2012
Electronics and
Telecommunications
Trends

사용자 중심
차세대콘텐츠기술 특집

3D 얼굴 복원 기술 분석 및 연구 동향

A Survey and Trends on 3D Face Reconstruction Technologies

윤승욱 (S.-U. Yoon) CG기반기술연구팀 선임연구원
황본우 (B.-W. Hwang) CG기반기술연구팀 선임연구원
김갑기 (K.-K Kim) CG기반기술연구팀 선임연구원
임성재 (S.-J. Lim) CG기반기술연구팀 선임연구원
최진성 (J.-S. Choi) CG기반기술연구팀 팀장
구본기 (B.-K. Koo) 영상콘텐츠연구부 부장

* 본 연구는 문화체육관광부/지식경제부 및 한국산업기술평가관리원의 산업원천기술개발사업의 일환으로 수행하였음(KI001798, 방통융합형 Full 3D 복원 기술 개발).

최근 3DTV, 입체 모니터, 입체 노트북 등이 출시되고, 3D 영화, 게임 등 3D 관련 산업이 성장하면서 관련 콘텐츠의 요구사항이 증가하고 있다. 특히, 3D 콘텐츠의 주요 요소 중 하나인 인체는 전통적으로 고가의 3D 스캐너를 이용해 모델링하는 방식을 주로 사용해 왔다. 하지만 근래에는 광학 기술 및 컴퓨팅 성능의 향상으로 구조광과 같은 능동 센서나 카메라로부터 획득한 영상을 기반으로 3D 인체 외형을 복원하는 연구가 각광을 받고 있다. 이런 추세에 발맞춰 본고에서는 인체 중에서도 사용자의 민감도가 높은 얼굴의 3D 복원 기술 및 연구 동향을 살펴보고, 다양한 응용을 목적으로 ETRI에서 개발 중인 3D 얼굴 복원 기술을 소개하고자 한다.

Ⅰ. 개요

최근 광학 기술 및 컴퓨팅 성능의 향상으로 광학 장치를 이용해 인체를 3D로 복원하려는 시도가 증가하고 있다. 광학 장치를 이용한 인체의 3D 외형 복원은 레이저나 구조광을 이용하는 능동 센서 방식과 카메라로부터 획득한 영상을 기반으로 하는 수동 센서 방식으로 분류할 수 있다. 능동 방식은 정확도가 높아 전부터 많이 활용돼 왔으나 가격이 비싸고 사용이 어려운 단점이 있고, 수동 방식은 상대적으로 저가지만 영상의 해상도가 낮고, 처리에 많은 시간이 소요되는 문제점이 있다[1]. 하지만 최근에는 광학 소자 기술 발달이 가속화되면서 능동 센서의 가격이 낮아지고 보급이 확산되고 있다. 또한, 컴퓨팅 성능의 증가와 병렬처리 기술의 발달로 카메라의 가격 대비 해상도 및 영상 처리 속도가 향상되면서 인체의 3D 외형 복원에 관한 연구가 각광을 받고 있다.

인체의 3D 외형 복원은 크게 얼굴 복원과 몸 복원으로 나눌 수 있는데, 일반적으로 인간은 얼굴 외형의 변화에 민감한 반면 몸은 의복 착용으로 가려지는 부분이 많아 복원 품질에 덜 민감하다. 따라서 몸과 관련해서는 몸 전체의 정확한 3D 외형 복원 보다는 신체 치수 측정과 움직임 추정 및 인식 등에 관한 연구가 활발한 편이다.

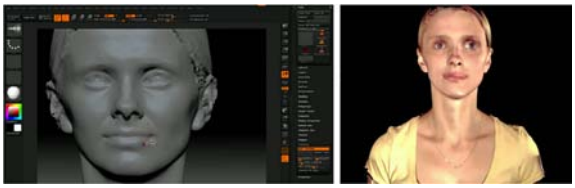
본고에서는 인체 부위 중 얼굴을 대상으로 하는 3D 얼굴 복원 기술 및 관련 연구 동향을 살펴보고, ETRI에서 개발 중인 3D 얼굴 복원 기술을 소개한다.

Ⅱ. 3D 얼굴 복원 기술 분석 및 연구 동향

1. 능동 센서 기반 3D 얼굴 복원 기술

가. 3D 스캐너를 이용한 얼굴 복원

능동 센서 방식 중 전통적으로 많이 사용해 왔고



(a) (b)

(그림 1) 스캐너를 통해 획득한 얼굴 외형

요즘에도 산업계에서 주로 사용하고 있는 것이 3D 스캐너를 이용한 방식이다. 3D 스캐너는 레이저를 물체에 투사하고 삼각측량을 통해 물체의 3D 정보를 획득하는 광학 장치로 영화, 게임과 같은 엔터테인먼트, 의학, 제품 디자인, 문화유산 보존 등 넓은 분야에서 이용되고 있다. 대상 물체의 크기에 따라 다양한 형태의 3D 스캐너가 존재하며, 기하 정보와 동시에 컬러 정보도 획득이 가능하므로 주로 물체의 정교한 3D 데이터를 얻기 위한 목적으로 사용된다[1].

최근에는 스캐너 장비 자체의 무게와 크기도 감소해 직접 손에 들고 스캔이 가능한 장비도 출시됐다. (그림 1)은 상용 핸드헬드 스캐너인 Artec MHT™ 장비를 이용해 스캔한 3D 얼굴 데이터이다[2].

일반적으로 3D 스캔한 데이터는 3D 좌표를 갖는 정점의 집합, 즉, 포인트 클라우드 형태이며, 레이저만으로는 컬러 정보를 획득할 수 없으므로 스캐너에 부착된 카메라를 사용해 스캔과 동시에 컬러 정보를 획득한다. 이렇게 획득된 포인트 클라우드 데이터는 스캐너 전용 소프트웨어를 통해 다각형 메시 모델로 변환된다. 그 후 (그림 1a)와 같이 불필요한 오류나 구멍이 생긴 부분을 처리하는 등 편집과정을 거치고, 여기에 앞서 획득한 텍스처 정보를 추가하면 최종 결과가 완성된다.

이러한 3D 스캐닝 방식은 한 번의 스캔으로는 정제된 데이터를 획득하기 어렵기 때문에 같은 물체를 여러 번 스캔하고 결과를 목적에 적합하도록 다듬는 후처리 과정이 필수적이다. 따라서 스캐너 장비뿐만 아

니라 전용 소프트웨어의 가격이 높고 이를 다룰 수 있는 전문 인력이 필요하다. 또한, 정지된 물체일 경우에는 스캔의 정확도가 높지만 움직임이 있는 경우나, 검은 머리나 눈동자 등 검은색이 존재하는 곳은 레이저가 흡수되어 스캔이 어려운 단점도 존재한다[1].

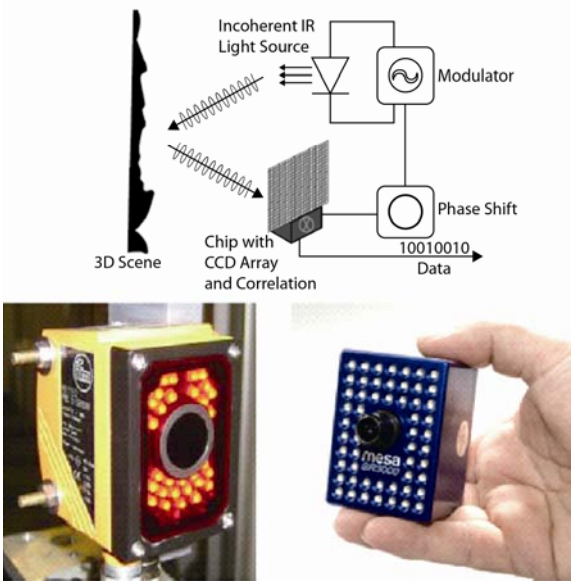
나. 깊이 카메라를 이용한 3D 얼굴 복원

광학 신호를 물체에 투사한다는 측면에서는 3D 스캐너와 유사하지만 레이저가 아닌 적외선이나 초음파 등을 이용하는 방식에 관한 연구도 지속적으로 수행돼 왔다. 이 방식은 광선을 물체에 투사하고 이들이 물체에 부딪힌 후 반사되어 돌아오면 이를 카메라 등을 통해 측정함으로써 물체의 3D 정보를 실시간으로 획득하는 접근법을 사용하며, 이를 'ToF(Time-of-Flight)' 기반 방식 또는 깊이 카메라 방식이라고 부른다. (그림 2)는 적외선의 시간에 따른 위상차(phase shift)를 이용하는 깊이 카메라의 구동방식 및 제품의 예를 나타낸다[3].

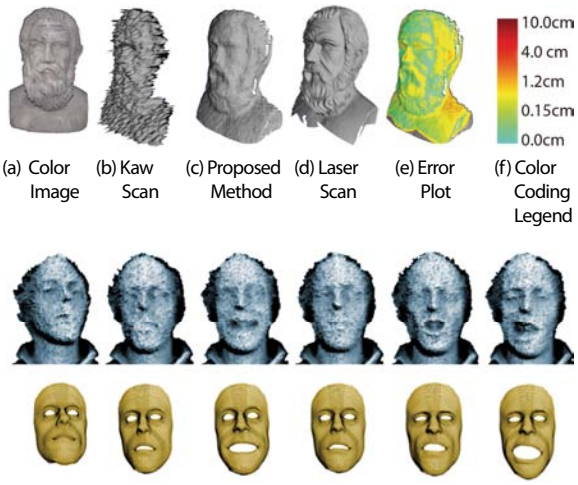
수년간 깊이 카메라의 가격이 수천만 원 이상의 고가였기 때문에 이를 활용한 연구가 범용적으로 수행되지는 않았지만, ToF 방식은 실시간으로 물체의 깊이 정보를 추출할 수 있어 3D 장면 복원 등에 사용돼 왔다. ToF 방식은 환경 제약이 상대적으로 적고, 동기신호를 받을 수 있어 송출신호 간 간섭이 없는 범위 내에서는 몇 대의 카메라를 동시에 설치해 장면을 획득할 수 있다. 하지만 잡음에 매우 민감해 후처리가 필요하며, 최근 출시된 깊이 카메라들은 대부분 해상도가 176×144 등으로 매우 낮은 단점이 있다[1].

깊이 카메라는 인체의 3D 외형 복원보다는 주로 물체 사이의 깊이 관계를 추출하거나 물체의 실시간 움직임을 유추하는 데 많이 이용돼 왔다. 특히, 근거리 물체 촬영 시에는 깊이값의 왜곡, 잡음에 민감한 센서 문제가 있어 왜곡 보정 및 후처리가 필요하다. 최근에는 (그림 3)에서 보는 바와 같이 이런 단점들을 어느 정도 해소하면서 깊이 카메라를 이용해 인체의 얼굴 외형을 3D로 복원하는 연구가 발표됐다[4],[5].

그 밖에도 깊이 카메라는 실시간 깊이 정보 획득이 가능함과 동시에 동기신호를 받을 수 있으므로, 이를 범용 산업용 카메라와 결합해 하이브리드 시스템을 구성할 수 있다. 하이브리드 시스템을 이용해 깊이 카



(그림 2) 깊이 카메라의 작동 원리 및 제품



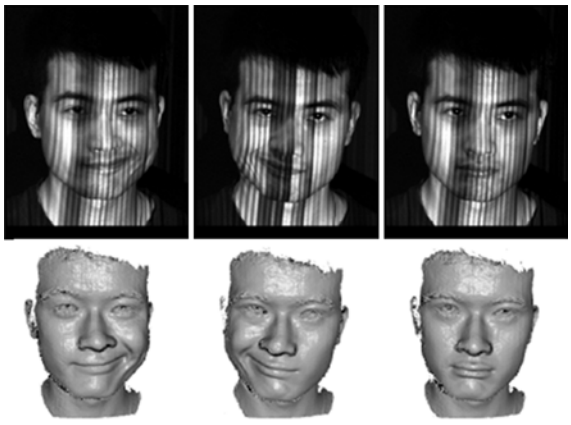
(그림 3) 깊이 카메라를 이용한 3D 얼굴 복원 결과

메라에서 획득한 깊이값을 스테레오 정합의 초기값으로 사용함으로써 스테레오 정합의 성능과 속도를 향상시키려는 연구도 진행되고 있다[6],[7].

다. 구조광 방식을 이용한 얼굴 복원

능동 센서 기반 접근 중 최근 가장 주목을 받고 있는 방법은 구조광(structured light)을 이용하는 방식이다. 이 방식은 미리 고안된 2D 패턴을 정지된 물체에 투사하고 이를 카메라로 촬영해 패턴의 변형과 왜곡을 분석해 물체의 3D 정보를 추출한다. 이때 사용되는 패턴은 이진, 컬러, 줄무늬 등 색상과 모양이 다양하며, 구조광을 이용하면 3D 스캐너에 버금가는 정확한 모델 획득 및 비교적 저가의 시스템 구성이 가능하다. 하지만 움직이는 물체는 복원이 어렵고, 패턴 영상의 크기와 해상도가 복원 정밀도에 영향을 미치므로 고성능 프로젝터가 필요하다. 또한, 프로젝터와 카메라들 사이의 카메라 보정이 필수적이다[1].

이러한 단점을 보완하기 위해 (그림 4)와 같이 패턴이 투사된 여러 프레임 정보를 사용하는 연구도 발표되었다. 시간축으로 누적된 프레임을 모두 고려해 계산하므로 많은 처리 시간이 소요되지만 정지된 물체뿐 아니라 움직이는 얼굴도 복원이 가능하다[8].



(그림 4) 구조광 기반 3D 얼굴 복원 결과

구조광 방식 중 정밀도 측면에서 가장 성능이 우수한 방식은 편광 패턴(polarized light patterns)을 이용하는 접근이다. 대표적인 연구로는 2008년 발표된 디지털 에밀리 프로젝트('The Digital Emily Project')가 있다[9]. 모델의 기본적인 외형은 줄무늬 모양의 구조광을 사용해 복원하고, 얼굴 표면의 미세한 외형 복원을 위해 LED(Light-Emitting Diode), 편광 조명 하에서 획득한 영상으로부터 정반사 및 난반사 성분을 분리해 낸다. 정반사 성분으로부터 각 화소 위치에서 빛의 법선 벡터 방향을 계산하면 얼굴의 주름까지도 복원이 가능하다[9]. 이 단계까지의 결과는 얼굴의 3D 형상만을 복원한 결과이며 여기에 정교한 텍스처를 얻기 위한 후처리를 거치고 최종 렌더링을 하게 되면 (그림 5)와 같은 디지털 액터가 탄생하게 된다. 실사 인물과의 차이를 구별하기 힘들 정도로 정밀한 결과이다.

이 방식은 입력 데이터 획득을 위한 복잡하고 정교한 구조물과 이의 운용을 위한 전문 인력이 필요하며, 각 단계별로 최종 결과물의 품질을 높이기 위한 많은 시간과 노하우가 필요하다. 따라서 복원 결과의 정확도와 품질이 가장 중요하게 취급되는 영화나 고해상도 비디오 게임 등의 응용 분야에서 주로 사용되고 있다. 결과를 통해 확인할 수 있듯이 이 연구는 능동

(그림 5) 디지털 에밀리: 최종 렌더링 결과

센서 방식을 이용한 3D 얼굴 복원 분야에서 한 획을 그었다고 볼 수 있으며, 발표 이래 많은 연구의 비교 기준으로서의 역할도 수행하고 있다[1].

한편, 3D 복원의 정확도를 최대로 높이려는 연구방향과는 달리 최근에는 Microsoft사의 Kinect가 출시되면서 실시간으로 물체의 깊이 정보를 추출할 수 있게 되었다. Kinect 역시 구조광 방식의 일종으로 랜덤 패턴을 물체에 투사하고 이를 읽어 들여 미리 정해진 규칙에 따라 분석함으로써 실재 물체의 3D 정보를 유추한다. 초당 30 프레임의 속도로 640×480 해상도의 깊이 영상과 컬러 영상을 출력하는 Kinect의 등장으로 저가의 깊이 센서 시대가 열렸다고 해도 과언이 아니다[1]. Kinect는 Microsoft사의 게임기인 Xbox 360의 부가 장치로 출시되었지만, 가격대비 성능이 우수해 2010년 제품 출시 이후 최근까지 각종 판매 기록을 갱신하며 컴퓨터 비전 분야를 비롯한 다양한 영역의 연구를 가속화시키고 있다.

Kinect는 패턴을 투사하는 프로젝터와, 투사된 패턴 영상을 획득하는 카메라, 그리고 컬러정보를 획득할 수 있는 RGB 카메라로 구성된다[10]. 하지만 인체 동작의 인식을 기본 목적으로 하므로 Kinect만으로 3D 복원을 수행하기 위해서는 이들 사이에 카메라 보정을 별도로 수행해야 한다. 또한, 측정 거리의 제약이 있어 약 1미터 이상 3.5미터 이내일 때만 깊이 정보 추출이 가능하므로, 근거리에서 사람의 얼굴에 대한 3D 정보를 획득하는 데는 어려움이 있다[1].

최근에는 앞서 언급한 이동식 3D 스캐너와 유사하게 Kinect를 손에 들고 움직이면서 정지된 물체를 여러 번 스캐닝하면 실시간으로 물체의 외형을 3D로 복원하는 연구가 발표됐다[11]. KinectFusion이라 불리는 이 방식은 부정확한 깊이값을 보상하기 위해 물체를 반복적으로 스캐닝한다. 물체의 깊이 정보를 초당 30 프레임의 속도로 얻을 수 있으므로 물체에 작은

(a) (b) (c)

(그림 6) Kinect를 통해 실시간으로 복원한 인체 상반신

움직임이 있어도 프레임 정보를 누적해 이를 일부 보상할 수 있다. 또한, 복원 결과를 사용자의 움직임에 따라 바로 확인할 수 있으므로 복원이 불완전하거나 스캔이 부족한 부분에서는 스캔 횟수를 늘림으로써 최종 복원 결과의 품질을 향상시킬 수 있다[1]. 이러한 과정을 거쳐 인체의 상반신을 3D로 복원한 결과를 (그림 6)에 나타내었다. (a)는 키넥트의 입력 데이터이며, (b)는 복원과정을 통해 계산한 노멀 맵, (c)는 복원된 모델의 쉐이딩 결과이다[11].

기존의 Kinect는 센서의 제약으로 측정 거리에 한계가 존재하지만, 2012년 출시된 Kinect for Windows는 하드웨어 성능이 향상되어 근접 모드 기능이 추가되었고, RGB 카메라의 해상도도 높아져 관련 분야의 연구가 더욱 가속화될 것으로 전망된다.

2. 수동 센서 기반 3D 얼굴 복원 기술

수동 방식은 광학 신호를 능동적으로 발신하는 것이 아니라 반대로 수신한다는 측면에서 붙여진 이름이며, 주로 카메라로 촬영한 여러 장의 영상을 분석해 얼굴의 3D 정보를 추출한다. 즉, 카메라를 이동하거나 여러 대의 카메라를 사용해 얼굴을 촬영하고 각 영상 사이의 대응 관계를 계산해 3D 정보를 유추한다. 카메라의 움직임과 위치뿐만 아니라, 화소 간 대응 관계를 추적해야 하므로 영상의 해상도와 특성에 따라 결과가 달라지며, 계산 시간도 많이 소요된다.

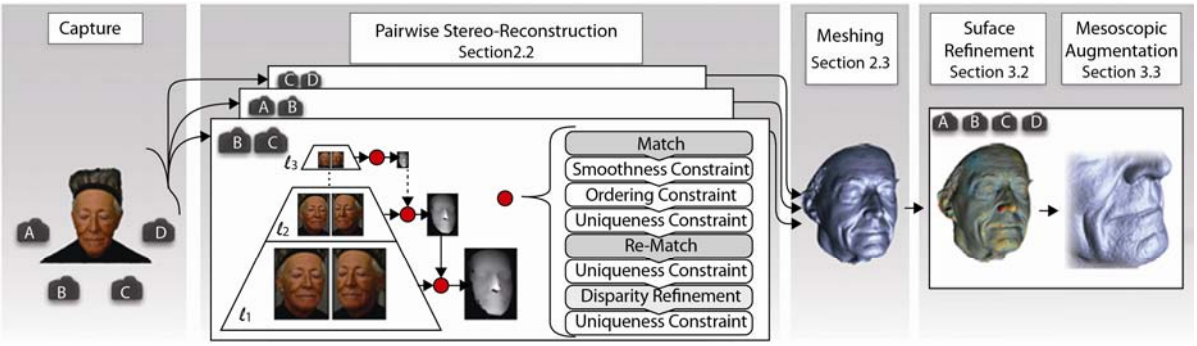
전통적으로는 주로 인체보다는 일반 물체나 건물

등을 대상으로 한 structure from motion이나 스테레오 정합, 다시점 스테레오 정합, 복셀 조각(voxel carving) 방식 등이 사용되어 왔다. 이중 3D 얼굴 복원과 관련해서는 스테레오 정합 방식이 최근 다시 각광을 받고 있다. 그 동안은 스테레오 정합 계산에 많은 부하가 소요되고, 영상의 해상도가 낮아 좌우영상 간 차이를 구별하기 어려운 점 등 문제점이 많았다. 하지만 최근 카메라 성능 향상으로 고해상도 영상 획득이 가능해지고, 병렬처리 및 GPU(Graphics Processing Unit) 기술의 발달로 계산 성능 또한 향상돼 스테레오 기반으로 인체의 얼굴을 3D로 복원하는 연구 결과가 발표되고 있다.

대표적인 연구로 여러 대의 DSLR(Digital Single-Lens Reflex) 카메라를 사용해 사용자의 얼굴을 촬영하고 인접한 두 장의 영상 단위로 스테레오 정합을 수행한 후, 이를 3D 포인트 클라우드 형태로 변환해 시점별 정합 결과를 하나로 합성(merging)함으로써 얼굴을 복원하는 방식이 있다[12]. 이 방식은 DSLR 카메라를 사용해 약 2K급 이상 해상도의 입력 영상을 획득하고, 촬영환경을 일반적인 실내로 설정해 조명 조건을 제어했으며, 얼굴에 적합한 카메라 보정용 기구를 고안해 카메라의 정확한 위치와 방향을 계산했다[1].

구체적으로 살펴보면 우선 여러 대의 카메라들로부터 획득한 두 장의 영상마다 스테레오 정합을 수행한다. 이때, 한 번만 정합을 수행하는 것이 아니라 영상 해상도를 미리 정한 몇 개의 단계로 나누고 가장 작은 해상도에서 큰 해상도로 정합을 수행하면서 정합 조건(constraint)을 확인하고 정제(refinement)하는 과정을 반복한다. 특히, 사람의 얼굴은 깊이값의 변화가 부드럽게 변하는 영역이 많기 때문에 피라미드식 다단계 정합을 통해 낮은 해상도에서의 계산 결과를 높은 해상도 계산 시 초기값으로 이용함으로써 정합 오류를 최소화할 수 있다. 전체적인 3D 얼굴 복원 과정을 (그림 7)에 나타냈다[12].

이렇게 시점별로 얼굴에 대한 깊이 맵을 계산한 후에는 카메라 보정 인자를 사용해 이를 3D 포인트 클라우드로 변환한다. 시점별로 복원된 포인트 클라우드를 정합한 후, 합성을 통해 하나의 메시 모델을 생성한다. 생성된 메시 모델의 각 정점 위치를 텍스처 정보를 기반으로 법선 방향으로 조정한다. 이와 같은 표면 정제 과정을 반복 수행하게 되면 편광 패턴을 이용한 방식과 유사하게 얼굴의 주름이나 땀구멍 수준까지 표현이 가능하다. 이때, 복원된 미세한 주름의 기하 정보는 텍스처를 기반으로 의사적으로(pseudo) 유추한 결과이므로 물리적으로 정확하다고 보기는 어렵지만, (그림 8)의 결과에서 확인할 수 있듯이 복원된 3D 얼굴의 주관적인 품질은 상당히 우수하다.



(그림 7) 다단계 스테레오 정합 기반 3D 얼굴 복원 과정

(그림 8) 다단계 스테레오 정합 기반 3D 얼굴 복원 결과

2011년 발표된 이 연구는 수동 방식을 이용해 3D 얼굴 외형을 밀리미터 수준까지 정밀하게 복원한 대표적인 예로, 스테레오 기반 방식으로도 능동 방식과 유사한 복원 결과를 얻을 수 있다는 것을 보여준다. 하지만 고해상도 영상을 사용하고 3D 표면 정제 등 복잡한 단계를 거쳐야 하므로 처리 시간은 수십 분 정도 소요된다. 또한, 여러 대의 하이엔드 DSLR 카메라의 사용으로 전체적인 시스템 구성에 필요한 가격도 수천만 원 수준으로 아직은 고가이다[1].

3. 3D 얼굴 표정 및 퍼포먼스 복원

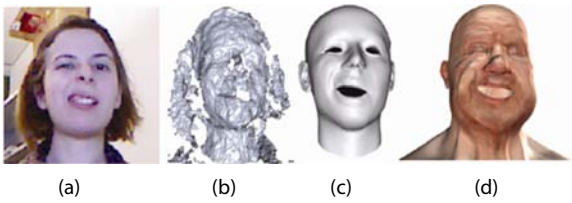
앞서 언급한 연구를 확장해 최근에는 정지된 얼굴뿐만 아니라 움직이는 얼굴 표정이나 퍼포먼스(performance)를 3D로 복원하려는 연구가 수행되고 있다[13],[14]. 기존 방식의 단점 중 하나는 얼굴 표정이 움직이는 경우 마커를 사용해 이를 추적하는 방식을 사용하지 않으면 복원이 매우 어렵다는 점인데, 제안 방식은 마커 없이 프레임 단위로 얼굴 외형을 복원함으로써 움직이는 표정을 3D로 복원한다.

참조(reference) 프레임을 선택하고 복원된 참조 프레임 3D 메시의 각 정점을 시간에 따라 추적하면서 표정의 움직임을 복원하거나, 표정이 급격하게 변하는 위치에서 기준(anchor) 프레임들을 선택해 3D로 복원하고, 중간 프레임은 보간으로 생성하는 방식이 사용된다. 3D 정점뿐 아니라 텍스처도 프레임 기반 추적을 통해 생성한다. (그림 9)는 두 방식을 사용해 얼굴 표정을 3D로 복원한 결과이다.

(그림 9) 얼굴 표정의 3D 복원 결과

한편, 사용자 얼굴 외형의 사실적 복원보다는 표정의 움직임에 초점을 맞춰 아바타가 피촬영자의 퍼포먼스를 따라하는 얼굴 애니메이션 연구도 진행 중이다[15]. 이를 위해 3D 공간에서 얼굴의 특징점들을 추적하고 아바타의 표정을 자연스럽게 생성하는 기술이 사용된다. 입력 데이터로부터 3D 얼굴을 복원한 후 템플릿 모델로 복원한 모델의 표정을 추적하고, 이를 통해 아바타의 얼굴 애니메이션을 생성한다. (그림 10)의 (a)와 (b)는 Kinect를 통해 획득한 입력, (c)는 템플릿 모델의 복원 모델 표정을 추적 결과, (d)는 이를 통해 생성된 아바타의 표정이다[15].

이렇게 3D 얼굴 복원 연구가 새롭게 각광을 받게 되면서, 스테레오 기반으로 인체 얼굴 및 상반신의 3D 외형, 움직임을 복원하는 상용 제품도 출시되었다. 영국의 Dimensional Imaging사는 여러 대의 다시점 DSLR 카메라와 산업용 카메라 기반으로 스테레오 정합, 시점별 복원 결과 정합, 합성 과정을 거쳐



(그림 10) 얼굴 표정 복원 및 애니메이션

3D 얼굴을 복원하는 제품과 복원 모델을 다른 모델로 전이하고 편집할 수 있는 제품을 판매 중이다[16].

Ⅲ. ETRI 3D 얼굴 복원 기술

앞서 설명한 바와 같이 스테레오 기반 3D 얼굴 복원 기술이 새롭게 관심을 받고 있지만, 대부분 여러 대의 DSLR 또는 산업용 카메라 기반 고가의 시스템을 사용한다. 또한, 배경과 조명 조건의 제약으로 적어도 $2m^2$ 이상의 설치 공간이 필요해 이동이 쉽지 않은 단점이 있다. 이는 관련 연구가 주로 높은 해상도와 정밀도를 갖는 3D 복원 결과 획득, 즉, 영화나 게임 산업에서 사용 중인 3D 스캐너나 구조광 기반 시스템을 대체하기 위해 진행되어 온 이유도 있다.

최근에는 이런 연구방향과는 달리 모바일폰이나 디지털 컴팩트 카메라 등을 통해 영상을 획득하고 이로부터 간편하게 3D 얼굴을 복원하려는 시도도 많다. 하지만 흥미 위주의 시도를 넘어 다양한 산업 분야에서의 쓰임새를 고려하면 응용에 맞는 품질 조절이 필요하다. 또한, 시스템의 설치 제약이 적고 이동이 용이하며, 작업의 간편성을 확보할 수 있는 저가형 3D 얼굴 복원 시스템에 대한 요구사항도 존재한다.

예로 미용 분야에서는 아직까지 사진을 중심으로 피부 분석 등을 수행하고 고객에게 정보를 제공했으나, 요즘에는 이를 3D로 확장하려는 움직임이 증가하고 있다. 성형 등 의료 분야에서도 성형 전후 환자의 모습을 짧은 시간에 3D로 복원해 상담에 이용하려는 시도가 늘고 있다. 이러한 다양한 응용을 목적으로 ETRI에서는 웹캠, 산업용/CMOS(Complementary Metal-Oxide Semiconductor) 카메라 등 입력 장치를 다변화하고 응용에 따라 3D 얼굴 복원 모델의 해상도를 조절할 수 있는 기술을 연구 중이다[1].

ETRI에서 연구 중인 3D 얼굴 복원 기술은 스테레오 영상을 입력으로 사용한다. 시점의 개수, 즉, 스테레오 카메라의 개수는 응용에 따라 달라질 수 있으며, 두 시점 이상일 경우 전체적인 3D 얼굴 복원 과정을 (그림 11)에 나타냈다. 우선 각 시점별로 스테레오 정합을 통해 변위 맵을 계산하고, 이를 카메라 보정 인자를 사용해 격자형 3D 메시로 변환한다. 모든 변환된 메시를 공통 좌표계에 위치시킨 후 이들을 합성해 하나의 3D 모델을 생성한다. 여기에 입력 영상과 카메라의 위치 및 방향 정보 등을 이용해 텍스처 맵을 추가함으로써 3D 얼굴 복원 결과를 얻게 된다.

(그림 11) 스테레오 기반 얼굴 복원 과정

단계별로 살펴보면 우선 입력 영상은 웹캠, 산업용 카메라 혹은 CMOS 카메라 등 다양한 카메라로 촬영할 수 있으며, 제안 시스템에서는 최종 텍스처 품질을 고려해 약 2K급 정도 해상도를 갖는 영상을 사용한다. 해상도가 너무 작으면 모델의 세부를 복원하기가 어렵고, 해상도가 너무 크면 계산 시간이 증가한다.

입력 영상에서 얼굴 영역만을 분리하기 위해 크로마키 배경이 사용될 수 있다. 제안 시스템에 사용된 스테레오 정합 기술은 조명 변화에 강인해 실내 자연조명 환경에서도 구동이 가능하며, 필요에 따라 적절한 지속광 또는 스트로브 조명 이용도 가능하다. (그림 12a)는 제안 시스템을 이용해 자연조명 환경에서 얼굴 데이터를 획득하는 예를 나타낸다. 카메라의 속

(a) (b)

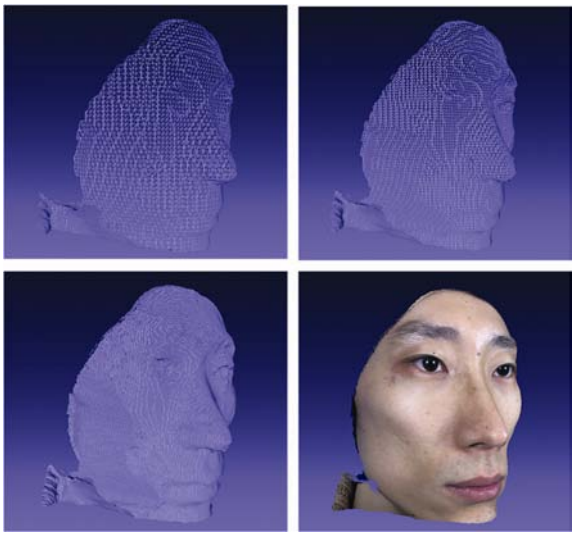
(그림 12) 자연조명 환경에서 얼굴 데이터 획득 예(a) 및 카메라 보정 구조물(b)

성 및 위치와 방향 정보를 계산하는 카메라 보정 작업은 일반적인 격자 패턴 또는 동심원 패턴 등을 사용해 수행할 수 있다. 이 경우는 보정 영상 촬영과 처리에 많은 시간이 소요되는 단점이 있어, ETRI에서는 (그림 12b)와 같이 자체적으로 26면체 형태의 동심원 패턴을 개발하였다. 카메라별로 세 장의 영상만 획득하면 수작업 없이 일분 내외의 짧은 시간에 카메라 보정 인자를 얻을 수 있다.

변위 맵 생성은 스테레오 정합을 통해 좌우 영상의 화소별 대응 관계를 계산하는 과정이다. 영상 해상도에 따라 단계별로 정합을 수행하면 저해상도에서의 정합 결과를 고해상도 정합의 초기값으로 사용함으로써 검색 범위를 줄이고, 정확도를 높일 수 있으며 계산 시간도 단축할 수 있다[12].

시점별 복원은 카메라의 보정 인자와 계산한 변위 맵을 입력으로 삼각법을 이용하여 영상 위의 화소를 3D 카메라 좌표계 위의 한 점으로 변환해 격자형 3D 메시를 구하는 과정이다[1].

정합 및 합성 단계에서는 우선 생성된 격자형 메시를 카메라 외부 인자를 이용하여 공동 좌표계 위에 대략적으로 위치시키고 겹치는 영역이 최대가 되도록 위치와 방향을 조절한다. 그 후 정합을 통해 한 좌표계로 이동된 격자형 메시들을 합성해 하나의 3D 메시 모델을 생성한다. 제안 시스템에서는 오류에 강인하면서도 단순한 체적 합성 방법을 사용한다. 통합된 3D 메시를 생성할 공간을 복셀로 구성하고, 각 복셀 내에 존재하는 객체의 등가면(iso-surface)을 찾는다. 얻어진 등가면에 Marching Cubes 알고리즘을 적용해 체적 모델을 3D 메시로 변환한다[1],[17].



(그림 13) ETRI 3D 얼굴 외형 복원 결과

3D 외형 데이터 출력은 합성된 메시에 텍스처를 맵핑하여 일반적으로 사용되는 그래픽 데이터 파일로 출력하는 과정이다. 각 정점을 모든 카메라 시점으로 투사해 가장 적절한 시점을 선택하고 선택된 시점에서 텍스처를 가져오는 방식을 사용한다. 최종 생성된 데이터는 3D 그래픽스 파일 형식으로 저장된다. (그림 13)의 상단과 좌하단의 얼굴 메시 모델은 제안 시스템을 통해 생성된 다양한 품질별 복원 결과이며, 우하단은 좌하단 고품질 메시 모델에 텍스처를 추가한 최종 3D 얼굴 복원 결과이다.

IV. 맺음말

본고에서는 전통적으로 많은 연구가 수행되어온 능동 센서 방식과 최근 다시금 각광을 받고 있는 스테

레오 기반 3D 얼굴 복원 기술 및 연구 동향에 대해 소개하였다. 또한, 다양한 응용을 위해 ETRI에서 연구 중인 3D 얼굴 복원 시스템에 대해 기술하였다. 현 추세대로 광학 소자 및 카메라의 성능이 지속적으로 향상되고, 대용량 데이터의 처리 속도가 계속 증가하면, 향후에는 훨씬 저가의 예산으로 시스템 구축이 가능해져 관련 응용 분야가 더욱 확대될 것으로 전망된다.

용어해설

스테레오 정합 좌우 영상의 각 화소 간 대응 관계를 계산하는 기법

변위 맵 스테레오 정합을 통해 계산한 화소별 변위값을 이차원 영상 포맷으로 저장한 결과

정합 두 시점 이상의 포인트 클라우드 또는 메시 데이터를 공통 좌표계로 이동시켜 겹치는 영역이 최대가 되도록 위치 또는 방향을 조절하는 기법

퍼포먼스 캡처 얼굴 표정의 변화를 센서나 비전 기술을 이용하여 감지해낸 뒤 디지털로 옮기는 기술

템플릿 모델 사용 용도에 맞게 정규화된 기하 모델과 텍스처 맵으로 구성된 3D 메시 모델

전이 3D 메시 모델의 기하 모델 및 텍스처 맵을 목표로 하는 3D 모델에 정합 및 피팅시키는 기법

약어 정리

CMOS	Complementary Metal-Oxide Semiconductor
DSLR	Digital Single-Lens Reflex
GPU	Graphics Processing Unit
LED	Light-Emitting Diode
ToF	Time-of-Flight

참고문헌

[1] 윤승욱, 황본우, “다시점 영상을 이용한 3D 복원 기술,” 방송과 기술, 2012. 3, pp. 136-145.

[2] http://www.artec3d.com

[3] A. Kolb, E. Barth, and R. Koch, “ToF-Sensors: New Dimensions for Realism and Interactivity,” *Comput. Vision Pattern Recognit., Workshops*, June 2008, pp. 1-6.

[4] Y. Cui et al., “3D Shape Scanning with a Time-of-Flight Camera,” *Comput. Vision Pattern Recognit.*, June 2010, pp. 1173-1180.

[5] M. Breidt, H.H. Bulhoff, and C. Curio, “Face Models from Noisy 3D Cameras,” *ACM SIGGRAPH ASIA Sketches*, 2010.

[6] E.-K. Lee and Y.-S. Ho, “Generation of High-Quality Depth Maps Using Hybrid Camera System for 3-D Video,” *J. Visual Commun. Image Representation*, vol. 22, no. 1, 2011, pp. 73-84.

[7] J. Zhu et al., “Reliability Fusion of Time-of-Flight Depth and Stereo Geometry for High Quality Depth Maps,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, July 2011, pp. 1400-1414.

[8] L. Zhang, B. Curless, and S.M. Seitz, “Spacetime Stereo: Shape Recovery for Dynamic Scenes,” *Comput. Vision Pattern Recognit*, June 2003, pp. 367-374.

[9] O. Alexander et al., “Creating a Photoreal Digital Actor: the Digital Emily Project,” *Eur. Conf. Visual Media Production*, 2009.

[10] http://www.microsoft.com

[11] R.A. Newcombe et al., “KinectFusion: Real-Time Dense Surface Mapping and Tracking,” *Int. Symp. Mixed and Augmented Reality*, Oct. 2011.

[12] T. Beeler et al., “High-Quality Single-Shot Capture of Facial Geometry,” *ACM Trans. Graphics*, vol. 29, no. 4, July 2010, pp. 40:1-40:9.

[13] D. Bradley et al., “High Resolution Passive Facial Performance Capture,” *ACM Trans. Graphics*, vol. 29, no. 4, July 2010, pp. 41:1-41:10.

[14] T. Beeler et al., “High-Quality Passive Facial Performance Capture using Anchor Frames,” *ACM Trans. Graphics*, vol. 30, no. 4, July 2011, pp. 75:1-75:10

[15] T. Weise et al., “Realtime Performance-Based Facial Animation,” *ACM Trans. Graphics*, vol. 30, no. 4, July 2011, pp. 77:1-77:10.

[16] http://www.di3d.com

[17] W.E. Lorensen and H.E. Cline, “Marching Cubes: A High Resolution 3D Surface Construction Algorithm,” *ACM SIGGRAPH Comput. Graphics*, vol. 21, no. 4, 1987.