

논문 2011-3-18

# 학습기반의 객체분할과 Optical Flow를 활용한 2D 동영상의 3D 변환

## 2D to 3D Conversion Using The Machine Learning-Based Segmentation And Optical Flow

이상학\*

Sang-Hak Lee

**요약** 본 논문에서는 2D 동영상을 3D 입체영상으로 변환하기 위해서 머신러닝에 의한 학습기반의 객체분할과 객체의 optical flow를 활용하는 방법을 제안한다. 성공적인 3D 변환을 가능하게 하는 객체분할을 위해서, 객체의 칼라 및 텍스처 정보는 학습을 통해 반영하고 움직임이 있는 영역 위주로 객체분할을 수행할 수 있도록 optical flow를 도입한 새로운 에너지함수를 설계하도록 한다. 분할된 객체들에 대해 optical flow 크기에 따른 깊이맵을 추출하여 입체 영상에 필요한 좌우 영상을 합성하여 생성하도록 한다. 제안한 기법으로 인해 효과적인 객체분할과 깊이맵을 생성하여 2D 동영상에서 3D 입체동영상으로 변환됨을 실험결과들이 보여준다.

**Abstract** In this paper, we propose the algorithm using optical flow and machine learning-based segmentation for the 3D conversion of 2D video. For the segmentation allowing the successful 3D conversion, we design a new energy function, where color/texture features are included through machine learning method and the optical flow is also introduced in order to focus on the regions with the motion. The depth map are then calculated according to the optical flow of segmented regions, and left/right images for the 3D conversion are produced. Experiment on various video shows that the proposed method yields the reliable segmentation result and depth map for the 3D conversion of 2D video.

**Key Words** : 2D/3D Conversion, Optical Flow, Segmentation, Depth Map, Machine Learning

### 1. 서론

2009년 3D 영화 ‘아바타’의 열풍으로 3D 콘텐츠에 대한 관심이 확산되고 있으며, 방송, 엔터테인먼트 등 다양한 분야로 3D 미디어가 서비스 되거나 시행될 계획이다. 이처럼 대중과 미디어에 의해 형성된 큰 관심에도 불구하고 3D 콘텐츠의 공급은 원활하게 이루어지고 있지 못하다. 일반적으로 3D 콘텐츠를 제작하기 위해서는 두 가지 정도의 방법이 있는데, 첫 번째는 입체영상 콘텐츠 촬

영용 카메라 ‘리그’를 이용해 동일한 사물을 촬영하여 인간의 양안시차로 인해 느낄 수 있는 입체영상을 제작하는 것이다. 이러한 방식은 알려져 있는 것처럼 촬영 방법이 상당히 까다롭고 많은 초기 투자 및 제작비용을 필요로 한다. 따라서, 이와 같은 입체카메라를 이용한 3D 콘텐츠 제작을 통해 공급될 수 있는 콘텐츠의 수는 상당히 부족할 수밖에 없고 앞으로도 많은 시간과 노력이 필요할 것으로 보인다. 이런 이유로 인해 입체영상 시장에 대한 대중의 관심과 요구에도 불구하고 폭발적인 성장을 기대하기에는 나름의 문제를 내포하고 있다. 입체 콘텐츠 제작의 두 번째 방법은 3D 컨버팅 방법을 활용하는 것이다. 이것은 2D로 촬영한 콘텐츠를 컴퓨터비전의 기

\*정회원, SK c&c 3D 솔루션 사업팀  
접수일자 2011.5.9, 수정일자 2011.6.3  
게재확정일자 2011.6.10

술들을 활용하여 입체감을 제공할 수 있는 3D 콘텐츠로 변환하는 기법으로 상당한 정도의 영상처리 기술을 필요로 한다. 그러나, 3D 카메라로 촬영해야 하는 방법에 비해 기존의 2D 콘텐츠 제작 방법을 그대로 사용할 수 있기 때문에 제작 시간과 기술 및 비용을 감안하면 훨씬 많은 3D 콘텐츠를 제작할 수 있는 장점을 가지고 있다. 물론, 실제 3D 카메라로 제작된 콘텐츠에 비해 입체품질은 부족하지만, 현재의 3D 카메라 촬영기술도 시각적인 피로감이나 일관된 입체감을 담보할 수 없는 여러 문제점들을 가지고 있다. 더구나, 3D 컨버팅 기술을 활용하면 기존에 제작된 양질의 무수히 많은 2D 콘텐츠를 3D로 변환하여 서비스할 수 있는 중요한 효과를 창출할 수 있다. 이러한 이유로 인해 3D 변환 기법에 관한 다양한 연구들이 이루어지고 있으며, 다수의 변환된 콘텐츠들이 서비스되면서 3D 입체영상 시대의 도래를 재촉하고 있다. 이와 같은 3D 변환 기법에는 운동시차를 이용하는 방법<sup>[1]</sup>과 압축영상의 움직임 벡터를 활용하여 제작하는 방법<sup>[2]</sup>, 그리고 깊이정보를 추출하여 입체로 변환하는 방법<sup>[3]</sup> 등이 있다. 본 논문에서는 깊이정보를 포함하는 깊이맵을 생성하여 입체로 변환하는 방법을 제안할 것이다. 본 논문의 구성은 다음과 같다. 2장에서는 입체영상의 원리와 일반적인 3D 변환에 대해 간략하게 설명하고, 3장과 4장에서 머신러닝 기반의 객체분할 방법과 optical flow를 활용한 3D 변환기법에 대해 제안할 것이다. 그리고 5장과 6장에서 실험결과와 결론을 적시하면서 논문을 끝맺음 할 것이다.

## II. 3D 입체 변환

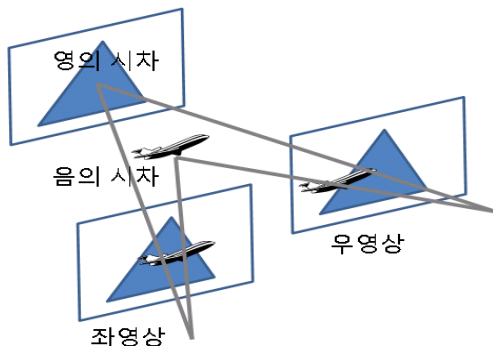


그림 1. 입체영상 원리 (Ross)  
Fig. 1. Basic principle of stereopsis (Ross)

3D 입체영상은 영의 시차, 음의 시차, 양의 시차 등 세 가지 종류의 시차를 보는 사람에게 제공하여 입체감을 느끼도록 한다. 일반적으로 입체감을 설명하기 위해 사용하는 Ross 현상<sup>[4]</sup>이 그림 1에 설명되어 있다. 그림에서 배경이 되는 산은 정지해 있고 비행기는 좌에서 우로 이동하고 있다. 이 영상을 왼쪽과 오른쪽 눈으로 보았을 때 시차가 발생하게 되며, 배경인 산과 이동하면서 앞쪽에 위치한 비행기는 다른 시차를 가지게 된다. 그림에서 보는 바와 같이 산은 영의 시차를 가지고 비행기는 화면 앞쪽에서 수렴점을 생성하는 음의 시차를 가지게 된다. 이와 같이 서로 다른 시차에 의해서 물체가 화면 앞으로 돌출되어 있거나 들어가 보이는 입체감은 입체영상을 통해 느낄 수 있게 된다.



그림 2. 2D/3D 변환방법  
Fig. 2. Workflow of 2D to 3D conversion

2D 영상을 3D 입체영상으로 변환하기 위해서, 앞서 설명한 시차의 개념을 도입하여 2D 영상에 있는 각 객체들에 서로 다른 시차 값을 부여하고 이들을 각각 좌/우 영상에 투영하여 3D 입체영상 변환의 목적을 달성한다. 일반적인 2D/3D 변환 기법의 블록다이어그램을 그림 2에 도시하였다<sup>[5]</sup>. 여기서 객체들은 3D 변환시 입체감을 느낄 수 있도록 서로 다른 시차값을 가지는데, 입체감을 표현하고 싶은 객체의 정확한 추출이 변환과정에서 중요함을 알 수 있다. 특히, 유사한 특성을 가지는 영역들을 하나의 객체로 분류하여 동일한 깊이정보를 취할 수 있도록 하는 과정이 입체 영상 생성시 잡음을 감소시킨다 (의미론적 관점에서 한 개로 보이는 객체를 여러 개의 작은 영역들로 구분하여 시차를 다르게 보이도록 하면 입체 효과 감소 등의 문제를 발생시킬 수 있다). 따라서, 본 논문은 움직임을 보이는 객체에 대해 가능한 외곽선에 해당하는 영역을 검출하여 깊이 정보를 부여하는 방법에 초점을 맞추고자 한다.

## III. 머신러닝 기반의 객체 분할

정지영상에서의 객체분할은 칼라 또는 텍스처 등의

특징들과 조건부 랜덤필드를 기반으로 구성된 에너지함수를 활용하여 상당한 정도의 발전을 실현하였다. 따라서, 본 장에서는 이러한 특징들을 영역의 특성 학습을 위해 사용하고 동영상에 적합한 특징인 optical flow 정보와 결합하여 새로운 에너지함수를 구성한 후 동영상에 대한 객체분할 방법을 설계한다.

### 1. 영역 특성의 머신러닝

하나의 영역을 다른 영역과 구별하기 위해 현저히 두드러진 특징(예, 칼라, 텍스처 등)을 활용하는 것은 당연한 인지이다. 그러나, 이와 같은 우세한 특징들은 주어진 영역에 대해 항상 다르게 선택되어야 효과적이므로, 이들을 미리 정의하여 사용하는 것은 바람직하지 않은 방법이다. 따라서, 주어진 영역을 가장 잘 묘사할 수 있는 특징들을 학습 기반의 머신러닝을 이용하여 적응적으로 선택하고 조합하도록 한다. 본 논문에서는 아다부스트<sup>[6]</sup>를 활용하여 머신러닝을 수행하도록 한다. 자세히 설명하면, 아다부스트는 각 단계별로 에러를 최소화하는 약분류기(칼라 또는 텍스처와 같은 특징들을 의미)를 선택하고, 이들을 적응적으로 조합하여 최종 강분류기를 훈련시키는 머신러닝 기법이다. 특징 벡터  $\vec{x}$ 에 대한 약분류기는 로그우도테스트 방법을 이용하여 다음과 같이 나타낸다.

$$h(\vec{x}) = \log \frac{h^+(x)}{h^-(x)} \tag{1}$$

여기서,  $h^+(\bullet)$ 과  $h^-(\bullet)$ 은 각각 양의 샘플과 음의 샘플에 대한 히스토그램을 의미하고, 최종 강분류기는 적응계수  $\alpha$ 와 약분류기를 결합하여 식 (2)와 같이 나타낸다.

$$F(\vec{x}) = \sum_{t=1}^T \alpha_t h_t(\vec{x}) \tag{2}$$

이와 같은 아다부스트 러닝 방법의 과정은 일반적인 것으로, 본 논문에서는 영역의 특성을 러닝의 대상으로 설정하고 러닝을 위한 약분류기들은 칼라, 텍스처 등의 특징들로 정의하여 사용한다. 결과적으로, 한 개의 주어진 영역에 대해 다른 영역들과 구별할 수 있는 가장 적절한 특징들을 선택하고 이들을 조합하여 학습된 최종 강분류기를 얻을 수 있다.

### 2. 에너지함수 설계

기존 영역분할에 많이 사용한 마르코프 랜덤필드(MRF)<sup>[7]</sup>는 관측 데이터들이 상호 독립이라는 가정을 이용하기 때문에 적절한 영역분할 결과를 보이지 못하는 경우가 자주 발생한다. 이를 보완한 조건부 랜덤필드(CRF)<sup>[8]</sup>는 전체 관측데이터들( $Y$ )과 라벨( $L_p$ )간의 상호 연관성을 포함할 수 있으며, 일반적으로 식 (3)과 같이 정의한다.

$$E(L, Y) = \sum_{p \in P} V_p(L_p, Y) + \sum_{p \in N_p} V_{p,q}(L_p, L_q, Y) \tag{3}$$

여기서,  $V_p$ 와  $V_{p,q}$ 는 각각 데이터 항과 pairwise 항이며,  $N_p$ 는 사이트  $p$ 의 네이버후드를 의미한다. 따라서, 함수  $V_p$ 와  $V_{p,q}$ 를 설계하는 것이 필요하다. 더불어, 본 논문에서는 동영상에서 각 객체의 움직임을 반영하는 optical flow<sup>[9]</sup>를 활용하여 위의 에너지함수를 새롭게 정의하고자 한다. 그림 3은 연속된 두 개의 프레임에서 optical flow의 예를 보여주고 있다. 그림에서 보는 바와 같이 optical flow는 동영상에서 각 객체들의 움직임을 잘 나타내고, 이러한 optical flow의 특징은 동영상에 대한 객체의 영역분할에 효과적인 정보로 사용될 수 있다. CRF 기반의 에너지함수에서 optical flow 정보를 포함하는  $V_o(\bullet)$ 를 사용하여 식 (4)에 정의하였다.

$$E(L, Y) = \sum_{p \in P} V_m(L_p, Y) + \sum_{p \in N_p} [\lambda_1 V_s(L_p, L_q, Y) + \lambda_2 V_o(L_p, L_q, Y)] \tag{4}$$

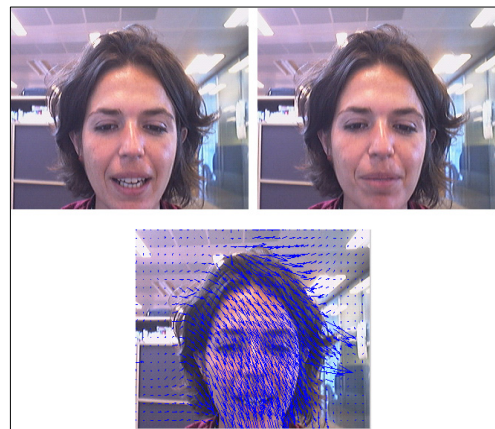


그림 3. 연속된 두 프레임과 해당하는 optical flow  
Fig. 3. Two consecutive frames and their optical flow

지금부터, 식 (4)의 각 항을 설계하고자 한다. 먼저, 사이트  $p$ 에 라벨  $L_p$ 를 할당했을 때, 머신러닝의 결과값이 클수록  $V_m$ 의 값이 작아지도록 식 (2)를 이용하여 다음과 같이 정의한다.

$$V_m(L_p, y) = \frac{e^{-F(y)}}{e^{-F(y)} + e^{F(y)}} \quad (5)$$

이것은 칼라와 텍스처 특징들을 이용하여 영상에 대해 영역분할을 수행하는 것으로, 머신러닝한 강분류기 내에 특징들의 정보가 약분류기로서 포함되어 있다. 그리고, 인접한 두 개의 사이트  $p$ 와  $q$ 에 해당하는 optical flow를 활용하여  $V_o$ 를 다음과 같이 정의한다.

$$V_o = \Phi_{p,q} \cdot \delta(L_p, L_q) \quad (6)$$

$$\Phi_{p,q} = \begin{cases} 1, & \text{if } \begin{cases} |mag(\Phi_p) - mag(\Phi_q)| < T_1 \\ |atan(\Phi_p) - atan(\Phi_q)| < T_2 \end{cases} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$\delta(L_p, L_q) = \begin{cases} 0, & L_p = L_q \\ 1, & L_p \neq L_q \end{cases} \quad (8)$$

$mag(\cdot)$ 는 optical flow의 크기이고  $atan(\cdot)$ 는 optical flow의 방향을 의미한다. 따라서, 식 (7)은 이웃한 두 개의 optical flow가 유사한 크기와 방향을 가질수록 에너지함수에서 높은 비용을 나타내도록 하는 성질을 나타내며, 에너지함수의 최소화 과정에서 인접한 두 사이트  $p$ 와  $q$ 에 동일한 라벨을 할당하는 역할을 담당하게 된다. 마지막으로,  $V_s$ 는 에너지함수의 정의에서 통상적으로 많이 사용하는 Potts 모델을 사용하여 다음과 같이 정의한다<sup>[10]</sup>.

$$V_s = \frac{1}{dist(p,q)} \exp\left(-\frac{(I_p - I_q)^2}{\sigma^2}\right) \delta(L_p, L_q) \quad (9)$$

여기서  $dist(p,q)$ 는 사이트  $p$ 와  $q$ 사이의 Euclidean 공간상 거리이며,  $\sigma$ 는 통상적인 카메라 노이즈 모델에 해당한다.

### 3. 에너지최소화를 통한 영역분할

주어진 영상을 일정한 간격으로 초기화 한 후 각 영역에 대해서 3.1절의 머신러닝과 3.2절의 에너지 최소화 과

정을 반복적으로 수행하여 최종 영역분할을 달성한다. 에너지 최소화는 그래프컷<sup>[7]</sup> 기법을 사용하였고, 반복 횟수는 10회 이내로 제한하도록 한다. 실험에서는 2~3회 이내에서 수렴하는 결과를 나타내고 있다. 그림 4는 제안한 방법을 적용했을 때, 영역분할한 결과를 보여주고 있다.

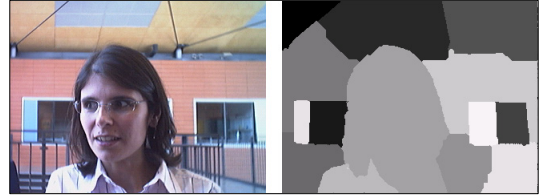


그림 4. 제안한 방법을 이용한 영역분할  
Fig. 4. Segmentation result of the proposed method

## IV. 3D 변환

본 장에서는 앞서 설명한 영역분할의 결과와 optical flow를 활용한 깊이맵 생성 기법과 이를 기반으로 3D로 변환하는 방법을 설명한다.

### 1. 영역분할과 optical flow 기반의 깊이맵 생성

영역분할의 결과에 의해 주어진 프레임의 영상은 다음과 같이 표현할 수 있다.

$$S = \{s_1, \dots, s_M\} \quad (10)$$

$$s_i \cap s_j = \emptyset \text{ for } (i, j)$$

분할된 각 영역내의 특정 화소  $(x,y)$ 에 대해서 optical flow를 x방향의  $\Phi_x$ 와 y방향의  $\Phi_y$ 로 표현할 수 있다. 분할 영역내의 모든 화소들은 대체로 비슷한 optical flow 값을 가지지만 때로는 잡음과 같은 값을 가질 수도 있고 optical flow를 구하는 과정에서 잘못된 정합으로 인해 주변과는 상이한 값을 가질 수도 있다. 따라서, 식 (11)와 같이 분할 영역내의 평균 optical flow를 검출하여 해당 영역의 전반적인 움직임을 분석할 필요가 있다.

$$\bar{\Phi}_i = \frac{1}{N_i} \sum_{(x,y) \in S_i} (|\Phi_x(x,y)| + |\Phi_y(x,y)|) \quad (11)$$

여기서  $N_i$ 는 영역  $S_i$  내의 픽셀들의 개수를 의미하고  $|\Phi|$ 는 optical flow의 크기를 나타내며,  $S_i$  영역 내

의 모든 화소들은 동일한 평균 optical flow  $\overline{\Phi}_i$ 를 이용하여 나타낸다.

이와 같은 평균 optical flow는 해당 영역이 다음 프레임에서 어느 정도 이동하였는지를 나타내고, 앞서 2장에서 설명한 바와 같이 움직임이 큰 영역일수록 인간의 시각 체계에 의해 앞으로 돌출되어 있는 것과 같은 느낌을 주게 되고, 움직임이 없거나 아주 작은 경우에는 영의 시차를 가지는 영역으로 인식되는 것이 자연스런 깊이감을 형성할 수 있다. 따라서, 평균 optical flow가 클수록 돌출되는 느낌을 가지도록 큰 음의 시차로 설정하고 작을수록 작은 음의 시차를 가지도록 하며, 가장 작은 optical flow를 가지는 영역을 배경, 즉 영의 시차를 가지도록 깊이맵을 형성해야 된다. 이를 도식화하면 식 (12)로 나타낼 수 있다.

$$D(x,y) = \overline{\Phi}(x,y) - \Phi_{\min} \quad (12)$$

그림 5는 이와 같은 방법에 의해 형성된 깊이맵을 보이고 있으며, 밝은 영역이 큰 음의 시차, 즉 보는 이에게 앞으로 돌출되는 효과를 제공하고 어두울수록 상대적으로 뒤로 위치하는 효과를 나타낸다.

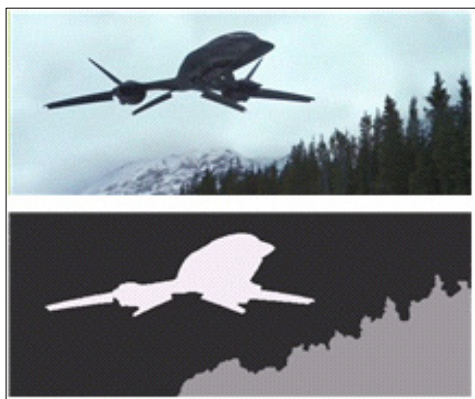


그림 5. 깊이맵 생성 (밝기로 표시)  
Fig. 5. Generation of depth map. Intensity of region shows its depth value

## 2. 입체영상 생성

앞서 2장에서 설명한 바와 같이 입체감을 제공하기 위해서 좌/우 두 개의 영상이 필요하고, 이들의 생성은 산출한 깊이맵에 의해서 이루어진다. 자세히 설명하면, 좌영상은 원래의 영상을 활용하여 제작하고, 우영상은 깊

이맵이 제공하는 값에 따라서 작은 깊이를 가지는 영역을 깊이값 만큼 먼저 이동하고 차례대로 다음 깊이값을 가지는 영역들을 이동하여 생성하도록 한다. 이와 같은 우영상 생성의 경우에 각 영역들간 깊이값의 차이로 인해 그림 6에 보는 바와 같이 천공(hole)이 필연적으로 발생한다. 천공을 메우기 위해서 인페인팅 기법중 기초적인 방법에 해당하는 주변 픽셀의 확산방법을 적용한다. 이럴 경우, 인위적으로 생성하는 우영상은 좌영상과 비교하여 에피폴라 선을 유지하도록 생성해야 하기 때문에 천공의 수평 좌측에 위치하는 픽셀의 값을 활용하여 채우도록 한다. 결과적으로, 한 개의 원영상에 대해 입체감을 실현하도록 하기 위해서 원영상과 동일한 좌영상을 생성하고 깊이맵에 따라 좌영상과 수평시차를 가지는 우영상을 생성하여 3D 영상으로 변환될 수 있도록 한다.

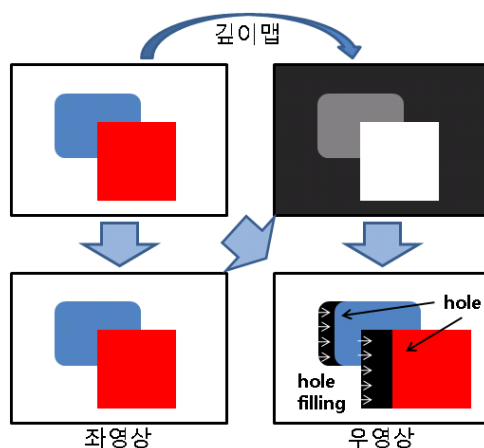


그림 6. 깊이맵을 활용한 3D 영상 생성  
Fig. 6. Generation of 3D images using depth map

## V. 실험 및 결과

3D 입체영상으로 변환하는 작업에는 동영상에 사용되기 때문에 각 프레임의 영상들을 입체영상으로 변환하기 위해서는 동영상내에서 움직임의 정보를 이용하여 프레임들을 그룹화 할 필요가 있다. 즉, 장면전환과 같은 새로운 프레임이 나타나면 이전 프레임들에서 현재 프레임으로의 전환에 대해 검출하고 추정된 영역분할 결과와 optical flow 등의 정보에 대한 신뢰도를 확보할 수 없게 되고, 이들을 이용한 입체영상 변환은 수행되지 않도록



그림 7. 제안한 방법을 활용한 2D 동영상의 3D 변환 실험결과 (왼쪽 열부터 좌영상, 우영상, 인터레이스 영상)  
 Fig. 7. Experimental results of the proposed method on the 3D conversion of 2D video (left image, right image, interlaced image)

해야 한다. 따라서, 신뢰도를 확보할 수 없는 프레임이 발생할 때 이전 프레임까지 추정하고 적용하였던 결과들을 사용하지 않고 새 프레임부터 재탐색과 검출을 시작하도록 설정하여 3D 입체동영상으로 변환해야 한다. 이를 위한 간단한 동영상내 프레임 그룹핑 과정은 다음과 같다.

Step 1. 동영상의 각 프레임에 해당하는 정지영상들을 밝기(intensity)값만을 가지는 흑백영상으로 변환한다.

Step 2. 흑백영상에 대해 히스토그램 평활화 (histogram equalization)를 수행하여 조명 변화로 인한 프레임간 평균 밝기 차이의 영향을 제거한다.

Step 3. 연속된 두 개의 프레임( $I_t, I_{t+1}$ )에 대해서 히스토그램간의 거리를 측정하는 식 (13)의  $\chi^2$  측정방법을 이용하여 프레임간의 유사도를 판단하고, 결과값이 일정 임계값 보다 작을 때는 동일한 프레임 그룹으로 설정하고 그렇지 못할 경우에는 다른 프레임 그룹으로 분류하면서 이전까지의 프레임들을 동일한 프레임 그룹으로 지정한다.

$$\chi^2(h_i, h_j) = \frac{1}{2} \sum_{n=1}^{255} \frac{(h_i - h_j)^2}{h_i + h_j} < t \quad (13)$$

새롭게 나타나는 프레임부터 다시 프레임 그룹핑 과정을 수행하도록 한다.

위와 같은 프레임 그룹핑 이후, 각 그룹 내의 영상들에 대해 본 논문에서 제안한 3D 변환 과정을 수행하도록 한다. 그림 7은 몇 가지 동영상에 대해 실험한 결과를 나타내고 있다. 그림의 첫 번째 열은 원영상과 동일하게 생성한 좌영상이고 두 번째 열은 깊이맵을 이용하여 원영상으로부터 생성한 우영상에 해당한다. 그리고, 좌/우 영상의 효과적인 비교를 위해 두 개의 영상에 대해 인터레이싱 방법을 활용하여 세 번째 열에 도시하였다. 보는 바와 같이 돌출되어 보이도록 영역별로 시차가 발생함을 알 수 있고, 이로 인해 사용자들에게 입체감을 전해줄 수 있게 된다. 제안하는 방법은 객체분할시 optical flow의 정보를 에너지함수에 활용하기 때문에 실제 움직이는 영역에 대해서 집중할 수 있고 이는 기존에 많이 사용하는 영역분할의 단점인 다수의 작은 영역 생성(오버세그멘테이션)

선) 문제를 해결한다. 그리고, 움직이는 영역 전체에 대해 동일한 시차를 발생시키기 때문에 입체영상 생성시 작은 영역에 의한 빈번한 깊이 정보의 변화가 발생하지 않고 이것으로 인한 시각적인 피로감을 감소시키면서 입체감을 전달할 수 있다.

## VI. 결론

본 논문에서는 머신러닝 기반의 영역분할과 optical flow를 활용하여 2D 동영상을 3D 입체영상으로 변환하는 방법을 제안하였다. 영역분할을 위해 조건부 랜덤필드를 활용한 새로운 에너지함수를 제안하였고, 분할 영역이 객체의 내부요소인 칼라, 텍스처 특징 뿐 아니라 프레임간 움직임 정보인 optical flow를 채용하여 지나치게 작은 영역으로 분할되는 문제를 경감시켰으며, 움직임이 존재하는 객체 위주로 영역분할을 달성할 수 있었다. 그리고, 분할된 영역에 대해서 optical flow 기반의 깊이맵 생성방법을 제안하였으며, 이후 생성된 깊이맵을 활용하여 입체영상 구현에 필요한 좌,우 영상을 생성할 수 있었고, 제안한 방법에 의해 생성된 입체영상은 움직임이 존재하는 단위 객체별로 다른 시차를 가지고, 움직임이 큰 영역에 상대적으로 큰 시차를 할당하여 입체감을 충분히 제공할 수 있었다. 더불어, 객체를 한 개의 영역으로 분할하였기 때문에 여러 개의 작은 영역으로 분할하여 시차를 발생시킨 것 보다 좀더 사용자 친화적인 입체영상을 생성할 수 있었다. 여러 개의 2D 동영상에 대한 실험에서 제안한 방법이 신뢰성 있는 객체분할을 수행하고 깊이맵을 생성하여 3D 동영상으로 변환할 수 있음을 보였다.

## 참고 문헌

[1] Y. Matsumoto, H. Terasaki, K. Sugimoto, and T. Arakawa, "Conversion system of monocular image sequence to stereo using motion parallax," Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems vol. 3012, pp.108-115, 1997.

[2] 홍호기, 백윤기, 이승현, 김동욱, 유지상, "2D H.264 동영상의 3D 입체변환," 한국통신학회, 제31권, 제12C호, 1208-1215쪽, 2005년 12월.

[3] S. Battiato, S. Curti, M. LaCascia, E. Scordato, and M. Tortora, "Depth-Map generation by image classification," Proc. of SPIE Electronic Imaging2004, Three-Dimensional Image Capture and Application VI, vol. 5302, pp.95-104, 2004.

[4] Ross, J., "Stereopsis by binocular delay," Nature, vol. 248, pp.354-364, 1974.

[5] 이요섭, "2D-3D 변환 기술의 동향 및 전망," 전자공학회, 제38권, 제2호, 37-43쪽, 2011년 2월.

[6] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistica view of boosting," The Annual of Statistics, vol. 28, pp.377-386, 2000.

[7] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," IEEE Trans. PAMI, vol. 30, no. 6, pp.1068-1080, 2008.

[8] J. Lafferty, A. McCallumm, and F. Pereira, "Conditional random fields: Probabilistic models and for segmenting and labeling sequence data," Proc. of ICML, 2001.

[9] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l1 optical flow," Proc. of the 29th DAGM Conference on Pattern Recognition, 2007.

[10] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans. PAMI, vol. 23, no. 11, pp.1222-1239, 2001.

## 저자 소개

### 이 상 학(정회원)



- 1996년 연세대학교 학사
- 1998년 연세대학교 석사
- 2004년 삼성전자 디지털미디어
- 2011년 서울대학교 박사
- 현재 SK c&c 3D 솔루션 사업팀