

정규논문 (Regular Paper)

방송공학회논문지 제20권 제4호, 2015년 7월 (JBE Vol. 20, No. 4, July 2015)

<http://dx.doi.org/10.5909/JBE.2015.20.4.606>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 다른 조명 환경을 갖는 영상 간의 얼굴 교체 기술

송 중 석<sup>a)</sup>, Xingjie Zhang<sup>a)</sup>, 박 종 일<sup>a)†</sup>

### Face Replacement under Different Illumination Condition

Joongseok Song<sup>a)</sup>, Xingjie Zhang<sup>a)</sup>, and Jong-Il Park<sup>a)†</sup>

#### 요 약

컴퓨터 그래픽스(CG: computer graphics) 기술은 영화나 TV와 같은 미디어 콘텐츠에서 중요한 요소 기술로 자리매김하고 있다. 영상 속 특정 인물의 얼굴을 다른 사람의 얼굴로 교체하는 얼굴 교체 기술은 대표적인 CG 기술 중의 하나로써 오래전부터 많이 연구되어 왔다. 본 논문에서는 3D 모델을 사용하지 않고 특정 조명 환경을 갖고 있는 목표 동영상 속의 얼굴을 다른 조명 환경을 갖고 있는 참조 정지 영상 속의 얼굴로 자연스럽게 교체할 수 있는 기술을 제안한다. 실험 결과, 제안하는 방법이 서로 다른 조명을 갖고 있는 두 얼굴 간의 얼굴 교체를 자연스럽게 수행할 수 있는 것을 확인할 수 있었다.

#### Abstract

Computer graphics(CG) is being important technique in media contents such as movie and TV. Especially, face replacement technique which replaces the faces between different images have been studied as a typical technology of CG by academia and researchers for a long time. In this paper, we propose the face replacement method between target and reference images under different illumination environment without 3D model. In experiments, we verified that the proposed method could naturally replace the faces between reference and target images under different illumination condition.

Keyword : Face replacement, face swapping, face synthesis

a) 한양대학교 컴퓨터소프트웨어학과(Department of Computer Software, Hanyang University)

† Corresponding Author : 박종일(Jong-Il Park)

E-mail: [jjpark@hanyang.ac.kr](mailto:jjpark@hanyang.ac.kr)

Tel: +82-2-2220-4368

ORCID: <http://orcid.org/0000-0003-1000-4067>

※ 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 ICT융합고급인력과정지원사업의 연구결과로 수행되었음(IITP-2015-H8601-15-1005).

· Manuscript received May 8, 2015; Revised July 13, 2015; Accepted July 13, 2015.

## 1. 서 론

컴퓨터 그래픽스(CG: computer graphics) 기술은 영화나 드라마와 같은 미디어 콘텐츠들의 품질에 영향을 줄 수 있는 중요한 요소로 자리 잡고 있다. CG 기술의 대표적인 기술 중 하나인 얼굴 교체 기술은 다양한 분야에서 쓰일 수 있는 이점이 있어 점차 주목을 받고 있다. 영화나 드라마를 촬영할 경우, 주인공을 대신해 스티트 맨이나 대역이 대신

촬영을 할 경우가 있는데, 이 기술을 사용하면 대역의 얼굴을 주인공의 얼굴로 자연스럽게 교체할 수 있다. 또한, 구글 로드 뷰와 같이 온라인 상에서 개인의 얼굴이 노출되는 문제도 이 기술을 이용하면 해결할 수 있다. 최근, 소셜 네트워크 서비스 (SNS: social network service)가 급속도로 발달하면서, 재미있는 동영상이 있으면 그 영상 속의 등장인물의 얼굴을 임의의 얼굴을 합성하고 그 결과를 SNS를 통해 타인과 공유함으로써 대중의 관심을 유도하는 등 민간 분야에서도 얼굴 교체 기술에 대한 관심이 자연스럽게 증가하고 있다. 이러한 이유로, 학계에서는 자연스럽게 얼굴을 교체하는 기술에 대해 많은 연구를 진행하고 있다<sup>1-5)</sup>. 영상 내의 얼굴을 교체하는 기술은 3D 모델을 사용하는 방법과 사용하지 않는 방법으로 나눌 수 있다. 3D 모델을 사용할 경우에 얼굴의 과도한 회전이나 이동에도 적응적으로 얼굴을 교체할 수 있다는 장점이 있으나, 영상에 등장하는 얼굴의 3D 모델 정보는 영상을 제작할 때 같이 획득해야 한다는 문제가 있기 때문에 실용성이 떨어진다는 단점이 있다. 반면, 3D 모델을 사용하지 않을 경우, 영상 속의 얼굴이 과도한 회전이나 이동할 때에 얼굴 교체가 제대로 되지 않는다는 단점이 있으나, 영상의 2D 정보만을 갖고 얼굴을 교체하기 때문에 전자보다는 실용성이 상대적으로 높다는 장점이 있다.

Columbia 대학에서는 3D 모델을 사용하지 않고 정지 영상 간의 얼굴을 자연스럽게 교체하는 방법에 대해서 제안하였다<sup>6)</sup>. 기존의 방법들은 얼굴을 교체할 때 초기화 변수를

수동적으로 설정해야 하는 단점이 있었으나, 이 방법은 얼굴 교체 과정이 전부 자동으로 수행된다는 장점이 있다. Harvard 대학에서는 동영상 간의 얼굴을 자연스럽게 교체하는 방법에 대해서 제안하였다. 이 방법은 얼굴을 추적할 때 사용자가 수동적으로 변수를 초기화해야하는 과정이 요구된다는 단점이 있다<sup>7)</sup>. 이와 달리, Max Planck 연구소에서는 3D 모델을 사용하지 않고 동영상 간에 얼굴을 완전히 자동으로 교체하는 방법에 대해 제안하였다<sup>8)</sup>. 지금까지 제안된 얼굴 교체 방법들은 주로 정지 영상 간, 혹은 동영상 간의 얼굴을 교체하는 방법에 대한 연구이기 때문에 실용성이 떨어진다는 단점이 있다. 본 논문에서는 3D 모델 없이 동영상 (목표 영상)과 정지 영상(참조 영상) 간의 서로 다른 조명 환경에서도 자연스럽게 얼굴을 교체하는 방법에 대해서 제안한다. 제안하는 방법을 사용하면, 기존의 동영상에 등장하는 임의의 얼굴을 사용자가 갖고 있는 정지 영상 속의 얼굴로 자연스럽게 교체할 수 있다. 또한, 얼굴 교체 과정 전체가 완전히 자동으로 수행되기 때문에 실용적으로 사용할 수 있다. 그림 1은 목표 영상  $T$ 와 참조 영상  $R$ 을 나타내고 있다.

## II. 얼굴 교체 시스템

본 논문에서 제안하는 방법은 그림 1-(가)와 같은 동영상 속의 목표 얼굴을 그림 1-(나)와 같은 정지 영상 속의 참조



(가) 목표 영상



(나) 참조 영상

그림 1. 목표 영상과 참조 영상

Fig. 1. Target image  $T$  and reference image  $R$

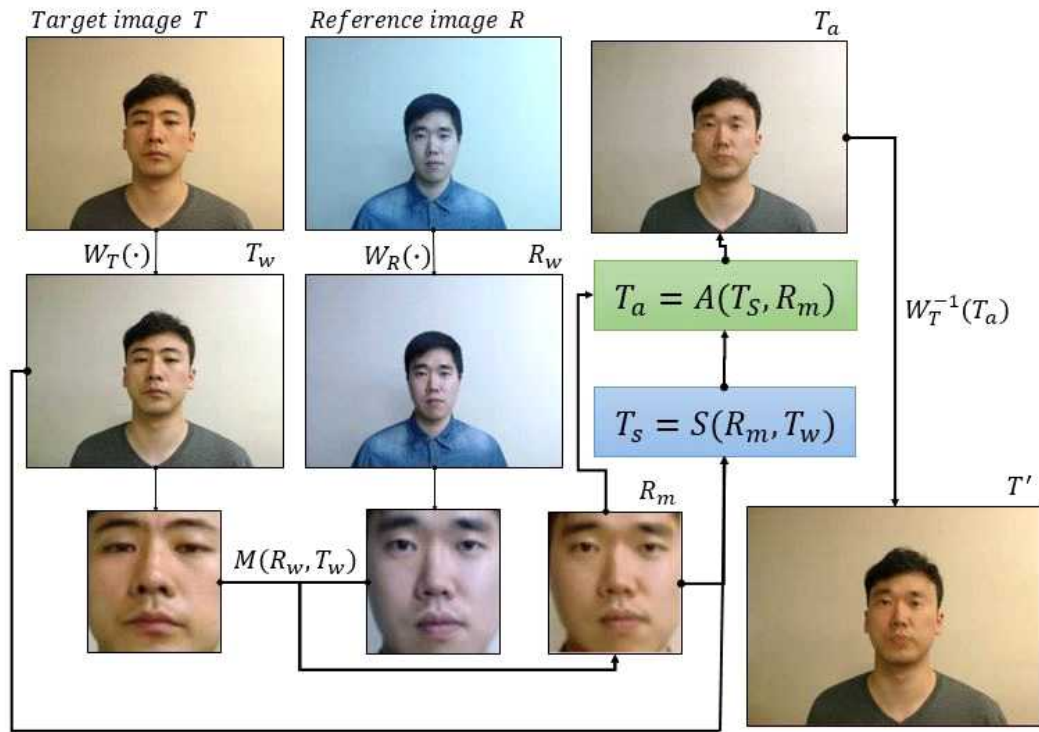


그림 2. 제안하는 얼굴 교체 시스템 흐름도  
 Fig. 2. The flowchart of proposed face replacement system

얼굴로 자연스럽게 교체하는 방법에 대해 제안하고 있다. 다음 그림 2는 본 논문에서 제안하는 얼굴 교체 기술의 전체 흐름을 나타낸다. 얼굴 교체의 과정은 다음과 같다. 얼굴을 교체하고자 하는 두 영상의 조명 환경은 같은 환경에서 촬영된 영상이 아니면 다를 수 밖에 없기 때문에, 참조 영상의 조명 환경을 목표 영상의 조명 환경에 맞춰주는 과정이 필요하다. 본 논문에서는 화이트 밸런스(white balance) 기법  $W$ 를 사용해 참조 영상의 조명 환경과 목표 영상의 조명 환경을 근사시켜준다. 두 영상의 조명 환경이 근사되면, 히스토그램 명세화(histogram specification) 기반의 피부 톤 일치 기법  $M$ 을 사용하여 두 얼굴의 피부 톤을 일치시켜주고<sup>[15]</sup>, 두 얼굴에 대해 각각 ASM(active shape model)을 적용하여 77개의 얼굴 특징점들을 추출한다. 추출된 얼굴 특징점들 중에서 얼굴의 경계선에 해당하는 16개의 특징점은 다른 특징점에 비해 정확도가 상대적으로 낮기 때문에 이들을 제거하고, 이들을 대체하기 위한 18개의 가상 얼굴 특징점들을 생성한다.

목표 얼굴의 경우에는 동영상이기 때문에, 얼굴의 움직임에 따라 참조 얼굴의 포즈도 변화되어야 한다. 두 얼굴 영상에서 기준이 되는 5개의 얼굴 특징점을 선택하고, 이 5쌍의 특징점들을 사용하여 유사 변환 행렬(similarity transform matrix)  $S$ 를 생성한다. 참조 얼굴의 특징점들의 좌표들을 추출된 유사 변환 행렬  $S$ 에 투영하면, 목표 얼굴에 맞게 투영된 참조 얼굴의 특징점들의 좌표가 계산된다. 이 과정을 목표 영상의 매 프레임마다 수행하여 유사 변환 행렬  $S$ 를 매번 생성하고, 이를 통해 연산된 좌표와 참조 영상의 얼굴 특징점 좌표들 간의 어파인 변환(affine transform)  $A$ 를 사용하여 목표 영상의 얼굴을 참조 영상의 얼굴로 교체한다. 교체된 두 얼굴의 경계 부분은 두 얼굴의 차이로 부자연스럽기 때문에 적응적 가중치 기반 블렌딩 기법을 사용하여 이들을 자연스럽게 제거한다<sup>[16]</sup>. 마지막으로 교체된 영상에 목표 영상의 이득값(gain value)으로 역 화이트 밸런스  $W^{-1}$ 를 수행하여, 목표 영상의 조명 환경으로 재조명 시켜준다.

### 1. 화이트 밸런스를 이용한 두 영상의 조명 환경 근사화

본 논문에서는 목표 영상과 참조 영상의 조명 환경을 근사시켜주기 위해 화이트 밸런스 기법을 사용하였다. 화이트 밸런스는 특정 조명 환경에서 영상의 색상을 정색으로 보정해주는 기술을 의미한다. 화이트 밸런스를 적용할 경우, 백색의 피사체를 촬영했을 때, 영상 내 광원의 색 온도와 상관없이 해당 피사체를 백색으로 재현해야 한다. 이때, 영상의 R, G, B(red, green, blue) 채널의 레벨을 조정하게 되는데, 이 과정을 화이트 밸런스라고 한다. 본 논문에서는 영상 내에서 백색 피사체라 여겨지는 후보를 자동적으로 설정하고, 화이트 밸런스를 수행하는 자동 화이트 밸런스 기법을 사용하였다<sup>[8]</sup>. 목표 영상을  $T$ , 참조 영상을  $R$ 이라고 했을 때, 화이트 밸런스는 다음 수식 1과 같이 표현할 수 있다.

$$R_w = W_R(R) = G_R R, \quad T_w = W_T(T) = G_T T. \quad (1)$$

위 수식에서  $G_R$ 과  $G_T$ 는 참조 영상과 목표 영상의 조명

성분을 포함하는 이득값(gain value)을 나타낸다.  $W$ 는 화이트 밸런스 함수를 의미한다.  $R_w$ 와  $T_w$ 는 참조 영상  $R$ 과 목표 영상  $T$ 의 화이트 밸런스 결과를 나타낸다. 다음 그림 3은 목표 영상과 참조 영상의 화이트 밸런스 결과를 나타내고 있다. 그림을 보면 참조 영상과 목표 영상의 조명이 근사화 되어 있음을 볼 수 있다.

### 2. 히스토그램 명세화 기반의 피부 톤 매칭

화이트 밸런스를 사용하여 두 영상의 조명 환경을 일치시킨 후, 참조 영상 얼굴의 피부 톤을 목표 영상 얼굴의 피부 톤에 일치시켜주는 과정을 수행한다. 이를 위해, 본 논문에서는 히스토그램 명세화 기반의 피부 톤 매칭 기법을 사용하였다<sup>[15]</sup>. 그 과정을 설명하면 다음과 같다. 우선, 참조 얼굴과 목표 얼굴의 컬러 히스토그램을 각각 추출한다. 두 얼굴의 히스토그램은 각각  $h_r(r)$ 과  $h_t(t)$ 로 나타낸다. 여기서  $r$ 과  $t$ 는 참조 영상과 목표 영상의 얼굴을 각각 나타낸다. 히스토그램을 추출한 후, 수식 2과 같이 이들을 누적분포함수  $H_r$ 과  $H_t$ 로 변환한다.

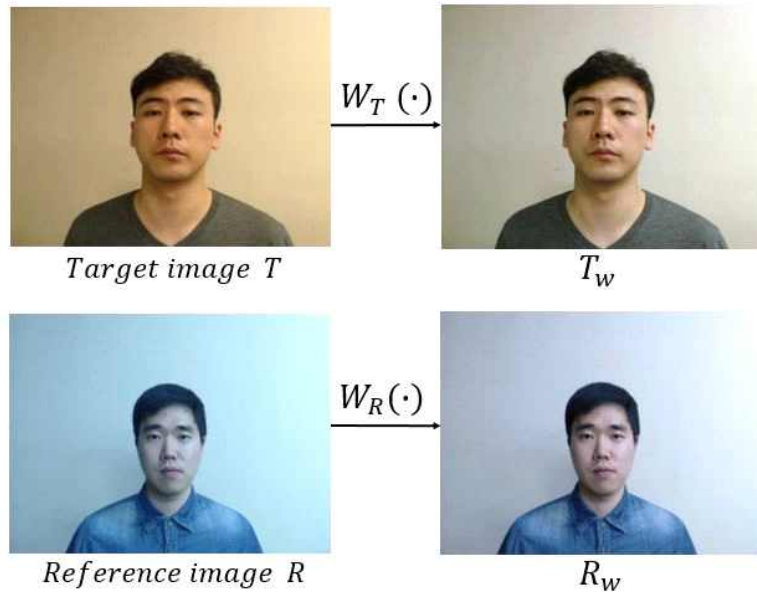


그림 3. 화이트 밸런스를 이용한 두 영상의 조명 환경 근사화  
 Fig. 3. The matching of illumination condition of both images by using the white balance

$$H_r[j] = \sum_{i=0}^j h_r[i], \quad H_t[j] = \sum_{i=0}^j h_t[i]. \quad (2)$$

누적분포함수를 추출하고, 이를 이용해 다음 수식 3과 같이 명암 변환표  $\tau$ 를 얻을 수 있다.

$$\tau(n_i) = n_o = \operatorname{argmin}_k |H_r[i] - H_t[k]| \quad (3)$$

위 수식에서  $n_i$ 는 입력 명암 값,  $n_o$ 는 출력 명암 값을 나타낸다. 참조 얼굴의 임의의 픽셀에서, 해당 픽셀의 명암 값  $n_i$ 를 명암 변환표  $\tau$ 에 입력 값으로 대입하면, 표는  $n_i$ 에 해당하는 출력 명암 값  $n_o$ 을 출력한다. 위의 과정은 R, G, B 채널에 대해서 각각 개별적으로 수행된다. 위 과정에서 참조 영상과 목표 영상의 얼굴 영역을 각각 추출하기 위해 Viola 얼굴 감지기를 사용하였다<sup>[13]</sup>. 피부 톤 매칭 과정은 다음 수식 (4)와 같이 나타낼 수 있다.

$$R_m = M(R_w, T_w) \quad (4)$$

위 수식에서  $M$ 은 피부 톤 매칭 함수를 의미하고<sup>[15]</sup>,  $R_m$ 은  $R_w$ 를  $T_w$ 에 맞게 피부 톤 매칭을 수행한 결과를 나타낸다.

### 3. 얼굴 특징점을 이용한 얼굴 교체

화이트 밸런스를 이용한 조명 환경 일치와 피부 톤 매칭 과정이 끝나면, ASM을 사용하여 목표 얼굴과 참조 얼굴의 얼굴 특징점들을 각각 추출한다<sup>[9]</sup>. ASM을 사용하면, 다음 그림 4-(가)와 같이 77개의 특징점을 추출할 수 있다. 77개 중에서 1번부터 16번에 해당하는 16개의 특징점들은 얼굴의 윤곽선을 나타내는데, 다른 61개의 특징점에 비해 상대적으로 검출 정확도가 떨어진다. 그래서 본 논문에서는 윤곽선에 해당하는 16개의 특징점들은 사용하지 않는다. 하지만 나머지 61개의 특징점들만 사용하면 다음 그림 4-(나)와 같이 얼굴을 제대로 표현할 수가 없기 때문에, 본 논문에서는 그림 4-(다)와 같이 34개의 가상 특징점들을 추가적으로 생성하여 사용한다. 이렇게 되면 특징점들이 얼굴의 대다수를 포함하기 때문에, 얼굴 교체를 좀 더 자연스럽게 수행할 수 있게 된다.

가상 특징점들은 1차와 2차 특징점들로 분류할 수 있는데, 다음 그림 5와 같은 원리에 의해 생성할 수 있다. 1차 특징점들은 초기의 61개 특징점 중에서, 얼굴의 중심에 해당하는 코의 특징점(52번)을 기준으로 최외곽에 해당하는 특징점들까지의 거리  $d_c$ 를 측정하고,  $d_c$ 에 실험적 가중치  $\beta_1$ 을 곱한 값만큼  $d_c$ 로부터 더 연장한 거리  $d_{v1}$ 에 생성된

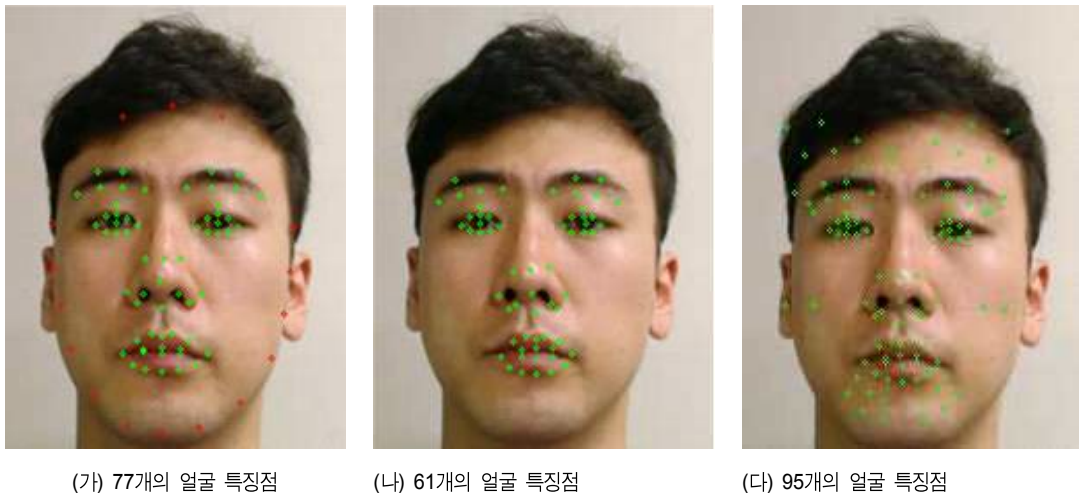


그림 4. ASM으로 추출한 얼굴 특징점  
Fig. 4. The facial features generated by ASM

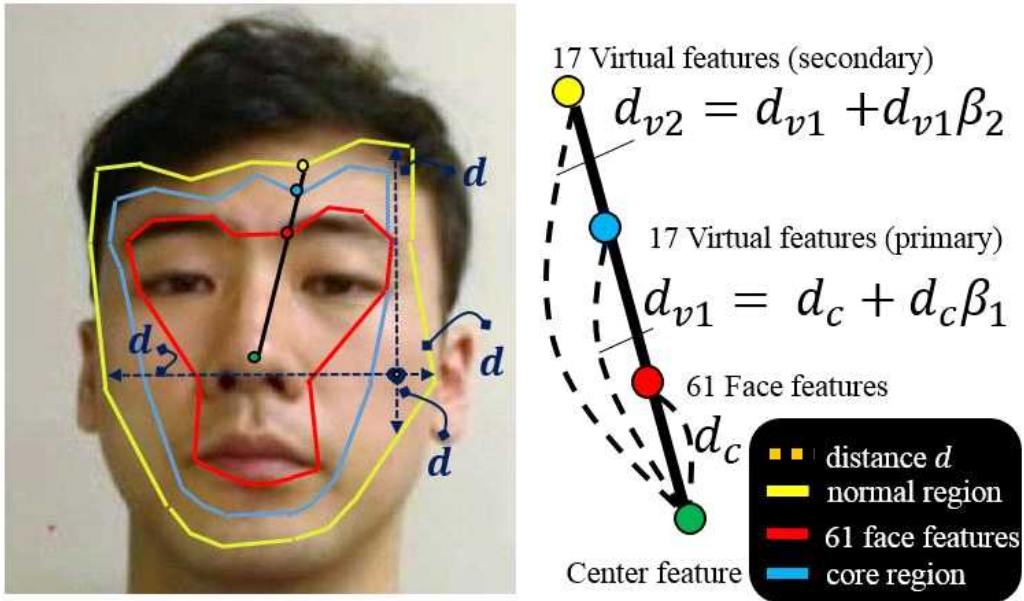


그림 5. 가상 특징점의 생성 원리  
 Fig. 5. The principle of generation of virtual features

다. 2차 특징점들은 코의 중심으로부터 1차 특징점까지의 거리  $d_{v1}$ 에 실험적 가중치  $\beta_2$ 를 곱한 값만큼  $d_{v1}$ 로부터 더 연장한 거리  $d_{v2}$ 에 생성된다.

본 논문에서는 동영상 속의 움직이는 목표 얼굴을 정지 영상 속의 참조 얼굴로 자연스럽게 교체하는 것이 목표이

기 때문에, 정지 영상 속의 참조 얼굴을 동영상 속의 목표 얼굴의 움직임에 따라 자연스럽게 움직이게 하는 것이 중요하다. 다음 그림 6과 같이 목표 얼굴과 참조 얼굴에서 기준이 되는 5개의 특징점들의 좌표들을 각각 추출하고, 이들을 사용하여 두 얼굴 사이의 유사 변환(similarity trans-



(가) 목표 얼굴

(나) 참조 얼굴

그림 6. 유사 변환 행렬을 생성하기 위한 기준 특징점  
 Fig. 6. The standard features for generation of similarity transformation matrix

form) 관계를 계산한다<sup>[4]</sup>. 유사 변환은 하나의 형태를 다른 형태로 변환 시키는 변환 관계들 중 하나이며, 변환 관계를 회전각과 이동량, 그리고 균등 크기 조절을 통해 정의할 수 있다. 유사 변환은 정점간의 거리가 일정한 비율로 유지되는 특성이 있어서, 참조 얼굴의 갖는 눈, 코, 입 등의 고유한 비율을 유지한 상태에서 목표 얼굴을 교체시키기 위해 사용되었다. 다음 수식 5에서  $S$ 는 유사 변환을 나타내고,  $T_s$ 는 피부 톤이 일치된 참조 얼굴  $R_m$ 과 화이트 밸런스가 수행된 목표 얼굴  $T_w$ 에 대해, 각각 5개의 얼굴 특징점들의 좌표들을 추출하여 유사 행렬을 구한 후, 참조 얼굴의 95개 특징점들의 좌표들을 유사 행렬에 투영하여 얻어낸 좌표들을 나타낸다. 참조 얼굴  $R_m$ 의 특징점 좌표들을 유사 변환 행렬  $S$ 에 투영하게 되면, 투영된 좌표들이 참조 얼굴의 고유한 비율을 유지한 상태에서 목표 얼굴  $T_w$ 의 좌표계에 맞게 변형된다.

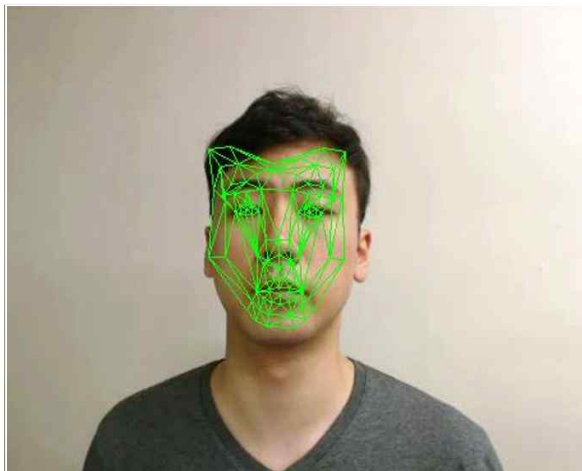
$$T_s = S(R_m, T_w) \tag{5}$$

유사 변환을 통해 계산한 특징점들과 참조 얼굴의 특징점들은 각각 그림 7과 같이 목표 얼굴과 참조 얼굴의 메쉬 모델을 각각 형성하는데 폴리곤의 정점으로써 사용된다.

두 얼굴의 메쉬 모델들은 삼각형 형태의 폴리곤들로 구성되어 있으며, 두 얼굴 사이에 대응되는 폴리곤들끼리 어파인 변환(affine transform)을 수행하여 얼굴을 교체한다<sup>[14]</sup>. 어파인 변환은 선형 변환과 이동 변환의 합성으로 이루어져 있으며, 하나의 벡터 공간을 다른 벡터 공간에 대응시켜 줄 수 있는 변환이다. 두 벡터 사이에는 최소 3개의 대응점



그림 8. 어파인 변환을 통해 참조 얼굴로 교체된 목표 얼굴  
Fig. 8. The target face replaced with the reference face using affine transform



(가) 목표 얼굴



(나) 참조 얼굴

그림 7. 목표 얼굴과 참조 얼굴의 메쉬 모델  
Fig. 7. The mesh models of target and reference face

이 있어야 둘 사이의 변환을 수행할 수 있기 때문에 메쉬 모델에서 삼각형 폴리곤을 이루고 있는 3개의 정점들을 어파인 변환의 대응점으로 활용한다. 다음 그림 8은 어파인 변환을 통해, 목표 얼굴을 참조 얼굴로 변환한 모습을 나타낸다. 다음 수식 6은 유사 변환을 이용해 추출한 얼굴 특징 점  $T_s$ 와 피부 톤이 일치된 참조 얼굴  $R_m$  사이의 어파인 변환을 구하고, 이 어파인 변환을 통해 참조 얼굴로 교체된 목표 얼굴  $T_a$ 를 구하는 과정을 의미한다.

$$T_a = A(T_w, R_m) \quad (6)$$

#### 4. 적응적 가중치 기반 블렌딩 기법

어파인 변환을 통해 목표 얼굴을 참조 얼굴로 교체 시키면, 위의 그림 8과 같이 두 얼굴의 경계에 부자연스러운 부분이 생기는 것을 볼 수 있다. 이런 경우, 부자연스러운 부분을 부드럽게 만들어주기 위해 일반적으로 알파 블렌딩 기법을 사용한다. 다음 수식 (7)은 알파 블렌딩을 나타낸다.

$$\gamma = I_1\alpha + I_2(1 - \alpha). \quad (7)$$

$\gamma$ 는 두 영상에서, 동일한 위치의 각각의 명암 값  $I_1$ ,  $I_2$ 가 가중치  $\alpha$ 에 따라 선형적으로 조절되어 합산된 명암 값을 의미한다. 본 논문에서는 위 수식 (7)에서 가중치  $\alpha$ 를 적응적으로 조절하는 적응적 가중치 기반 블렌딩 기법을 사용하였다<sup>16)</sup>. 이 방법은 우선 교체된 얼굴을 눈, 코, 입 등의 얼굴 주요 요소가 포함되어 있는 핵심 영역과 그 외의 영역인 일반 영역으로 나누는데, 그림 5에 나와 있듯이, 61개의 얼굴 특징점들과 1차 가상 특징점들로 구성된 영역을 핵심 영역으로 지정하고, 2차 가상 특징점까지 확장된 그 외의 영역들을 일반 영역으로 지정한다. 적응적 가중치 기반 블렌딩은 다음 수식 (8)처럼 나타낼 수 있다.

$$\gamma = \begin{cases} r \frac{d}{d_l} + t(1 - \frac{d}{d_l}), & (0 < d \leq d_l) \\ r, & (d > d_l) \end{cases} \quad (8)$$

위 수식에서, 수식 (7)의 가중치  $\alpha$ 가  $d/d_l$ 에 의해 결정된

다는 것을 볼 수 있다.  $d$ 는 그림 5에 나와 있듯이, 교체된 얼굴 내의 임의의 위치에서 일반 영역의 경계까지 상, 하, 좌, 우 방향으로 각각 거리들을 구하고, 그 거리들 중에서 가장 작은 거리 값을 의미한다.  $d_l$ 은 실험적 값으로 본 논문에서는 30으로 지정하여 사용하였다. 이 방법을 통해, 핵심 영역에서는 적응적 가중치  $\alpha$ 가 항상 1로 설정되어, 참조 얼굴의 명암 값을 사용하고 있는 것을 확인할 수 있고, 일반 영역에 적응적 가중치 기반 블렌딩이 적용되는 것을 확인할 수 있다. 다음 그림 9는 적응적 가중치 기반 블렌딩을 사용하여 부자연스러웠던 경계를 자연스럽게 제거한 결과를 나타낸다.



그림 9. 적응적 가중치 기반 블렌딩 기법 적용 결과  
 Fig. 9. The result of adaptive-weight-based blending

#### 5. 역 화이트 밸런스를 이용한 영상 재조명

적응적 가중치 기반 블렌딩을 적용한 후에, 그 결과의 조명 환경을 목표 영상의 원래 조명 환경으로 다시 재조명해 줘야 한다. 본 논문에서는 역 화이트 밸런스를 통해 재조명 과정을 수행한다. 수식 (1)에서 볼 수 있듯이, 목표 영상에 대해서 화이트 밸런스를 수행하면 목표 영상의 조명 성분이 포함되어 있는 대안 이득 값을 얻어올 수 있다. 이때, 교체된 결과 영상에 대해 목표 영상의 이득 값으로 역 화이트 밸런스를 수행하면 원래 목표 영상이 갖고 있던 조명





(가) 목표 영상

(나) 재조명 된 교체 영상

그림 10. 역 화이트 밸런스를 이용해 재조명 된 결과  
 Fig. 10. The result of relighting using inverse white balance

환경으로 다시 재조명 된다. 다음 수식 (9)에서  $T'$ 는 목표 영상의 역 화이트 밸런스  $W_T^{-1}$ 를 통해 블렌딩 된 영상  $T_a$ 을 다시 재조명 한 결과를 나타낸다. 다음 그림 10는 역 화이트 밸런스를 통해 재조명 된 결과를 나타낸다. 재조명 영상 결과를 보면 목표 영상의 조명 환경과 거의 일치함을 확인할 수 있다.

$$T' = W_T^{-1}(T_a) \tag{9}$$

### III. 실험 및 결과

본 논문에서는 다음 그림 1과 같이 목표 영상은 동영상, 참조 영상은 정지 영상을 사용하였다. 두 영상의 해상도는  $640 \times 480$ 이며, 촬영 장비는 일반 USB 웹캠을 사용하였다. 목표 영상과 참조 영상의 촬영 환경은 일반 형광등이 장착된 실내 공간에서 촬영하였으며, 목표 영상을 촬영 할 때는 할로겐 투광기를 추가적으로 사용하였다. 프로그램은 C++기반으로 구현하였으며, 유사 변환을 계산하기 위해 OpenCV를, 어과인 변환을 수행하기 위해 OpenGL 라이브러리를 사용하였다. 그리고 얼굴 특징점을 추출하기 위해서 STASM 라이브러리를 사용하였다<sup>[9]</sup>. 가상 특징점을 생성하기 위한 변수  $\beta_1, \beta_2$  과 블렌딩을 위한 변수  $d_i$  는 각각

실험적으로 얻은 수치인 0.25, 0.25, 30으로 설정하였으며, 자동적 화이트 밸런스를 수행하기 위해 Weng이 제안한 automatic white balance 방법을 사용하였다<sup>[8]</sup>.

#### 1. ASM 기반 얼굴 특징점 정확도 분석

본 논문의 2.3장에서 STASM을 사용하여 추출한 77개의 얼굴 특징점들 중, 1번부터 16번까지의 얼굴 특징점(얼굴의 윤곽선)들이 검출 정확도가 낮아 제안하는 방법에서는 사용하지 않는다고 하였다. 이 부분을 확인하기 위해, 본 실험에서는 초당 30프레임의 속도로 촬영한 100 프레임 길이의 목표 영상에 대해 매 프레임 STASM을 사용하여 77개의 얼굴 특징점들을 추출하고, 각 특징점들의 프레임 간 좌표 변화량을 측정하였다. 다음 그림 11는 각 특징점들에 대해서 100 프레임 동안 프레임 간 좌표 변화량을 누적한 결과이다. 프레임 간 얼굴의 변화량은 미세하기 때문에, 특징점 추출기가 매 프레임 정확하게 특징점을 찾아낸다면, 각 특징점들의 프레임 간 좌표 변화량은 작을 것이라는 가정을 하고 이와 같은 실험을 설계하였다. 실험을 통해, 1번에서 16번에 해당하는 얼굴 특징점들이 다른 특징점들보다 프레임 간 좌표 변화량이 상대적으로 큰 것을 확인할 수 있었다.

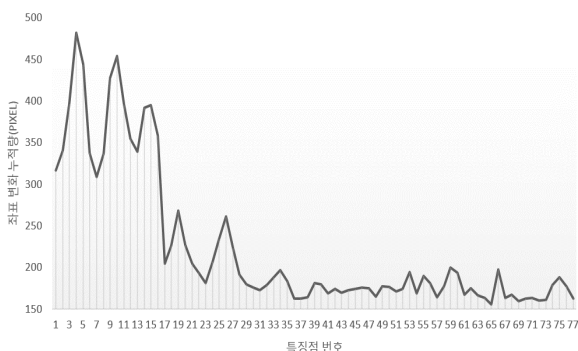


그림 11. 동영상에서 얼굴 특징점 추출의 정확도 실험 결과  
 Fig. 11. The result of accuracy of detection of face features in movie

## 2. 다른 조명 환경을 갖는 영상 간의 얼굴 교체 실험

본 실험에서는 그림 2의 흐름에 따라, 목표 영상의 얼굴을 참조 영상의 얼굴로 자연스럽게 교체하는 과정을 순차

적으로 수행하였다. 다음 그림 12는 그림 1과 달리 참조 얼굴의 샘플을 바꿔서 얼굴 교체를 수행한 실험 결과를 나타낸다. 그림 12-(가),(나)는 목표 얼굴과 참조 얼굴을 나타내고, 목표 얼굴의 경우 할로젠 투광기를 추가적으로 배치하여 조명 환경을 조성하였다. 그림 12-(다),(라)는 블렌딩 전 후 결과를 나타내며, 그림 12-(바)는 역 화이트 밸런스를 이용한 재조명 결과를 나타낸다. 그림 12-(가)와 (바)를 보면, 제안하는 방법을 통해 참조 얼굴이 목표 얼굴의 조명에 맞게 재조명 되어 자연스럽게 얼굴이 교체되고 있음을 확인할 수 있었다.

## 3. 얼굴 각도와 가중치에 따른 얼굴 교체 결과

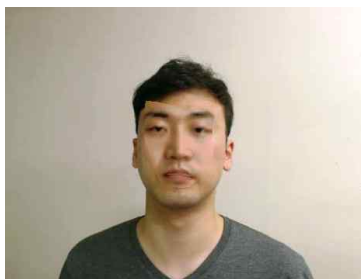
다음 그림 13은 얼굴 각도에 따른 얼굴 교체에 대한 실험 결과를 나타낸다. 목표 얼굴의 회전각을 조절해 가면서 교



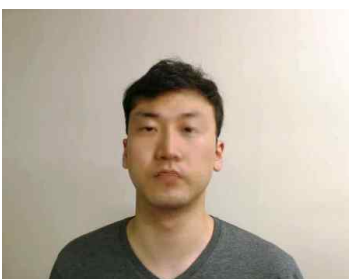
(가) 목표 얼굴



(나) 참조 얼굴



(다) 블렌딩 전



(라) 블렌딩 후



(바) 재조명 결과

그림 12. 제안하는 방법의 얼굴 교체 결과  
 Fig. 12. The result of proposed face replacement method

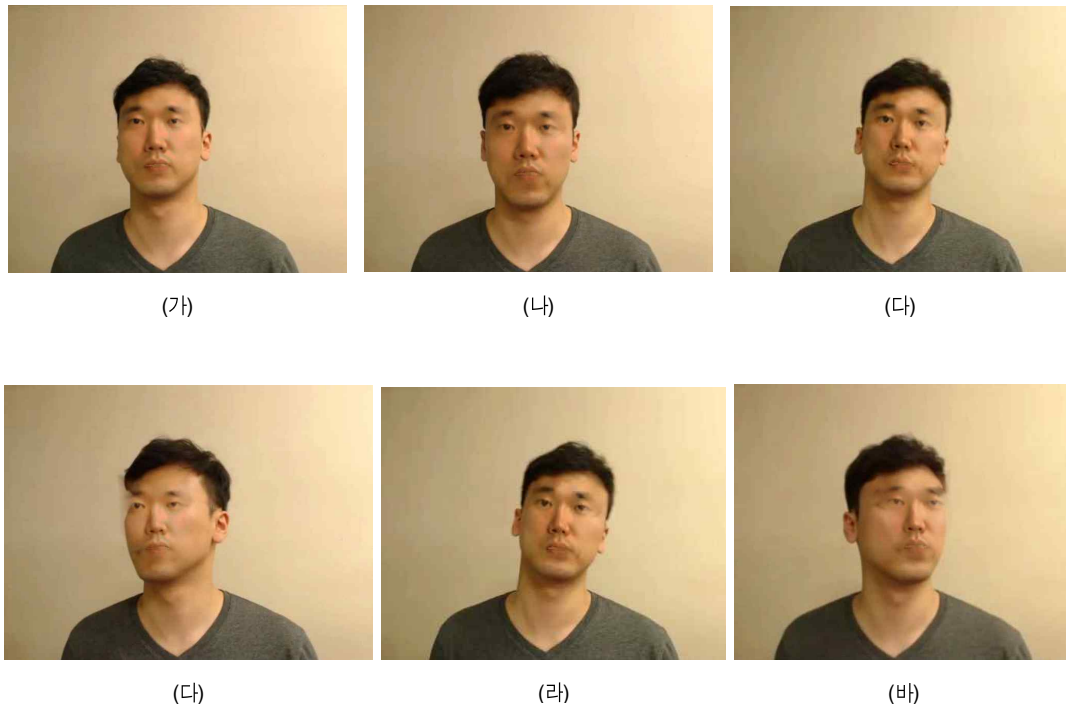


그림 13. 목표 얼굴의 회전에 따른 교체 결과  
 Fig. 13. The face replacement results depending on the rotation angle of target face

체 실험을 수행한 결과, 그림 13-(가),(나),(다)처럼 30도 내의 회전에는 자연스러운 교체를 보였으나, 그림 13-(다),(라),(바)처럼 30도를 벗어난 회전에는 자연스럽게 교체되지 못하는 결과를 볼 수 있었다. 그 이유는 본 논문에서 사용하고 있는 STASM 알고리즘의 경우<sup>9)</sup>, 과도한 얼굴 회전에 대해서는 얼굴 특징점을 제대로 추출하지 못한다는 단점이 있기 때문이다. 일반적으로, 얼굴이 30도 이상으로 과

도한 움직임을 보일 경우, 얼굴의 3D 모델 없이 얼굴 특징점들을 정확하게 추출하는 것은 매우 어려운 일이다.

다음 그림 14는 블렌딩 가중치  $d_i$ 를 30으로 고정하고, 가상 특징점을 생성할 때 사용되는 변수  $\beta_1, \beta_2$ 를 0.1, 0.25, 0.5, 0.75로 변경해가면서 영상을 합성한 결과를 나타낸다. 실험결과, 0.1로 했을 때는 얼굴의 교체 범위가 좁아 부자연스럽고, 0.5와 0.75로 했을 때는 가상 특징점이 얼굴을 한참

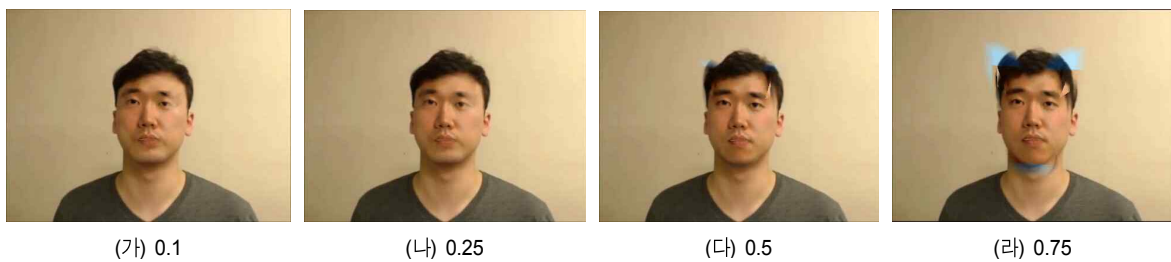


그림 14.  $\beta_1, \beta_2$ 에 따른 얼굴 교체 결과  
 Fig. 14. The result of face replacement depending on  $\beta_1, \beta_2$

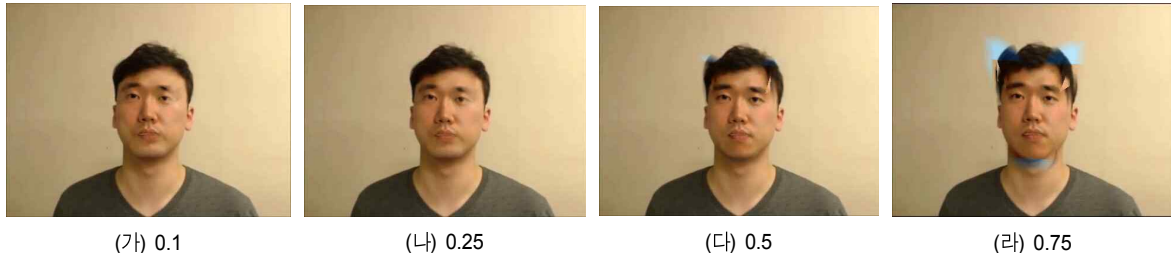


그림 14.  $\beta_1, \beta_2$ 에 따른 얼굴 교체 결과  
 Fig. 14. The result of face replacement depending on  $\beta_1, \beta_2$



그림 15.  $d_i$ 에 따른 얼굴 교체 결과  
 Fig. 15. The result of face replacement depending on  $d_i$

벗어나 있기 때문에 교체가 제대로 수행되지 못하는 것을 볼 수 있었다. 0.25로 했을 때의 가상 특징점이 얼굴을 제대로 표현할 수 있음을 확인할 수 있었다.

다음 그림 15는 가상 특징점 변수  $\beta_1, \beta_2$ 를 0.25로 고정하고, 블렌딩 가중치  $d_i$ 를 10부터 50까지 10씩 증가시켜가면서 실험한 결과를 나타낸다. 10과 20은 참조 얼굴이 많이 강조되어 블렌딩이 부자연스럽게 된 반면, 40과 50은 참조 영상의 얼굴이 목표 영상의 얼굴에 너무 많이 섞이는 현상을 볼 수 있었다. 30으로 설정했을 때, 참조 얼굴과 목표 얼굴이 자연스럽게 블렌딩되는 것을 볼 수 있었다.

#### IV. 결론

본 논문에서는 조명 환경이 다른 두 영상의 얼굴을 자연스럽게 교체하는 기술을 제안하였다. 동영상 속의 목표 얼굴을 정지 영상 속의 참조 얼굴로 교체할 수 있으며, 얼굴 교체의 모든 과정이 자동적으로 수행된다. 본 논문에서는

화이트 밸런스를 사용하여 두 영상의 조명 환경을 근사시킨 후, 히스토그램 명세화 기반의 피부 톤 일치 기술을 통해, 참조 얼굴의 피부 톤을 목표 얼굴의 피부 톤에 맞게 변경하였다. 이후, STASM을 사용하여 두 얼굴의 얼굴 특징점들을 생성한 후, 이 중에서 정확도가 낮은 특징점들을 제거하고 가상 특징점들을 추가적으로 생성하였다. 생성된 얼굴 특징점들을 사용하여 얼굴의 메쉬 모델을 각각 형성한 후, 유사 변환과 어파인 변환을 통해 목표 얼굴을 참조 얼굴로 교체하였다. 교체를 한 후, 교체된 얼굴의 부자연스러운 경계 부분을 적응적 가중치 기반 블렌딩 기법으로 자연스럽게 제거하였다. 이후 역 화이트 밸런스를 사용하여 교체된 얼굴의 조명을 목표 영상의 조명으로 재조명하였다. 실험 결과, 제안하는 방법을 사용하여 조명 환경이 다른 두 얼굴을 자연스럽게 교체할 수 있음을 확인할 수 있었으나, 3D 얼굴 모델 없이 교체를 수행하기 때문에 과도한 회전에 다소 부자연스럽게 합성되는 결과를 확인할 수 있었다. 얼굴의 과도한 회전에도 얼굴 특징점을 정확하게 추출할 수 있는 기술에 대해서 연구를 지속할 계획이다.

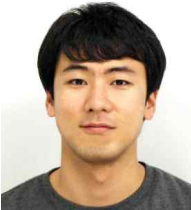
### 참 고 문 헌 (References)

- [1] V. Blanz, K. Scherbaum, T. Vetter, and H.-P. Seidel, "Exchanging faces in images," *Comp. Graph. Forum*, 23(3):669676, 2004.
- [2] O. Alexander, M. Rogers, W. Lambeth, M. Chiang, and P. Debevec, "The Digital Emily Project: photoreal facial modeling and animation," *ACM SIGGRAPH Courses*, pages 12:1-12:15, 2009.
- [3] K. Dale, K. Sunkavalli, M. K. Johnson, D. Vlastic, W. Matusik, and H. Pfister, "Video face replacement," *Proc. Siggraph Asia*, 30(6):130:1130:10, 2011.
- [4] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: Automatically replacing faces in photographs," *Proc. SIGGRAPH*, 27(3):39:139:8, 2008
- [5] D. Vlastic, M. Brand, H. Pfister, and J. Popovic, "Face transfer with multilinear models," *Proc. SIGGRAPH*, 4(3):426433, 2005.
- [6] Garrido, Pablo, et al. "Automatic face reenactment." *Computer Vision and Pattern Recognition (CVPR)*, 2014 IEEE Conference on. IEEE, 2014.
- [7] W. Matusik et al., "Image-based 3D photography using opacity hulls," *ACM Transactions on Graphics*, pp. 427-437, 2002.
- [8] C. C. Weng, H. Chen, and C.S. Fuh, "A novel automatic white balance method for digital still cameras," *Proc. of IEEE International Symposium on Circuits and Systems*, 2005.
- [9] S. Milborrow and F. Nicolls, "Active shape models with SIFT descriptors and MARS," *Proc. of VISAPP*, 2014.
- [10] L. G. Shapiro and G. C. Stockman, *Computer Vision*, Prentice Hall, pp.326-340, 2000.
- [11] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, 2001.
- [12] T. Porter and T. Duff, "Compositing Digital Images," *Proc. SIGGRAPH*, pp.253-259, 1984.
- [13] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, 57(2):137-154, 2004.
- [14] Hartley, Richard, and Andrew Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [15] J. Song, X. Zhang, H. Kim, J. Park, "Face Skin Tone Conversion Method Robust to Illumination Changes," *2014 Korean Society of Broadcast Engineers Summer Conference*, pp. 71-72, 2014.
- [16] X. Zhang, J. Song, D. Han, J. Park, "The image blending method for face swapping," *2014 Korean Society of Broadcast Engineers Summer Conference*, pp. 73-74, 2014.

---

### 저 자 소 개

#### 송 중 석



- 2010년 : 한국산업기술대학교 전자공학과 공학사
- 2010년 ~ 2012년 : 한양대학교 전자컴퓨터통신공학과 공학석사
- 2012년 ~ 현재 : 한양대학교 컴퓨터소프트웨어학과 공학박사 과정
- ORCID : <http://orcid.org/0000-0001-5639-1184>
- 주관심분야 : 3차원 영상처리/비전, 디지털 홀로그램, GPGPU

#### Xingjie Zhang



- 2012년 : 장춘대학교 컴퓨터과학과 기술학과 학사
- 2013년 ~ 현재 : 한양대학교 컴퓨터소프트웨어학과 공학석사 과정
- ORCID : <http://orcid.org/0000-0001-7393-5067>
- 주관심분야 : 영상처리/비전, 얼굴 정합, 트래킹

#### 박 종 일



- 1987년 : 서울대학교 전자공학과 공학사
- 1989년 : 서울대학교 전자공학과 공학석사
- 1995년 : 서울대학교 전자공학과 공학박사
- 1992년 ~ 1994년 : 일본 NHK방송기술연구소 객원연구원
- 1995년 ~ 1996년 : 한국방송개발원 선임연구원
- 1996년 ~ 1999년 : 일본 ATR지능영상통신연구소 연구원
- 1999년 ~ 현재 : 한양대학교 공과대학 컴퓨터공학부 교수
- ORCID : <http://orcid.org/0000-0003-1000-4067>
- 주관심분야 : 증강현실, 계산사진학, 3차원 컴퓨터비전, 인간컴퓨터상호작용