성 대표영상들을 이용한 얼굴 인식

### 3. 실험 및 결과 분석

본 논문에서 제안한 방법의 얼굴 인식 성능을 검증하기 위하여 Eigenface WO3 방법 [1]과 이중선형 분석을 통한 선형 조합 방법 [5], 그리고 제안한 방법을 비교, 분석하였다. 여기서 Eigenface WO3 방법이란 조명 문제를 해결하기 위해 최상위 주성분 3개를 제거한 방법을 의미한다.

#### 3.1 실험 데이터

제안된 방법의 실험을 위해서 MPI DB(Max-Planck-Institute Database)를 사용하였다. MPI DB의 영상들은 그림 5에서와 같이 30°씩 변하는  $\theta$ 와  $\phi$ 의 집합,  $(\theta, \phi)$ 에 의해 25가지 방향 조명으로 표현 되고, 제안 방법의 성능 검증을 위해 조명 방향마다의 인식을 실험을 하였다. 각 데이터베이스에서 학습 영상으로 100명, 한 사람당 3장의 영상으로 300장의 영상을 사용하였으며 등록 및 인식 영상은 학습 영상에 포함되지 않은 영상으로 각 조명 방향에 대해 100명의 영상들을 사용하였다. 이때 학습 데이터를 구성하는 3장의 조명 영상은 임의의 방향의 조명 영상을 사용하였으며 사전에 조명 정보를 가지고 있지 않는다.

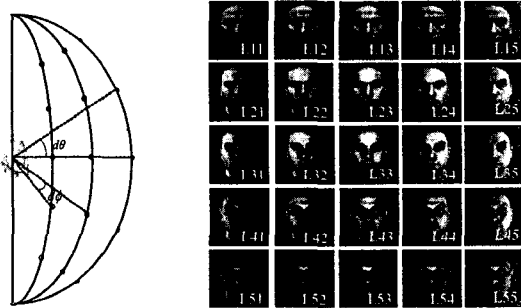


그림 5. 25가지의 조명 방향에 대한 영상의 예

#### 3.2 인식률 비교

그림 6은 등록 영상의 조명 방향을 고정하고 인식 영상들의 여러 조명 방향에 대한 인식 실험 결과이다. 등록 영상은 11번 조명(L11) 방향의 100장 영상을 사용하였으며 그 외 24방향 조명에 대해 인식률 실험을 수행하였다.

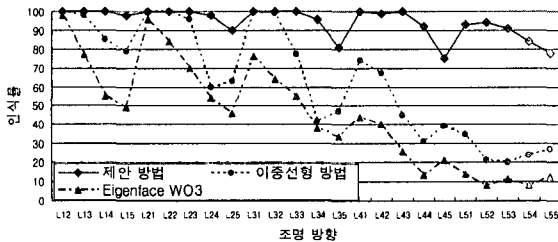


그림 6. 한 조명 방향의 등록 영상에 대한 인식률

그림에서 보여지듯이 기존 방법들은 등록 영상과 인식 영상의 조명 방향이 멀어질수록 인식률이 저하됨을 볼 수 있다. 반면 본 논문에서 제안한 방법은 등록 영상과 인식 영상의 조명 방향이 멀어지더라도 인식률 저하의 폭이 크지 않음을 알 수 있다.

좌우 5방향 조명에 대한 인식 실험 결과는 표 1과 같다. 왼쪽이 제안한 방법의 인식률이고 오른쪽이 이중 선형 분석 방법의 인식률이다. 여기서 G는 등록 영상들을 의미하고 P는 인식 영상들을 의미한다. 다양한 조명 변화에도 불구하고 평균 인식률이 97%로 안정된 성능이 유지되는 것을 확인할 수 있었다.

표 1. 좌우 5방향 조명에 대한 영상들의 인식률

$\begin{matrix} P \\ \backslash \\ G \end{matrix}$	L31	L32	L33	L34	L35	평균
L31	—	100/100	100/88	94/91	76/78	94/91.4
L32	100/100	—	100/98	100/88	97/82	99.4/93.6
L33	100/100	100/100	—	100/100	100/99	100/99.8
L34	94/84	100/84	100/98	—	100/100	98.8/93.2
L35	79/75	96/89	100/92	100/100	—	95/91.2
평균	94.6/91.8	99.2/94.6	100/95.2	98.8/95.8	94.6/91.8	97.4/93.8

#### 4. 결론 및 향후 과제

본 논문에서는 photometric stereo 영상들로부터 조명 대표영상들을 합성하고 대표영상들의 선형 분석을 통해 추출된 선형 계수를 가지고 인식을 수행하는 방법론을 제안하였다. 적은 데이터베이스를 가지고 한 장의 영상으로도 등록이 가능하고 학습 데이터들의 조명 정보 없이도 인식이 가능하다는 점에서 기존 연구와 차별성을 갖는다. 본 논문에서는 조명 변화로 인한 문제점을 해결함으로써 실제 얼굴 인식 시스템, 보안 시스템 등에 유용하게 사용될 수 있는 가능성을 보여주었다.

향후 연구로는 대표영상을 합성할 때에 각 영상의 형태 정규화 알고리즘이 필요하고 MPI DB 뿐 아니라 여러 DB의 다양한 조명 영상을 사용한 실험 및 분석이 따라야 한다.

#### 감사의 말씀

본 연구에 사용된 얼굴 영상 데이터베이스는 독일 Max-Planck-Institute로부터 제공 받았음.

#### 참고 문헌

- [1] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 84-91, June 1994.
- [2] R. Basri and D. Jacobs, "Photometric Stereo with General, Unknown Lighting," Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Vol. 2, pp. 374-381, 2001.
- [3] A. Georghiadis, P. Belhumeur, and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 23, pp. 643-660, 2001.
- [4] A. Shashua and T. Raviv, "The Quotient Image: Class Based Rendering and Recognition with Varying Illuminations," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 23, pp. 129-139, 2001.
- [5] S. Zhou and R. Chellappa, "Rank Constrained Recognition under Unknown Illuminations," Proc. of IEEE Intl. Workshop on Analysis and Modeling of Faces and Gestures, 2003.
- [6] V. Blanz and T. Vetter, "Face Recognition based on Fitting a 3D Morphable Model," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 25, pp. 1063-1074, 2003.