

입체영상 제작 및 사용자 상호작용 기술

A Technical Trend of Stereoscopic Content Production and User Interaction

정혁 (H. Jeong)	인터랙티브입체영상연구팀 책임연구원
유정재 (J.J. You)	인터랙티브입체영상연구팀 선임연구원
장경호 (K.H. Jang)	인터랙티브입체영상연구팀 선임연구원
김혜선 (H.S. Kim)	인터랙티브입체영상연구팀 선임연구원
장호욱 (H.W. Jang)	인터랙티브입체영상연구팀 책임연구원
박지영 (J.Y. Park)	인터랙티브입체영상연구팀 선임연구원
이재호 (J.H. Lee)	인터랙티브입체영상연구팀 선임연구원
남승우 (S.W. Nam)	인터랙티브입체영상연구팀 팀장

* 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2011년도 콘텐츠산업기술지원사업(과제명1: 주-부 카메라 기반 실시간 입체영상 획득 및 보정, 합성, 제작 기술, 과제명2: 3D 장면 분석 및 모델 재구성 기법에 의한 입체 3D 영상 생성/합성 기술) 및 지식경제부 산업융합원천기술개발 사업(과제명: 차세대 체감형 콘텐츠를 위한 사용자 동작기반 플랫폼 및 입체 상호 작용 기술 개발)의 연구결과로 수행되었음.

영화 ‘아바타’를 계기로 다양한 입체영화의 등장과 3D TV의 대중화로 인해 입체영화를 비롯한 입체 콘텐츠에 대한 관심이 증가되고 있다. 또한 MS사의 ‘키넥트’가 인기를 끌면서 사용자와 영상 콘텐츠 간의 상호작용 기술이 게임을 중심으로 다양한 분야에 적용되고 있다. 본고에서는 입체영상을 제작하는데 있어서 필요한 촬영 기술 및 관련 장비의 현황 및 스테레오 입체영상의 색감일치 기술의 현황을 소개하고자 한다. 아울러 영상 콘텐츠와 사용자 간의 상호작용을 위한 비전 기반의 상호작용 기술 및 장비의 현황에 대해서도 살펴본다.

2012
Electronics and
Telecommunications
Trends

사용자 중심 차세대콘텐츠
기술 특집

- I. 서론
- II. 입체영상 제작기술
- III. 영상과의 상호작용 기술
- IV. 결론

I. 서론

인간은 두 눈을 이용하여 물체를 인식하며, 입체영상은 이를 모방하여 좌우 두 개의 영상을 좌우 양안에 각각 보여줌으로써 입체감(깊이감)을 느끼게 하는 것을 의미한다. 따라서 입체영상 콘텐츠를 제작하기 위해서는 좌우 양안에 각각 보여지게 될 두 개의 영상이 필요하다. 특히 입체영상 콘텐츠가 잘못 제작되는 경우에는 어지러움증 및 눈의 통증을 유발할 수도 있기 때문에, 2D 영상을 제작하는 일반 콘텐츠 제작보다 입체 3D 영상으로 제작하는 경우에 2배 이상의 예산과 노력이 소요된다.

입체 콘텐츠 제작을 위해서는 프리프로덕션(pre-production) 단계에서 좌우 영상 촬영을 위한 연출이 구상되어야 하고, 프로덕션 단계에서 좌우 영상을 각각 촬영할 수 있도록 두 개의 카메라를 배치하여야 한다. 포스트프로덕션(post-production) 단계에서도 입체영상 색감을 일치시키거나, 깊이감을 재조정하는 등의 작업이 추가적으로 필요하게 된다.

한편, 사용자의 움직임을 실시간으로 감지하여 이를 영상 콘텐츠와 상호작용이 가능하도록 하는 콘텐츠를 인터랙티브 영상 콘텐츠라고 정의할 수 있으며, 이를 위해서는 카메라 등을 통해 사용자의 3차원 정보를 파악하는 기술과 사용자의 3차원 정보를 통해 움직임 정보를 파악하는 기술이 필요하다. 이러한 기술을 통해 영상 콘텐츠와 사용자 간에 자연스러운 상호작용이 이루어 질 수 있게 된다.

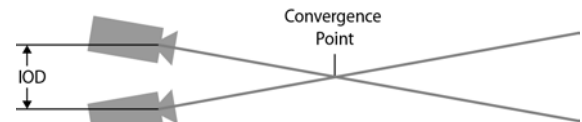
본고에서는 입체영상 제작 및 사용자 상호작용 기술에 대해 관련 장비와 기술 현황을 살펴보고자 한다. 구체적으로는 입체 촬영의 원리 및 장비와 입체영상의 색감일치 기술에 대한 동향과 아울러, 입체영상과 사용자 간의 상호작용을 위한 비전 기반의 상호작용 기술 및 장비의 현황에 대해서 살펴본다.

II. 입체영상 제작 기술

1. 입체 촬영의 원리 및 장비

입체 콘텐츠를 촬영하기 위해서는 사람의 양쪽 눈 역할을 하는 두 개의 실사 캡처(capture) 장치가 필요하다. 이를 위해 일반적으로 두 개의 카메라를 사용하기도 하고 두 개의 렌즈 또는 센서가 달린 특수 카메라를 사용하기도 한다.

사람의 눈이 일정 간격으로 떨어져 있고 눈동자를 움직여 초점을 맞출 수 있듯이, 촬영 장비도 두 개 카메라 사이의 간격과, 바라보는 방향을 움직여 입체영상의 깊이감을 조절할 수 있다. (그림 1)에서 볼 수 있듯이 두 카메라 사이의 간격을 IOD(inter-ocular distance)라고 하고, 수렴점을 영점(convergence point)이라고 하는데, 이 두 가지를 이용하여 콘텐츠의 입체감을 결정한다[1].



(그림 1) 입체 카메라 간격과 초점 방향

IOD는 간격 크기에 따라 좌우 영상의 가로 편차를 크게 또는 작게 만들어서 전체적인 입체량을 좌우하며, 영점은 깊이가 0으로 인지되는 지점이 된다. IOD와 영점 계산은 입체 촬영 준비 단계 및 실제 촬영 단



(그림 2) IOD Calculator

계에서 항상 필요하므로, (그림 2)와 같이 스마트폰용 애플리케이션으로도 많이 개발되어 간편하게 사용할 수 있다[2],[3].

가. 입체 카메라

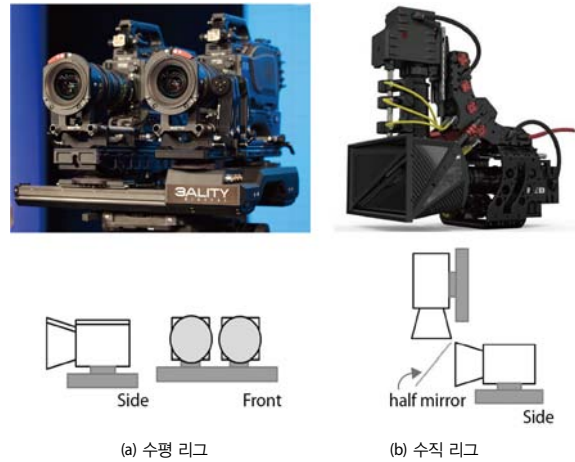
입체 촬영용 카메라는 두 개의 카메라가 동시에 동일한 영상을 촬영할 수 있어야 한다. 카메라의 화각이나 색감이 달라지거나, 동기가 맞지 않는 경우에는 양쪽 눈에 보이게 될 영상 정보가 틀려져서 인지 과정에 어려움이 생기기 때문이다. 이 외에도 두 개의 카메라를 동시에 운영해야 하는 점 때문에, 가볍고 작을수록 우수하다. 최근에는, 한 대의 카메라에 두 개의 렌즈와 센서가 장착되어 동시에 좌우 영상을 촬영하는 카메라가 출시되기도 하였다. 가볍고 조작성이 간단하기 때문에 이동이 많은 촬영에 사용되고 있다. (그림 3)은 Sony, Silicon Image, 파나소닉에서 제작한 입체용 카메라들을 보여주고 있다.



(그림 3) 입체 카메라

나. 입체 리그

두 개의 카메라를 사용하는 경우, 기존의 카메라를 그대로 사용하는 대신 카메라를 공간적으로 배치하기



(그림 4) 입체 리그

위해서 리그(rig)라는 특별한 장치를 추가로 사용해야 한다. 리그는 구조상 크게 수평형과 수직형 타입으로 나뉘어 진다.

(그림 4a)와 같이 수평형 리그는 사람의 눈과 비슷하게 카메라 두 개가 좌우로 나란히 배치되어 있는 구조이다. 하지만, 두 카메라 사이의 간격(IOD)을 좁히는데 한계가 있기 때문에 멀리 있는 물체를 촬영할 때 사용되고 있다.

수직형 리그는 수평형의 단점을 보완하기 위해 고안되었다. (그림 4)의 (b)에서와 같이 두 개의 카메라를 수직으로 배치하고 양쪽 렌즈 사이에 하프미러(half-mirror)를 끼워 넣어, 들어오는 빛을 50%씩 위와 아래의 카메라로 나눠서 보내준다. 카메라 간격(IOD)을 최대 0cm까지도 줄일 수 있기 때문에 근접 촬영은 물론 원경까지 촬영 가능하나, 부피가 크고 무게가 무겁다는 단점이 있다. 또한, 하프미러를 통해 빛이 양쪽으로 분배되기 때문에 영상이 전체적으로 어둡고 좌우 영상의 색감차이가 많이 나기도 한다. 이러한 단점에도 불구하고 자유로운 IOD 값 조절의 장점으로, 대부분의 입체 콘텐츠는 수직형 리그로 제작되고 있다.

다. 이미지 프로세서

입체 촬영의 여러 가지 어려움 때문에, 입체 촬영 장에는 카메라와 리그 외에 이미지 프로세서가 동원되기도 한다. 좌우 영상을 믹싱(muxing)하여 두 영상의 차이를 보기 좋게 디스플레이해주는 가장 기본적인 역할에서부터, 입체 촬영 영상을 분석하여 입체영상의 오류를 분석해서 알려주는 좀 더 적극적인 역할을 하기도 한다. 또한, 분석된 결과에 따라 현장에서 실시간으로 좌우 입체영상을 보정해서 저장 또는 전송할 수도 있다.

(그림 5)의 아래쪽 이미지는 실제 많이 사용되고 있는 이미지 프로세서인 3Ality SIP2100 모델과 Sony 3D Box 모델을 보여주고 있다[4],[5].



(그림 5) 이미지 프로세서의 기능

2. 스테레오 입체영상의 색감일치 기술

좌, 우 카메라로 촬영된 스테레오 입체영상은 좌, 우 영상 간의 색감차이가 존재하고 이 차이가 클 경우 보는 이에게 피로감을 유발할 수 있다. 개발 기술은 스테레오 입체영상의 좌, 우 영상이 갖는 특성을 활용하여 자동으로 좌, 우 영상 간의 색감차이를 일치시키는 기술이다.

가. 기존 기술

영상 간의 색감차이를 일치시키는 기술은 색 변환(color transfer)라는 이름으로 오래 전부터 학계에서 연구되어 왔다. 이 기술은 입력 영상과 참조 영상의 색상 분포를 비교하여 입력 영상의 분포형태를 참조



(그림 6) 색 변환 적용 예

영상의 분포형태로 변환하여 입력 영상이 참조 영상과 유사한 색감을 갖도록 하는 기술이다. (그림 6)은 색 변환을 적용한 결과이다.

다시점 입체영상에서의 색감일치 역시 색상 회귀모델 분석을 이용하여 다른 시점 영상 간의 색감을 일치시키는 기술이 연구되었지만 이 경우 역시 카메라 간의 중심 거리(baseline)가 큰 경우에 집중되어 왔다.

한편, 업계에서는 최근 스테레오 입체영상에서의 색감일치 기능을 제공하는 상용 SW가 개발되었다. (그림 7)은 Quantel사의 Pablo 시스템 구성 예와 Foundary사의 NukeX의 플러그인 세트인 Occula의 화면 예이다. Quantel사의 Pablo는 전문적인 DI 작업을 위한 색상보정 기능을 제공하지만 자동 일치 기능 보다는 소프트웨어와 하드웨어를 접목시킨 시스템 구성으로 정밀한 보정을 위한 수작업 기반 작업환경을 제공한다. Foundary사, NukeX의 플러그인 세트인 Occula는 좌, 우 영상 간에 전역적, 국부적 색감차이를 보정하는 기능을 제공한다. 전역적 차이를 보정하기 위해서 영상 간의 색상, 밝기 분포를 일치시키고, 국부적 차이는 입력 영상을 다수의 블록으로 나누고

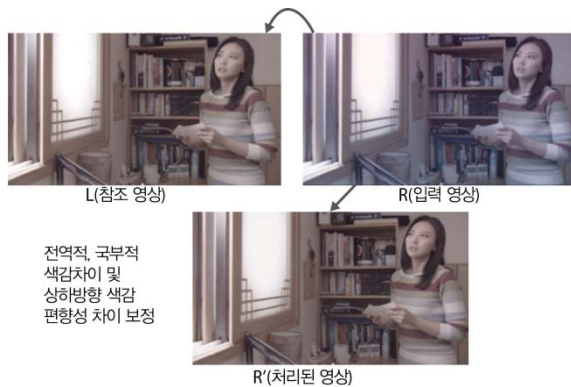


(그림 7) 색감일치 기능 제공 상용 SW

참조 영상에서의 대응영역을 찾아 국부적으로 색상분포를 일치시킴으로써 보정한다. 하지만 이 과정에서 블록 간의 이질감과 색상변짐 현상이 발생하고 상용 콘텐츠 제작에 활용하기 위해서는 다수의 색상보정 노드를 생성하여 이들을 융합하여 사용하는 복잡한 작업 설계과정이 필요하다.

나. ETRI 개발 기술

ETRI에서는 수직 리그로 촬영된 스테레오 입체영상에 특화된 자동 색감일치 기술을 개발하였다. 우선 기본적으로 좌, 우 영상 간의 색상분포 일치를 통하여 전역적인 색감일치를 수행한다. 수직 리그로 촬영된 입체영상은 수직 리그에 사용되는 하프미러의 광학적 특성으로 전체적인 색감차이 외에 상하방향 색감 편향성 현상이 관찰된다. 개발기술은 상하방향 색상분포 비교 및 보정을 통해서 이러한 문제점을 해결하였다. 마지막으로 이러한 보정을 거친 후에도 남아있는 일부 국부적인 영역에서 색감차이를 보정하기 위해서 좌, 우 영상 간의 특징점 매칭을 수행하고 특징점 매칭 주변영역에서 샘플 포인트를 추출하여 LUT(Look Up Table)를 구성하여 위의 처리결과와 융합하여 최종적으로 색감일치가 완료된 영상을 얻는다. (그림 8)은 ETRI에서 개발된 색감일치 기술이 적용된 예이다.

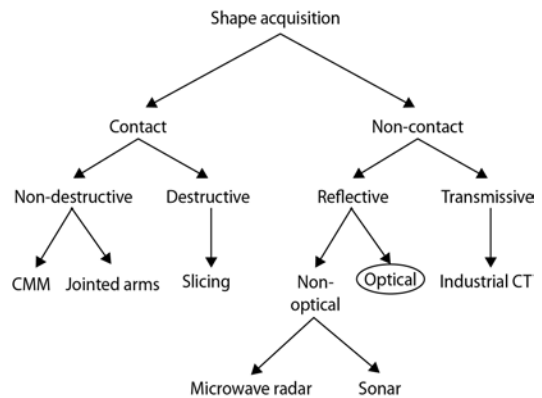


(그림 8) ETRI에서 개발된 색감일치 기술 적용 예

III. 영상과의 상호작용 기술

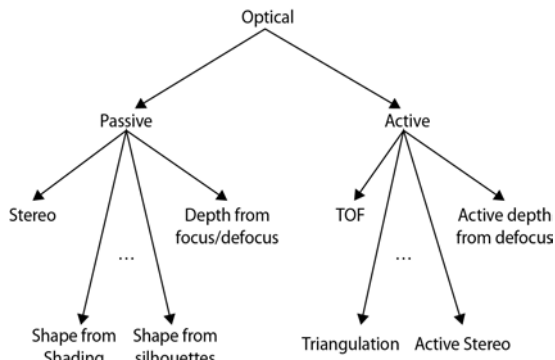
1. 3차원 정보 획득 기법

3차원 정보를 획득하는 방법은 (그림 9)의 분류와 같이 크게 접촉식과 비접촉식 방법으로 나누어 볼 수 있다[6]. 접촉식 방법은 물체 외형을 따라가는 형태인 비파괴 형식과 파괴 형식으로 나뉘지며, 비접촉식 방법은 소리나 전자기파가 물체에 반사되어 되돌아오는 반사 방식과 물체에 투과되는 투과 방식으로 나눌 수 있다. 현재 3차원 측정에 주로 사용되는 기법은 광학 방식으로써 빠르고 높은 정확도를 가지고 물체에 어떠한 변형 없이 측정할 수 있다. 본 절에서는 광학 방식을 사용한 3차원 측정 기법들을 중점적으로 다루기로 한다.

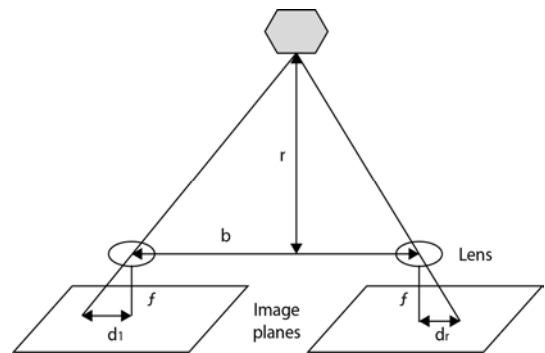


(그림 9) 3차원 획득 기술의 분류

광학 방식은 (그림 10)과 같이 카메라에서 획득한 영상 정보를 바탕으로 3차원 거리를 측정하는 수동 방식과 특정 전자기파 등을 투사하는 능동 방식으로 분류할 수 있다. 3차원 거리 측정에서 최근 주로 사용되고 있는 기법은 다른 방법들에 비해 빠른 속도와 우수한 정확도를 보이는 수동 방식의 스테레오 방식과 능동 방식의 TOF(Time-Of-Flight), 능동형 구조광 방식이다.



(그림 10) 광학 방식의 3차원 획득 기술



(그림 12) 스테레오 매칭 시스템의 기하학적 구조

가. TOF 방식

TOF 방식은 3차원 카메라에서 특정 전자기파를 쏘고 되돌아 오는 시간을 측정(TOF principle)하여 3차원 거리 정보를 획득하는 기법이다. TOF 센싱 방식은 정확한 거리 측정이 가능한 장점이 있으나, 센서의 가격이 고가이며, 최대해상도가 QCIF급인 단점이 있다. 최근 SoftKinetic사에서 출시된 DS311 모델은 QQVGA급의 해상도를 지원하며 499달러의 가격으로, 기존의 고가 장비로만 인식되어 왔던 TOF 카메라에 새로운 가능성을 제시하고 있다. (그림 11)은 TOF 방식의 3차원 인식 카메라 외형이다.



(a) SoftKinetic, DS311

(b) MESA Imaging

(그림 11) TOF 방식의 3차원 인식 카메라

나. 스테레오 방식

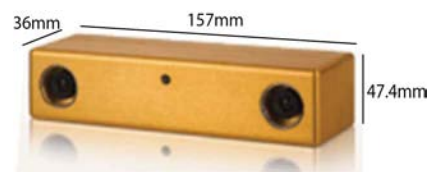
스테레오 방식은 인간의 두 눈과 같이 한 쌍의 카메라 혹은 그 이상의 카메라로부터 들어 오는 영상을 이용하여 환경/물체의 삼차원 형태를 측정하는 방법

이다. 삼차원 거리 측정(그림 12)과 같이 삼각 측량으로 이루어지며, 삼각 측량을 위한 매칭 기법은 코너, 에지 등 다양한 방법들이 시도되어 왔다.

임의의 베이스라인 b 를 가진 두 카메라에서 영상을 획득하고 두 영상에서 같은 지점을 찾았을 때, 알고 있는 초점거리 f 가 있다면 다음과 같은 식을 통해 거리 정보 r 을 알 수 있다.

$$r = f(b/d), \text{ 단 } d = d_l - d_r$$

상용 스테레오 카메라 제품은 3차원 복원 과정을 컴퓨터에서 연산하는 소프트웨어 방식과 FPGA와 같은 하드웨어에서 연산하는 하드웨어 방식으로 나눌 수 있다. 전자의 제품으로는 Point Grey Research의 제품과 Videre Design의 제품들이 있고 후자의 제품으로는 TYZX의 스테레오 카메라가 있다(그림 13)



(a) Point Grey, Bumblebee2



(b) Videre Design, STH-MDCS

(c) TYZX, DeepSea G2

(그림 13) 스테레오 방식의 카메라

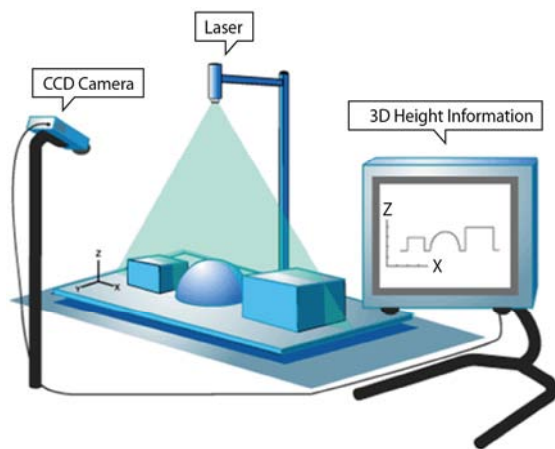
참조). 두 방식의 제품들의 해상도는 비슷하나 3차원 복원 결과는 하드웨어 방식의 스테레오 카메라 경우에 실시간으로 3차원 거리 정보를 획득할 수 있다.

스테레오 방식의 카메라들은 양쪽 카메라의 특징점 일치 성능을 좌우하게 되며, 따라서 특징점을 찾기 어려운 상황 즉, 영상 내에 컬러 변화가 없는 영역이 많이 존재하거나, 조명이 지나치게 어둡거나 밝은 경우 사용할 수 없는 단점이 있다.

다. 구조광 방식

구조광(structured light) 방식은 능동 방식의 3차원 카메라로서 스테레오 카메라의 한 부분을 투사장치로 대체한 형태로 구성되어 있다. 투사장치는 점, 선, 면 형태로 빛을 투사할 수 있으며 투사된 빛은 물체 형태에 따라 변형이 일어난다. 구조광 방식의 3차원 카메라는 이 변형된 정보를 이용하여 3차원 거리 정보를 계산한다.

(그림 14)는 구조광 방식의 기본적인 원리를 보이는 것으로서, 선 형태의 빛을 투사하여 물체에 맺힌 선에 대한 3차원 거리 정보를 획득할 수 있다. 때문에 카메라 영역 전체에 대한 3차원 거리 정보를 얻기 위해서는 많은 수의 라인이 필요하게 되는데, 이를 보안



(그림 14) 라인을 이용한 3차원 거리 정보 획득 개요[6]

하는 방법으로 면 형태의 빛을 투사하여 빠르게 3차원 정보를 얻는 방법을 사용한다.

삼각 측량을 위해서는 월드 좌표상에 표시되는 한 점을 두 카메라 좌표상에 어디에 맺히는지를 알아야 한다. 이때 스테레오 카메라는 각 카메라 영상 이미지 평면의 U, V 좌표 값으로 두 위치를 알 수 있으나, 구조광 방식의 3차원 카메라는 투사부 파트의 좌표 값을 바로 얻기가 어렵다. 투사부 파트의 좌표 값을 카메라로 획득되는 영상 정보로 계산하게 되는데, 이를 위해 투사하는 빛에 정보를 담아 패턴을 형성하고 물체에 투사한다. 이때 정보를 담는 방법에 따라 (그림 15)에서 볼 수 있는 direct code, spatial code, temporal code로 나눌 수 있다. Direct code는 색상 값

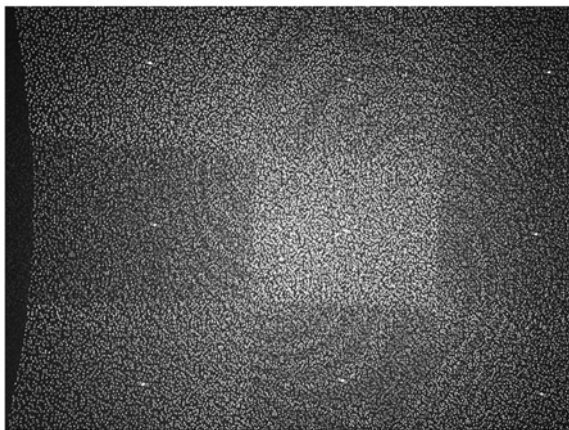
Table 1: The proposed classification

	Binary codes	Posdamer et al.	Inokuchi et al.	Minou et al.	Trobina	Valkenburg and McIvor	Skocaj and Leonardis	Rocchini et al.	
Time-multiplexing	n-ary codes	✓	✓	✓	✓	✓	✓	✓	
	Gray code + Phase shifting	✓	✓	✓	✓	✓	✓	✓	
	Hybrid methods	Kosuke sato	✓	✓	✓	✓	✓	✓	✓
		Hall-Holt and Rusinkiewicz	✓	✓	✓	✓	✓	✓	✓
Spatial Neighborhood	Non-formal codification	Maruyama and Abe	✓	✓	✓	✓	✓	✓	
		Durdle et al.	✓	✓	✓	✓	✓	✓	
		Ito and Ishii	✓	✓	✓	✓	✓	✓	
	De Bruijn sequences	Boyer and Kak	✓	✓	✓	✓	✓	✓	
		Chen et al.	✓	✓	✓	✓	✓	✓	
Direct coding	M-arrays	Hügli and Maitre	✓	✓	✓	✓	✓	✓	
		Monks et al.	✓	✓	✓	✓	✓	✓	
		Vuytckc and Oosterlinck	✓	✓	✓	✓	✓	✓	
	Grey levels	Salvi et al.	✓	✓	✓	✓	✓	✓	
		Lavoie et al.	✓	✓	✓	✓	✓	✓	
		Zhang et al.	✓	✓	✓	✓	✓	✓	
		Morita et al.	✓	✓	✓	✓	✓	✓	
Colour	Petriu et al.	✓	✓	✓	✓	✓	✓		
	Kiyasu et al.	✓	✓	✓	✓	✓	✓		
	Spoelder et al.	✓	✓	✓	✓	✓	✓		
	Griffin and Yee	✓	✓	✓	✓	✓	✓		
Scene applicability	Pixel depth	Davies and Nixon	✓	✓	✓	✓	✓		
		Morano et al.	✓	✓	✓	✓	✓		
		Carrhill and Hummel	✓	✓	✓	✓	✓		
Coding strategy	Absolute	Chazan and Kiryati	✓	✓	✓	✓	✓		
		Hung	✓	✓	✓	✓	✓		
Coding strategy	Periodical	Tajima and Iwakawa	✓	✓	✓	✓	✓		
		Smutny and Pajdla	✓	✓	✓	✓	✓		
		Geng	✓	✓	✓	✓	✓		
		Wust and Capson	✓	✓	✓	✓	✓		
Coding strategy	Absolute	Tatsuo Sato	✓	✓	✓	✓	✓		
		Static	✓	✓	✓	✓	✓		
Coding strategy	Absolute	Moving	✓	✓	✓	✓	✓		
		Binary	✓	✓	✓	✓	✓		
Coding strategy	Absolute	Grey levels	✓	✓	✓	✓	✓		
		Colour	✓	✓	✓	✓	✓		

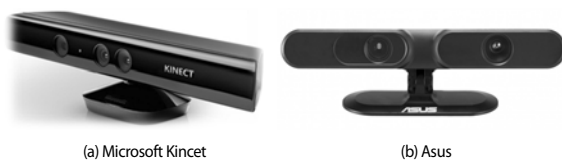
(그림 15) Coding 방법에 대한 pattern 분류

등을 이용하여 프로젝터 좌표 정보를 한 장의 패턴에서 바로 계산할 수 있는 기법이다. Spatial code는 주변 값의 정보를 이용하여 좌표를 계산하는 기법이고 temporal code는 좌표 값을 시간 축으로 여러 장의 패턴을 이용하여 계산한다. 패턴에 대한 분류는 (그림 15)에 소개되어 있다[7].

최근 Primesense사에서 개발한 3차원 카메라 모듈은 (그림 16)에서 보이는 것과 같이 IR 광원을 사용하여 투사되는 speckle 패턴을 도입하였다. 이 모듈은 640×480급 해상도의 3차원 정보를 가진 맵스 맵을 30fps로 획득하며, 출시된 제품군은 150~300달러의 가격으로 판매 되고 있다. Primesense의 카메라 모듈은 (그림 17)에서 보이는 것과 같이 Microsoft사의 Kinect와 Asus 카메라에 탑재하여 공급되고 있으며, 빠른 처리속도와 높은 해상도, 저렴한 가격으로 인해 최근 가장 많이 사용되고 있는 3차원 정보 획득 장치이다. 또한 장치에서 나오는 3차원 정보를 기반으로



(그림 16) Primesense 제품군에 사용된 speckle 패턴



(그림 17) Primesense의 모듈로 제작된 카메라 제품

다양한 제스처 인식, 공간 인식 기반의 응용 기술들이 활발히 발표되고 있다.

2. 깊이 영상에서의 사용자 동작인식

깊이 센서로부터 만들어진 공간정보를 가지고 있는 영상은 화면상에 나타나는 사용자를 실제의 공간에서 인지하고 그 사용자의 동작에 대한 인식을 수행하여 응용 프로그램에 적용하기까지의 과정이 기존의 2차원 영상에서 같은 작업을 진행하는 것보다 훨씬 수월하게 적용될 수 있는 이점을 가지고 있다. 간단히 그 장점을 열거하여 보면 다음과 같다

사용자와 배경의 분리가 수월하다: 기존의 2차원 영상에서 배경과 사용자 분리를 수행할 때보다 3차원 깊이 센서에서 나오는 영상의 경우 조명의 영향에서 보다 자유로울 수 있기 때문에 기존의 2차원 영상에서 오류로 판명되기 쉬운 그림자의 처리에 따른 부담과 조명의 변화에 따른 잘못된 인식이 현저히 감소한다.

사용자들 간의 추적과 분리가 용이하다: 깊이 정보가 포함된 영상에서는 사용자들 간의 겹침을 판단하기가 기존의 2차원 영상보다 수월하여 화면상에 존재하는 사용자들 간의 분리 및 추적에 대한 알고리즘을 실시간으로 구현이 가능하다.

사용자의 골격구조를 찾는 데 보다 많은 정보가 존재한다: 하나의 사용자에 대하여 인식을 수행하는데 필요한 기본정보가 인간의 골격정보를 추출해 내는 것이다. 깊이 영상에서는 사람의 주어진 골격조건을 기본조건을 이용하여 스켈레톤 정보를 추출하는데 많은 장점을 가지게 된다.

이러한 장점으로 인하여 현재 많은 요구가 형성되고 있는 사용자의 동작인식을 통한 인터랙션이 가능한 콘텐츠 기술을 적용하기에 3차원 깊이 센서를 통한 접근이 활발히 진행되고 있는 현실이다.

최근 가장 대표적인 3차원 깊이 센싱 기술은 프라임센서에서 원천 기술을 개발하여 만들어진 3차원 깊이 센서로 마이크로소프트에서 Xbox에 실용화함에 따라 그 파급효과가 거의 전 세계에 이를 정도로 넓게 확장되어 있다. 현재 가장 대중적으로 사용하는 동작인식 라이브러리는 크게 2종류로 구분되고 있다. 하나는 프라임센서에서 개발한 OpenNI[8]이고, 다른 하나는 마이크로소프트에서 개발한 Kinect SDK[9]이다.

OpenNI는 프라임센서에서 개발과 동시에 발표되었으며, 최초로 센서의 구입자에게만 오픈되었던 것을 Xbox의 대중화를 타고 모두에게 공개하는 형태로 변형되었다. 크게 마이크로소프트에서 개발한 Kinect SDK와의 차이점을 살펴보면 OpenNI의 경우 컬러 카메라의 입력을 사용하지 않고 골격추출의 경우에도 템플릿 영상의 자료를 가지고 제약조건을 가지고 골격구조를 추출하는 방식을 취했으며, 이와 반대로 Kinect SDK의 경우에는 인식의 응용에 있어 컬러 카메라의 입력을 사용하는 경우가 많아, 얼굴인식과 음성인식들을 결합하여 사용하고 있다. 또한, Kinect SDK의 경우 골격구조의 추출에서 미리 학습된 다량의 데이터에서 학습을 방법을 통하여 골격구조를 추출하는 방식을 취하고 있다[10]. Kinect의 대중화에 따라 OpenNI를 Kinect에 연동하여 사용할 수 있는 라이브러리[11]도 발표되어, 개발자들이 자유롭게 Kinect를 사용하여 OpenNI를 이용할 수 있도록 지원하고 있다.

최근 들어 이러한 라이브러리를 이용하여 여러 응용에 적용하고 이를 웹에 발표하는 사례들이 늘고 있다. 가장 대표적인 사이트로는 Kinect Hacks[12]가 있다. 이 사이트는 세계 각지에서 여러 응용 분야에 적용한 실제 사례들이 빠른 시간 안에 전파되고 공유되는데 중요한 역할을 하고 있다. 이러한 3차원 깊이



(그림 18) 3차원 깊이 영상에서의 사용자 골격추출과 동작 인식에서의 적용 예(가상수족관: 물고기 낚시, 물고기 몰기, 먹이 주기 등의 인터랙션 콘텐츠)

센서를 이용한 동작인식 및 응용 분야에 대한 관심의 증폭에 따라, 현재의 주요 관심을 받는 분야에 대하여 소개하는 TED[13]라는 기술 소개사이트에서도 교육이나, 로봇, 발표 등에 사용하는 기술들을 소개하고 있다. 국내에서도 활발한 연구가 진행 중이며, 여러 콘텐츠에 접목한 기술들을 발표하고 있으며, 계속해서 교육, 국방, 엔터테인먼트 등의 분야에 확장시켜 가고 있다[14]. 다음 (그림 18)은 3차원 깊이 영상에서의 사용자 골격추출 과정과 골격 정보로부터 동작 인식을 인터랙티브 콘텐츠에 적용한 예이다.

IV. 결론

본고에서는 입체 촬영의 원리 및 장비와 입체영상의 색감일치 기술에 대한 동향 분석을 통해 입체영상

제작에 필요한 기술의 전반적인 현황을 살펴보았다. 또한 영상 콘텐츠와 사용자 간의 자연스러운 상호작용을 가능케 하는 비전 기반 장비 및 기술 동향분석을 통해 상호작용 기술의 현황도 살펴보았다.

입체영상의 제작에는 색감일치 및 깊이 조정 등 고려해야 할 점이 기존의 영상 제작에 비해 훨씬 많다. 이를 해결해 주는 촬영 솔루션에 대한 개발이 지속적으로 이루어지고 있으며, 앞으로 보다 후반 작업의 노력이 많이 줄어들게 개발될 것으로 예상된다. 아울러 깊이 정보를 보다 넓은 범위에서 빠르고 정확하게 인식하여 사용자의 움직임을 실시간으로 파악하는 기술 개발이 진행되고 있어서, 방 전체의 범위에서 사용자-영상 콘텐츠 또는 다수의 사용자 간의 자연스러운 상호작용을 제공하는 제품 및 영상 콘텐츠가 등장할 것으로 기대된다.

용어해설

Stereoscopic 3D 'Stereo'(두 개의)+'scopic'(to see, 보다). 두 눈으로 보는 양안식 3D를 의미함. 인간이 두 눈을 이용하여 물체를 인식하는 것을 모방하여 좌우 두 개의 영상을 좌우 양안에 각각 보여줌으로써 입체감(깊이감)을 느낄 수 있게 하는 것

IOD(Interocular distance) 눈 사이의 거리를 뜻함. 입체 영상 제작 시에는 좌우 카메라 렌즈 중심점 사이의 거리를 의미함.

Convergence point 수렴점, 집합점. 좌우 양쪽 눈의 바라보는 방향이 수렴되는 점을 의미함. 깊이값이 0으로 인지되는 영점이 되며, 입체영상을 시청할 경우 스크린과 같은 깊이로 인지되는 깊이 지점이 됨.

약어 정리

CG	Computer Graphics
DI	Digital Intermediate
GPU	Graphics Processing Unit

IOD	Interocular distance
LUT	Look Up Table
S3D	Stereoscopic 3D
TOF	Time-Of-Flight
VFX	Visual Effects

참고문헌

- [1] <http://e3dcreative.com>
- [2] <http://www.vizworld.com/2010/08/stereowedge-iod-calculator-iphone>
- [3] <http://itunes.apple.com/us/app/reald-professional-stereo3d/id362539528?mt=8>
- [4] Sony Korea, Sony's 3D Live Production Solution Workshop. 2010.
- [5] <http://www.3alitytechnica.com/3D-rigs/products.php#sip>
- [6] B. Curless, "Overview of Active Vision Techniques," SIGGRAPH 2000 Courses on 3D Photography, SIGGRAPH, New Orleans, Louisiana, July 2000.
- [7] J. Salvi, J. Pages, and J. Batlle "Pattern Codification Strategies in Structured Light Systems," Pattern Recognition, Elsevier, vol. 37, no. 4, 2004, pp. 827-849.
- [8] OpenNI. <http://www.openni.org>
- [9] KinectSDK. <http://www.microsoft.com/en-us/kinectforwindows/>
- [10] J. Shotton et al., "Real-Time Human Pose Recognition in Parts from a Single Depth Image," CVPR, 2011.
- [11] Kinect Sensor for OpenNI. <https://github.com/avin2/SensorKinect>
- [12] Kinect Hacks. <http://www.kinecthacks.net/>
- [13] TED. <http://www.ted.com>
- [14] 김희권, 이재호, 남승우, "3D 사용자 깊이정보를 이용한 제스처 인식 시스템," 대한전자공학회 하계학술대회, 2011.