

일반논문-11-16-1-05

Depth layer partition을 이용한 2D 동영상의 3D 변환 기법

김수동^{a)}, 유지상^{a)‡}

3D conversion of 2D video using depth layer partition

Sudong Kim^{a)} and Jisang Yoo^{a)‡}

요 약

본 논문에서는 depth layer partition을 이용한 2D 동영상의 자동 3D 변환 기법을 제안한다. 제안하는 기법에서는 먼저 2D 동영상의 장면 전환점을 검출하여 각각의 프레임 그룹을 설정하여 움직임 연산 과정에서의 오류 확산을 방지하여 깊이맵(depth map) 생성과정에서 오차를 줄여준다. 깊이정보는 두 가지 방법으로 생성되는데 하나는 영역 분할과 움직임 정보를 이용하여 깊이맵을 추출하는 것이고 다른 하나는 에지 방향성 히스토그램(edge directional histogram)을 이용하는 방법이다. 제안하는 기법에서는 객체와 배경을 분리하는 depth layer partition 과정을 수행한 후 생성된 두 개의 깊이맵을 원 영상에 최적이 되도록 병합하게 된다. 제안된 기법으로 신뢰도 높은 깊이맵과 결과 영상을 생성할 수 있다는 것을 다양한 실험 결과를 통해 알 수 있다.

Abstract

In this paper, we propose a 3D conversion algorithm of 2D video using depth layer partition method. In the proposed algorithm, we first set frame groups using cut detection algorithm. Each divided frame groups will reduce the possibility of error propagation in the process of motion estimation. Depth image generation is the core technique in 2D/3D conversion algorithm. Therefore, we use two depth map generation algorithms. In the first, segmentation and motion information are used, and in the other, edge directional histogram is used. After applying depth layer partition algorithm which separates objects(foreground) and the background from the original image, the extracted two depth maps are properly merged. Through experiments, we verify that the proposed algorithm generates reliable depth map and good conversion results.

Keywords : 2D/3D conversion, 3D conversion, 2D video, segmentation, motion information, depth layer partition, depth map generation

1. 서론

영화 “아바타” 개봉 이후 최근 3D 콘텐츠에 대한 관심

이 매우 높아지고 있고, 3D 미디어가 방송, 게임, 엔터테인먼트 등 다양한 분야에서 서비스되는 등 큰 관심을 받고 있다. 삼성, LG, 파나소닉 등의 TV 제작사들도 다양한 종류의 3D 기능이 탑재된 TV를 출시하고, 노트북, 디지털 카메라 등 다양한 개인용 3D 기기도 보급되고 있다. 따라서 3D 콘텐츠에 대한 수요는 앞으로 계속 증가할 것으로 예측되지만 공급은 그에 비해 매우 부족한 것이 현실이다. 초기 투자비용 때문에 3D 콘텐츠 제작비용이 아직 고가인

a) 광운대학교
Electronic Engineering Dept. Kwangwoon University
‡ 교신저자 : 유지상 (jsyoo@kw.ac.kr)
※ 본 연구는 2010년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 구입(한국연구재단-2010-0026245)
· 접수일(2010년9월10일), 수정일(1차:2010년11월10일,2차:12월2일), 게재확정일(2010년12월20일)

편이며, 이런 현상은 당분간 계속될 것으로 전망된다. 3D 콘텐츠의 공급 부족 문제를 해결할 수 있는 방법 중의 하나가 기존의 2D 콘텐츠를 3D 입체 형태로 변환하는 방법이다. 3D 변환 기법은 여러 가지가 있지만 크게 수동기법과 자동기법 두 가지로 분류할 수 있다. 물론 모두 3D 카메라를 이용하여 직접 제작하는 것보다는 비용을 절감할 수 있으며 특히 자동 기법인 경우 비용과 시간을 많이 절약할 수 있다. 하지만 변환의 한계성으로 인한 성능 부족이 문제로 지적되고 있다.

3D 변환 연구는 90년대부터 지금까지 활발히 진행되고 있다. 자동 변환기술은 기본적으로 Ross 현상에 기반을 두고 있다. Ross 현상은 정신물리학 이론으로써 양안의 시간적인 지연이 입체감을 느끼는 원인이 된다는 것이다^[1]. 많은 변환기법들이 제안되었으며 대표적인 자동변환 기법으로는 T. Okino의 MTD(modified time difference) 기법^[2], Y. Matsumoto의 운동 시차(motion parallax)를 이용한 스테레오 동영상 변환기법^[3] 등이 있다. 압축 영상의 움직임 정보를 이용한 연구도 진행되었고^[4], 영상의 깊이 정보를 추출하고 그에 따른 깊이감을 부여하여 변환하는 방법들도 제안되었다^[5,6,7,8]. 2차원 평면 영상으로부터 영상의 입체적인 특징을 파악하는 것은 쉬운 문제가 아니다. 특히 깊이가 정보가 존재하지 않는 2차원 영상으로부터 영상의 특징을 분석하여 깊이 정보를 생성하는 것은 매우 어려운 작업이며 2D/3D 변환의 핵심 과정 중의 하나이다. 영역 분할과 움직임 정보를 이용한 깊이맵 생성 기법^[9]과 에지 방향성 히스토그램을 이용한 깊이맵 생성 기법^[10] 등도 제안되었다. 기존의 방법들 중 MTD 기법은 원본 영상을 그대로 좌우 영상으로 이용하기 때문에 화면간 떨림 현상이 있다는 문제점이 있다. 영역 분할과 움직임 정보를 이용한 변환 기법^[9]은 부정확한 움직임 정보에 대해서 취약한 문제가 있고, 에지 방향성 히스토그램을 이용한 기법^[10]은 프레임간의 상관도를 고려하지 않아 영상이 흔들리며 이로 인해 시청자의 눈이 피곤할 수 있다는 단점이 있다. 또한, 두 가지 기법은 영상의 객체 부분과 배경 부분에 각각 좋은 성능을 보여주지만 반대의 경우에는 취약한 문제가 있다.

이와 같은 문제점들을 보완하고자 본 논문에서는 depth

layer partition을 이용한 변환 기법을 제안한다. 우선 부정확한 움직임 정보에 의한 오류를 줄이고자 장면 전환점을 검출 하여^[11] 각각의 프레임 그룹을 설정한다^[12]. 프레임 그룹을 설정하게 되면 움직임 연산 과정에서의 오류가 확산되는 것을 방지하여 깊이맵과 결과 영상 생성과정에서의 오차를 줄여준다. 제안하는 기법에서는 먼저 두 가지 방법으로 깊이맵을 생성하고 이 두 개의 깊이맵을 적당히 조합하여 최종 깊이맵을 생성하게 된다. 즉, 각각 다른 두 가지의 깊이맵을 획득한 뒤, 객체와 배경을 분리하는 depth layer partition 과정을 수행하여 두 가지의 깊이맵을 병합하고, 최종 3D 변환 영상을 재생한다. 다양한 실험을 통해 제안하는 변환 기법의 성능이 기존의 다른 기법들의 성능보다 우수하다는 것을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 일반적인 입체 변환의 원리에 대해 간단하게 설명하고, 3장에서는 새로 제안하는 3D 변환 기법에 대하여 설명한다. 4장에서는 실험결과에 대한 분석을 하고 마지막으로 5장에서 결론을 내린다.

II. 2D/3D 입체 변환의 원리

1. 입체 영상의 기본 원리

3D 입체영상은 양의 시차, 음의 시차, 영의 시차 등 크게 세 가지 종류의 양안 시차 (binocular disparity)를 제공함으로써 사람들에게 영상 내에서 깊이감을 느낄 수 있게 한다. 시차의 종류는 사람이 스크린을 통해 입체 영상을 보았을 때 수렴점이 어느 위치에 생기느냐에 따라서 결정되며 음의 시차는 스크린의 앞쪽에, 양의 시차는 스크린의 뒤쪽에, 영의 시차는 스크린 상에 수렴점이 존재하는 경우이다.

그림 1은 Ross 현상을 이용한 입체 영상의 원리를 설명하기 위한 예를 보여준다^[1]. 영상내의 비행기는 좌에서 우로 이동하고 있으며 배경이 되는 산은 정지하고 있다. 좌안과 우안은 모두 배경이 되는 산과 비행기를 바라보게 되며 그때 좌영상과 우영상의 차이인 시차가 발생하게 된다. 따라서 비행기는 스크린보다 앞쪽에 수렴점이 생기는 음의

시차를 얻을 수 있고, 산은 스크린에 수렴점이 생기게 되어 영의 시차를 얻을 수 있다.

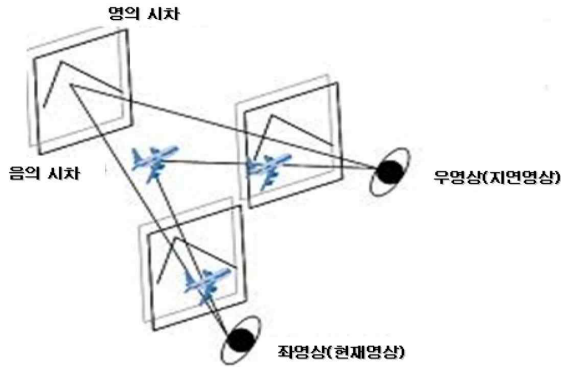


그림 1. Ross 현상 기반의 3D 변환 원리
Fig. 1. Basic principle of 3D conversion based Ross phenomenon

이와 같은 Ross 현상에 기반을 둔 입체영상에서 시청자는 음의 시차와 양의 시차에 따라 영상 내 객체가 화면 바깥쪽으로 나와 보이거나 화면 안쪽으로 들어가 보이는 현상 등으로 입체감을 느낄 수 있게 된다.

2. 2D/3D 변환 원리

보통 3차원 영상은 2대의 카메라로 구성된 3D 카메라를 이용하여 직접 획득되지만 2D/3D 변환 작업은 상대적으로 복잡한 원리 및 영상 처리 과정을 필요로 한다. 특히 자동변환의 경우에는 다른 깊이감을 갖는 서로 다른 객체를 정확하게 분리하여야 하는 등 고난이도의 영상처리 기술을 필요로 한다. 수동 방법은 객체 분리 등의 문제를 직접 사람의 손으로 해결하는 것이 가능하지만 자동 방법에서는 모

든 작업이 자동으로 처리되어야 한다.

그림 2는 2D/3D 변환의 기본 원리를 보여준다. 일반적으로 TV, DVD 등에서 재현되는 영상은 2D 영상이다. 이러한 2D 영상을 가지고 좌영상과 우영상으로 구성되는 3D 영상을 생성하는 것이 2D/3D 변환의 기본 개념이다. 구현 방법은 크게 세 가지로 분류된다. 첫째, 현재 영상이 좌, 우 영상 중 하나가 되면, 다른 하나는 이전 영상(delayed image) 중에서 하나를 선택한다. 둘째, 현재 영상은 좌, 우 영상 중 하나를 선택하고 다른 영상은 합성 영상으로 구현한다. 셋째 좌영상과 우영상을 모두 합성하여 재구성한다. 본 논문에서는 두 번째 방법을 통해 3D 영상을 생성한다.

3D 영상을 생성하기 위해서는 먼저 주어진 2D 영상에서 장면전환점과 운동 정보를 추출하게 된다. 이 정보들을 가지고 객체의 원근 정도(깊이감)를 구하여 두 눈을 위한 좌, 우영상을 생성하는 것이다. 이때 두 눈에 투시되는 영상이 미묘한 차이가 있듯이, 동일점이 좌영상과 우영상에 투시될 때 두 시점 간의 거리인 시차(parallax)를 구하는 것이 필요하다.

III. 새로 제안하는 3D 변환기법

본 논문에서 제안하는 변환 기법에서는 그림 3과 같은 순서로 2D/3D 변환 과정을 수행한다. 먼저 주어진 2D 동영상의 장면 전환점을 검출하여 프레임 그룹을 설정한다. 설정된 각각의 프레임 그룹은 다음 단계에서 수행될 움직임 연산 과정에서의 오류가 확산되는 것을 방지하여 깊이맵(depth map) 생성과정에서의 오차를 줄여주는 역할을 한다.

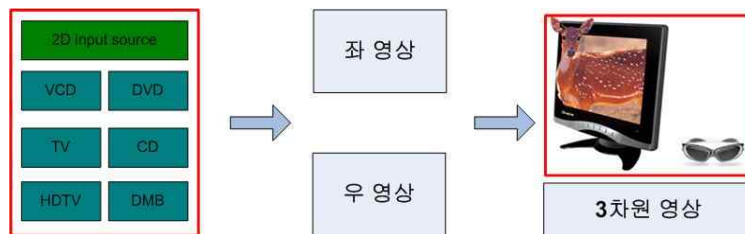


그림 2. 2D/3D 입체 변환 원리
Fig. 2. Principle of 2D/3D conversion

다음으로 두 가지 방법을 이용하여 깊이맵을 추출하는데, 첫 번째 방법은 영역 분할과 움직임 정보를 이용하여 깊이맵을 추출하는 것이고 다른 하나는 에지 방향성 히스토그램을 이용하는 방법이다. 영역 분할과 움직임 정보를 이용하여 란을 깊이맵은 영상의 객체 부분에서 좋은 성능을 보여주고, 에지 방향성 히스토그램을 이용하여 란을 깊이맵은 영상의 배경 부분에서 좋은 성능을 보여준다. 따라서 객체와 배경을 토그램을 각각에 대램을 최적의 깊이맵을 적용하는 것이 새로이 제안하는 기법의 핵심내용이다. 따라서 제안하는 기법에서는 depth layer partition 과정을 수행하여 객체와 배경을 분리하고 이미 앞에서 생성한 두 가지의 깊이맵을 주어진 영상의 객체와 배경의 특징에 최적이 되도록 병합하여 최종 깊이맵을 생성하고, 이것을 이용하여 최종 3D 입체 영상을 생성하게 된다.

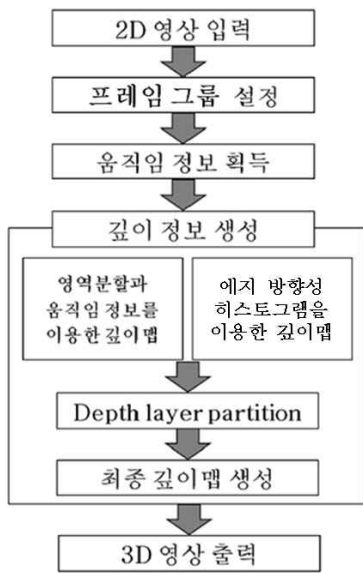


그림 3. 제안하는 2D/3D 변환 기법의 흐름도
 Fig. 3. Flow chart of proposed 2D/3D conversion algorithm

1. 프레임 그룹 설정 및 움직임 정보 획득

영상의 움직임 정보를 추정하기 위하여 일반적으로 전, 후 프레임에 이용한 움직임 추정 연산을 수행하지만, 전후 프레임의 움직임 정보가 거의 존재하지 않는 영상에서는

잘못된 결과값으로 인해 변환 과정에서 오차가 발생할 수 있다. 제안하는 기법에서는 이를 방지하기 위해 연산할 프레임들을 그룹화하여 양방향 탐색을 수행한다^[12].

프레임을 유사한 특성을 갖는 그룹으로 나누기 위해 식 (1)을 이용하여 장면전환점 검출연산을 먼저 수행하게 된다^[11].

$$\sum_{j=1}^N |H_i(j) - H_{i+1}(j)| > T \quad (1)$$

여기서 $H_i(j)$ 는 i 번째 영상에서의 히스토그램 값 j ($0 \sim 255$)의 누적 히스토그램 분포를 나타내고 T 는 허용 오차값이다. 히스토그램 분포의 차이가 허용오차 이상이라면 장면 전환점이라고 판단한다. 장면 전환점이라고 판단된 i 번째 영상은 그룹의 마지막 영상으로 설정되고, 설정된 시작 영상부터 마지막 영상까지의 그룹화된 입력 영상에서 움직임 추정 연산을 수행하게 된다.

$$C_{SAD}(x,y) = \sum_{i=0}^6 \sum_{j=0}^6 |Y_n(x+i,y+j) - Y_{n+1}(d_x+i,d_y+j)| \quad (2)$$

식 (2)의 SAD(sum of absolute difference) 값을 비용값으로 하고 밝기 정보 Y 를 사용하여 비용값을 구한다. 식 (2)에서 Y_n 은 n 번째 프레임의 Y 값을 의미하고 Y_{n+1} 은 $n+1$ 번째 프레임의 Y 값을 의미한다. dx 와 dy 는 탐색 위치에 따라서 변경되는 정합창 내에서의 시작점을 의미한다. 움직임 추정 연산은 화소 단위로 수행한다. 화소 단위의 움직임 추정 방법은 모든 화소가 변이를 갖게 됨으로 블록 기반의 움직임 예측보다 신뢰도 있는 세밀한 움직임 예측이 가능하다^[13]. 7×7 크기의 정합창을 사용하며, 움직임 예측 과정은 spiral search를 통해^[14], 정합창 내의 화소값을 비교함으로써 이루어진다.

그림 4와 같이 i 번째 프레임과 $(i+1)$ 번째 프레임간의 움직임 추정 연산 수행 중 움직임 값이 존재하는 화소수가 전체 화소수의 일정비율 이하로 존재한다면 움직임 정보가 없다고 판단하고, i 번째 영상은 $(i+2)$ 번째 영상과 움직임 추정 연산을 수행하게 한다. 마찬가지로 $(i+2)$ 영상과의 움직임 추정 결과 역시 움직임 정보가 없다면 다음 프레임과



그림 4. 움직임 정보의 유무에 따른 탐색 영상 변경
 Fig. 4. Change of search frame depending on existence of motion information

비교하게 된다. 설정된 프레임 그룹내의 마지막 영상이거나, 또는 연속할 프레임 그룹이 마지막 영상과 인접하여 연속 가능한 다음 프레임이 존재하지 않는다면, 프레임 그룹의 탐색을 역방향으로 수행하게 된다. 즉, i번째 영상과 (i+1)번째 영상 또는 그 다음 영상에서의 움직임 정보가 존재하지 않는다고 판단되면, (i-1)번째 프레임 방향과 움직임 추정을 수행하게 된다.

카메라 움직임으로 인한 영상 전체의 움직임 정보가 존재하는 경우는 영상 전체에서 분포하는 전역 움직임 값(global motion intensity)과 객체 내에서 분포하는 지역 움직임 값(local motion intensity)을 모두 구하게 된다. 각 프레임 별로 연산을 수행하여 전체 움직임 값과 방향의 평균 값을 구한 뒤 이전에 얻은 움직임 값에서 상쇄시킴으로써 지역 움직임 값을 얻을 수 있다^[15].

2. 깊이정보 생성

2.1 영역 분할과 움직임 정보를 이용한 깊이맵 생성

본 논문에서는 영상의 색상 정보를 이용하여 영역을 분할한다^[9]. 색상 정보는 동질의 영역에 대한 정보를 포함하고 있기 때문에 영역 분할을 위한 척도로 자주 사용된다. 영역 분할은 초기화 단계, 영역 병합, 최소 화소 이하의 영역 처리 단계로 나누어진다. 초기화 단계에서는 영상의 화소 자체가 하나의 영역으로 초기화되면 영역병합(region

merging) 과정에서 유사한 색상을 가진 영역들을 합하여 하나의 영역으로 만들게 된다. 본 논문에서는 Y, U, V 색 공간을 사용하여 식 (3)과 같이 색상분리를 수행한다.

$$\begin{aligned} C_Y(R_i, R_j) &= \{Y(R_i) - Y(R_j)\} \\ C_U(R_i, R_j) &= \{U(R_i) - U(R_j)\} \\ C_V(R_i, R_j) &= \{V(R_i) - V(R_j)\} \end{aligned} \quad (3)$$

여기서 R_i 는 i번째 영역을 의미하고, $Y(R_i)$ 는 영역 R_i 의 Y 성분의 평균값, $C_Y(R_i, R_j)$ 는 영역 R_i 와 R_j 사이의 Y 성분의 평균값 차이를 의미한다. U, V 성분에 대해서도 적용한다. 식 (3)에서 구한 Y, U, V 성분의 비용값 모두가 설정된 허용 오차보다 작다면 그 영역은 병합된다. 허용 오차 값은 잘못된 병합의 발생을 방지하기 위하여 임의의 작은 값 (0 ~ 5)으로 설정한다. 이 과정을 수행한 후 여전히 영역이 과도하게 분할된 결과를 얻을 수 있다. 과분할 영역의 결과는 차후 깊이 값을 부여하는 과정에서 오차를 발생하게 할 수 있으므로 화소수가 적은 영역들도 병합하게 된다. 즉 식 (4)를 이용하여 이웃한 영역과의 유사도를 비교하여 가장 작은 비용값을 가진 영역과 병합하게 되며, 이 과정은 모든 영역이 최소 화소수 이상이 될 때까지 반복한다.

$$C(R_i, R_j) = C_Y(R_i, R_j) + C_U(R_i, R_j) + C_V(R_i, R_j) \quad (4)$$

각 분할된 영역에서 전 단계에서 구한 움직임 정보를 식 (5)를 이용하여 깊이 정보로 변환한다.

$$D(x, y) = Motion_x + Motion_y \quad (5)$$

여기서 $D(x, y)$ 는 (x, y) 좌표에서의 깊이값을 의미한다. $Motion_x$ 는 x 방향으로의 움직임 크기이고, $Motion_y$ 는 y 방향으로의 움직임 크기이다. 깊이정보 $D(x, y)$ 는 각 영역 안에서 화소들 간 편차가 크므로, 영역 내 화소에 대하여 평균을 구해서 그 영역의 깊이값으로 정의한다^[9]. 그림 5에서 각 과정의 결과 영상을 보여주고 있다.

2.2 에지 방향성 히스토그램을 이용한 깊이맵 생성

입력 영상으로부터 에지를 추출하고, 보다 정확한 에지 방향성 히스토그램 (edge directional histogram) 데이터를 얻기 위해 에지를 세선화한 후, 에지 방향성 히스토그램 데이터를 얻는다. 이를 기준으로 초기 깊이맵을 선택하고, 물체와 물체 사이의 수직 경계정보를 이용한 깊이맵을 합성하여 최종 깊이맵을 생성한다^[10].

영상 내에서의 소실점(vanishing point)의 위치는 에지 방

향성 히스토그램 데이터를 이용하여 결정한다^[10]. 그에 따라 소실점이 영상의 왼쪽에 위치한 경우, 소실점이 영상의 위쪽에 위치한 경우, 소실점이 영상의 오른쪽에 위치한 경우, 그리고 객체가 영상의 아래쪽 중심에 위치한 경우, 이렇게 4가지의 초기 깊이맵을 생성한다. 그림 6에서 위의 네 가지 경우의 초기 깊이맵을 보였다.

에지 방향성 히스토그램 데이터를 구하기 위하여 식 (6)과 같이 Sobel 에지 마스크를 이용한다.

$$\begin{aligned} dx_{i,j} &= p_{i-1,j+1} + 2 \times p_{i,j+1} + p_{i+1,j+1} - p_{i-1,j-1} - 2 \times p_{i,j-1} - p_{i+1,j-1} \\ dy_{i,j} &= p_{i+1,j-1} + 2 \times p_{i+1,j} + p_{i+1,j+1} - p_{i-1,j-1} - 2 \times p_{i-1,j} - p_{i-1,j+1} \end{aligned} \quad (6)$$

여기서 $dx_{i,j}$ 와 $dy_{i,j}$ 는 에지의 수직, 수평방향 성분을 의미한다. 에지의 방향성은 $dx_{i,j}$ 와 $dy_{i,j}$ 를 이용하여 식 (7)과 같이 구할 수 있다.

$$Ang(D_{i,j}) = \frac{180^\circ}{\pi} \times \arctan\left(\frac{dy_{i,j}}{dx_{i,j}}\right), \quad |Ang(D_{i,j})| < 90^\circ \quad (7)$$

에지의 방향성에 따라 각 화소는 그림 7과 같이 5개의

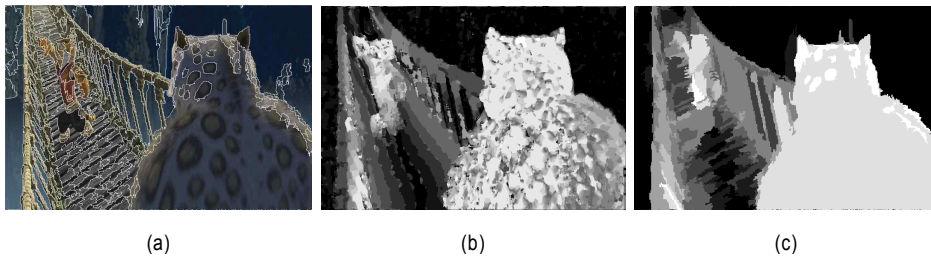


그림 5. 영역을 분할하고 깊이값을 부여하는 과정 (a) 영역 분할 결과 (b) 깊이정보로 변환 (c) 영역별 평균값 부여
Fig. 5. Process of segmentation and depth value aligning (a) segmentation (b) depth information (c) aligning mean value

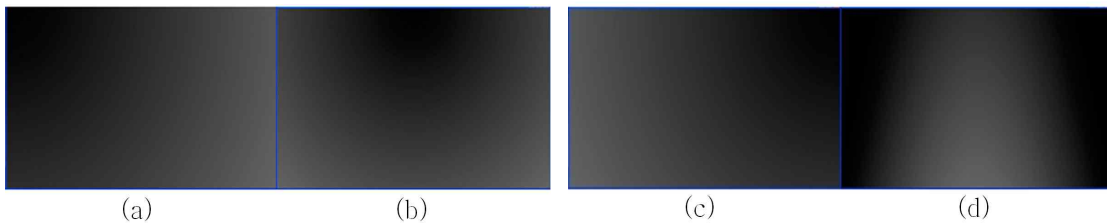


그림 6. 영상에서 소실점의 위치에 따른 영상의 초기 깊이맵 영상 - 소실점의 위치 (a) 왼쪽 (b) 위쪽 (c) 오른쪽 (d) 객체가 아래쪽에 위치한 경우
Fig. 6. Initial depth map image follow position of vanishing point - Position of vanishing point : (a) left (b) up (c) right (d) Object image

범위로 분리된다^[10].

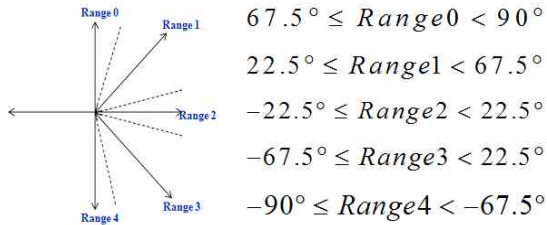


그림 7. 에지 방향성 히스토그램의 범위
Fig. 7. Range of edge direction histogram

가장 많은 화소가 축적된 범위의 방향에 따라 주어진 영상의 초기 깊이맵이 다음과 같이 결정된다^[10].

- Range0 : 소실점이 영상의 위쪽에 위치한 깊이맵
- Range1 : 소실점이 영상의 오른쪽에 위치한 깊이맵
- Range2 : 객체가 영상의 아래쪽 중심에 위치한 깊이맵
- Range3 : 소실점이 영상의 왼쪽에 위치한 깊이맵
- Range4 : 객체가 영상의 아래쪽 중심에 위치한 깊이맵

2.3 Depth layer partition

최종 깊이맵은 위에서 획득한 두 가지의 깊이맵을 병합하여 생성한다. 이 과정에서 먼저 주어진 영상을 분석하여 배경과 객체 영역을 분리한 후 각 영역에 최적의 깊이값을 부여하게 된다. 배경과 객체를 분리하기 위하여 각 화소별 연산을 수행하는데, 연속된 프레임 그룹 내에서 전후 프레임간의 화소별 색상 차이를 이용한다. 프레임 그룹내에서 모든 프레임과 비교하여 일정값 이상 차이가 나는 프레임의 수가

허용오차 이상이라면 1차적으로 배경 영역으로 판단한다.

1차적으로 객체라고 판별된 영역은 그 프레임 내에서 다시 한 번 객체 영역 확인 과정을 거치게 된다. 앞에서 획득한 움직임 정보를 이용하는데 일정 값 이상의 움직임 값을 갖는 화소만을 최종적으로 객체 영역으로 판단한다. 그림 8은 위 과정을 적용하여 객체와 배경이 분리된 예를 보여준다. 빨간색 부분이 객체로 판단된 영역이고, 파란색 부분이 배경으로 판단된 영역이다.



그림 8. 객체와 배경으로 분리된 영상
Fig. 8. Result image of separation to object and back ground

최종적으로 객체로 판단된 영역에는 영역 분할과 움직임 정보를 이용하여 얻은 깊이정보를 적용하고, 배경으로 판단된 영역에는 에지 방향성 히스토그램을 이용하여 얻은 깊이정보를 적용하여 최종 깊이맵을 생성하게 된다.

3. 3D 입체영상 생성

그림 9와 같이 좌영상은 주어진 원 영상을 이용하고 우영

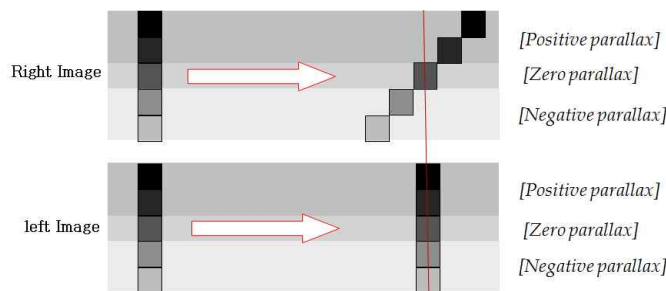


그림 9. 좌, 우영상 생성 방법
Fig. 9. Method of Left and right image generation

상은 앞에서 획득한 깊이 정보의 세기에 따라 화소별 이동 연산을 수행하여 생성한다. 최대 이동 크기는 사용자에게 의해 설정된 값을 사용한다. 영의 시차보다 적은 깊이 값을 갖는 화소는 양의 시차라고 판단하고 그 값이 적을수록 이동되는 값을 크게 하여 시차의 크기를 조절한다. 마찬가지로 영의 시차보다 큰 깊이 값을 갖는 화소는 음의 시차라고 판단하고 값이 클수록 음의 시차의 방향으로 이동을 크게 하여 시차의 크기를 조절한다. 생성된 좌, 우 영상은 인터레이스 방식으로 합성되어, 3D 입체영상으로 재현된다^[16].

IV. 실험 결과 및 분석

제안된 변환 기법의 성능을 객관적으로 비교하는 것은 어려움이 있다. 따라서 본 논문에서는 주관적 화질 평가를 수행하였다. 기존 기법으로는 대표적인 변환 기법인 MTD 기법^[2]과 H.264로 압축된 영상으로부터 복호화 과정에서 얻게 되는 움직임 정보를 이용한 기법^[4], 에지 방향성 히스토그램을 이용한 기법^[10]을 이용하였다. 주관적 화질 평가는 DSCQS 방법을 이용하였고 정상으로 생각되는 10명의 평가자를 선별하여 입체감, 화면떨림 현상, 눈의 피로도 등의 항목에 대하여 평가를 진행하였다^[17]. 720x480 크기의

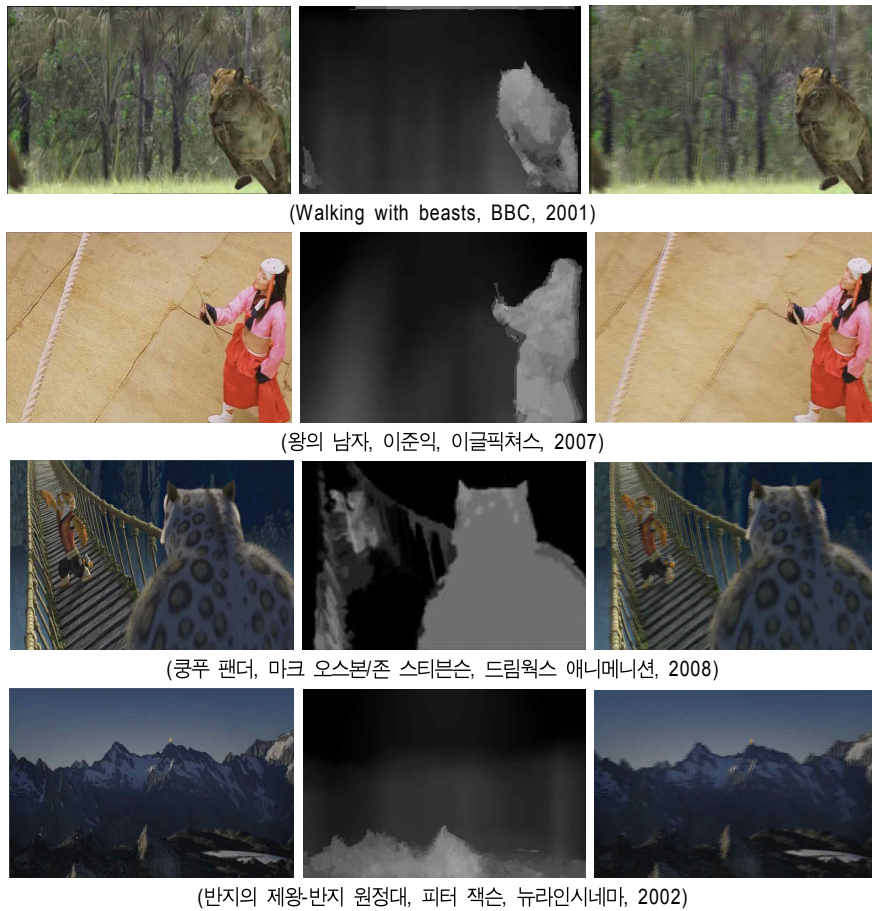


그림 10. 최종 결과 영상 (a) 원 영상 (b) 깊이맵 영상 (c) 재현된 3D 변환 영상
 Fig. 10. Final result image (a) Original image (b) Depth map image (c) Generated 3D image

4가지 실험 영상을 이용하였고, 시차를 위한 최대 이동량은 화소수 21로 설정하였다. 32° 편광 안경식 3D 입체모니터를 사용하였고 시청 거리는 1m로 설정하였다. 기존의 기법과 제안한 기법의 순서는 무작위로 설정하였고 평가 결과는 입체감이 좋을수록, 화면 떨림 현상과 눈의 피로도가 덜할수록 높은 점수를 책정하게 하였다. 1~5점까지의 점수로 평가하여 각 영상에 대한 평균값을 취하여 표 1에 보였다.

표 1에서 보듯 기존의 기법에 비해 제안한 기법의 성능이 입체감, 화면떨림 현상, 눈의 피로도 등의 모든 항목에 대하여 전반적으로 우수하다는 것을 알 수 있다. MTD 기법은 원본 영상을 그대로 사용함에 따라 발생하는 눈의 피로도와 화면 떨림 현상으로 인해 낮은 점수를 받았고, H.264로 압축된 영상으로부터 복호화 과정에서 얻게 되는 움직임 정보를 이용한 기법은 움직임 정보를 이용해 영상의 특징을 판별하고 그에 따라 각기 다른 변환 방법을 적용함으로써 안정적인 변환을 보여주고 눈의 피로도가 적다는 평가를 받았지만, 입체감과 화면 떨림 현상은 다소 떨어지는 것을 알 수 있다. 에지 방향성 히스토그램을 활용한 기법은 입체감은 준수하지만 프레임간 상관도를 고려하지 않아 화면 떨리는 현상이 발생하고 그에 기인한 눈의 피로도로 인해 낮은 평가를 받은 것으로 분석된다. 제안한 변환기법은 depth layer partition을 이용하여 2D/3D 변환 과정에서 가장 중요한 정보인 깊이 정보의 신뢰도를 향상시켜서 좋은 성능을 보인 것으로 판단된다. 결과적으로 재현된 3D 입체 영상의 품질이 향상되었고 결과 영상의 완성도를 높일 수 있었다. 그림 10은 재현된 3D 변환 영상을 보여주고 있다.

표 1. 주관적 평가 결과 (5점 만점)
Table 1. Result of DSCQS

평가 기준 \ 기법	MTD 기법	H.264 움직임정보를 이용한 기법	에지 방향성 히스토그램을 이용한 기법	제안하는 기법
입체감	2.3	2.2	2.6	2.9
화면 떨림 현상	2	2.1	1.9	3.3
눈의 피로도	1.9	2.1	1.5	2.4
합 계	2.1	2.1	2	2.7

V. 결 론

본 논문에서는 depth layer partition을 이용한 2D 동영상의 3D 변환 기법을 제안하였다. 영역을 분할하고, 움직임 정보를 이용하여 분할된 영역에 깊이 값을 부여하는 과정에서 여러 변수들에 의해 많은 오차가 발생하게 된다. 그 중 부정확한 움직임 정보로 인한 오차는 깊이맵을 제대로 생성할 수 없게 하고, 생성하더라도 신뢰도가 매우 낮은 깊이맵이므로 추후의 3D 변환 과정에서 좋은 성능을 기대하기 어렵다. 이러한 문제점을 해결하기 위해 본 논문에서는 영상의 프레임을 같은 특성을 갖는 그룹으로 분리하고 정확한 움직임 정보를 찾으려고 하였다. 제안된 기법은 움직임 정보가 존재하지 않았던 영상에서 잘못된 깊이맵이 부여되던 문제점을 해결하고, 움직임 정보의 신뢰도도 향상시켰다. 또한, 객체와 배경의 두 가지 패턴에 적합한 서로 다른 두 가지 깊이맵 생성 기법을 활용하여 깊이맵을 생성하고, depth layer partition을 통하여 분리된 객체와 배경에 최적의 깊이 값을 부여함으로써 최종적으로 신뢰도 높은 깊이맵과 결과 영상을 생성할 수 있었다.

하지만 3D 콘텐츠를 시청할 때 영상의 입체감이 늘어날수록 눈의 피로도는 증가하게 되는데, 적절한 입체감을 주면서도 눈의 피로도를 줄여주는 방법은 앞으로도 꾸준히 연구가 진행되어야 할 것이다. 또한 움직임 정보가 존재하지 않는 영상의 경우에는 배경 영역으로만 인식이 되어 제안하는 기법의 효과를 보기 어렵다. 이와 같이 움직임 정보가 존재하지 않는 영상에 대한 신뢰도 높은 깊이맵 생성 기법에 관한 연구도 꾸준히 진행되어야 할 것이다.

참 고 문 헌

- [1] Ross, J., "Stereopsis by binocular delay," Nature, vol. 248, pp. 354-364, 1974.
- [2] T. Okino and et. al, "New television with 2D/3D image conversion techniques," Proceedings of SPIE, vol. 2653, pp. 96-103, 1995.
- [3] Y. matsumoto, H. Terasaki, K. Sugimoto and T. Arakawa, "Conversion system of monocular image sequence to stereo using motion parallax," Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems, vol. 3012, pp. 108-115, May 1997.
- [4] 홍호기, 백윤기, 이승현, 김동욱, 유지상, "2D H.264 동영상의 3D 입체 변

- 환”, 한국통신학회논문지, 제31권 제12C호, pp. 1208-1215, 2006년 12월.
- [5] D. F. McAllister, Stereo computer graphics and other true 3D technologies, Princeton, NJ:Princeton University Press, 1993.
- [6] S. Battiato, S. Curti, M. LaCascia, E. Scordato, M. Tortora, "Depth-Map Generation by Image Classification," Proceedings of SPIE Electronic Imaging 2004, Three-Dimensional Image Capture and Applications VI, vol. 5302, pp 95-104, San Jose, California, USA, Jan. 2004.
- [7] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV," Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI, pp. 93-104, San Jose, CA, USA, Jan. 2004.
- [8] Man-Bae Kim, Mun-Sup Song, Do-Kyoon Kim, Kwang-Chul Choi, "Stereoscopic conversion of monoscopic video by the transformation of vertical-to-horizontal disparity," Proceedings of SPIE, vol. 3295, pp. 65-75, 1998.
- [9] 백윤기, 최미남, 박세환, 유지상, "2차원 동영상으로부터 다시점 동영상 생성 기법," 한국통신학회논문지, 제33권 제1호, pp. 53-61, 2008년 1월.
- [10] 김지홍, 유지상, "에지 방향성 히스토그램을 이용한 2차원 동영상의 3차원 입체변환기법," 한국통신학회 제 34권 제8호, pp. 782-789, 2009년 8월.
- [11] 노정진, 문영호, 유지상, "자기 상관과 에지 히스토그램을 이용한 동영상 전환점 검출," 한국통신학회, 제 29권 제 9C호, pp. 1269-1278, 2004년 9월.
- [12] 김수동, 안재우, 서영호, 김동욱, 유지상, "영역분할과 움직임 정보를 이용한 깊이맵 생성 기법," 한국방송공학회 하계학술대회, 제주도, 2010년 7월.
- [13] Y. Wang, J. Ostermann, Y. Zhang, Video processing and communication, Prentice Hall, 2002.
- [14] Y. Fisher, Fractal Image Compression: Theory & Application, Springer, 1995.
- [15] KuK-Jin Yoon, In-So Kweon, "Moving Object Segmentation with Accurate Boundary Using Color and Motion," in Workshop on Image Processing and Image Understanding, pp. 150-155, 2000.
- [16] Lenny Lipton, "Stereo-vision formats for video and computer graphics," Proceedings of SPIE, vol. 3012, pp. 239-244, 1997.
- [17] "Methodology for subjective assessment of the quality of television picture," ITU-R Recommendation BT.500-11.

저 자 소 개



김 수 동

- 2009년 2월 : 광운대학교 전자통신공학과 학사
- 2009년 3월 ~ 현재 : 광운대학교 전자공학과 석사 과정
- 주관심분야 : 3차원 영상 처리, 스테레오 매칭, 2D/3D 변환



유 지 상

- 1985년 : 서울대학교 전자공학과 학사 졸업
- 1987년 : 서울대학교 전자공학과 석사 졸업
- 1993년 : Purdue대학교 전기공학과 박사 졸업
- 1997년 ~ 현재 : 광운대학교 전자공학과 교수
- 주관심분야 : 3D 입체영상처리/압축, 웨이블릿 기반 영상처리, 비선형 디지털 신호처리