

가상의 학습데이터 생성 및 Pix2pix GAN 학습을 이용한 얼굴 그림자 제거

이수현¹ · 최영우^{2*}

¹숙명여자대학교 컴퓨터과학과 학사

²숙명여자대학교 컴퓨터과학과 교수

Face Shadow Removal by Training Pix2pix GAN with Generated Virtual Training Data

Soohyeon Lee¹ · Yeongwoo Choi^{2*}

¹Student, Department of Computer Science, Sookmyung Women's University, Seoul, Korea

²Professor, Department of Computer Science, Sookmyung Women's University, Seoul, Korea

[요약]

다양한 조명 변화는 얼굴 이미지의 인식 성능을 떨어뜨리는 요인 중 하나이며, 특히 조명 또는 주변 환경에 의해 얼굴 이미지 위에 그림자가 생기면 인식 성능이 크게 저하되는 것이 일반적인 경향이다. 따라서 그림자가 발생한 얼굴 이미지를 원래의 상태로 복원할 수 있으면 얼굴 인식 성능의 향상을 기대할 수 있다. 이 연구에서는 적대적 생성 신경망인 GAN의 대표적 모델 중 하나인 Pix2pix를 사용하여 그림자를 완화하고 제거하는 방법을 제안한다. Pix2pix GAN 모델은 학습에 대응되는 이미지 쌍(Pair)을 필요로 하기 때문에 이를 위해 정상적인 얼굴 이미지로부터 그에 대응되는 가상의 학습 이미지를 다양한 이미지 블렌딩 방식으로 생성하여 학습 쌍으로 활용하는 아이디어를 제안한다. 제안한 방법으로 생성한 데이터를 이용해서 학습한 모델을 테스트한 결과 얼굴 이미지의 그림자가 자연스럽게 완화되는 것과 이를 통해 얼굴 인식 성능이 향상된 것을 확인할 수 있다.

[Abstract]

Various lighting changes are one of the factors that degrade the recognition performance of face images, in particular, when a shadow is formed on a face image due to lighting or surrounding environment, it is a general tendency that recognition performance is greatly degraded. Therefore, if the face image in which the shadow has occurred can be restored to its original state, improvement in face recognition performance can be expected. In this study, we propose a method of mitigating and removing shadows using Pix2pix, one of the representative models of GAN, a hostile generating neural network. Since the Pix2pix GAN model requires a pair of images corresponding to training, for this, we propose an idea to create a virtual training image corresponding to it from a normal face image using various image blending methods and use it as a training pair. Results of testing the model trained using the data generated by the proposed method, it can be seen that the shadows of the face image are naturally reduced and that the facial recognition performance is improved.

색인어 : 얼굴 그림자 제거, Pix2pix GAN, 가상 학습데이터, 이미지 전처리 등

Key word : Face shadow removal, Pix2pix GAN, Virtual training data, Image preprocessing, etc

<http://dx.doi.org/10.9728/dcs.2021.22.2.347>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 31 December 2020; **Revised** 01 February 2021

Accepted 01 February 2021

***Corresponding Author; Yeongwoo Choi**

Tel: [REDACTED]

E-mail: ywchoi@sookmyung.ac.kr

1. 서론

다양한 조명 및 주변 환경에 따라 사람 얼굴에 그림자가 발생하며 이로부터 얻은 얼굴 이미지를 인식하는 것은 그림자로 인해서 인식 성능이 일반적으로 크게 저하된다. 이는 UHDB31 데이터 셋을 사용한 Le[1]의 연구 결과에서도 확인할 수 있다. 그림 1은 Yale 데이터에서[2] 다양한 방향 및 조명 강도에 따라 그림자가 얼굴에 드리워진 이미지 데이터를 보여 준다.

얼굴 이미지의 그림자를 완화하거나 제거하는 연구로서 영상처리에 기반한 방법[3,4,5]과 신경망 모델을 이용한 방법 [6,7,8] 등으로 나누어 생각할 수 있다. W. Zhang 등은[3] 얼굴 이미지의 조명 정규화 및 반사도 모델 (reflectance model)을 사용하여 그림자를 완화하고자 하였다. Choi 등은[4] 얼굴의 구조적 특징과 영상처리 접근 방식을 이용해 얼굴 위에 드리운 그림자를 완화하는 방법을 제시하였다[4]. X. Zhang 등은[8] 신경망을 사용해 그림자로 인해 만족스럽지 않게 촬영된 얼굴 이미지에서 그림자를 완화하거나 제거하여 자연스러운 이미지로 변환하는 방법을 제안하였다. 또한 일반 카메라 대신 적외선 카메라를 이용하여 조명변화에 강한 영상을 취득하여 활용할 수 있다. 이를 위해 다량의 적외선 얼굴 이미지 취득이 필요하며, [6]는 Cycle GAN 딥러닝 모델을 이용하여 RGB 얼굴 이미지로부터 적외선 얼굴 이미지를 가상으로 생성하여 인식에 활용하는 방법을 제안하였다.

본 연구는 최근 데이터 생성에 활발하게 사용되고 있는 머신러닝 모델인 GAN(Generative Adversarial Network)을[9] 이용한다. 다양한 GAN 모델 중 CycleGAN[10]은 쌍(Pair)이 없는 데이터를 학습 데이터로 삼아서 새로운 데이터를 생성할 수 있으며 활용 범위가 넓다. 반면 본 연구에서는 쌍을 이루고 있는 데이터를 사용하는 모델인 Pix2pix[11]를 이용하여 얼굴의 그림자를 제거 또는 완화하는 방법을 제안한다. Pix2pix 모델은 학습 및 결과의 데이터 쌍이 필요한 생성 모델로서 이 연구에서는 CASIA 얼굴 데이터로부터[12] 다양한 방향 및 농도의 얼굴 그림자를 가상으로 만들어서 이를 학습 데이터로 사용하여 그림자를 제거하는 방법을 제안한다.



그림 1. 조명 변화에 따른 얼굴 이미지
Fig. 1. Face images with various illuminations

CASIA 얼굴 데이터로부터의 그림자가 있는 학습 데이터 생성은 주어진 얼굴 이미지의 우측 반쪽에 이미지 블렌딩 (blending) 방법으로 인위적으로 그림자를 만들어 넣는 것이다. 이를 위해 다양한 농도의 그림자를 일률적으로 얼굴의 우측 반쪽 이미지에 삽입하는 방법과 중심으로부터의 우측으로 거리에 따라 그림자 농도가 점점 진해지는 점진적 방식을 사용한다. 또한 이와 같이 만들어진 얼굴 이미지들을 좌우 플립(flip) 즉 대칭을 통해서 얼굴의 나머지 반쪽에 해당하는 가상의 학습 데이터들도 만들어 함께 사용한다. 제안하는 방법으로 생성한 가상의 학습 데이터를 이용해서 학습한 GAN 모델을 통해서 테스트 이미지의 얼굴에 포함되어 있는 그림자가 효과적으로 완화되거나 제거되는 것을 확인할 수 있었다.

논문의 구성은 2장에서는 관련 연구, 3장은 이 연구에서 제안하는 방법을 기술한다. 4장에서는 학습된 모델의 테스트 결과를 제시하고, 5장에서 결론을 맺는다.

II. 관련 연구

Tsai 등은[13] 도시 환경의 컬러 항공 사진에 대해서 휘도 (luminance)와 색도(chromaticity)와 같은 속성들을 사용해 그림자를 완화하는(de-shadowing) 방법을 제시하였다. 제안한 방법은 그림자를 탐지하기 위해 자동으로 색도 임계값을 설정하며, 이미지의 특성 간 대비를 증가시켜 특정 영역을 탐지하는 영상 분할 기법을 사용하였다. 이 방법을 통해 촬영된 지역의 정보나 광원에 대한 지식 없이도 그림자를 자동으로 완화할 수 있었다. 이 방법에서 자동화와 그림자 탐지 성능은 서로 상충관계에 있지만, 제안한 영상처리 방법으로 항공 사진에서의 그림자 완화 가능성을 보여줬다.

2차원 얼굴 이미지의 인식은 통제하기 어렵고 예측 불가능한 실제세계의 조명 변화에 따라 성능이 크게 좌우된다. 이를 위해 W. Zhang 등은[3] 얼굴 이미지의 조명 정규화 방식을 제안하여 얼굴의 피부 표면과 광원, 카메라 센서 간의 관계를 특징하는 반사도 모델을 구축하였다. 여기서 제안한 조명 처리 방법은 구축한 반사도 모델을 기반으로, 로그 색도 공간(log chromaticity space)에서 조명 환경에 강인한 색도 영상(chromaticity image)을 생성하게 한다. 또한 사실적인 얼굴 이미지의 생성으로 다양한 형태의 그림자를 제거하는 동시에 색상 정보와 개인적 특성을 유지하는 것이 가능하며, 실험을 통해서 다양한 조명에서 얼굴 인식 성능이 향상된 것을 보여줬다.

Hsieh 등은[5] 여러 조명 환경에서 얼굴의 대칭성을 사용해 얼굴의 그림자를 보정하는 얼굴 인식 접근 방법을 제안하였다. 이는 얼굴 이미지를 대칭적으로 분할하여 이미지의 평균을 도출해냄으로써 얼굴을 인식하는 절차를 제안하였으며, 제안한 방법으로 인식 성능이 향상됨을 보여줬다.

Choi 등은[4] 얼굴의 구조적 특징과 영상처리 접근 방법을 사용해서 얼굴 위에 드리워진 그림자를 완화하는 방법을 제안

하였다. 얼굴의 눈, 코 등의 구성 요소들이 만들어 내는 그림자는 빛의 방향에 따라 여러가지 특성을 갖게 되는데, 이러한 특성을 이용해서 그림자를 완화시켰다. 또한 히스토그램 평활화 방법을 적용하여 조명 변화에 둔감한, 인식 성능을 높이는 특징을 추출하여 사용하였다.

X. Zhang 등은[8] 신경망을 이용해서 그림자로 인해 만족스럽지 않게 촬영된 얼굴 이미지에서 그림자를 완화하거나 제거하여 자연스러운 이미지로 변환하는 방법을 제안하였다. 주요 접근 방식에서는 한 쌍의 신경망을 기반으로, 하나는 외부 물체에 의해 발생한 그림자를 제거하고, 다른 하나는 얼굴 구조에 의해 드리워진 그림자를 완화시켰다. 첫 번째 신경망의 학습에는 실세계의 얼굴 이미지 원본과 원본 이미지에 외부 물체의 그림자를 가상으로 합성한 이미지 쌍을 데이터 셋으로 사용하였다. 이를 통해 외부 그림자를 제거하는 학습을 수행하였다. 또한 두 번째 신경망의 학습에는 여러 특수한 조명 환경에서 촬영된 Light Stage 스캔 데이터 셋을 사용해 실세계의 이미지를 완만하게 정규화하는 모델을 제안하고, 실세계에서 촬영된 얼굴 이미지의 그림자와 조명을 사실적이고 자연스럽게 변환시켰다.

Ramaiah 등은[7] 합성곱 신경망인 CNN을 사용해 일정하지 않은 여러 조명 환경에서의 얼굴 인식 접근법을 제시하였다. 특히 왼쪽 얼굴을 이미지의 오른쪽에 비추게끔 변환한 이미지를 학습 데이터로 사용하는 등 얼굴의 대칭 정보를 학습 데이터셋에 반영함으로써 분류기에 추가 정보를 제공하여 제안 방법으로 인식 성능이 향상된 것을 확인하였다.

조명 변화에 민감한 얼굴 인식의 성능 향상을 위해, 근적외선(Near Infrared) 카메라로 촬영한 적외선 영상을 사용할 수 있다. 근적외선 카메라로 촬영한 얼굴 영상은 조명 변화에 둔감하며, 최근에 제안된 머신러닝 모델로서 얼굴 인식 성능을 향상시킬 수 있다. 그러나 다량의 얼굴 근적외선 영상을 취득하는데 어려움이 있어서 이를 해소하기 위해 이소정 등은[6] CycleGAN을 사용해 일반 카메라 영상을 근적외선 영상으로 변환시키는 방법을 제안하였다. 간단한 얼굴 이미지 전처리를 수행한 후에 일반 카메라의 얼굴 이미지들과 근적외선 카메라의 얼굴 이미지들을 서로 무관한 쌍을(Unpaired) 데이터 셋으로 구성하여 학습을 수행한다. 또한 Cycle GAN의 손실함수에 Perceptual 손실 함수를 추가하여 생성된 이미지의 품질을 높이고 인식 성능을 향상시키는 방법을 제안하였다.

본 논문에서는 Isola 등이[11] 제안한 GAN 모델인 Pix2pix를 통해 얼굴 이미지의 그림자를 완화시키는 방법을 제안한다. [3], [4]에서 제안한 방식은 그림자에 가려진 얼굴 이미지의 인식 또는 복원에 주로 3차원 얼굴 모델을 사용했지만, 제안하는 방법은 3차원의 얼굴 모델이 없이도 간단한 영상처리 방식을 이용해 가상의 학습데이터를 생성한 것과 이를 딥러닝 신경망을 사용한 점에서 차이가 있다. 또한 [11]에서 신경망을 통해 그림자를 완화하지만, 실세계 이미지로 가상의 학습 데이터 셋을 구성한 본 연구에서와는 달리 특수한 조명 환경에서 촬영된 데이터 셋인 Light Stage 스캔 데이터 셋을 사용하여 구분된다.

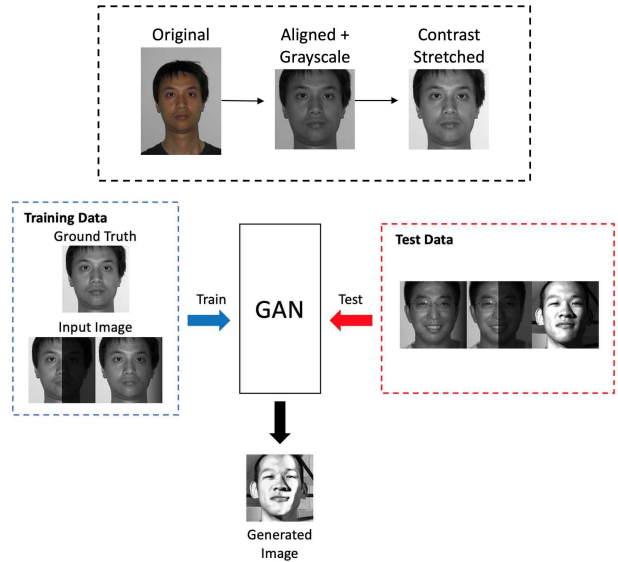


그림 2. 제안 방법 구조도
Fig. 2. Overview of the proposed method

본 논문은 딥러닝 생성 모델을 이용해 얼굴 이미지의 그림자를 제거하는 방법을 제안한다. 그림 2는 제안하는 방법의 전체적인 요약이다. 여기서 이미지 전처리 과정은 먼저 얼굴 이미지의 랜드마크를 추출하여 이를 기반으로 얼굴을 정렬하고, CASIA[12] 얼굴 이미지 데이터가 전반적으로 어둡기 때문에 간단히 Contrast stretching을 수행하여 명암을 보정한다. 다음으로 GAN 모델을 학습하기 위해서 얼굴의 정면 이미지들로 이루어진 CAISA 데이터를 사용해 가상의 학습 데이터 셋을 만든다. 그림 2에서의 학습 데이터는 원본 그레이스케일 얼굴 이미지와 그림자 효과를 주기 위한 이미지 블렌딩을 이용해 가상으로 그림자가 드리워진 얼굴 이미지를 보여 준다. 이와 같이 쌍으로 구성된 데이터를 사용해 Pix2pix GAN 모델을 학습시켰으며, 이 모델로 그림자 완화 및 제거 성능을 검증하였다.

III. 제안하는 방법

본 논문에서 제안하는 방법은 얼굴 이미지 전처리, 그림자 얼굴 학습 데이터 생성, GAN 모델 학습으로 진행된다.

3-1 얼굴 이미지 전처리

그림 3은 주어진 480x640 해상도의 얼굴 이미지에서 랜드마크를 찾고, 눈에 해당하는 랜드마크를 이용해서 먼저 얼굴 이미지를 정렬하고 얼굴을 기준으로 128x128 크기의 해상도로 자른 이미지를 보여 준다. 얼굴 이미지에서 눈, 코, 입과 같은 얼굴의 구조적인 특징 부분들이 어느 정도 일정한 위치에서 입력되면 데이터의 일관성이 생겨 손실함수 계산 시에 이점이 있다.

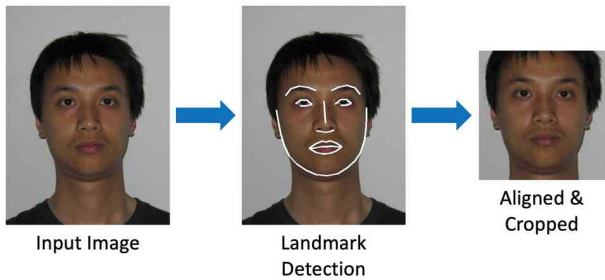


그림 3. Landmark 탐지를 이용한 얼굴 이미지 정렬
 Fig. 3. Face image alignment with landmark detection

얼굴 정렬은 얼굴의 랜드마크에서 찾아진 양쪽 눈의 좌표를 이용하여 눈 높이를 맞추고 얼굴의 비틀어짐을 교정하는 방식으로 진행된다. 마찬가지로 배경을 최대한 제거하고 얼굴 영역만이 포착되도록 찾아진 랜드마크를 이용해서 얼굴의 상, 하, 좌, 우 기준을 설정해서 Cropping한다.

그림 4(A)는 정렬되기 전의 얼굴 이미지이며, 이미지 정렬 및 Cropping을 마친 이미지는 그림 4(B)와 같다. 정렬된 컬러 이미지를 명도(Grayscale) 이미지로 바꾸고 히스토그램 스트레칭을 적용하여 명암 대비를 향상시킨다. 연구에서 사용한 CASIA NIR-VIS 2.0[12] 얼굴 이미지 데이터와 구글의 이미지 검색으로 찾은 얼굴 이미지 등은 명도 이미지로 변환하였을 때 명암 대비가 작아서 전반적으로 어두운 영상으로 변환되는 경향이 있다. 이미지의 가시도 즉 명암대비를 높이기 위해 간단한 히스토그램 스트레칭(contrast stretching)[14]을 적용하였다. 그림 5는 연구에 사용한 이미지에 히스토그램 스트레칭을 적용한 결과를 보여준다. 이 과정을 거친 이미지가 GAN 모델의 입력으로 최종 사용된다.

3-2 그림자 얼굴 학습 데이터 생성

연구에서 사용한 Pix2pix는 입력 이미지와 목표 이미지를 쌍(Pair)으로 이용해 학습한다. 얼굴 영역에 있는 그림자를 완화하기 위해, 입력 영상으로는 얼굴에 그림자가 있는 이미지를 사용하고, 목표 이미지의 경우 얼굴 영역의 그림자가 없는 이미지를 사용해 학습 데이터 쌍을 구성한다. 모델 학습에 사용한 데이터 셋(CASIA NIR-VIS 2.0)의 경우 얼굴 구조나 외부 물체에 의한 뚜렷한 그림자가 이미지에 없기 때문에 그림자가 있는 이미지를 임의로 생성해 가상의 학습 데이터를 구성해 사용한다. 이를 위해 입력 이미지의 우측 영역에는 원본 이미지와 검색된 이미지의 농도를 변화하여 블랜딩(Blending)으로 만든다. 블랜딩은 매 픽셀 일관된 농도를 일률적으로 적용한 그림 6의 데이터 셋과, 매 픽셀 블랜딩 농도를 점진적으로 변화시킨 그림 7의 데이터 셋, 그리고 일률적, 점진적 blending을 적용한 이미지를 통합하고 이를 랜덤하게 좌우로 대칭시킨 flip된 이미지도 포함하는 그림 8과 같은 데이터 셋으로 나누어 각각 학습을 진행하고 비교하였다.

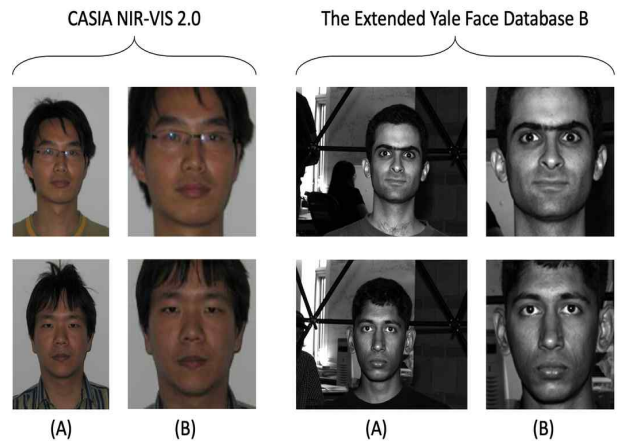


그림 4. (A) CASIA 데이터(좌), Yale B 데이터(우) 원본 이미지; (B) 랜드마크 탐지를 통해 정규화된 이미지
 Fig. 4. (A) Raw image in CASIA data(left), and Yale B data(right); (B) Normalized images by landmark detection

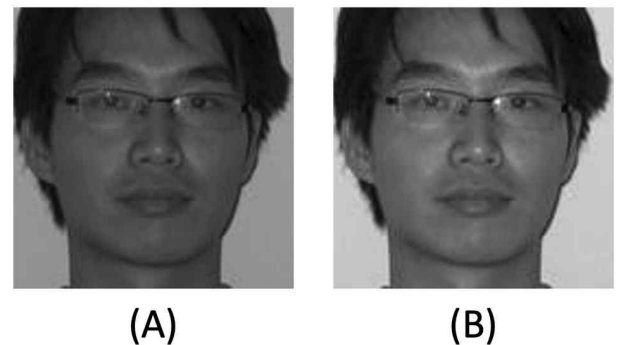


그림 5. (A) CASIA 데이터 그레이스케일 원본 이미지; (B) 히스토그램 스트레칭 적용 결과
 Fig. 5. (A) Grayscale image from the CASIA database; (B) result image by applying histogram stretching

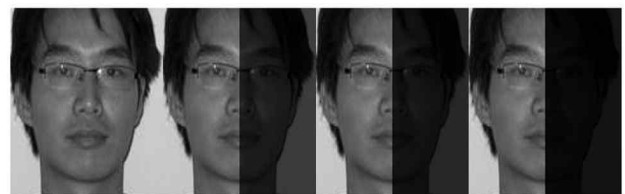


그림 6. 일률적 블랜딩을 이용한 학습 데이터 예
 Fig. 6. Example training images of uniform blending

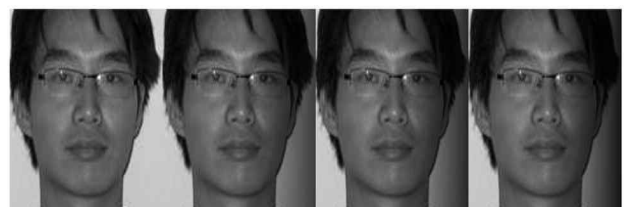


그림 7. 점진적 블랜딩을 이용한 학습 데이터 예
 Fig. 7. Example training images of gradual blending

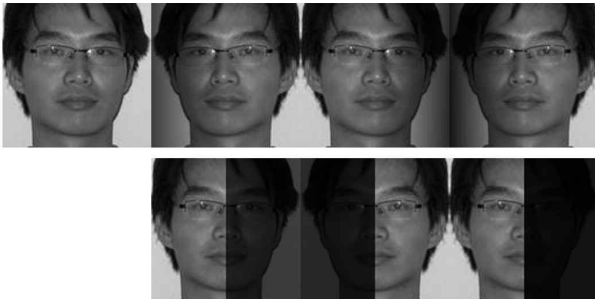


그림 8. 좌우 대칭을 이용해 확장하고 통합한 학습 데이터 예
 Fig. 8. Example training images of randomly flipped, all combined

3-3 이미지 복원 GAN 모델

딥러닝 생성 모델 중 하나인 Pix2pix는 이미지의 입력 및 목표가 쌍(pair)을 이루는 학습 데이터셋을 이용한 이미지 변환에 용이하다. 따라서 이 연구 또한 그림자에 의해 얼굴이 가려진 이미지와 그림자에 가려지지 않은 본래의 얼굴 이미지를 쌍을 Pix2pix에 적용한다. 특히 Pix2pix에서는 이미지 상의 변화가 적은 저주파 영역의 내용을 학습하는 경향이 있는 L1 손실함수와 cGAN(conditional GAN) 손실함수를 추가로 사용하여 더욱 실재와 유사한 이미지를 생성하는 장점이 있다[11]. 따라서 본 연구에서도 아래 그림 9의 L1 손실과 cGAN 손실을 함께 사용해 학습에 사용한다. 본 연구에서 사용한 Generator와 Discriminator는 각각 U-Net 네트워크와 [11]에서 제안한 PatchGAN을 사용하였다.

그림 9는 Pix2pix를 사용하여 모델이 학습되는 과정을 보여주며 이미지 사이의 매핑 관계를 학습하는 것은 아래의 수식을 최적화하는 것과 같다. 여기서 L_{cGAN} 은 cGAN의 손실함수로서 G (Generator)로 인해 최소화, D (Discriminator)로 인해 최대화하는 방향으로 최적화된다. 또한 L_1 손실함수인 $L_{L1}(G)$ 을 사용해서 L_{cGAN} 과 함께 Generator는 검증 정오표(ground truth)에 더 가까운 출력을 생성하는 역할을 하며, λ 는 두 손실함수 사이의 비중을 조정하는 파라미터이다.

$$G^* = \arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G)$$

Pix2pix의 또 다른 주요한 특징은 U-Net과 PatchGAN을 사용한다는 점이다. Bottleneck 레이어를 통과해 정보의 손실이 발생하는 Encoder-decoder 네트워크 대신, Pix2pix에서는 Generator에 Skip-connection을 추가한 U-Net을 사용하여 이미지 생성 시 정보 손실을 줄이고 전체의 맥락 정보를 유지하여 Encoder-decoder를 사용할 때 보다 성능이 향상된 것을 확인할 수 있다[11]. PatchGAN 기반의 Discriminator는 이미지 전체 영역 대신 특정 크기의 patch 단위로 이미지 진위 여부를 판별하여 이미지 픽셀의 지역적인 특성뿐만 아니라 전체 맥락상의 손실을 측정할 수 있도록 하는 장점이 있다.

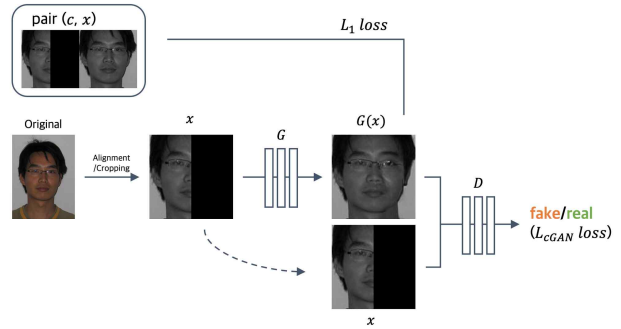


그림 9. Pix2pix generator, discriminator 학습
 Fig. 9. Training process of each generator and discriminator in pix2pix

IV. 실험 및 결과

4-1 실험 데이터 셋

실험에서는 CASIA NIR-VIS 2.0 Face Database[12](이하 CASIA), Extended Yale Face Database B[2](이하 Yale B), 그리고 구글 이미지 검색으로 찾은 얼굴 이미지를 사용하였다.

CASIA 데이터셋은 근적외선 카메라로 촬영한 NIR(Near Infrared) 얼굴 이미지와 일반 카메라로 촬영한 얼굴 이미지인 VIS(Visual) 이미지로 이루어져 있다. 본 연구에서는 VIS 이미지 데이터 셋만을 이용하였으며, VIS 이미지는 총 2,998 장으로서 534명의 묶음(identity)으로 구성되어 있다. CASIA 데이터셋은 1명의 묶음 당 1장 이상의 이미지가 있으며, 총 534명의 묶음을 촬영한 이미지 중 80%인 426명의 묶음 이미지 총 2,401장의 이미지를 학습에 사용하였고, 나머지 20%인 107명의 이미지를 모델 검증을 위한 테스트 데이터로 사용하였다. 그림 10은 CASIA 데이터의 예를 보여주며, 얼굴의 크기 차이가 다양하며 이에 대한 보정이 필요하다.

Yale B 데이터 셋은 한 명의 묶음 당 9개의 자세와 64종류의 조명 환경에서 촬영한 총 28명의 16,128장의 이미지 데이터로 구성되어 있다. 이 중에서 얼굴의 우측에 서로 다른 농도로 그림자가 드리워진 5개의 조명 환경에서 촬영한 135장의 이미지들을 테스트 데이터로서만 사용하였다. 그림 11은 Yale B 데이터 셋에서 사용한 이미지의 예를 보여준다.

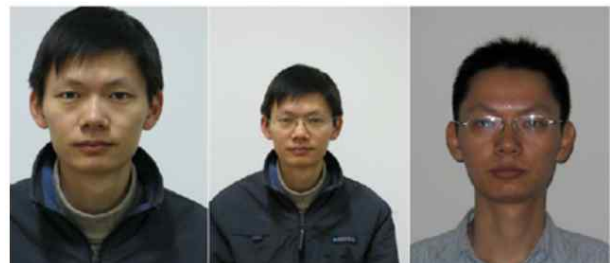


그림 10. CASIA NIR-VIS 2.0 Face Database
 Fig. 10. CASIA NIR-VIS 2.0 Face Database



그림 11. The Extended Yale Face Database B
 Fig. 11. The Extended Yale Face Database B

구글 이미지 검색으로 찾은 얼굴 이미지들은 측면광으로 인해 얼굴의 우측 또는 좌측에 그림자가 발생한 31장의 이미지들이다. 이 얼굴 이미지들은 동양인 및 서양인의 얼굴을 포함하고 있는데, 이는 CASIA 데이터가 동양인 얼굴로 구성되어 학습되었지만 서양인 얼굴에 대해서도 제안한 모델의 적용 결과를 확인하기 위함이다. 그림 12는 구글 검색 이미지의 예이다.

실험에 사용한 모든 이미지들은 얼굴의 랜드마크를 추출하여 이를 이용해 얼굴 영역을 추출하고 정렬시킨다. 모델의 학습에는 CASIA 데이터 셋에서 임의로 추출한 이미지 2,401장의 원본과 그림자를 삽입한 이미지를 쌍으로 구성하여 사용하였다. 모델의 테스트는 나머지 CASIA 데이터 597장, 다양한 조명 환경으로 촬영한 Yale B 데이터, 그리고 구글 검색 이미지를 각각 사용하였다.



그림 12. 구글 검색 이미지 예
 Fig. 12. Image samples by Google image search

4-2 일률적 및 점진적 학습 데이터 생성 결과

표 1은 각 모델의 블랜딩 방식과 농도를 나타내며, 일률적(Uniform) 블랜딩 농도 70%의 의미는 해당 영역을 검은색 70%, 원본 30%로 섞은 것을 의미한다. 점진적(Gradual) 블랜딩 70%는 얼굴 이미지의 중간 수직선을 기준으로 우측 끝으로 가면서 검은색 블랜딩 농도가 0%에서 70%까지 변하는 것이다.

그림 13의 (A)~(F)는 각 모델에 일률적 또는 점진적 블랜딩으로 학습한 모델에 테스트 이미지를 적용하여 생성한 결과의 예를 보여준다. 각 결과 이미지를 검증 정오표(Ground Truth) 이미지와 비교할 때 거의 차이가 없을 정도로 정확하게 복원한 것을 확인할 수 있다. 그림 14는 Yale B 데이터로 이미지를 생성한 테스트 결과이다. 그림 14 (A)~(F) 일률적, 점진적인 블랜딩 강도로 변환한 이미지를 모델에 입력으로 사용한 것이다. 그림 14 (A)~(C)의 경우, 변환한 이미지의 일부 영역이 과하게 밝아지거나 왜곡되어 있으며 이는 자연스러운 이미지를 생성하는 그림 13 (A)~(C)의 결과와는 차이를 보인다. 그림 14 (D)~(F)에서는 블랜딩 강도가 작을수록 좀 더 자연스럽게 보이지만, 블랜딩 강도가 크면 부자연스러운 모습으로 변환되기도 한다.

표 1. 생성 모델 별 블랜딩 방식 및 블랜딩 농도

Table 1. Blending method and density for each model

	Blending	Blending Density (Black %)
Model 1	Uniform	70
Model 2		80
Model 3		90
Model 4	Gradual	0 (Middle) ~ 70 (Rightmost)
Model 5		0 (Middle) ~ 80 (Rightmost)
Model 6		0 (Middle) ~ 90 (Rightmost)

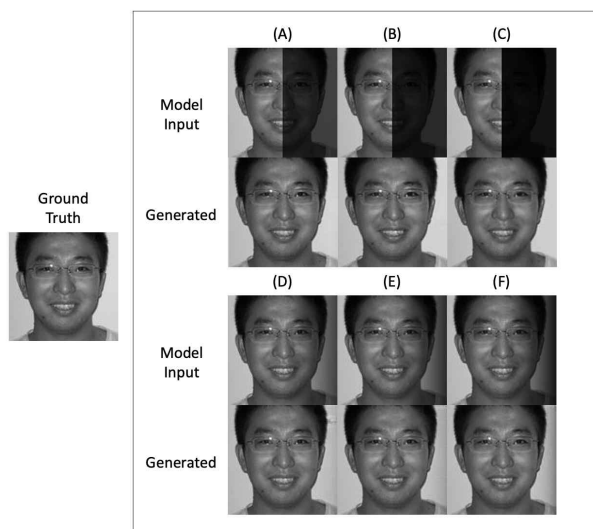


그림 13. CASIA 테스트 데이터에 대한 각 모델별 생성 결과
 Fig. 13. Generated results for each model for CASIA test data

그림 15는 측면광으로 오른쪽 영역에 그림자가 있는 구글 검색 이미지를 대상으로 테스트한 결과이다. 그림 14(A)~(C)에서와 마찬가지로 변환된 이미지의 일부 우측 영역에 과하게 밝아지는 점 등을 확인할 수 있지만, 블렌딩 강도가 낮을수록 그림자가 더 자연스럽게 완화되는 모습을 확인할 수 있다.

4-3 혼합 데이터 학습 결과

그림 16과 17은 앞에서 블렌딩 방식과 농도에 따라 6개의 모델을 각각 학습시켰던 것과는 달리, 변환한 모든 종류의 이미지를 모아서 하나의 모델(이하 모델 7)로 학습한 결과를 보여준다. 모델 7의 학습에는 일률적 또는 점진적인 방식으로 3가지 블렌딩 강도를 달리해 생성한 6종류의 데이터 셋에서 각각 100명씩 임의로 추출해 총 3,366장의 이미지를 사용하였다. 한 명의 묶음 당 이미지 수가 1장에서 12장으로 상이해서 하나의 데이터 셋에서 약 500장의 이미지를 추출하였으며, 그 안에서 임의로 50%인 250장은 좌우 플립하여 학습에 사용하였다. 그림 16은 CASIA 데이터로, 그림 17은 Yale B 데이터와 구글 검색 이미지로 테스트한 결과이다. 각각의 모델을 별도로 구축하여 테스트한 결과와 큰 차이는 없지만, 6개의 개별 모델에서는 그림자 방향이 일정한 이미지만 사용할 수 있었지만 다양한 방향의 그림자에 대해서 사용할 수 있는 장점이 있다.

4-4 정량적 인식 평가

본 연구에서 제안한 모델이 실제 얼굴 인식 성능을 개선할 수 있는지의 여부를 확인하기 위해, CASIA 데이터 셋의 VIS 이미지로 학습한 얼굴 인식을 만들고 이를 통해 성능을 평가하였다. CASIA 데이터셋의 VIS 이미지들은 한 명의 묶음 당 최소 1장에서 최대 12장의 이미지로 구성되어 있으며, 이 연구에서는 이미지 수가 3장 미만인 묶음을 제외한 총 533명의 이미지를 정량 평가에 사용하였다. 한 명의 묶음 이미지를 대략 2:1의 비율로 나눠 학습 이미지와 테스트 이미지로 사용하며 평가를 진행하였으며, 학습에는 원본 이미지 총 2,001장, 테스트에는 996장을 원본 그대로 사용하거나 그림자를 삽입하는 등 변환하여 모델의 입력으로 이용하였다. 정량적 평가에 사용한 인식기는 CASIA Webface Dataset[15]으로 사전 학습된 Inception Resnet (V1)[15]의 파라미터를 전이학습하여 사용하였다.

표 2에서 원본 이미지의 인식 정확도가 99.5%인데 반해, 그림자를 인위적으로 추가하여 변환한 이미지의 인식 정확도는 평균 86.95%로 원본 이미지의 정확도와 비교하였을 때 상대적으로 크게 낮아진 것을 확인할 수 있다. 특히 일률적으로 블렌딩한 이미지의 인식률이 아주 낮음을 확인할 수 있다.

이 연구에서 제안한 모델로 생성한 이미지의 인식 정확도는 평균 99.55%로서 원본 이미지의 인식 정확도와 거의 차이가 없이 개선된 것을 확인할 수 있다. 특히 일률적으로 그림자를 블렌딩한 이미지에서 크게 개선된 것을 확인할 수 있다.

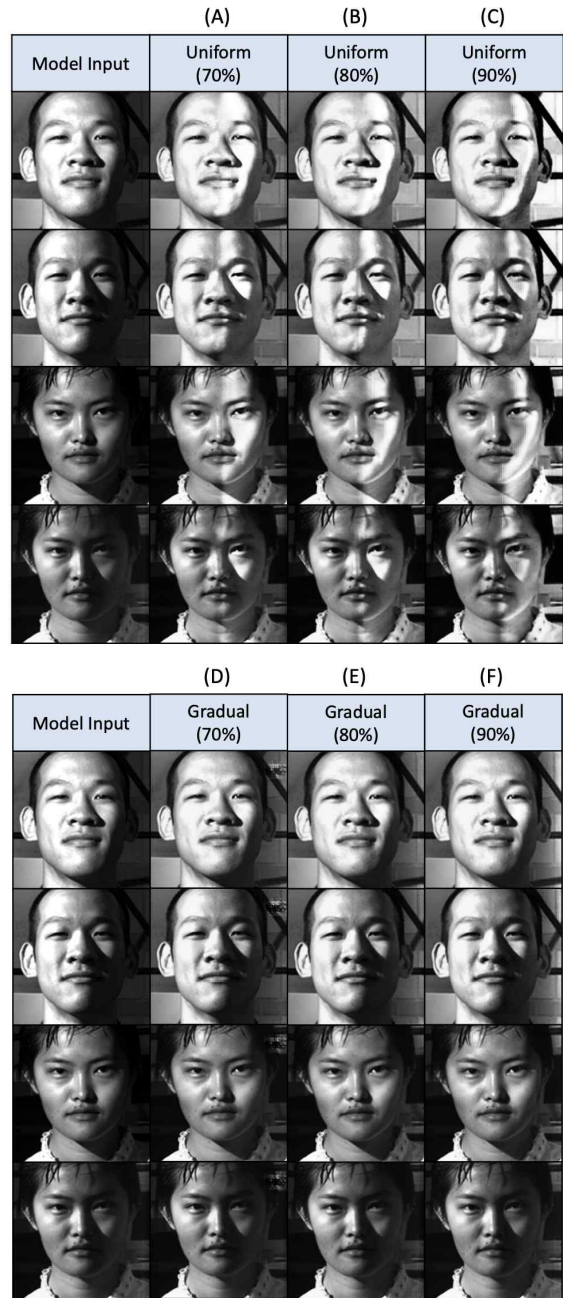


그림 14. Yale B 테스트 데이터에 대한 각 모델별 생성 결과
Fig. 14. Generated results for each model for Yale B test data

조명에 의해 가려지거나 그림자가 있는 동일한 목적의 얼굴 동일한 데이터를 인식하는 방법으로서 [6]에서는 CycleGAN 모델을 이용해서 근적외선 이미지를 생성하여 학습한 후 얼굴 인식을 수행하였다. 근적외선 이미지를 사용하는 것은 조명 변화에 둔감한 장점이 있기 때문이며, 또한 [6]은 본 연구와 동일한 CASIA 데이터셋[12]을 사용했기 때문에 결과를 비교한다.

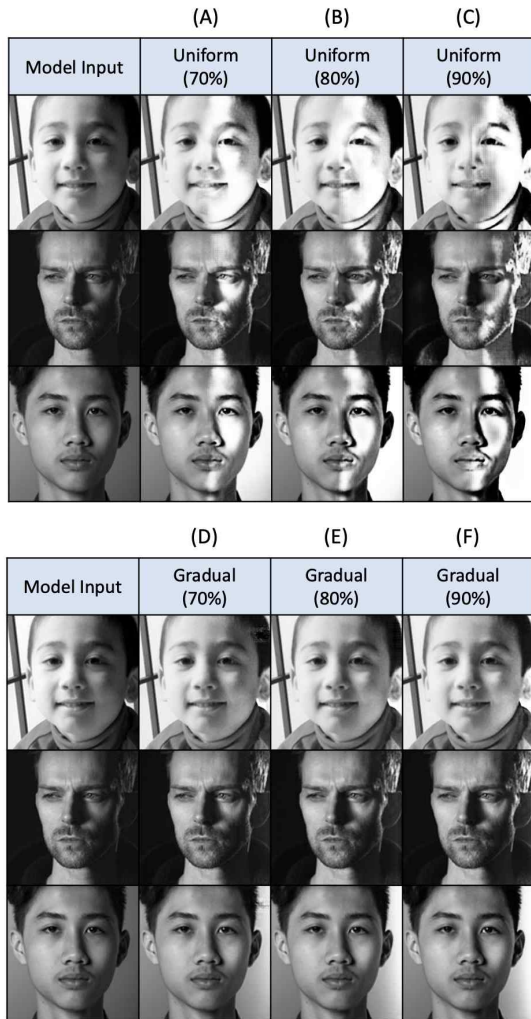


그림 15. 구글 테스트 데이터에 대한 각 모델별 생성 결과
Fig. 15. Generated results for each model for Google image test data

생성한 근적외선 얼굴 이미지를 포함하여 학습한 얼굴 인식 모델 중에서 성능이 가장 우수한 모델의 얼굴 인식 정확도는 96.23%이다. 이와 비교하여 본 논문에서 제안한 모델로 생성한 그림자를 완화하거나 제거한 이미지로 학습한 인식 모델의 평균 인식률은 99.55%로서 [6]보다 우수한 것을 확인하였다.

V. 결론

이 연구에서는 Pix2pix GAN과 이를 학습하기 위한 가상의 학습 데이터를 생성하고 이를 쌍으로 이용하여 얼굴 이미지의 그림자를 완화하고 복원하는 방법을 제안하였다. 제안한 방법으로 생성한, 그림자가 제거된 얼굴 이미지를 정성적, 정량적 관점에서 평가했을 때 개선 전보다 많이 향상된 것을 확인할 수 있었다. 이 과정에서 간단한 이미지 전처리와 특징 기반의 Perceptual 손실함수를 추가하는 것이 생성한 얼굴 이미지의 품질 개선에 도움이 된다는 것도 확인하였다.

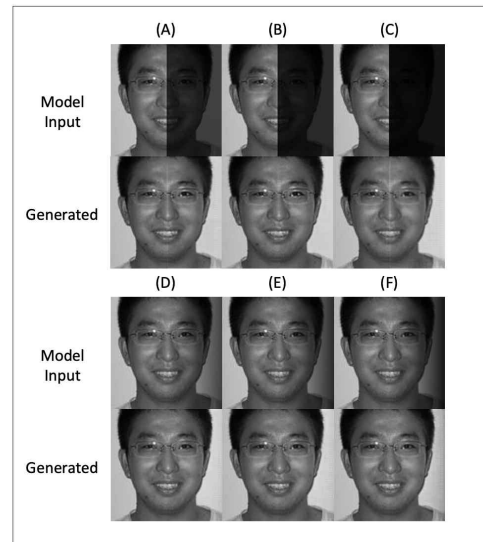


그림 16. CASIA 테스트 데이터에 대한 모델 7 생성 결과
Fig. 16. Generated results for model 7 for CASIA test data

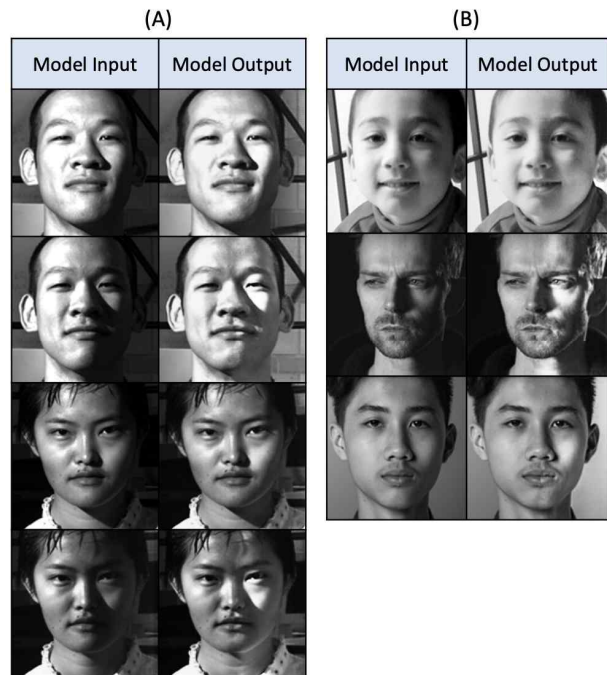


그림 17. Yale B(좌), 구글(우) 테스트 데이터에 대한 모델 7 생성 결과
Fig. 17. Generated results for model 7 for Yale B(left) and Google(right) test data

제안한 방법은 현재는 주로 동양인 얼굴을 대상으로 했지만 다양한 인종이 혼합된 데이터를 사용하여 지역에 상관없이 조명 변화에 강인한 얼굴 이미지를 생성할 수 있는 시스템으로 확장하고자 한다.

표 2. 복원 전, 후 데이터에 따른 인식률 비교

Table 2. Recognition accuracy comparison before and after data reconstruction

Data		Recognition Accuracy (%)	
Original Images		99.50	
Transformed Images	Uniform 70%	82.63	86.95 (Average)
	Uniform 90%	70.08	
	Gradual 70%	99.10	
	Gradual 90%	99.10	
Generated Images	Uniform 70% (Model 1)	99.70	99.55 (Average)
	Uniform 90% (Model 3)	99.50	
	Gradual 70% (Model 4)	99.50	
	Gradual 90% (Model 6)	99.50	
	All mixed (Model 7)	99.56	

감사의 글

본 연구는 한국연구재단의 지원으로 이루어진 연구로서 (No. NRF-2017R1D1A1B04035633) 관계부처에 감사 드립니다.

참고문헌

[1] H. A. Le and I. A. Kakadiaris, "UHDB31: A dataset for better understanding face recognition across pose and illumination variation", Proc. IEEE International Conference on Computer Vision Workshops, pp. 2555-2563, Oct. 22-29 2017.

[2] 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 Transactions on Pattern Analysis and Machine Intelligence, 23(6), pp. 643-660, 2001.

[3] W. Zhang, X. Zhao, J. M. Morvan and L. Chen, "Improving shadow suppression for illumination robust face recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(3), pp. 611-624, 2018.

[4] S. I. Choi, C. Kim and C. H. Choi, "Shadow compensation in 2D images for face recognition", Pattern Recognition, 40(7), pp. 2118-2125, 2007.

[5] P. C. Hsieh and P. C. Tung, "Shadow compensation based on facial symmetry and image average for robust face recognition", Neurocomputing. Vol. 73, Issues 13-15, pp. 2708-2717, 2010.

[6] S. J. Lee and Y. W. Choi, "Generation and Evaluation of Unpaired Near-Infrared Face Images Using CycleGAN", Journal of Digital Contents Society Vol. 21, No. 3, pp. 593-600, Mar. 2020.

[7] N. P. Ramaiah, E. P. Ijjina and C. K. Mohan, "Illumination invariant face recognition using convolutional neural networks.", 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems, 2015.

[8] X. C. Zhang, Y. T. Tsai, R. Pandey, X. Zhang, R. Ng and D. E. Jacobs, "Portrait Shadow Manipulation", arXiv:2005.08925, 2020.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair and Y. Bengio, "Generative adversarial nets", In Advances in Neural Information Processing Systems, pp. 2672-2680, 2014.

[10] J. Y. Zhu and T. Park, "Unpaired image-to-image translation using cycle-consistent adversarial networks", IEEE International Conference on Computer Vision, pp. 2223-2232, 2017.

[11] P. Isola, J. Y. Zhu, T. Zhou and A. A. Efros, "Image-to-image translation with conditional adversarial networks", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125-1134, 2017.

[12] S. Li, D. Yi, Z. Lei and S. Liao, "The casia nir-vis 2.0 face database", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 348-353, 2013.

[13] V. J. D. Tsai, "A comparative study on shadow compensation of color aerial images in invariant color models", IEEE Transactions on Geoscience and Remote Sensing, Vol 44, Issue 6, pp. 1661-1671, 2006.

[14] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Addison-Wesley Publishing Company, 1993.

[15] D. Yi, Z. Lei, S. Liao and S. Z. Li, "CASIAWebface: Learning Face Representation from Scratch", arXiv:1411.7923, 2014.

이수현(Sooheon Lee)



2021년 : 숙명여자대학교 컴퓨터과학전공 학사

2017년~현 재: 숙명여자대학교 컴퓨터과학전공

※ 관심분야 : 인공지능(AI), 머신러닝(ML), 영상처리(IP), 얼굴 인식(FR) 등

최영우(Yeongwoo Choi)



1985년 : 연세대학교 전자공학과(학사)

1986년 : University of Southern California 컴퓨터공학과(석사)

1994년 : University of Southern California 컴퓨터공학과(박사)

1994년~1997년: LG전자기술원 선임연구원

1997년~현 재: 숙명여자대학교 컴퓨터과학과 교수

※ 관심분야 : 인공지능(AI), 머신러닝(ML), 영상처리(IP), 지능형자동차(IV) 등