

## 서포트 벡터 데이터 서술을 이용한 대표 얼굴 영상 합성

이상웅<sup>0,†</sup> 박주영\* 이성환<sup>†</sup>

<sup>†</sup>고려대학교 정보통신대학 컴퓨터학과

\*고려대학교 과학기술대학 제어계측공학과

sangwlee@image.korea.ac.kr<sup>0</sup>, parkj@korea.ac.kr, swlee@image.korea.ac.kr

### Synthesis of Face Exemplars using Support Vector Data Description

Sang-Woong Lee<sup>0,†</sup> Jooyoung Park\* Seong-Whan Lee<sup>†</sup>

<sup>†</sup>Department of Computer Science and Engineering, Korea University

\*Department of Control and Instrumentation Engineering, Korea University

#### 요 약

최근 얼굴 인식은 사용자의 편의성을 포함한 다양한 장점으로 인하여 생체 인식 시장에서 주요 기술로 대두되고 있다. 그러나 조명 변화에 기인한 얼굴 인식 성능의 저하는 실용화에 걸림돌이 되고 있는 실정이다. 따라서 조명 변화에 따른 얼굴의 외형 변화를 분석하는 연구들이 세계적으로 활발히 진행되고 있다. 그러나 기존 방법들은 다수의 등록 영상이나 조명에 대한 사전 정보가 필요하거나 실시간으로 구현되기 어렵기 때문에 실용 시스템에 적용하기는 어려운 실정이다. 따라서, 본 논문에서는, 여러 조명 영상들로 구성된 학습 데이터를 이용하여, 조명에 대한 정보가 없는 한 장의 입력 영상을 분석하는 방법을 제안한다. 제안된 방법은 SVDD를 이용하여 학습 데이터의 여러 조명 영상들로부터 입력 영상의 조명과 같은 대표영상을 합성하고 이 대표영상들의 선형 조합을 이용하여 입력 영상을 표현한다. 제안 방법의 효율성을 검증하기 위하여 공인 얼굴 데이터베이스들을 이용하여, 기존 방법들과 비교 실험을 수행하였으며, 조명 변화가 큰 영상에서도 안정된 조명 변화의 분석이 가능하였다.

#### 1. 서 론

얼굴 인식 분야에서 조명 변화에 기인한 얼굴의 변화는 인식 성능에 많은 영향을 주고 있다. 얼굴 인식기에 개인의 얼굴을 등록하는 시점과 인식하는 시점에서의 조명 변화가 다르기 때문에 발생하는 얼굴의 차이는 얼굴 인식의 정확도를 낮게 하기 때문이다.

따라서 세계적으로 이러한 조명 변화가 얼굴에 미치는 영향을 분석하는 연구가 진행되어 왔다. 얼굴 인식 방법으로 가장 널리 알려진 Eigenface 방법에서는 조명 변화가 얼굴 특징들 보다 우선하기 때문에, 최상위 3개의 주성분을 제외하고 얼굴 영상을 분석할 것을 제안한다. 이 방법은 조명 변화에 해당하는 특징도 제거되지만 얼굴 고유 특징도 제거되기 때문에 인식 성능이 저하되는 단점을 가지고 있다[1]. 이와는 달리, 최근에는 조명 변화 문제를 처리하기 위해 얼굴 인식 시점에서의 입력 영상을 분석하여 등록 시점의 조명 변화를 가지는 얼굴 영상을 합성하는 방법을 시도하고 있다[2,3,4,5].

이 방법들은 얼굴이 램버시안 표면(Lambertian surface)을 가진다고 가정하고, 얼굴 표면의 반사도, 반사 계수, 조명 정보들을 각기 분리하여 추정한다. 이는 미지의 3가지 조건을 추정하는 문제로서, 일반적으로 반사 계수는 상수로서 가정하고 있다. 이 중에서 측광 스테레오(photometric stereo) 방법[2][3]은 조명에 대한 근본적인 분석을 통하여 신뢰할 수 있는 분석이 가능하다. 임의의 조명 영상을 합성하기 위해 3장 이상의 등록 영상이 필요한 한계점이 존재한다. 최근에 한 장의 등록 영상으로도 인식이 가능한 학습 기반 선형 분석 방법

[4][5]이 발표되었으며, 이러한 접근 방법에서는 학습 데이터를 구성하는 객체를 3장 이상의 조명으로 구성하고, 이를 이용하여 1장의 등록 영상을 이중선형 분석(bilinear analysis)을 통해 분석한다. 하지만 객체가 가지는 조명 집합이 모두 동일해야 하는 제약 조건이 따르고 얼굴이 가지는 고유 특성과 조명 특성을 분리하기가 쉽지 않다는 한계점을 가지고 있다.

#### 2. 대표 영상을 이용한 얼굴 인식

앞에서 기술된 문제점을 해결하기 위해, 대표 영상 합성에 기반한 새로운 얼굴 인식 방법이 제안되었다[6]. 이 방법은 조명을 알 수 없는 학습 영상들을 이용하여 한 장의 새로운 입력 영상을 등록하고 인식할 수 있는 방법이다. 이 과정은 그림 1에 표현되었으며, 주성분 분석 방법을 이용하여 얼굴들의 기저벡터를 추출한 후, 측

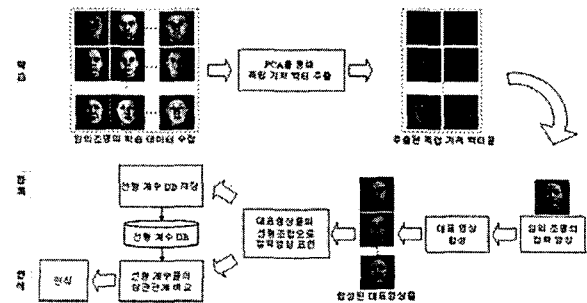


그림 1. 대표 얼굴 영상을 이용한 얼굴 인식 방법

광 스테레오(photometric stereo)방법을 사용하여 새로운 조명 환경의 대표 얼굴 영상을 합성한다.

미리 저장되어 있는 개인 마다 이러한 대표 얼굴 영상을 합성한 후, 각 대표 얼굴 영상의 선형 조합으로 입력 영상과 최대한 유사한 얼굴을 합성한다. 이 과정에서 추출된 선형 계수들은 입력 영상의 고유한 특징으로 사용되며, 등록과 인식 과정에서 얼굴 인식의 유사도 측정에 이용된다.

이 논문에서는 새로운 조명 환경하의 얼굴 영상으로 합성하기 위하여, 측광 스테레오 방법을 도입하였으며, 이는 필수적으로 영상 벡터들이 직교(orthogonal)해야 한다는 제약 조건을 가지고 있다. 이러한 제약 조건을 만족시키기 위하여, 주성분 분석을 이용하여 학습 영상들의 기저벡터들을 생성하였으며, 이를 측광 스테레오에 이용하였다. 이러한 과정에서 주성분 분석 방법과 최소자승 오차 방법 등의 사용으로 복잡한 단계를 거치게 되며, 또한 선형 분석 방법의 한계성을 가지는 단점이 발견되었다.

본 논문에서는 이러한 선형 분석 방법의 한계성을 극복하고, 얼굴 인식 단계를 간소화할 수 있는 새로운 대표 영상 얼굴 합성 방법을 제안한다.

### 3. SVDD 기반 대표 영상 합성 방법

#### 3.1 서포트 벡터 데이터 서술

본 논문에서는 서포트 벡터 데이터 서술(Support Vector Data Description: SVDD)을 이용하여, 대표 영상을 합성하는 방법을 제안한다. 측광 스테레오 방법은 학습 데이터의 존재 영역을 선형 부분 공간에 투영하여 표시하지만, SVDD는 학습 데이터의 존재 영역을 볼(Ball)을 이용하여 직접 근사하는 방법을 사용한다. SVDD는 특징 공간에서 최대한 많은 학습 데이터를 포함하는 최소 크기의 볼을 찾는 것을 목적으로 한다. 중심이  $a$ 이며, 반지름이  $R$ 인 볼의 크기는 슬랙 변수(slack variable)에 의해 조절되며, 최적의 볼의 크기를 구하기 위하여 식 (1)이 사용된다.

$$\min L_0(R^2, \alpha, \xi) = R^2 + C \sum_{i=1}^N \xi_i \quad (1)$$

$$s. t. \|x_i - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \quad i=1, \dots, N$$

여기서, 상수  $C$ (trade-off constant)는 각 항의 중요도를 조절한다.

일반적으로 입력 공간보다 주어진 입력 영상을 더 잘 표현할 수 있는 특징 공간(Feature space)으로 변환한다면 SVDD는 더욱 좋은 성능을 발휘한다. 이를 위해 커널을 적용하고, 얻어진 2차 프로그래밍(Quadratic Programming)을 풀기 위해 라그랑지(Lagrangian) 방법을 적용하여, 새로운 라그랑지 계수 알파에 대하여 식을 전개하면 식 (2)를 얻는다[7].

얻어진 식 (2)를 수치해석에 의해 풀이하면 우리는 원하는 구의 반지름  $R$ 을 얻게 된다.

$$\min_{\alpha} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i K(x_i, x_i) \quad (2)$$

$$s. t. \sum_{i=1}^N \alpha_i = 1, \quad \alpha_i \in [0, c], \quad i=1, \dots, N$$

#### 3.2 대표 영상 합성 방법

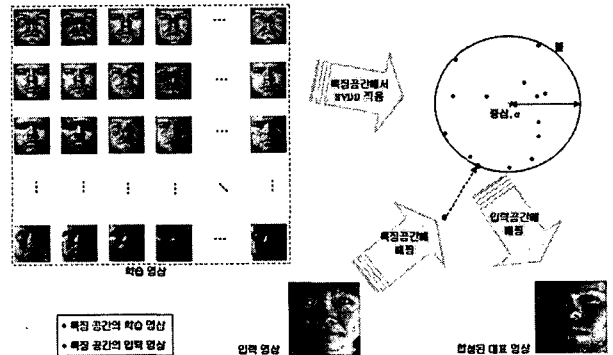


그림 2. SVDD를 이용한 대표 영상 합성 방법

그림 2에서 보여 지는 한 사람에 대한 다양한 조명 변화 얼굴 영상을 커널에 의한 특징 공간으로 이동하여 SVDD를 이용한 학습하면, 이 영상들을 최대한 감싸는 볼의 형성되며, 이 볼의 내부는 같은 사람의 영상으로 간주된다. 한편 볼 밖의 데이터는 다른 사람의 영상으로 간주된다.

또한 식 (2)에서 SVDD의 Mercer 커널을 식 (3)과 같은 가우시안 커널을 사용한다면, SVDD에 의해 학습되어 특징 공간에 옮겨진 데이터들은 단위구의 표면에 존재하게 된다.

$$K(x, y) = \exp(-\|x - y\|^2 / \delta^2) \quad (3)$$

만약 학습된 영상과 다른 사람의 얼굴 영상을 특징 공간으로 이동한다면, 대부분의 경우는 볼 바깥에 위치할 것이다. 특징 공간상의 이 얼굴 벡터를 볼 표면으로 사영한다면, 이는 볼 바깥의 얼굴 벡터가 볼 내부의 얼굴 중에 유사한 얼굴 쪽으로 변형된다는 것을 의미한다.

본 논문에서는 볼 내부의 학습 데이터들은 동일한 외형을 가지기 때문에, 유사한 얼굴로의 변형이라는 것은 유사한 조명 변화로의 변형을 의미하게 된다.

볼 표면에 사영된 특징 벡터를 주변에 근접해 있는 학습 데이터들을 이용하여, 유추하고 입력 공간으로 재이동을 하면, 그림 4와 같이 입력 영상의 조명 환경과 유사한 조명 영향을 가진 볼 내부의 사람의 영상을 얻을 수 있다.

본 논문에서는 특징 공간상의 볼 표면에 사영된 벡터를 유추하고, 입력공간에 재 이동하는 경우에 생성되는 pre-image를 J. T. Kwok에 의해 제안된 방법[8]으로 생성하였다.

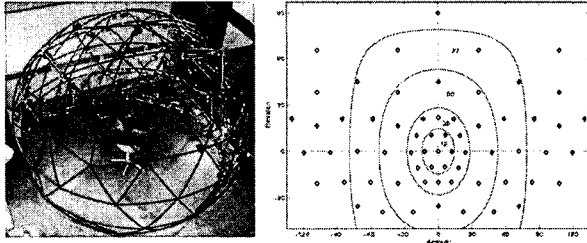


그림 3. Yale Database B의 조명 환경



그림 4. 대표 영상 합성의 예

#### 4. 실험 및 결과

제안된 방법의 실험을 위해서 Yale Database B 데이터를 사용하였다[9].

Yale Database B의 영상들은 10명에 대한 그림 3에서와 같이 돔 형태의 구조물에 64개의 스트로보를 장착하여 저장되었다. 또한 포즈 변화를 위하여 9개의 카메라를 이용하였으며, 전체 조명과 포즈의 조합으로 이루어져 있다.

본 논문에서는 제안 방법의 성능 검증을 위해, 한 명당 각기 다른 조명을 가지는 64장의 정면 얼굴 영상들을 사용하였으며, 입력 영상을 랜덤하게 추출하여 10회 실험하였으며, 이 때, SVDD에는 가우시안 커널을 사용하였으며, pre-image 유추를 위한 근점 벡터의 경우 3개를 사용하였다.

실험데이터의 경우 조명에 대한 정보를 미리 가지고 있으며, 같은 조명하의 정답 영상을 알고 있으므로, 픽셀 간의 평균 오차를 계산하여 정답 영상과 가장 근접한 결과를 얻었는지를 알 수 있다.

	정답률	평균 픽셀 오차
측광 스테레오 방법	89%	10.76
제안 방법	94%	6.73

표 1. 합성된 대표 영상의 복원 실험 결과

표 1은 제안된 방법과 기존 측광 스테레오를 이용한 선형 분석 방법의 비교 결과를 보여주고 있다. 제안된 방법은 기존 방법에 비해 좋은 성능을 보였으며, 평균 픽셀 오차 역시 상대적으로 낮은 수치를 보여주었다. 또한 SVDD를 이용하여 복원된 결과 영상을 그림 4에서 보여주고 있다. 입력 영상의 조명 방향을 참조하여, 합성된 대표 영상은 정답 영상과 매우 유사한 결과를 보였다. 이러한 결과는 대표 영상을 이용한 얼굴 인식에서도 좋은 성능을 보여줄 것으로 예측된다.

본 논문에서는 조명에 대한 사전 정보가 없는 얼굴 영상들을 사용하여, 임의의 조명 환경하의 얼굴 영상을 합성하고 표현하는 방법을 제안하였다.

제안된 방법은 조명을 알 수 없는 데이터들을 SVDD를 이용하여 학습하여, 볼 표면으로의 사영된 입력 영상 벡터를 입력 공간으로 재 이동하여 대표 영상을 합성해낸다. SVDD를 사용한 대표 영상 방법은 기존 방법의 선형

적 한계성을 극복하고, 얼굴의 비선형적인 조명 변화를 잘 반영할 수 있으며, 알 수 없는 조명 정보를 이용하여 간단한 절차를 통하여, 입력 영상과 유사한 얼굴을 얻어 낼 수 있었다.

#### 참고문헌

- [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 Re-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, pp. 11-18, 2003.
- [6] S.-H. Moon, S.-W. Lee, and S.-W. Lee, "Illumination Invariant Face Recognition Based on the Synthesized Exemplars," Audio- and Video-based Biometric Person Authentication, Lecture Notes in Computer Science, Vol. 3546, pp. 112-121, 2005.
- [7] D. Tax and R. Duin, "Support Vector Domain Description," Pattern Recognition Letter, Vol. 20, p. 1191-1199, 1999.
- [8] J. Park, D. Kang, J. Kim, J. T. Kwok and I. W. Tsang, "Pattern De-Noising Based on Support Vector Data Description," Proc. of Intl. Joint Conf. on Neural Networks, Montreal, Canada, pp. 949-953, 2005
- [9] K.-C. Lee, J. Ho and D. Kriegman, "Acquiring Linear Subspaces for Face Recognition under Variable Lighting," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 27, No. 5, pp. 684-698, 2005.