실감형 원격 영상회의를 위한 시선 맞춤 기술

Ⅰ. 서 론

원격 영상회의는 자신의 모습을 카메라로 촬영하고 음성을 녹음하여 상대방에게 보내고, 디스플레이를 통해 멀리 떨어져 있는 상대방의 모습을 보고 음성을 들으며 마치 한 회의실에 함께 있는 분위기로 회의를 진행하는 것을 말한다. 원격 영상회의 시스템은 기본적으로 회의를 하는 양쪽의 회의실에 카메라, 모니터, 마이크, 스피커를 설치하고 양자간을 쌍방향 영상회선 및 음성회선, 또는 전용회선으로 연결한다. 최근에는 개인용 컴퓨터로도 원격 영상회의를 할 수 있는 시스템이 개발되었으며, 컴퓨터 내에 있는 자료를 화면상에 서로 공유해 가면서 회의를 진행할 수도 있고, 쌍방의 모니터 화면에 전자펜을 이용하여 글씨나 그림을 그릴 수 있는 기능도 활용되고 있다. 〈그림 1〉은 원격 영상회의 시스템을 개발한 대표적인 기관들의 원격 영상회의 솔루션을 보여준다.

실시간 영상, 음성 데이터의 교환으로 마치 한 회의실에 함께 있는 분위기로 회의를 진행하는 원격 영상회의에서 시선 맞춤(eye contact) 기술은 기술 구현에 있어서 가장 중요한 기술 중 하나이다.

시선 맞춤(eye contact) 기술은 원격 영상회의 시스템의 구현에 있어서 가장 중요한 기술 중 하나이다. 기존의 원격 영상회의 시스템은 일반적으로 카메라가 디스플레이의 위쪽이나 아래쪽에 위치하였고, 따라서 화자의 시선과 카메라의 렌즈의 위치가 달라 〈그림 2〉와 같이 상대방과의 시선 불일치가 발생한다. 시선 불일치 문제는 화자끼리의 대화의 집중을 떨어뜨리고 몰입감을 떨어뜨리기 때문에 이러한 문제를 해결하기 위해 많은 연구 기관들에 의해 시선 맞춤을 위한 연구가 진행되었다. 하지만, 현재까지 진행된 선행 연구들을 살펴보면, 성능에 비해 하드웨어 구성이 너무 복잡하고, 시스템 구축 비용이 많이 든다는 단점을

호 요 성
광주과학기술원
정보통신공학부

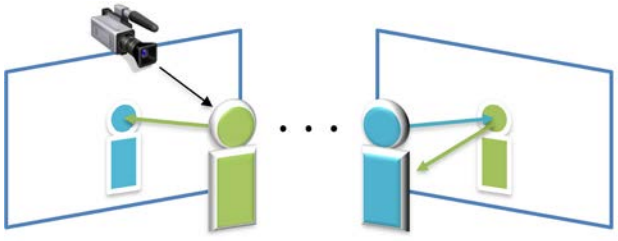
〈그림 1〉 대표적인 원격 영상회의 시스템

갖고 있다.

이와 같이 비용이 많이 드는 단점을 극복하면서 시선 맞춤을 해결하기 위해서 3차원 비디오 시스템 기술을 이용할 수 있다. MPEG (moving picture experts group)에서는 〈그림 3〉과 같은 3차원 비디오 시스템을 정의하고, 다시점 색상 영상과 깊이 영상을 포함하는 3차원 비디오의 부호화에 관한 국제 표준화 작업을 진행하고 있다. MPEG에서 정의한 3차원 비디오 시스템은 3시점 혹은 그 이상의 넓은 시야각을 제공하는 고해상도의 3차원 비디오 시스템을 말한다.

3차원 비디오 시스템의 구현을 위해서는 여러 대의 카메라로 획득한 넓은 시야각의 다시 점 영상을 이용해서 3차원 장면의 거리 정보를 표현하는 깊이 맵을 제작하는 기술과 제작된 깊이 맵을 이용하여 사용자가 원하는 시점에서 장면을 시청할 수 있게 하는 중간시점 영상 합성 기술이 사용된다. 즉, 3차원 비디오 시스템의 핵심 기술 중 하나인 중간시점 영상 합성을 이용하면 화자의 시선이 상대방과 일치하는 시점에서의 영상을 만들어 낼 수 있고, 이것이 실감형 원격 영상회의에 사용될 수 있다.

본 논문에서는 실감형 원격 영상회의를 위한 시선 맞춤 기술 동향에 대해 소개한다. 먼저 원격 영상회의 기술의 국내외 동향을 살펴보고, 3차원 비디오 시스템의 핵심 기술인 깊이 맵 생성과 중간시점 영상 합성 기술을 이용한 시선 맞춤 기술을 설명한다.

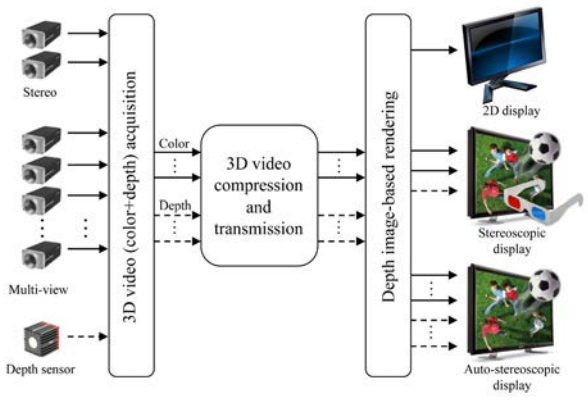


〈그림 2〉 시선 불일치 문제

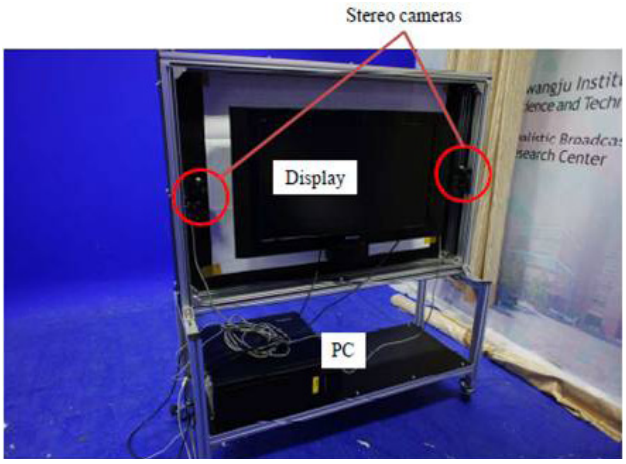
Ⅱ. 실감형 원격 영상회의 시스템의 국내외 기술 동향

1. 국내 기술 동향

시선 맞춤 기술과 관련하여 1997년 삼성전자에서는 3대의 카메라를 이용하여 촬영된 영상들에 대한 이미지 모델을 생성하고 디스플레이의 제어 하에 각 회의 참가자들에게 시선 맞춤이 가능한 영상을 출력하는 기술을 개발하였다.[1] 2007년에는 이화여자대학교에서 촬영된 얼굴 영상을 타원으로 근사화시키고 이를 3차원 공간에서의 타원체로 모델링하여 임의의 가상 시점의 위치에서 영상을 합성하는 알고리즘을 개발하였다.[2] 2011년 광주과학기술원에서는 깊이 맵과 중간시점 영상 합성 기술을 이용



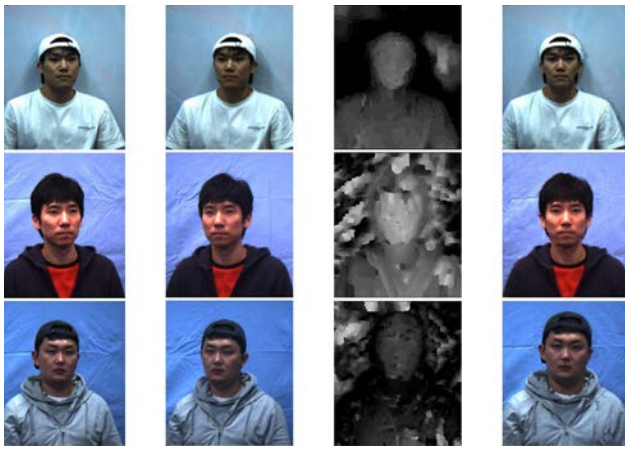
〈그림 3〉 3차원 비디오 시스템



〈그림 4〉 광주과학기술원의 시선 맞춤 시스템

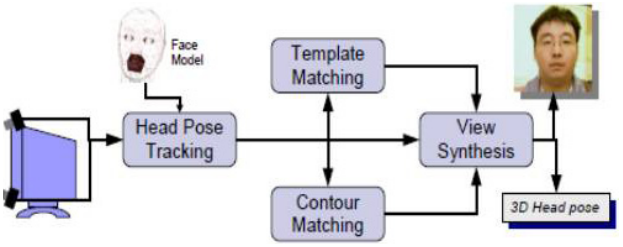
한 시선 맞춤 기술을 제안하였다.[3] 〈그림 4〉와 같이 두 대의 카메라를 디스플레이의 왼쪽과 오른쪽에 각각 수렴 형태로 배치하고, 여기서 촬영된 두 시점의 영상을 이용하여 시선이 맞추어진 중간의 임의 시점에서의 깊이 맵을 생성한다. 생성된 깊이 맵을 이용하여 좌우 양쪽 시점으로부터 색상 정보를 중간 시점으로 가져옴으로써 시선이 맞추어진 영상을 획득한다.

〈그림 5〉는 광주과학기술원에서 제안한 시선 맞춤 기술을 이용하여 얻은 중시점 영상을 보여준다. 좌영상과 우영상으로부터 중간 시점의 깊이 정보를 획득하고, 이



〈그림 5〉 시선 맞춤 결과

복잡 구현도가 비교적 간단하고 높은 시선 맞춤 성능을 보여주는 3차원 비디오 시스템 기술은 국내의 광주과기원, 미국의 Microsoft, 독일의 Fraunhofer HHI와 같은 기관/회사를 중심으로 연구/개발되고 있다.



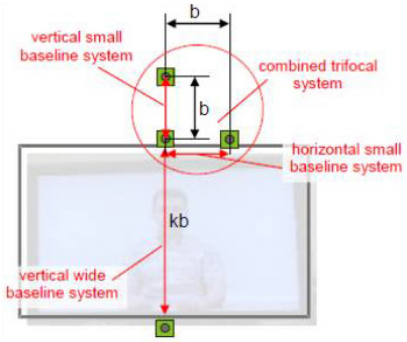
〈그림 6〉 Microsoft 연구소의 시선 맞춤 기술

깊이 정보를 이용하여 합성되어 시선이 교정된 영상을 각각 보여준다.

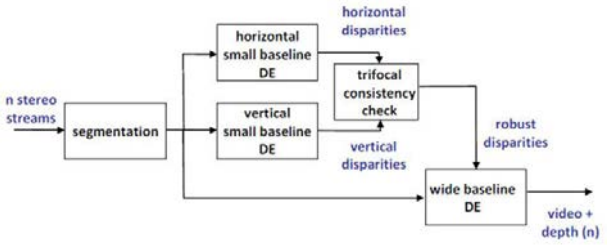
이 방법은 3차원 비디오 시스템의 핵심 기술들을 이용하여 구현이 용이한 장점이 있으나, 같은 카메라 배열은 카메라 간 간격이 너무 멀고 각도 차이가 크게 발생하기 때문에 영상에서 폐색 영역의 범위가 넓게 나타나고, 상대적으로 깊이 맵의 품질이 떨어지게 된다. 또한 스테레오 정합(stereo matching)을 이용하여 깊이 맵을 생성하기 때문에 원격 영상회의에서 가장 중요한 정보인 얼굴의 깊이 정보를 세밀하게 구할 수 없다는 단점이 있다.

2. 해외 기술 동향

해외에서도 실감형 원격 영상회의를 위한 깊이 정보 기반의 시선 맞춤 기술이 활발하게 개발되고 있다. 먼저 Microsoft 연구소에서는 2004년 개인화된 얼굴 모델과 얼굴을 중심으로 촬영된 양안식 영상을 이용하여 중간시점 영상을 합성하는 기술을 개발하였다.[4]



〈그림 7〉 HHI 시선 맞춤 시스템



〈그림 8〉 HHI 시선 맞춤 기술

Microsoft 연구소의 기술은 〈그림 6〉과 같이 배경 모델 획득, 얼굴 추적 초기화, 배경 분리, 시간 특성 추적, 머리 위치 추적, 스테레오 정합, 실루엣 정합, 그리고 시선 맞춤된 시점 합성의 단계로 구성된다.

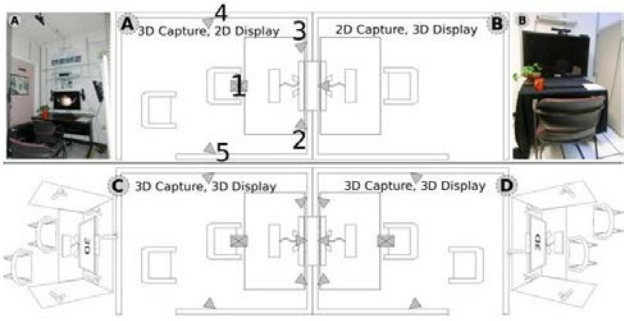
독일의 Fraunhofer HHI는 2009년 〈그림 7〉과 같이 L자형으로 배치된 3대의 카메라와 디스플레이 하단에 위치한 한 대의 카메라에서 촬영된 영상으로부터 스테레오 정합과 시각체(visual hull) 알고리즘을 사용하여 영상을 합성하는 시선 조정 기술을 개발하였다.[5]

위 시스템에서는 〈그림 8〉과 같이 먼저 입력 영상을 배경으로부터 분리하고, 가까운 카메라 간격에 대해 각 시점별로 깊이 정보를 구한 후 깊이의 상관도를 측정한다. 먼 카메라 간격에 대해서도 각 시점별로 깊이 정보를 구한 후, 최종 깊이 맵을 업데이트한다.

한편, HHI에서는 최근 〈그림 9〉와 같이 4대의 카메라를 배치하고, 화자의 시선을 추적하여 실시간으로 사용자가 바라보는 위치의 영상을 합성하는 기술을 개발하였다.

미국 노스캐롤라이나대학에서는 다수의 깊이 카메라를 이용하여 3차원 공간을 복원한 후, 화자 간 양방향 통신을 통한 영상회의 시스템을 개발하였다.[6] 〈그림 10〉은 노스캐롤라이나대학에서 개발한 원격 영상회의 시스템을

〈그림 9〉 HHI에서 새롭게 제안한 시선 맞춤 기술 시연



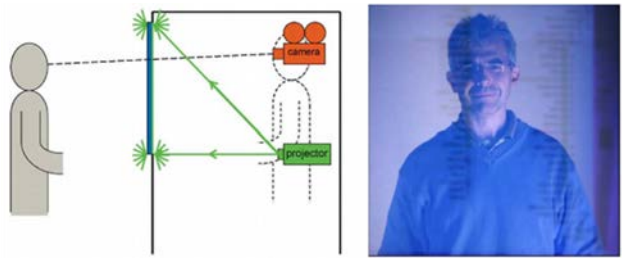
〈그림 10〉 노스캐롤라이나대학의 원격 영상회의 시스템

보여준다. 장면 내의 모든 화자를 3차원 모델로 구성하기 때문에, 복잡도가 높은 부분은 GPU를 사용하여 실시간으로 처리한다. 이렇게 제작된 3차원 배경과 화자는 3차원 디스플레이를 통해 상대방에게 보여진다.

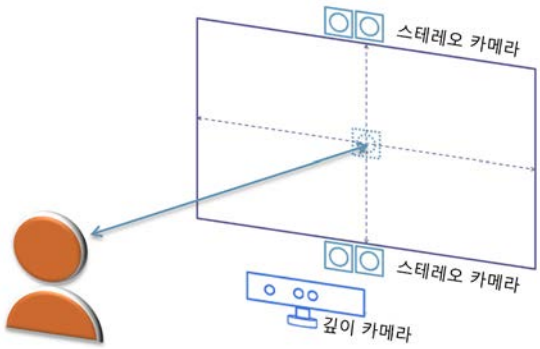
스위스의 ETH에서는 2011년 Kinect를 이용하여 얼굴을 3차원으로 모델링하고, 생성된 모델을 이용하여 정면을 응시하도록 하는 기술을 개발하였다.[7] 〈그림 11〉은 ETH의 시선 맞춤 기술로써 Kinect로부터 촬영된 색상 영상과 깊이 맵으로 3차원 모델을 구성하고, 타원형 3차원 얼굴 모델에 변환을 수행하여 시선이 정면을 향하는 영상을 원래의 영상에 합성한다.

한편, 깊이 정보 기반의 합성이 아닌 하드웨어적인 구현으로 시선 맞춤을 수행하기도 하는데, 그 중 하나는 〈그림 12〉에 보인 것과 같은 반투명 투과 디스플레이 기반의 원격 협력 시스템이다.[8] 광 필드 렌더링 기법을 사용하는 방법도 제안되었다.[9] 하지만 이와 같은 시스템은 구축 비용이 높고, 화질의 저하가 나타난다. 〈그림 1〉에 보인 Cisco, Polycom, HP의 원격 영상회의 기술도 디스플레이 사이에 배치한 카메라로 촬영한 영상을 이용한 것으로써 기본적으로 하드웨어 기반의 방식이다.[10-12]

〈그림 11〉 ETH의 시선 맞춤 기술



〈그림 12〉 반투명 투과 디스플레이를 이용한 시선 맞춤



〈그림 13〉 원격 영상회의 시스템

Ⅲ. 양안식 카메라와 깊이 카메라를 이용한 시선 맞춤 기술

본 절에서는 양안식 카메라와 깊이 카메라를 이용하여 실감형 원격 영상회의를 위한 시선 맞춤 기술을 구현하는 방법을 소개한다. 먼저 영상획득을 위한 카메라 시스템이 설치되어야 하며, 촬영된 영상의 전처리와 깊이 정보 생성, 그리고 정면 시점 영상 합성의 단계를 거치면 시선 맞춤된 영상을 획득할 수 있다.

본 기고문에서는 양안식 카메라와 깊이 카메라를 이용하여 실감형 원격 영상회의를 위한 시선 맞춤 기술을 구현하는 방법을 소개한다. 먼저 영상획득을 위한 카메라 시스템이 설치되어야 하며, 촬영된 영상의 전처리와 깊이 정보 생성, 그리고 정면 시점 영상 합성의 단계를 거치면 시선 맞춤된 영상을 획득할 수 있다,

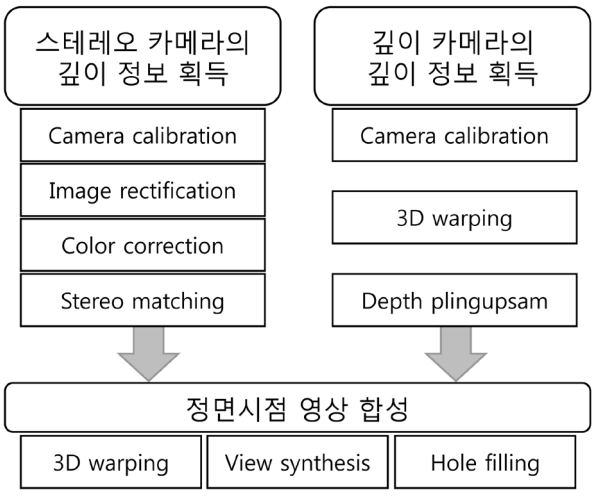
1. 시스템 구조 및 시선 맞춤

〈그림 13〉은 실감형 원격 영상회의 시스템에서의 시선 맞춤 기술 구현을 위한 시스템을 보여준다. 대형 디스플레이의 위쪽과 아래쪽 중앙에 각각 스테레오 카메라가 설치되어있고, 디스플레이보다 약간 앞쪽으로 깊이 카메라가 설치되어 있다. 디스플레이와 화자와의 거리는 약 2m이며, 실험에 사용한 디스플레이의 크기는 55인치이다. 두 세트의 스테레오 깊이 카메라와 함께 카메라를 사용한 이유는 조금 더 정밀한 깊이 정보를 예측하기 위함이다. 각 스테레오 카메라 세트 중에서 정면 시점에 사용하는 카메라는 한 대씩이며, 이들은 정면 시점과 동일한 수직선상에 위치한다.

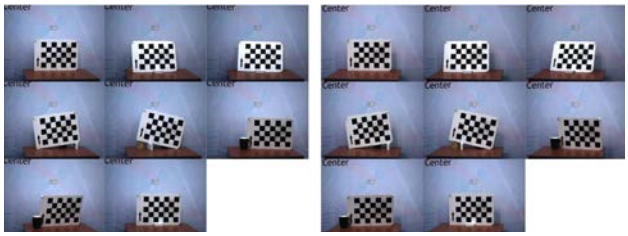
촬영에 있어서는 장면 내 화자를 한 명으로 제한하였으며, 화자의 전후 움직임 및 화자 주변 객체의 움직임도 제한하였다. 또한 배경의 변화가 없도록 촬영하였다.

〈그림 14〉는 시선 맞춤 기술의 전체 흐름을 보여준다.[13] 먼저 스테레오 카메라 세트와 깊이 카메라는 각각 독립적으로 깊이 정보를 계산한다. 깊이 정보의 계산을 위해서 각 카메라에서는 카메라 매개변수를 추출하고, 각 영상에 대한 교정 작업이 수행된다. 스테레오 카메라 세트에서는 스테레오 정합을 이용하여 깊이 정보를 계산하며, 획득된 깊이 정보는 3차원 워핑을 통해 정면 시점으로 이동된다.

깊이 카메라에서는 3차원 워핑을 이용하여 정면 시점으로 바로 깊이 정보를 이동시킨다. 이 때, 깊이 카메라의 해상도는 정면 시점의 해상도보다 낮기 때문에 깊이 맵을 업샘플링(upsampling)하는 과정이 반드시 필요하다. 깊이 카메라는 해상도가 낮기는 하지만 스테레오 정합



〈그림 14〉 스테레오 카메라와 깊이 카메라를 이용한시선 맞춤 기술



〈그림 15〉 스테레오 카메라에서 촬영된 격자 무늬 패턴

으로는 얻을 수 없는 세밀한 깊이 정보를 얻을 수 있기 때문에 촬영된 영상에서 가장 중요한 화자의 얼굴 부분 깊이 정보를 개선하는데 사용된다.

각 카메라에서 이와 같이 얻어진 깊이 정보들은 정면 시점에서 통합되고, 이를 바탕으로 텍스쳐(texure) 정보를 입히면 시선이 맞춰진 정면 시점의 영상을 얻을 수 있다.

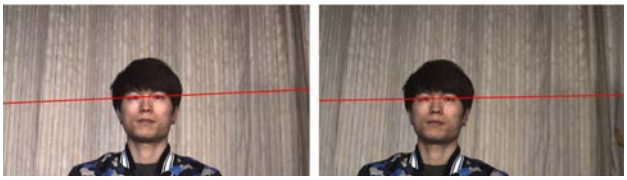
2. 카메라 보정

촬영에 사용된 각 카메라는 카메라 보정(camera calibration) 과정을 통해 카메라 매개변수(camera parameters)를 획득한다. 격자 무늬의 패턴을 다양한 자세로 변화시키며 〈그림 15〉와 같이 촬영하고, 그 영상에서 추출된 특징점을 기반으로 카매라의 내부 변수와 외부 변수를 예측한다. 카메라의 내부 변수는 초점거리나 주점 좌표와 같이 카메라 내부의 물리적 특성을 나타내는 값들로 이루어진 행렬로 표현되며, 카메라의 외부 변수는 3차원 공간상에서 카메라의 방향과 위치를 가리키는 회전 행렬과 이동 벡터로 이루어진다. 카메라의 내부 및 외부 변수를 이용하면 카메라의 투영 행렬을 계산할 수 있고, 이 투영 행렬은 3차원 공간의 한 점을 2차원 영상 평면의 한 화소로 옮겨오는 역할을 한다.

카메라 보정을 통해 얻은 카메라 매개변수는 3차원 영상처리 및 응용에서 가장 기본이 되는 필수적인 정보이다. [14]

3. 스테레오 영상 정렬

평행형 스테레오 카메라에서 촬영된 영상에는 일반적으로 기하학 오차가 존재한다. 이 오차는 카메라를 수동으로 배치할 때 발생하는 위치 및 방향의 오차 등으로 인한 것으로, 두 시점의 영상에서 대응점들의 수직 좌표의 위치 차이로 나타난다. 또한 카메라 보정을 통해 얻어진



〈그림 16〉 스테레오 영상에서의 에피폴라 선

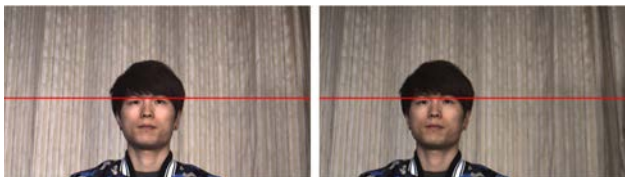
카메라의 내부 변수에도 차이가 발생한다. 이러한 오차는 깊이 정보를 생성하고 정면시점 영상을 합성함에 있어서 품질을 떨어뜨리는 원인이 된다. 따라서 영상 정렬(image rectification)을 통해 스테레오 영상에 존재하는 기하학 오차를 바로잡아야 한다.[15] 〈그림 16〉은 기하학 오차가 존재하는 스테레오 영상에서 구한 에피폴라 선(epipolar line)을 보여준다.

만약 두 시점의 영상이 정렬된 상태라면, 즉 두 시점의 모든 내부 변수가 동일하고 카메라의 위치 차가 수평 이동만 존재하며 카메라의 방향이 같다면 에피폴라 선은 서로 평행하며 동일한 직선을 지나야 한다. 따라서 〈그림 16〉은 스테레오 영상이 기하학 오차를 가지고 있음을 나타낸다. 스테레오 영상 정렬은 카메라 매개변수를 이용하여 구한 변환행렬을 각 시점에 적용함으로써 두 시점의 카메라의 내부 변수와 회전 행렬을 통일하고, 카메라가 동일 선상에 위치하도록 변환한다. 영상 정렬의 결과는 〈그림 17〉과 같이 두 시점에서의 에피폴라 선이 동일한 직선상에 위치하게 되며, 기하학 오차가 보정되어 대응점의 수직 좌표가 동일하게 맞춰지고 수평 방향으로의 변위만 가지게 된다.

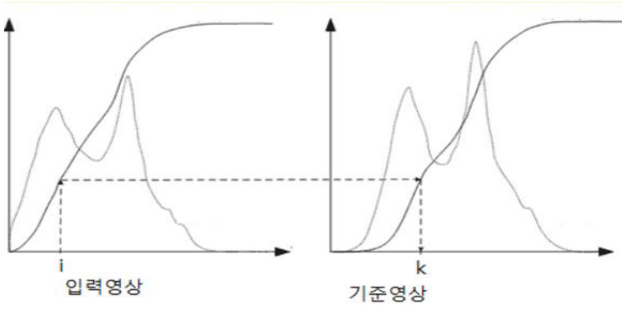
4. 스테레오 영상의 색상 보정

촬영된 스테레오 영상에는 시점 간 색상 불일치가 존재한다. 이는 단순히 디지털 카메라의 필름 역할을 하는 전하 결합소자(CCD)나 상보성 금속 산화막 반도체(CMOS)의 미세한 차이에서 기인한 것일 수도 있고, 카메라의 셔터 속도나 조리개, 혹은 초점거리의 차이에 의해 발생한 것일 수도 있다. 즉, 두 시점 같 색상 불일치 문제는 카메라의 반도체나 회로의 전자적 특성 차이 뿐 아니라 기계적 특성 차이로도 발생할 수 있는 것이다.[16]

스테레오 영상의 시점 간 색상 불일치는 사용자가 시점



〈그림 17〉 영상 정렬 결과

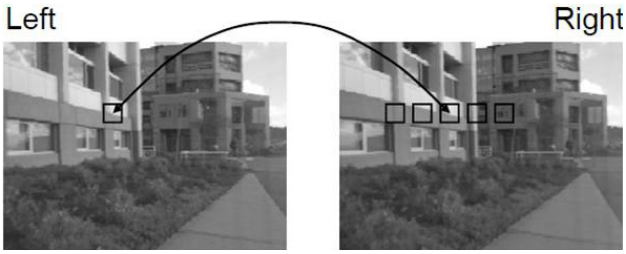


〈그림 18〉 히스토그램 매칭

을 이동하면서 볼 때 부자연스러움을 느끼는 것은 물론 색상 정보 기반으로 진행되는 영상처리의 성능 저하를 일으킨다. 즉, 깊이 정보 생성을 위한 스테레오 정합의 정확도를 감소시키며, 정면시점 영상을 합성할 때 색상 차이로 인한 얼룩을 발생시키기도 한다. 따라서 이러한 스테레오 영상에 존재하는 색상 차이를 반드시 제거해야 한다.

색상 불일치 문제를 해결하기 위한 대표적인 색상 보정으로는 색상 차트를 이용하는 방법이 있다. 이 방법은 주로 본 영상을 촬영하기에 앞서 색상 차트를 미리 촬영하고, 이를 기반으로 색상을 보정하는 방법이다.[17] 하지만 색상 차트가 촬영되지 않은 영상에는 적용할 수 없기 때문에 원격 영상회의 시스템에는 적합하지 않다고 볼 수 있다.

히스토그램 매칭(histogram matching) 방법은 시점별로 취도와 색차 성분이 서로 일치해야 한다는 가정 하에 누적 히스토그램을 매칭시킴으로써 색상을 보정하는 방법이다.[18, 19] 영상 정보만으로 색상 보정을 수행할 수 있는 장점이 있지만, 시점의 차이에 따른 폐색 및 비폐색 영역에 대한 고려가 없기 때문에 카메라 간 간격이나 촬영



〈그림 19〉 스테레오 정합

된 영상의 특성에 따라 성능 차이가 발생할 수 있다.

〈그림 18〉은 히스토그램 매칭의 그래프 관계를 보여준다. 그림에서 점선은 히스토그램을, 실선은 누적 히스토그램을 나타낸다. 영상에서 주어진 밝기 값이 기준영상의 어떤 값으로 바뀌어야 하는지를 결정할 때, 주어진 시점 영상의 i번째 밝기 값 까지 화소가 나타날 확률합과 같은 확률합을 주는 기준 영상의 밝기 값이 k일 때, i의 값을 k로 보내는 방식이다.[20]

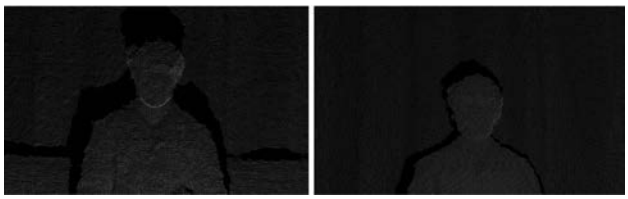
5. 스테레오 정합을 이용한 깊이 정보 생성

스테레오 정합은 기본적으로 양안 시차를 이용하여 장면의 깊이 정보를 획득한다. 멀리 있는 객체는 좌영상과 우영상에서 시차가 작게 발생하는 반면, 가까이 있는 객체는 큰 시차를 가진다. 따라서, 〈그림 19〉와 같이 기준 시점의 모든 화소가 참조 시점의 어느 위치에 존재하는지를 탐색하면 각 화소에 대한 시차 정보를 얻을 수 있고, 이를 화소 단위로 나타낸 것을 변위라고 한다. 변위는 정렬된 스테레오 영상에서 깊이 정보로 사용된다.

촬영된 스테레오 영상을 정합하여 양안 시차를 이용해 깊이 정보를 추출해 낸다. 멀리 있는 객체는 좌영상과 우영상에서 시차가 작게 발생하는 반면, 가까이 있는 객체는 큰 시차를 가진다.

스테레오 정합 알고리즘은 크게 지역적(local) 방식과 전역적(global) 방식으로 나뉜다. 지역적 방식은 화소를 기반으로 일정한 크기의 영역을 설정하고, 참조 시점으로의 대응점 탐색을 통해 매칭 비용이 최소가 되는 변위를 선택하는 것으로 수행속도는 빠르지만 희미한 영역이나 물체의 외곽과 같은 영역에서 성능 저하가 크게 나타나는 단점이 있다.

전역적 방식은 주어진 영상 전체에 걸쳐 데이터를 전달하고 통합하고 분할하는 과정을 반복하여 최적의 변위값



〈그림 20〉 깊이 카메라의 깊이 영상의 3차원 워핑
(위쪽 시점, 아래쪽 시점)

〈그림 21〉 JBU를 이용한 깊이 업샘플링
(위쪽 시점, 아래쪽 시점)

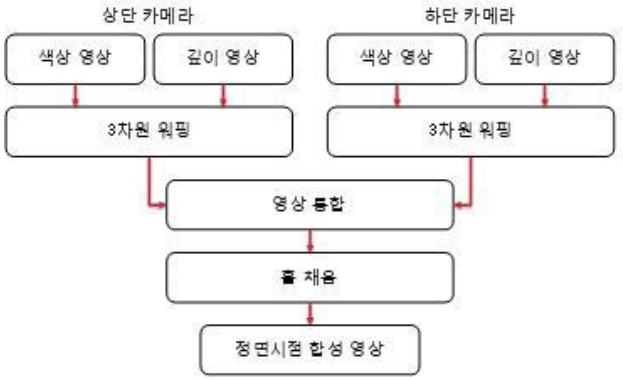
을 선택하는 것이다. 지역적 방식에 비해 성능은 우수하지만, 반복 연산으로 인해 수행 속도가 매우 느리다. 대표적으로 동적 프로그래밍(dynamic programming) [21], 그래프 절단(graph cut)[22], 신뢰 확산(belief propagation) [23] 등의 방법을 사용하며, 최근에는 계층적 접근방법을 사용하여 계산 속도를 향상시킨 계층적 신뢰 확산 방법도 제안되었다.[24] 또한 두 시점 간 폐색 및 비폐색 영역에서의 깊이 정보의 품질 향상을 위한 최적화 방법과, 객체의 깊이 경계부를 개선하기 위한 필터링 방법도 제안되었다.[25, 26]

6. 깊이 카메라에서의 깊이 업샘플링 기술

깊이 카메라에서 획득된 깊이 정보는 3차원 워핑을 통해 스테레오 정합으로 획득된 깊이 정보와 동일한 시점으로 이동된다. 먼저 깊이 카메라의 매개변수와 깊이 정보를 이용하여 모든 화소의 값을 3차원 세계 좌표계로 이동시키고, 이 값들을 색상 카메라의 매개변수를 이용하여 색상 영상 평면에 사상시킨다. 그리고 세계 좌표계로 이동되었던 모든 화소와 색상 카메라의 거리를 계산해서 새로운 깊이 영상을 생성한다. 이러한 과정을 통해 위쪽 카메라와 아래쪽 카메라로 시점 이동된 깊이 카메라의 깊이 영상이 〈그림 20〉에 나타나 있다.

〈그림 20〉에서 알 수 있듯이 깊이 카메라의 해상도는 색상 카메라의 해상도보다 작기 때문에 모든 화소마다 값

〈그림 22〉 객체 마스크와 JBU를 이용한 깊이 업샘플링
(위쪽 시점, 아래쪽 시점)



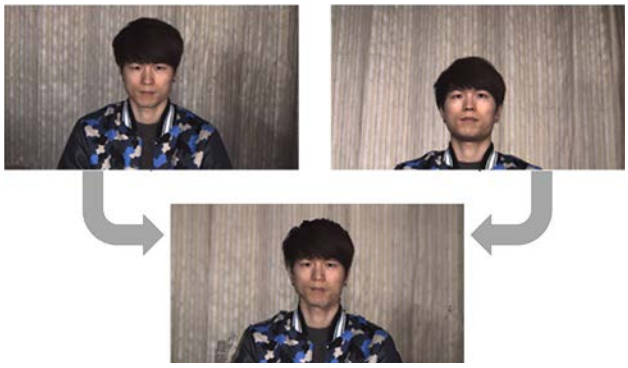
〈그림 23〉 정면 시점 영상 합성

을 가지지 못한다. 따라서 고해상도의 깊이 맵을 획득하기 위해서는 깊이 업샘플링 기술이 필요하다. 깊이 업샘플링 기술은 크게 필터 기반의 방법과 Markov random field (MRF) 기반의 방법으로 나눌 수 있는데, 계산량과 메모리 사용에 있어서 효율적인 필터 기반의 방법이 원격 영상회의 시스템에는 더 적합하다고 할 수 있다.

대표적인 필터 기반의 깊이 업샘플링 방법으로는 결합형 양방향 업샘플링(joint bilateral upsampling)을 들 수 있다.[27] JBU는 표적 화소와 그 주변의 화소들 사이의 유사도를 이용하는 범위 필터(range filter)와, 좌표 상 거리로부터 유도되는 공간 필터(spatial filter)로 구성되며, 고해상도의 색상 영상을 보조 영상으로 사용하여 색상 영상의 경계 정보, 색상 혹은 밝기 정보를 바탕으로 저해상도의 깊이 정보를 확장한다. 〈그림 21〉은 〈그림 20〉의 영상을 JBU를 이용하여 업샘플링 한 결과를 보여준다.

원격 영상회의를 위해서는 좀 더 정확한 객체 경계부가 필요하기 때문에, 스테레오 정합을 통해 얻은 깊이 맵으로부터 객체 마스크를 생성한 후, 마스크 내에서만 깊이 업샘플링을 수행하면 〈그림 22〉와 같이 더 정확한 결과를 얻을 수 있다.

〈그림 24〉 정면 시점으로 워핑된 영상



〈그림 25〉 시선 맞춤 결과

(a) 원본 영상

(b) 시선 조정된 영상

〈그림 27〉 3차원 모델을 이용한 시선 조정

7. 정면 시점 영상 합성

지금까지 설명한 촬영, 영상의 전처리, 깊이 정보 생성의 단계를 모두 거치면 정면 시점 영상을 합성할 수 있다. 〈그림 23〉은 정면 시점 영상을 합성하는 개념도를 보여준다. 상단과 하단 카메라에서의 색상 영상과 최종적으로 생성된 깊이 정보가 있을 때, 색상 화소는 3차원 워핑을 통해 시선 맞춤 시점의 위치로 이동되어 통합되고, 워핑 과정에서 발생한 홀을 채움으로써 정면 시점 영상이 만들어진다. 배경은 미리 촬영된 배경을 사용하고, 화자 부분만 배경 위에 합성한다.

깊이 정보 기반의 시선 맞춤 기술은 다양한 3차원 영상처리 알고리즘이 집약되어 이루어진 것으로 시스템의 구성이 비교적 쉽고 효율적으로 정면 시점을 얻을 수 있는 장점이 있다.

〈그림 26〉 텍스처가 적은 영상에서의 시선 맞춤 결과
(위쪽 시점, 합성 결과, 아래쪽 시점)

〈그림 24〉는 정면 시점 영상으로 이동된 깊이 맵과 색상 영상을 보여준다. 화자는 정면을 바라보고 있으나 영상에 많은 홀이 발생하였다. 이렇게 발생한 홀은 다른 시점으로부터 워핑되어 온 색상 값을 이용하여 채운다.

〈그림 25〉는 화자가 아래를 바라보는 영상과 위를 바라보는 영상으로부터 지금까지 설명한 모든 과정을 거쳐서 합성된 정면 시점 영상, 즉 시선 맞춤의 결과를 보여준다. 합성된 장면에서 화자가 정면을 응시하고 있음을 볼 수 있다.

〈그림 26〉은 지금까지의 설명한 시스템을 사용한 다른 결과를 보여준다. 이 결과는 얼굴 전체가 텍스처가 적어서 스테레오 영상만 가지고는 정확한 눈맞춤 영상을 생성할 수 없는 경우에도 합리적 결과를 보여 · 준다.

깊이 카메라를 이용하면 색상 카메라 없이도 시선 맞춤 영상의 생성이 가능하다.[28] 〈그림 27〉은 깊이카메라 정보만을 이용해서 3차원 모델을 생성하여 시선을 조정한 결과를 보여준다.

Ⅵ. 결론

본 논문에서는 실감형 원격 영상회의를 위한 시선 맞춤 기술을 소개했다. 시선 불일치 문제는 원격 영상회의의 몰입도를 떨어뜨리기 때문에 이를 해결하기 위해 국내외에서 활발한 연구가 진행되고 있다. 또한 스테레오 카메라와 깊이 카메라를 이용하여 정면 시점을 합성하여 시선 맞춤을 구현하는 기술을 자세히 설명하였다. 이 논문에서 설명한 깊이 정보 기반의 시선 맞춤 기술은 다양한 3차원 영상처리 알고리즘이 집약되어 이루어진 것으로 시스템의 구성이 비교적 쉽고 효율적으로 정면 시점을 얻을 수 있는 장점이 있다. 앞으로도 시선 맞춤 기술은지속적으로 개선될 것이며, 실제 원격 영상회의 상황에서 일어날 수 있는 많은 상황들까지 고려한 기술 개발이 필요하다.

감사의 글

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신 · 방송 연구개발사업의 일환으로 수행하였음.[11-921-05-001,다자간 협업을 위한 몰입형 스마트워크 핵심기술 개발]

참 고 문 헌

[1] 이재석, 김종탁, “참가자간의 아이-콘텍(eye-contact)을 고려한 3차원 화상회의시스템,” KR-A-1997-0057779, 1997.

[2] 윤나리, 이병욱, “타원체 MODEL을 사용한 얼굴 영상의 시점합성에 관한 연구,” 한국통신학회논문지, vol.32, no.6, pp. 572-578, 2007.

[3] S.B. Lee, I.Y. Shin, and Y.S. Ho, “Gaze-corrected View Generation using Stereo Camera System for Immersive Videoconferencing,” IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1033-1040, 2011.

[4] R. Yang and Z. Zhang, “Eye Gaze Correction With Stereovision for Video-Teleconferencing,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 7, pp. 956-960, 2004.

[5] O. Schreer, N. Atzapadin, and I. Feldmann, “Multi-baseline Disparity Fusion for Immersive Videoconferencing,” Proc. of International Conference on Immersive Telecommunications, pp. 27-29 2009.

[6] http://www.cs.unc.edu/~maimone/KinectPaper/kinect.html.

[7] J. Zhu, R. Yang, and X. Xiang, “Eye Contact in Video Conference via Fusion of Time-of-Flight Depth Sensor and Stereo,” 3D Research, vol. 2, no. 3, pp. 1-10, 2011.

[8] K. Tan, I. N. Robinson, B. Culbertson, and J. Apostolopoulos, “ConnectBoard: Enabling Genuine Eye Contact and Accurate Gaze in Remote Collaboration,” IEEE Transactions on Multimedia, vol. 13, no. 3, pp. 466-473, 2011.

[9] R. Yang, C. S. Kurashima, H. Towles, A. Nashel, and M. K. Zuffo, “Immersive Video Teleconferencing with User-Steerable Views,” Presence: Teleoperators and Virtual Environments, vol. 16, no. 2, pp. 188-205, 2007.

[10] http://h20338.www2.hp.com/enterprise/us/en/halo/index.html.

[11] http://www.polycom.com/products/telepresence_video/telepresence_solutions/immersive_telepresence/tpx.html.

[12] http://www.cisco.com/en/US/netsol/ns669/networking_solutions_solution_segment_home.html.

[13] Y. Ho and W. Jang, “Eye Contact Technique Using Depth

Image Based Rendering for Immersive Videoconferencing," International Conference on ICT Convergence, pp. 982–983, 2014.

[14] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330–1334, 2000.

[15] Y. Kang and Y. Ho, "An Efficient Image Rectification Method for Parallel Multi-camera Arrangement," IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1041–1048, 2011.

[16] 정재일, 호요성, "다시점 카메라 시스템을 위한 상대적 카메라 특성 기반 색상 보정법," Telecommunications Review, 제20권, 제6호, pp. 1004–1016, 2010.

[17] N. Joshi, B. Wilburn, V. Vaish, M. Levoy, and M. Horowitz, "Automatic Color Calibration for Large Camera Arrays," in UCSD CSE Technical Report CS2005–0821, 2005.

[18] U. Fecker, M. Barkowsky, and A. Kaup, "Improving the Prediction Efficiency for Multi-View Video Coding Using Histogram Matching," in Proc. of Picture Coding Symposium, pp. 2–16, 2006.

[19] Y. Chen, J. Chen, and C. Cai, "Luminance and Chrominance Correction for Multi-view Video using Simplified Color Error Model," in Proc. of Picture Coding Symposium, pp. 2–17, 2006.

[20] http://kipl.tistory.com/116

[21] Y. Ohta and T. Kanade, "Stereo by Intra and Interscanline Search using Dynamic Programming", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 7, no. 2, pp. 139–154, 1985.

[22] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222–1239, 2001.

[23] J. Sun, N. N. Zheng, and H. Y. Shum, "Stereo Matching using Belief Propagation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no.7, pp. 1222–1239, 2003.

[24] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Belief Propagation for Early Vision," in Proc. of International Conference on Computer Vision and Pattern Recognition, vol. 1, pp.261–268, 2004.

[25] W. Jang and Y. Ho, "Efficient Disparity Map Estimation Using Occlusion Handling for Various 3D Multimedia Applications," IEEE Transactions on Consumer Electronics, vol. 57, no. 4, pp. 1937–1943, 2011.

[26] P. Lai, D. Tian, and P. Lopez, "Depth Map Processing with Iterative Joint Multilateral Filtering," in. Proc. of Picture Coding Symposium, pp. 9–12, 2010.

[27] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint Bilateral Upsampling," ACM Transactions on Graphics, vol. 26, no. 3, pp. 1–5, 2007.

[28] S. Lee and Y. Ho, "Generation of Eye Contact Image Using Depth Camera for Realistic Telepresence," Asia-Pacific Signal and Information Processing Association, pp. OS.30–SPS.4.1(1–4), 2013.

호요성

- 1981년 2월 서울대학교 전자공학과 학사
- 1983년 2월 서울대학교 전자공학과 석사
- 1989년 12월 Univ. of California, Santa Barbara, Department of Electrical and Computer Engineering, 박사
- 1983년 3월~1995년 9월 한국전자통신연구소 선임연구원
- 1990년 1월~1993년 5월 미국 Philips 연구소, Senior Research Member
- 1995년 9월~현재 광주과학기술원 정보통신공학부 교수
- 2003년 8월~현재 광주과학기술원 실감방송연구센터 센터장

〈관심분야〉
디지털 신호처리, 영상신호 처리 및 압축, 멀티미디어 시스템, 디지털 TV와 고선명 TV, MPEG 표준, 3차원 TV, 실감방송