2차원 동영상의 3차원 변환을 위한 깊이 단서의 신뢰성 기반 적응적 깊이 융합

Adaptive Depth Fusion based on Reliability of Depth Cues for 2D-to-3D Video Conversion

한찬희, 최해철, 이시웅
한밭대학교 정보통신전문대학원

Chan-Hee Han(chani@hanbat.ac.kr), Hae-Chul Choi(hcchoi@hanbat.ac.kr),
Si-Woong Lee(swlee69@hanbat.ac.kr)

요약

3차원 동영상은 다양한 응용분야들에서 차세대 콘텐츠로 큰 주목을 받고 있다. 2D-to-3D 변환은 3차원 동영상의 시대로 넘어가는 과도기 동안에 3차원 동영상 콘텐츠의 부족현상을 해결하기 위한 강력한 기술로 여겨지고 있다. 일반적으로 2D-to-3D 변환을 위해서는 2차원 동영상 각 장면의 깊이영상을 추정/생성한 후 깊이 영상 기반 렌더링 (DIBR : Depth Image Based Rendering) 기술을 이용하여 스테레오 동영상을 합성한다. 본 논문은 2차원 동영상 내 존재하는 다양한 변환 단서들을 통합하는 새로운 깊이 융합 기법을 제안한다. 우선, 알맞은 깊이 융합을 위해 몇몇 단서가 현재 장면을 효과적으로 표현할 수 있는 지 아닌 지 검사된다. 그 후, 신뢰성 검사의 결과를 기반으로 현재 장면은 4개의 유형 중 하나로 분류된다. 마지막으로 최종 깊이 영상을 생성하기 위해 신뢰할 수 있는 깊이 단서들을 조합하는 장면 적응적 깊이 융합이 수행된다. 실험 결과를 통해 각각의 단서가 장면 유형에 따라 타당하게 활용되었고 최종 깊이 영상이 현재 장면을 효과적으로 표현할 수 있는 단서들에 의해 생성되었음을 관찰할 수 있다.

■ 중심어 : | 2차원 영상의 3차원 변환 | 깊이 추정 | 깊이 영상 기반 렌더링 |

Abstract

3D video is regarded as the next generation contents in numerous applications. The 2D-to-3D video conversion technologies are strongly required to resolve a lack of 3D videos during the period of transition to the full ripe 3D video era. In 2D-to-3D conversion methods, after the depth image of each scene in 2D video is estimated, stereoscopic video is synthesized using DIBR (Depth Image Based Rendering) technologies. This paper proposes a novel depth fusion algorithm that integrates multiple depth cues contained in 2D video to generate stereoscopic video. For the proper depth fusion, it is checked whether some cues are reliable or not in current scene. Based on the result of the reliability tests, current scene is classified into one of 4 scene types and scene-adaptive depth fusion is applied to combine those reliable depth cues to generate the final depth information. Simulation results show that each depth cue is reasonably utilized according to scene types and final depth is generated by cues which can effectively represent the current scene.

■ keyword : | 2D-to-3D Conversion | Depth Estimation | Depth Image based Rendering |

* 본 연구는 교육과학기술부와 한국연구재단의 지역혁신인력양성사업으로 수행된 연구결과임(No. 2012H1B8A2025982).

접수번호 : #120822-002

심사완료일 : 2012년 11월 30일

접수일자 : 2012년 08월 22일

교신저자 : 이시웅, e-mail : swlee69@hanbat.ac.kr

I. 서론

인간이 물체를 볼 때 좌안과 우안에 각각 다른 영상이 맺히게 되는 현상을 양안 시차(binocular disparity)라 하며, 뇌에서는 양안 시차가 있는 두 영상을 종합적으로 판단하여 입체감을 느끼게 된다. 스테레오 3D에서는 이러한 원리를 바탕으로 [그림 1]과 같이 동일한 장면에 대한 스테레오 영상을 동시에 디스플레이 함으로써 시청자가 입체감을 느낄 수 있도록 한다.

3D 디스플레이 기술은 장면의 사실적 인지를 가능케 하는 차세대 디스플레이 기술로서 큰 주목을 받고 있으며, 그 응용분야 또한 매우 방대할 것으로 예상된다. 그러나 3D 동영상 시대로의 완전한 진입을 위해서는 관련 기술의 발전 외에도 풍부한 3D 콘텐츠의 확보라는 과제가 반드시 해결되어야 한다. 현재의 상업용 3D 입체영상은 주로 스테레오 카메라나 영상편집 툴을 사용하여 제작된다. 따라서 스테레오 카메라를 이용한 높은 제작비용, 영상 편집에 따른 제작 기간의 증가 등으로 인해 아직까지는 방송이나 영화와 같은 영상매체 시장에서 2D로 콘텐츠를 제작하는 경우가 보다 보편적이다. 이러한 환경 하에서 3D 산업의 활성화를 위해 기존의 방대한 2D 콘텐츠를 높은 추가비용 없이 3D 콘텐츠로 확보할 수 있는 [그림 2]의 2D-to-3D 변환 기술이 최근 크게 주목 받고 있다.

그림 1. 스테레오 카메라에 의한 3차원 입체감 인지

그림 2. 2D-to-3D 변환에 의한 3차원 입체감 인지

사람은 양안시차 외에도 다양한 깊이 인식 요소를 통해 입체감을 인지한다. 실제로 인간은 한쪽 눈만으로도 입체감 즉, 현재 위치와 대상들 간의 상대적인 깊이를 인지할 수 있으며, 이를 가능케 하는 요인으로는 장면의 기하학적 구성, 폐색 구간, 객체의 크기나 질감, 인간의 경험적 사고, 객체의 흐릿한 정도, 객체의 움직임 등이 있다. 이러한 요소들은 2D 영상 내에 존재하며 이를 3D 변환을 위한 단안 단서(monocular cue)라 한다.

2D-to-3D 변환 기술은 이러한 2D 영상에 존재하는 다양한 깊이 단서를 추출하고 이로부터 각 영역의 상대적 깊이를 계산함으로써 스테레오 영상을 생성하는 기술이다. 관련 기술의 상업적 필요성으로 인해 현재까지 이에 대한 연구가 폭넓게 진행되어 왔다. 그러나 기존 연구들은 특정 하 나의 깊이 인식 단서를 해석함으로써 3D로 변환하는 연구가 대부분을 차지하고 있다[2-9]. 하지만 인간의 시각 시스템은 양안 시차 또는 움직임 시차 등의 양안 단서 뿐 아니라 다양한 단안 깊이 단서들의 복합적 인식을 통해 입체를 느끼게 된다. 따라서 다양한 깊이 단서들의 효율적인 융합을 통한 깊이 추정 은 인간 시각 시스템의 깊이 인식 원리에 부합하는 우수한 깊이 추정 방식이라고 할 수 있다. 이를 구현하기 위한 알고리즘 설계에 있어 핵심 문제는 어떻게 몇몇의 깊이 단서들을 효율적으로 선택하고 통합하느냐 하는 것이다.

최근 복수개의 깊이 단서로부터 개별 깊이 영상을 생성한 뒤 이들을 융합하여 최종 깊이를 생성하는 융합 기반의 3D 변환 기법들이 발표되고 있다[9-11]. 하지만 이들 알고리즘의 대부분은 가중 합 형태의 단순 융합 구조를 따르고 있다. 가중 합 형태의 융합 구조는 요소 깊이에 오류가 존재하는 경우 이 오류가 최종 깊이에 그대로 반영된다는 단점을 갖는다. 따라서 본 논문에서는 이러한 구조적 단점을 극복하기 위한 방안으로 장면에서 신뢰할 수 있는 단서들의 통합과 신뢰할 수 없는 단서들의 거부를 통한 적응적 깊이 융합 기법을 새로이 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구 및 기존 연구의 문제점을 기술하고, 3장에서는 제안하는 장면 적응적 깊이 융합 알고리즘을 설명한다. 그리고 4장에서는 실험 결과를 보이며, 5장에서 결론을 맺는다.

II. 관련 연구 및 문제점

1. 개별 단서를 통한 접근법

1.1 움직임 기반 접근법

동영상 내 동적객체들의 자동 분할 및 추적은 디지털 영상 분석 분야의 주요 연구주제 중 하나이며, 실시간 감시 시스템, 멀티미디어 서비스 시스템, 원격 탐사와 같은 매우 광범위한 응용분야를 가진다[1]. 2D-to-3D 변환에서도 움직임 분석은 주요한 도구로 사용된다.

움직임 정보를 이용한 2D-to-3D 변환 기법들은 지연 영상 기반의 기술들과 움직임 벡터 기반의 기술들로 분류된다. 지연영상 기반의 기법들은 양안 단서인 움직임 시차를 활용한 기법이라 할 수 있다. 지연영상 기반의 2D-to-3D 변환 기술들[2][3]은 2D 동영상에서 현재영상과 시간지연영상을 좌영상과 우영상으로 결정하여 3D 입체 영상을 생성한다. 좌우영상의 선택은 카메라와 객체의 운동유형을 분석하여 결정하게 된다. 하지만 일반적으로 카메라와 객체의 수평방향 운동만을 가정하는 경우는 영상이 수직운동을 포함할 때 객체의 변형이 발생할 수 있으며, 좌우영상 내 움직임의 일관성을 유지하는 것이 힘들다. 또한 수직운동을 인식할 수 있는 경우라 하더라도 영상 내 모든 객체의 운동을 한 방향으로 가정하기 때문에 영상 내에서 서로 다른 움직임 형태가 동시에 존재할 때 문제가 발생할 수 있다. 앞서 설명한 단안 단서 중에 하나인 객체의 움직임과 양안 단서인 움직임 시차를 종합하여보면 결국 2D 동영상 내에서 가까운 물체일수록 움직임 벡터가 크다는 결론에 도달할 수 있으며, 움직임 벡터 기반의 2D-to-3D 변환 기술들[4][5]은 이 가정을 기본으로 BMA(Block Matching Algorithm)나 feature tracking에 의한 움직임 벡터를 이용한다. [5]에서는 움직임 정보의 추정을 위해 색상 분할과 KLT(Kanade-Lucas-Tomasi)라는 특징 추적기를 이용하였고 카메라의 이동, 장면의 복잡도, 추정된 움직임의 크기와 같은 요소들을 고려하여 disparity 정보를 구해 입체 영상을 생성하였다. 이러한 접근법은 카메라 이동이 존재하지 않고 단일 객체가 움직임을 가지는 경우나 객체의 움직임이 없고 카메라 이동으로 인한 움직임이 존재하는 경우에는 우수한 성능

을 기대할 수 있다. 하지만 움직임 벡터 기반의 기법에서는 객체의 움직임 방향이 카메라 이동 방향과 같은 경우나 객체들 간의 상이한 움직임 속도를 보일 경우에는 문제가 발생할 수 있다. 또한, 객체와 카메라 움직임이 전혀 존재하지 않는 경우 3D 변환을 위한 어떠한 정보도 얻을 수 없다.

1.2 초점 기반 접근법

[그림 3]에서와 같이 영상을 촬영할 때 한 부분에 초점을 맞추는 경우가 있으며, 일반적으로 가까운 객체에 초점을 집중시켜 선명하게 표현하고 먼 영역일수록 흐릿하게 표현된다. 특히 객체와 배경 경계부분에서 이러한 현상은 더욱 두드러지게 나타난다. 이러한 초점 정보를 기반으로 한 깊이를 추출[6][7]을 depth-from-focus 또는 depth-from-texture 라고 한다. 기본 개념은 카메라로부터 가까운 영역일수록 그 질감은 더욱 세밀하고 초점이 맞추어진 영역 경계가 그렇지 않은 경계보다 날카롭다는 것이다. [6]에서는 초점연산자를 정의하여 영상의 각 영역에서 초점정도를 계산한다. 초점연산자의 종류에는 local scale 기반 연산, sobel mask 기반 연산, 표준편차 기반 연산 등이 있다. 이러한 접근법의 경우 실제로 가까운 영역에 초점을 맞추어 촬영한 영상이나 배경이 단조로운 영상에서는 깊이 추출의 성능이 좋지만, 영상 전체적으로 초점의 범위를 넓혀서 촬영한 영상이나 복잡한 배경을 가지는 영상에서는 성능이 저하된다.

그림 3. 초점정보에 의한 깊이인지

1.3 기하정보 기반 접근법

[그림 4]는 3D 변환을 위해 2D 영상 내 존재하는 기하학적 요소를 보여준다. 우선 하늘영역이 가장 멀리

있고 영상의 아래쪽으로 갈수록 카메라로부터의 깊이가 가깝다는 것을 알 수 있으며, 영상 내 존재하는 직선들이 만나는 한 점인 소실점(vanishing point)이 가장 멀리 있고 그 점을 지나는 직선들인 소실선(vanishing lines)을 따라 점점 가까워지는 것을 인지할 수 있다. 기하학적 정보를 통한 접근법[8][9]에서는 이러한 단서를 활용하고 있으며, 이를 통한 깊이 추출을 depth from geometric perspective 라고 한다.

그림 4. 3D 입체 인지를 위한 2D 영상 내 기하학적 요소

[8]에서는 우선 입력영상의 색상정보를 통해 영상을 세 가지 형태로 분류한다. 외부영상(outdoor)의 경우에는 색상분할 과정에서 얻은 정보를 기반으로 하늘영역을 가장 멀리하고, 그 다음으로 산 영역을 하늘보다는 가까운 깊이이면서 땅 영역보다는 먼 깊이로 할당하며, 땅 영역에 대해서는 화면 아래쪽으로 갈수록 가까운 깊이를 할당한다. 기하정보를 갖는 외부영상(Outdoor with geometry)과 내부영상(Indoor)의 경우에는 영상 내의 소실점을 검출하여 그 점으로부터 멀어질수록 가까운 깊이를 할당한다. 단, 기하정보를 갖는 외부영상의 경우에는 하늘영역의 깊이는 가장 멀리한다. 기하정보를 이용한 2D-to-3D 변환기법들에서는 정의된 각 영역의 색상정보가 올바르게 없을 경우, 첫 단계인 영상 분류에서부터 오류를 가지게 된다. 실제 자연영상에서는 하늘 또는 산의 색상이 다양하게 존재하기 때문에 이러한 색상정보를 일정한 범위로 한정하는 데에는 한계가 있다. 기하정보를 이용한 2D-to-3D 변환기법들의 또 다른 문제점으로는 소실점의 오검출로 인한 잘못된 깊이영상 형성이 있다. 소실점의 오검출은 깊이 영상의 심각한 오류를 초래한다. 영화나 드라마 등에서 풍경을 제외한 일반적인 자연영상의 경우 인물 또는 주 객체를 포함하고 있다. 하지만 소실점과 소실선들은 배경영역

을 기준으로 추출 되어야 하기 때문에 주 객체들과 배경영역의 경계에서 형성되는 직선성분들은 올바른 소실점 검출에 어려움으로 작용한다.

2. 깊이 융합을 통한 접근법

깊이 융합을 통한 접근법에서는 2D-to-3D 변환을 위해 한 가지 단서만을 이용하지 않고 몇몇의 단서를 함께 활용하여 각각의 깊이영상을 생성한 후 융합 알고리즘을 통해 최종 깊이영상을 추출한다. [10]에서는 기하정보를 이용한 깊이영상과 움직임정보를 통해 추출한 깊이영상을 합성하며, [11]에서는 추가적으로 질감정보를 통한 깊이 영상 또한 이용한다. [12]에서는 몇몇 인간의 시각적 인지 특성을 이용하여 전역 깊이 영상과 국부 깊이 영상을 추정하여 융합한다. 전역 깊이는 사람이 어떠한 광경을 바라볼 때 대부분 앞부분이 관찰자와 더 가깝다는 시각적 해석을 기반으로 한다. 국부 깊이는 붉은 계통이나 밝은 색상이 푸른 계통이나 어두운 색상보다 더 가까이 느껴지는 인간의 인지적 특성을 기반으로 한다.

개별 단서 기반의 접근법들은 오로지 하나의 단서에만 의존하기 때문에 그 단서가 가지는 가정이 만족하는 제한된 경우에만 성공적으로 깊이 영상을 추정할 수 있다. 그러므로 각각의 단서가 가지는 약점은 숨기고 장점은 살리기 위해 최근의 3D 변환 기법들은 대부분 깊이 융합 기반 접근법들이다. 하지만 장면 유형의 다양성 때문에 각 단서에 의한 깊이 영상들의 안정적인 융합은 매우 어려운 문제이다. [10-12]의 공통점은 각 단서의 깊이 영상들에 대한 가중치를 계산하는 가중치 기반의 합성 알고리즘이라는 것이다. 이는 활용하는 단서들 중 하나라도 깊이 맵이 올바르게 않게 형성된 경우 매우 부자연스러운 결과를 초래한다. 또한, 가중치 기반의 깊이 융합의 경우 아무리 정확하게 가중치를 계산한다 하더라도 기본적으로는 깊이 맵들의 합 연산 방식이기 때문에 인간이 각각의 깊이 단서들을 종합하는 과정과는 차이가 있다. 즉, 인간 두뇌에서 복수 깊이 단서들을 결합하는 과정은 가중치 합산과 같은 accumulation 방식 뿐 아니라, 신뢰성 있는 단서들 간의 cooperation, 다른 단서를 통한 특정 단서의 명확화 (disambiguation)

tion) 및 오류 단서의 거부 (veto) 등과 같은 다양한 형태의 과정으로 구성되며, 최종적인 결합 형태는 관찰되는 장면의 특성에 따라 달라질 수 있다는 것이다. 따라서 본 논문에서는 이러한 인간의 시각시스템에 부합하도록 현재 장면의 유형을 분류하여 그 유형에 적합하도록 최종 깊이 영상을 생성하는 장면 적응적 깊이 융합 기법을 제안한다.

III. 적응적 깊이 융합 기법

제안된 깊이 융합 알고리즘은 기하, 움직임, 초점의 3가지 깊이 단서를 활용하여 입력 영상의 깊이를 추정한다. 그러나 기존의 가중 합 형태의 융합 방식이 갖는 단점을 극복하기 위해 현재 장면에서 각 깊이 단서의 신뢰성을 추정하고, 이를 바탕으로 신뢰성이 검증된 깊이 단서만을 이용한 적응적 깊이 융합 구조를 제안한다. [그림 5]는 제안 시스템의 전체적인 구성도를 나타낸다. 그림에서 볼 수 있듯이 제안 방식은 크게 전역 깊이 생성 과정과 객체 깊이 생성 과정 및 이 두 깊이 정보를 융합하는 과정으로 구성된다.

그림 5. 제안 방식의 구성도

1. 전역 깊이 생성

전역 깊이는 객체 영역을 제외한 배경 영역의 전체적인 깊이 정보를 의미한다. 따라서 전역 깊이는 배경의 기하학적 구조로부터 생성되어야 하며, 제안 방식에서

는 이를 위한 깊이 단서로 소실점 기반의 기하 분석을 활용한다. 뿐만 아니라 가장 보편적 형태의 전역 깊이인 기본 깊이 패턴을 추가로 제공함으로써 신뢰성 있는 소실점 추출이 불가능한 영상에 대해서도 현실감 있는 전역 깊이가 생성될 수 있도록 하였다.

1.1 기하 분석에 따른 전역 깊이 생성

관찰자가 3D 공간의 한 장면을 바라볼 때 보통 인공 구조물들에 의한 평행한 직선들이 시점에서 멀어질수록 간격이 좁아져 한 점으로 수렴하는 특성을 이용한 소실점 및 소실선 추정에 의한 깊이 생성은 투영 기하의 원리에 따라 매우 현실적인 입체감 생성을 가능하게 한다.

제안 방식에서는 허프 변환 기반의 검출 방식을 이용하여 소실점을 추출하였다[8]. 일반적으로 소실점 및 소실선들은 배경영역을 기준으로 추출 되는 것이 바람직하므로 주 객체들과 배경영역의 경계에서 형성되는 에지 성분들은 오히려 정확한 소실점 검출에 잡음으로 작용할 수 있다. 따라서 제안 방식에서는 일반적으로 주 객체와 배경영역 사이에서는 수직방향의 경계가 강하게 형성되는 관측을 바탕으로 입체치 이상의 기울기를 갖는 수직 방향의 직선들은 소실점 추출 과정에서 제거함으로써 정확한 장면의 기하학적 재구성이 가능하도록 하였다. 또한, 영상을 촬영할 때 소실점 및 소실선들의 방향이 일정하더라도 카메라의 위치에 따라 장면을 구성하는 소실선들의 기울기가 달라 질 수 있기 때문에 이를 고려하여 소실점 기반 깊이 영상을 형성하여 더 자연스러운 결과를 얻을 수 있도록 하였다.

식 (1)은 소실점 기반 깊이 생성을 나타낸다.

$$Depth_{vp}(x, y) = D_{min} + \frac{(D_{max} - D_{min})}{dis_{max}} \times dis(x, y) \quad (1)$$

$$dis(x, y) = \sqrt{(x - VP_x)^2 + (y - VP_y)^2}$$

$$dis_{max} = 2\sqrt{Width^2 + Height^2}$$

여기서, $Depth_{vp}$ 는 소실점 기반 전역 깊이이며, x 와 y 는 각각 수평과 수직 위치이다. D_{min} 과 D_{max} 는 각각

최대 깊이 값과 최소 깊이 값으로 본 논문에서는 255와 0으로 한다. $dis(x,y)$ 는 각각 소실점과 (x,y) 위치와의 거리이며, dis_{max} 은 소실점과 (x,y) 위치와의 최대 거리로 영상 대각선 길이의 2배로 한다. VP_x 와 VP_y 는 각각 소실점의 수평 및 수직 좌표이며, $Width$ 와 $Height$ 은 각각 영상의 수평 및 수직 크기이다.

1.2 기본 깊이 패턴을 이용한 전역 깊이 생성

소실점 기반의 깊이가 장면의 전체적인 깊이를 표현하는데 매우 효과적이기는 하지만 모든 장면에서 소실점이 올바르게 추출될 수는 없기 때문에 추정된 기하단서의 신뢰성이 낮은 경우에는 소실점 기반 전역깊이를 대체할 수 있는 추가적인 전역깊이 패턴이 필요하다. 그림 6은 제안 방식에서 도입한 기본 전역깊이 패턴을 보여준다. 기본 깊이 패턴은 영상 윗부분에서 아랫부분으로 갈수록 깊이가 가까워지게 생성되었다. 이는 관찰자가 장면을 바라볼 때 일반적으로 장면의 윗부분보다는 아랫부분이 비교적 관찰자와 가까이 있다는 특성이 카메라로 촬영된 영상에서도 유사하게 적용되기 때문이다. 이러한 기본 패턴의 전역 깊이는 인공구조물을 포함하고 있지 않아 올바른 소실점 검출이 어려운 대부분의 장면에서 배경부분의 깊이 표현에 가장 적합하다. 기본 깊이 패턴에 따른 전역 깊이 생성은 식(2)와 같다.

$$Depth_{basic}(x,y) = D_{min} + \frac{(D_{max} - D_{min})}{Height} \times y \quad (2)$$

그림 6. 기본 패턴 전역 깊이

1.3 기하 분석의 신뢰성 검증 및 전역 깊이 선택

2D 동영상 내 모든 장면에 항상 인공구조물이 존재하는 것은 아니며 인공구조물이 존재하더라도 영상 내 직선성분들을 모두 인공구조물에 의한 것으로 볼 수는 없다. 이러한 장면의 다양한 특성들로 인해 많은 영상들에서 소실점이 부정확하게 추출될 수 있으며, 이로부

터 생성된 전역 깊이는 실제적인 깊이감과 다른 혼란스러운 깊이 인식을 초래할 수 있다. 특히 [그림 7](b) 및 (c)와 같이 자연풍경 장면이나 객체가 비교적 카메라와 가까운 장면에서는 이러한 현상이 더욱 심화된다.

본 논문에서는 현재 장면에서 추정된 소실점 및 소실선이 전역 깊이를 효율적으로 표현할 수 있는지를 판단하는 기하단서의 신뢰성 검사를 수행하여, 신뢰성이 있는 경우는 뛰어난 입체감의 소실점 기반 깊이를 전역 깊이로 선택하고, 그렇지 않은 경우는 가장 자연스러운 깊이 패턴인 기본 패턴의 깊이를 전역 깊이로 선택함으로써 전역 깊이 인식의 오류를 최소화하였다.

그림 7. (a) 올바른 소실점에 의한 전역깊이
(b),(c) 잘못된 소실점에 의한 전역깊이

그림 8. [그림 7] 영상들의 소실선 상 실제 에지 화소

소실점 검출의 신뢰성 검사 알고리즘은 검출된 소실선들의 강도를 기반으로 설계된다. [그림 8]에서 관찰할 수 있듯이, 소실점 및 소실선이 배경의 인공구조물에서 생성된 직선들로부터 추정된 경우에는 에지화소가 소

실선 상에 위치할 확률이 높은 반면, 자연풍경이나 카메라와 가까운 객체에서 생성된 에지성분에 의해 잘못 추정된 소실점의 경우에는 에지화소가 추정된 소실선 상에 위치할 확률이 낮아진다. 따라서 이 특성을 바탕으로 식 (3)을 이용하여 기하단서의 신뢰성을 검증한다.

$$R_{geometry} = \begin{cases} 1, & VL_{edge} > \alpha \\ 0, & otherwise \end{cases} \quad (3)$$

여기서, $R_{geometry}$ 는 기하단서의 신뢰성이며, VL_{edge} 는 소실선 상에 위치하는 실제 에지화소의 개수를 나타낸다. VL_{edge} 를 계산하기 위한 의사코드는 식(4)와 같다. 여기서, N 은 소실선의 개수이며, VL_i 는 i 번째 소실선 상에 위치하는 실제 에지화소의 개수이다. L_i 는 i 번째 소실선의 길이이며, P_i^j 와 P_{edge} 는 각각 i 번째 소실선 상의 j 번째 화소와 실제 에지화소를 뜻한다.

$$\begin{aligned} & \text{for } i = 0; i < N; i++ \\ & \quad VL_i = 0 \\ & \quad \text{for } j = 0; j < L_i; j++ \\ & \quad \quad \text{if } P_i^j \equiv P_{edge} \\ & \quad \quad \quad VL_i++ \\ & VL_{edge} = \text{argmin } VL_i, i \in N \end{aligned} \quad (4)$$

$R_{geometry}$ 를 기반으로 전역 깊이 $Depth_{global}$ 를 선택하는 과정은 식 (5)와 같다. 소실선 상에 위치하는 실제 에지화소의 개수가 임계치보다 작은 경우는 추정된 기하단서 즉, 소실점 및 소실선의 신뢰성이 낮은 것으로 판단하여 기본 패턴 전역 깊이가 선택된다.

$$Depth_{global} = \begin{cases} Depth_{vp}, & R_{geometry} = 1 \\ Depth_{basic}, & R_{geometry} = 0 \end{cases} \quad (5)$$

2. 객체 깊이 생성

앞 절에서 설명된 전역 깊이는 장면의 기하학적 구조로부터 추정된 배경 영역에 대한 깊이 정보이다. 따라서 전경 객체가 존재하는 경우, 이 객체 영역에 대한 깊이는 추가적인 깊이 단서를 활용하여 보정되어야 한다. 제안 방식에서는 전경 객체의 깊이 보정을 위한 깊이

단서로 초점과 움직임 정보를 선택적으로 활용한다. 이는 객체 영역은 일반적으로 in-focus이며 따라서 영상에서 비교적 뚜렷하게 나타난다는 점과, 움직임 객체의 경우 움직임 시차에 의해 상대적 깊이를 추정할 수 있다는 점을 바탕으로 한다. 단, 움직임 단서의 경우 전역 카메라 움직임에 따른 객체 깊이 생성의 오류를 방지하기 위해 카메라 이동에 의한 전역 움직임은 없고 객체의 움직임만 존재하는지 여부를 판정하고 이로부터 움직임 단서에 의한 깊이 정보의 신뢰성을 검증한다.

2.1 움직임 기반 깊이와 초점 기반 깊이 생성

객체영역의 깊이를 할당을 위해서는 의미 기반의 객체 분할이 이상적이지만, 일반적인 영상에서 의미 기반의 객체 인식은 매우 복잡한 영상 해석 과정을 필요로 할 뿐 아니라 인식 성능도 매우 제한적이다. 이러한 이유로 대부분의 영역 기반 깊이 추정 기법들에서는 특징 기반의 영역 분할을 이용한다. 본 논문에서는 객체 깊이 생성을 위해 색상특성을 이용하여 영상을 분할 한 후 분할된 각 영역의 깊이를 추정한다. 움직임 기반 깊이 생성을 위해서는 각 영역이 포함하는 8x8블록들의 평균 움직임 크기(magnitude)가 이용되고 초점 기반 깊이 생성을 위해서는 각 영역이 포함하는 화소들의 평균 에지 크기가 이용된다. [그림 9]는 카메라의 이동은 없으면서 객체의 움직임이 존재하는 장면에서의 색상분할 결과와 움직임 및 초점 단서로부터 추정된 깊이 정보들을 보여준다. 이 영상의 경우에는 두 단서 모두 깊이가 추정에 적합한 단서임을 보여준다.

그림 9. 객체 깊이 생성

2.2 움직임 단서의 신뢰성 검증 및 객체 깊이 선택
움직임 기반의 객체 깊이 추출은 전역 움직임이 없는

정적 배경에서 객체의 움직임이 존재하는 경우 뛰어난 성능을 보인다. 하지만 다수의 2D 동영상에서 카메라와 객체가 모두 움직이는 장면들이 존재하며, 이 경우 움직임 정보만을 이용한 깊이 추출은 [그림 10]과 같이 배경과 객체 사이의 상대적 깊이가 표현이 부정확하다. 따라서 본 논문에서는 장면 내 존재하는 움직임이 카메라에 의한 것이 아니고 오로지 객체에 의한 것일 때 움직임 기반 깊이를 신뢰할 수 있는 것으로 가정한다.

그림 10. 카메라와 객체의 복합적인 움직임 장면

[그림 11]은 카메라 움직임이 있는 장면과 그렇지 않은 장면에 대한 움직임 벡터 분석을 보여준다. 하나의 화면을 32개의 블록으로 나눈 후 움직임 벡터의 크기와 방향을 각각 밝기와 화살표로 표시하였다. [그림 11](a)에서는 카메라가 좌측에서 우측으로 움직이고 있으며 움직임의 방향이 거의 일정하고 움직임의 크기가 모든 블록에서 매우 큰 것이 관찰된다. [그림 11](b)의 카메라 줌인 장면에서는 움직임의 크기가 중앙 쪽 블록들보다 좌우측 경계 쪽 블록들에서 더 크다는 것이 관찰된다.

그림 11. 다양한 장면들에 대한 움직임 벡터 분석

반대로, [그림 11](c)의 카메라 움직임이 없고 움직이는 객체가 있는 경우에는 경계 쪽 블록의 움직임 크기가

가 더 작다는 것을 볼 수 있다.

이를 종합적으로 판단했을 때 카메라 움직임이 있는 장면에서는 경계 쪽 블록들의 움직임 크기가 크다는 것을 알 수 있고, 카메라 이동이 없이 객체부분에 움직임

이 존재하는 경우에는 경계 쪽 블록들의 움직임 크기가 작으면서 그 이외 영역의 평균 움직임 크기가 일정치 이상이라는 것을 알 수 있다. 이러한 움직임 벡터의 국부 특성을 이용하여 식 (6)과 같이 카메라 이동여부와 객체 부분의 움직임을 판단하여 움직임 단서의 신뢰성을 검사한다.

$$R_{Motion} = \begin{cases} 1, & \widetilde{M}_b < \alpha \text{ and } \widetilde{M}_i > \beta \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

여기서, R_{Motion} 은 움직임 단서의 신뢰성이며 α 와 β 는 검사를 위한 임계값들이다. $\widetilde{M}_b$ 와 $\widetilde{M}_i$ 는 각각 경계 블록들의 움직임 평균과 그 이외 블록들의 움직임 평균이며, 식 (7)과 같이 얻어진다.

$$\begin{aligned} \widetilde{M}_b &= \frac{1}{8} \sum_{B_i \in \{Boundary\ Blocks\}} M_{B_i} \\ \widetilde{M}_i &= \frac{1}{24} \sum_{B_i \in \{Inner\ Blocks\}} M_{B_i} \end{aligned} \quad (7)$$

여기서 M_{B_i} 은 각 블록의 움직임크기로 블록 크기는 가로가 영상 수평 크기의 1/8 이고 세로가 수직 크기의 1/4이다. 즉, $\widetilde{M}_b$ 는 영상의 오른쪽 및 왼쪽에 있는 8개의 경계 블록들에 대한 움직임 평균이며, $\widetilde{M}_i$ 는 경계 블록을 제외한 나머지 24개 내부 블록들에 대한 움직임 평균으로 계산된다.

식(6)에 의한 움직임 단서가 신뢰성 있는 것으로 판정되면, 움직임 단서만을 이용하여 객체 영역의 깊이를 추정한다. 이는 객체가 움직이면 움직임 영역에서 motion blur가 발생하여 초점 단서의 신뢰도가 떨어지기 때문이다. 반면에 $R_{Motion} = 0$ 으로 판정되면 초점 기반 깊이를 이용하여 객체 깊이를 생성한다. 식 (8)은 객체 깊이의 선택 과정을 보여준다.

$$Depth_{object} = \begin{cases} Depth_{motion}, R_{motion} = 1 \\ Depth_{focus}, R_{motion} = 0 \end{cases} \quad (8)$$

$$Depth_{motion}(x, y) = \alpha_m M_{R_i} + \beta_m, (x, y) \in R_i \quad (9)$$

$$M_{R_i} = \frac{1}{N_i} \sum_{(x, y) \in R_i} MotionMap(x, y)$$

$$Depth_{focus}(x, y) = \alpha_f F_{R_i} + \beta_f, (x, y) \in R_i \quad (10)$$

$$F_{R_i} = \frac{1}{N_i} \sum_{(x, y) \in R_i} FocusMap(x, y),$$

$$i = 0, \dots, l$$

여기서, $Depth_{object}$ 는 객체 깊이이며, $Depth_{motion}$ 과 $Depth_{focus}$ 는 각각 움직임 기반 깊이와 초점 기반 깊이를 나타낸다. M_{R_i} 및 F_{R_i} 는 색상분할[13] 후 얻어진 l 개의 영역들 중 i 번째 영역에 대한 평균 움직임 및 초점 정도이다. 그리고 $MotionMap$ 은 8x8 블록 움직임 맵이며, $FocusMap$ 은 [6]의 방식들 중에서 8x8 블록 표준편차 기반으로 얻어진 초점 맵이다.

3. 장면 유형 분류 및 적응적 깊이 융합

제안 기법에서 입력 장면은 [표 1]에 요약된 바와 같이 전역 및 객체 깊이를 생성을 위한 깊이 단서의 신뢰성 검사 결과를 기준으로 모두 4개의 유형으로 분류된다. 각 장면 유형 별로 신뢰성 있는 깊이 단서를 활용하여 전역 및 객체 깊이 정보가 생성되면 이를 토대로 최종 깊이를 생성을 위한 깊이 융합이 수행된다. 깊이 융합 알고리즘은 선택된 객체 깊이를 단서를 기준으로 크게 두개의 방식으로 나뉜다.

우선, 객체 깊이로 초점 기반 깊이가 선택된 경우에는 가중 합 방식의 융합 알고리즘이 사용된다. 영상 촬영시에 렌즈로 일정 거리의 객체에 초점을 맞추면 배경 영역이 이 객체로부터 멀어질수록 영상의 포터함의 정도는 점진적으로 낮아진다. 따라서 초점 단서로부터 추정된 깊이는 전경 영역뿐 아니라 배경 영역의 상대적 깊이도 포괄적으로 반영하게 된다. 그러므로 이 경우에는 가중합 형태의 융합 방식을 사용하여 두 단서로부터 추정된 깊이 정보를 병합함으로써 깊이 추정의 신뢰도를 향상시킨다.

객체 깊이로 움직임 단서가 선택된 경우에는 전경과 배경 영역에 대한 독립적 깊이 할당 방식의 융합 알고리즘을 사용한다. 이는 움직임 객체 영역은 이 객체에 의해 배경 영역이 가려지게 되며, 또 배경 영역에서는 움직임 단서에 따른 깊이 정보가 존재하지 않기 때문이다. 따라서 최종 깊이 융합은 전역 깊이 위에 객체 깊이를 움직임 크기에 따라 독립적으로 할당하는 방식으로 수행된다. 즉, 장면에 존재하는 움직임 크기의 구간을 나누어 움직임이 클수록 전역 깊이 스케일에서 가까운 깊이를 할당한다.

표 1. 장면 유형별 깊이 선정 및 깊이 융합 알고리즘

type	reliability		depth selection		fusion method
	Geo	motion	global	object	
1	○	○	depth from geometry	depth from motion	independent depth allocation for global & object depth
2	X	○	depth from basic pattern	depth from motion	
3	○	X	depth from geometry	depth from focus	accumulation of global & object depth
4	X	X	depth from basic pattern	depth from focus	

IV. 실험 결과

실험에서는 실제 응용을 고려하여 상업용 영화와 드라마에서 캡처된 비디오를 실험 영상으로 사용하였으며, 기존의 가중합 형태의 융합 방식[11]과의 성능 비교를 통해 제안 방식의 우수성을 검증하였다.

[그림 12]는 유형 1의 실험결과를 보여준다. [그림 12]의 원본영상은 객체의 움직임만 존재하며 인공 구조물에 의한 소실점 추출이 가능한 영상이다. 소실점 기반 깊이와 움직임 기반 깊이가 사실감 있게 추정 되었다는 것을 [그림 12](a)와 [그림 12](c)에서 확인할 수 있다. 반면 넓은 DOF로 인해 초점 기반 깊이는 [그림 12](b)와 같이 부정확하게 형성된다. 최종 깊이 융합에서 [그림 12](d)의 제안방식의 경우 올바른 소실점에 의한 전역깊이 위에 움직임 기반 깊이영상이 독립적인 객체깊

이로 할당되었음이 관찰된다. 반면 [그림 12](c)의 가중합에 의한 융합에서는 부정확한 초점 기반 깊이까지도 사용됨으로써 전체 융합 결과의 성능이 떨어진다. [그림 12](f)와 [그림 12](g)는 각각 제안방식과 가중합 방식에 의해 합성된 좌우영상을 보여준다. 제안방식에 의해 합성된 좌우영상에 객체부분이 가중합 방식에 의한 것 보다 더욱 옆으로 이동되어 있는 것을 관찰할 수 있다. 즉, 좌 영상에서는 더욱 오른쪽으로 우 영상에서는 더욱 왼쪽으로 이동되어 있는 것을 알 수 있다. 이는 제안방식에서는 객체부분에 대한 깊이가 움직임 깊이에 의해 독립적으로 할당되기 때문에 다른 깊이 영상에 영향을 받지 않기 때문이다.

그림 12. 유형1의 영상에 대한 실험 결과

[그림 13]은 유형 2의 실험결과를 보여준다. 원본영상은 기하 분석에 따른 전역 깊이가 추정이 불가능하며, 객체 움직임만 있는 특성을 갖는다. [그림 13](a), (b), (c)를 통해 이를 확인할 수 있으며, 이러한 영상은 유형 2로 분류되어야 한다. [그림 13](d)의 제안방식의 경우 신뢰성 검사에 의해 이 영상들을 유형 2로 정확히 분류하였으며 기본패턴 전역깊이 위에 움직임 깊이를 독립적으로 할당하여 최종 깊이를 사실감 있게 생성하였다.

반면 [그림 13](e)의 가중합 방식에서는 기하기반 깊이 추정의 오류가 최종 깊이 생성에 악영향을 미치고 있음을 알 수 있다. [그림 13](f)와 [그림 13](g)에서 각각 제안방식과 가중합 방식에 의해 합성된 좌우영상을 보여준다. 제안방식에 의한 영상들이 가중합 방식에 의해 합성된 영상들 보다 위쪽에서 아래쪽으로 갈수록 시차가 커지는 것이 관찰된다. 이를 통해 소실점 기반 전역깊이가 잘못 추정된 유형 2와 같은 영상에서 기본패턴 전역깊이가 효과적으로 작용한다는 것을 알 수 있다.

그림 13. 유형2의 영상에 대한 실험 결과

[그림 14]의 원본장면은 카메라 움직임이 존재하는 장면이며 소실점에 의한 깊이인지가 가능한 장면이다. 제안방식에서 이 영상은 유형 3으로 분류되었으며, 초점 기반 깊이와 소실점 기반 전역깊이가 가중 합산되어 [그림 14](d)의 최종 깊이가 생성되었다. [그림 14](a), (b), (c)에서 각각 움직임 기반 깊이, 초점 기반 깊이, 소실점 기반 깊이를 보여주고 있다. [그림 14](c)의 소실점 기반 깊이가 제대로 추정되었음을 알 수 있으며, 움직임 기반 깊이에서 보다는 초점 기반 깊이에서 객체부분의 전체적인 윤곽을 잘 표현하고 있음을 알 수 있다. [그림 13](f)와 [그림 13](g)는 각각 제안방식과 가중합 방식에 의해 합성된 좌우영상을 보여준다.

그림 14. 유형3의 영상에 대한 실험 결과

[그림 15]의 원본장면은 소실점과 움직임 단서 모두 신뢰할 수 없는 장면이다. 원본장면의 전역깊이를 위해 [그림 15](c)의 불확실한 소실점 정보에 의한 깊이형성보다 제안된 기본 패턴 깊이가 오히려 더 알맞다는 것을 알 수 있다. 또한, 카메라의 움직임으로 인해 [그림 15](a)의 움직임 기반 깊이보다는 [그림 15](b)의 초점 기반 깊이가 장면의 전체적인 깊이에 표현에 더 적당하다는 것을 알 수 있다. 제안방식에 의해 유형 4로 분류되어 기본 패턴 전역깊이와 초점 기반 깊이가 가중 합산된다. [그림 15](d)와 [그림 15](e)에서 각각 제안방식과 가중합 방식에 의해 융합된 최종 깊이 영상을 보여주고 있다. 제안방식에 의해 융합된 깊이가 장면을 더욱 효과적으로 표현하고 있음을 알 수 있다. [그림 15](f)와 [그림 15](g)는 각각 제안방식과 가중합 방식에 의해 합성된 좌우영상이다.

그림 15. 유형4의 영상에 대한 실험 결과

그림 15. 유형4의 영상에 대한 실험 결과

전체 실험 결과는 각 단서가 장면유형에 따라 타당하게 활용되었고 각 단서 기반의 깊이들을 가중 합산하는 경우보다 뛰어난 성능을 가진다는 것을 보여준다.

V. 결론

본 논문은 2D-to-3D 변환을 위해 장면 유형에 적응적인 깊이 융합 기법을 제안하였다. 현재장면의 최종 깊이영상은 신뢰할 수 있는 단서들의 통합과 쓸모없는 단서들의 제위를 통해 추정된다. 각 단서들의 신뢰성을 검사하여 현재 장면의 유형과 장면유형에 적응적인 융합 기법이 결정된다. 기하단서의 사용여부는 검출된 소실점 및 소실선 정보를 이용하여 결정되고 움직임 기반 깊이의 사용여부는 움직임 분석에 의해 결정된다. 제안된 깊이 융합 기법의 실험결과를 통해 각 장면 유형별로 신뢰성 있는 깊이 단서들이 깊이 융합에 사용됨으로써 보다 현실적이고 효과적인 깊이 융합이 가능함을 확인하였다.

참 고 문 헌

- [1] 한찬희, 송인환, 이시웅, “세포동영상의 자동분석을 위한 효율적인 세포추적방법”, 한국콘텐츠학회논문지, 제11권, 제5호, pp.32-40, 2011.
- [2] Y. Matsumoto, H. Terasaki, K. Sugimoto, and T. Arakawa, “Conversion system of monocular

- image sequence to stereo using motion parallax," Proc. of SPIE, Vol.3012, pp.108-115, 1997.
- [3] T. Okino, H. Murata, K. Taima, T. Iinuma, and K. Oketani, "New television with 2D/3D image conversion techniques," Proc. of SPIE, Vol.2653, pp.96-103, 1995.
- [4] M. T. Pourazad, P. Nasiopoulos, and R. K. Ward, "An H.264-based Scheme for 2D to 3D Video Conversion," IEEE Transactions on Consumer Electronics, Vol.55, No.2, pp.742-748, 2009(5).
- [5] D. Kim, D. Min, and K. Sohn, "A stereoscopic video generation method using stereoscopic display characterization and motion analysis," IEEE Transactions on broadcasting, Vol.54, No.2, pp.188-197, 2008(6).
- [6] W. J. Tam, A. S. Yee, J. Ferreira, S. Tariq, and F. Speranza, "Stereoscopic image rendering based on depth maps created from blur and edge information," Proc. of SPIE, Vol.5664, pp.104-115, 2005.
- [7] C. C. Cheng, C. T. Li, and L. G. Chen, "A 2D-to-3D Conversion System Using Edge Information," Digest of Technical Papers International Conference on Consumer Electronics, pp.377-378, 2010.
- [8] S. Battiatto, S. Curtib, M. L. Casciac, M. Tortorac, and E. Scordato, "Depth Map Generation by Image Classification," Proc. of SPIE, Vol.5302, pp.95-104, 2004.
- [9] S. Battiatto, A. Capra, S. Curti, and M. La Cascia, "3D stereoscopic image pairs by depth-map generation," Second International Symposium on 3D Data Processing, Visualization and Transmission, pp.124-131, 2004.
- [10] C. C. Cheng, C. T. Li, P. S. Huang, T. K. Lin, Y. M. Tsai, and L. G. Chen, "A Block-based 2D-to-3D Conversion System with Bilateral Filter," Digest of Technical Papers International Conference on Consumer Electronics, pp.1-2, 2009.
- [11] Y. L. Chang, J. Y. Chang, Y. M. Tsai, C. L. Lee, and L. G. Chen, "Priority Depth Fusion for the 2D-to-3D Conversion System," Proc. of SPIE, Vol.6805, pp.680513.1-680513.8, 2008.
- [12] S. F. Tsai, C. C. Cheng, C. T. Li, and L. G. Chen, "A real-time 1080p 2D-to-3D video conversion system," IEEE, Transactions on Consumer Electronics, Vol.57, No.2, pp.915-922, 2011(5).
- [13] D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.750-755, 1997.

저 자 소 개

한 찬 희(Chan-Hee Han)

준회원

- 2008년 2월 : 한밭대학교 멀티미디어공학전공(공학석사)
 - 2008년 3월 ~ 현재 : 한밭대학교 멀티미디어공학과 박사과정
- <관심분야> : 영상처리, 영상부호화

최 해 철(Hae-Chul Choi)

정회원

- 2004년 8월 : KAIST 전기및전자공학과(공학박사)
- 2004년 9월 ~ 2010년 4월 : 한국전자통신연구소 방송미디어연구부 선임연구원
- 2010년 3월 ~ 현재 : 한밭대학교

정보통신공학과 조교수

<관심분야> : 영상통신, 비디오부호화, 패턴인식

이 시 웅(Si-Woong Lee)

정회원

- 1997년 8월 : KAIST 전기및전자
공학과(공학박사)
- 1995 ~ 2000년 : 삼성전자 선임
연구원
- 2000년 4월 ~ 현재 : 한밭대학교
정보통신공학과 교수

<관심분야> : 컴퓨터비전, 영상처리, 영상압축