일반논문-08-13-3-01

단일 시점 축구 비디오의 3차원 영상 변환을 위한 깊이지도 생성 방법

고재승^{a)}, 김영우^{a)}, 정용주^{b)}, 김창익^{a)†}

2D-to-3D Stereoscopic conversion: Depth estimation in monoscopic soccer videos

Jaeseung Ko^{a)}, Youngwoo Kim^{a)}, Youngju Jung^{b)}, and Changick Kim^{a)}

요 약

본 논문에서는 일반 단일 시점의 축구 비디오를 스테레오스코픽 영상으로 변환하는 방법을 제안한다. 축구 비디오 분석 과정을 통하여 축구 비디오를 일정한 종류의 샷으로 분류하고, 분류된 샷 종류에 따른 깊이지도 생성 방법을 제안한다. 원거리 샷의 경우에는 운동장 영역 추출을 통하여 운동장 영역에 깊이지도 (*Depth Map*)를 생성하는 방법을 제안한다. 그리고 비 원거리 샷의 경우, 운동장 영역 블록 수와, 간단한 피부색 발견 알고리즘을 통해 생성한 스킨 블록의 수에 따라 다시 3가지로 샷을 분류하고, 각 종류의 샷에 따른 깊이지도 생성 방식 1) 오브젝트 영역 추출을 통한 깊이지도 생성, 2) 스킨 블록을 이용한 전경 영역 추출과 가우시안 함수를 이용한 깊이지도 생성, 그리고 3) 스킨블록이 없는 상황에서의 깊이지도 생성 방법을 제안한다. 제안한 방법을 통하여 생성한 깊이 지도를 이용하여, 스테레오스코픽 영상을 생성하는 방법을 소개하고, 생성한 실험 영상을 결과로 제공한다. 그리고 주관적 깊이감 품질 평가를 통해서, 제안된 방법을 통해 생성된 영상이 원본 영상에 비해 깊이감이 향상됨을 증명한다.

Abstract

This paper proposes a novel method to convert monoscopic soccer videos to stereoscopic videos. Through the soccer video analysis process, we detect shot boundaries and classify soccer frames into long shot or non-long shot. In the long shot case, the depth maps are generated relying on the size of the extracted ground region. For the non-long shot case, the shot is further partitioned into three types by considering the number of ground blocks and skin blocks which is obtained by a simple skin-color detection method. Then three different depth assignment methods are applied to each non-long shot types: 1) Depth estimation by object region extraction, 2) Foreground estimation by using the skin block and depth value computation by Gaussian function, and 3) the depth map generation for shots not containing the skin blocks. This depth assignment is followed by stereoscopic image generation. Subjective evaluation comparing generated depth maps and corresponding stereoscopic images indicate that the proposed algorithm can yield the sense of depth from a single view images.

Keyword: 2D-to-3D conversion, soccer video analysis, 3DTV, depth map

1. 서 론

a) 한국정보통신대학교, 시각정보처리연구실*

Visual Information Processing Lab., Information and Communications University*

b) 삼성종합기술원**

Samsung Advanced Institute of Technology**

† 교신저자 : 김창익(ckim@icu.ac.kr)

시청자에게 실재감과 입체감을 느끼도록 하기 위한 실감 방송 기술의 하나인 3차원 영상 기술이 멀티미디어 와 방송 기술의 발전과 함께 주목 받고 있다. 3차원 디스플레이 장

치의 발전과 함께 그에 따른 3차원 영상에 대한 수요가 꾸준히 증가하고 있다. 3차원 콘텐츠, 즉 스테레오 영상의 생성 방법은 2대 이상의 카메라를 사용하여 촬영을 통해서 좌우 영상을 생성하고 스테레오 영상을 생성하는 기술과 단일 시점의 영상으로부터 좌우 영상을 생성하여 복잡한 영상 처리와 해석 과정을 통해 좌, 우 시점 영상을 생성하여 스테레오 영상을 생성하는 스테레오 변환(2D-To-3D Conversion) 기술이 존재한다. 직접 스테레오 영상을 촬영하는 방법에 비해, 스테레오 변환 기술은 과거에 제작된 수 많은 단일 시점 영상을 차세대 디스플레이 장치에 적합한 3차원 콘텐츠로 활용 할 수 있다는 점이 장점이 될 수 있다.

그림 1. 일반적인 스테레오 변환 전체 과정

Fig 1. General structure of stereoscopic conversion

제안된 알고리즘에서 스테레오 변환은 크게 영상분석, 깊이 지도(depth map) 생성, 깊이 지도를 이용한 시차(Parallax) 생성, 좌우 영상 생성의 4단계의 과정을 통해 수행된다 (그림 1). 2차원 영상에서 영상 내의 3차원 정보를 추출하기 위한 영상 분석 단계를 거치고, 영상 분석을 통해 얻어진 단서들을 통하여 깊이 지도를 생성하는 단계를 수행하게 된다. 그리고 생성된 깊이 지도를 활용하여 영상 내의 모든 영역의 시차를 구할 수 있으며, 이를 시차를 이용하여 입력 영상을 좌, 우로 이동시킴으로써 새로운 시점의 좌, 우 영상을 생성할 수 있다.

스테레오 변환의 전체 과정에서 가장 중요하고 어려운 과정이 2차원 영상 내의 단서들을 활용하여 3차원 정보를 추출하고, 깊이지도 (depth map)를 생성하는 과정이다. 정지 영상의 경우, [1]에서는 영상 분류 과정(Image Classification)과 허프 변환을 이용한 영상의 소실점(vanishing point), 소실선(Vanishing line)을 예측하여, 이를 단서로 활

용하여 깊이 지도를 예측하였다. Valencia는 블록 단위의 웨이블릿 변환(Wavelet transform)을 사용한 영상 내의 공간 주파수 성분의 분석을 통해 초점 정보를 추출하여 전경과 배경을 분리하여 깊이 정보를 예측하는 방법을 제시하였다^[2]. 그리고 Tam 등은 [3]에서 Sobel Mask를 적용하여 추출한 Edge Map에 Blurring 과정을 수행하여 이를 깊이 지도로 활용하였다. 그리고 우리의 이전의 연구 [4]에서는 영상 분할 (Image Segmentation)을 통한 영역 정보와 분할된 각 영역에 대하여 고차 모멘트(Higher-order moment)를 구하여 공간 주파수 성분을 분석하여 깊이 정보를 예측하였다. 동영상의 경우, B.J. Garcia 등은 공간-시간적 보간법을 통한 스테레오 변환 기법^[5]을 제안 하였으며, [6]에서는 모션 벡터(Motion Vector)를 이용하여, 영상 분할 과정을 수행하고, 다시 모션 정보와 Overlap에 의한 occlusion 영역 발견을 통해서 깊이 지도를 생성하는 방법을 제안하였다. Okino 는 Modified Time Difference [7]를 이용하여, 스테레오스코픽 영상을 생성하는 방법을 제안하였다. Park 등은 영상의 모션 정보를 통하여 준-인, 준-아웃 영상을 찾고, 준 영상에 대한 깊이지도 생성과 스테레오스코픽 영상 생성 방법을 제안하였다^[8].

이전에 제안된 일반 동영상의 스테레오 변환 기법들의 경우, 대부분이 모션 정보를 활용하게 된다. 하지만 축구 비디오의 경우, 샷의 변환이 빠르고 영상 내의 다수의 작은 물체(예, 선수, 공)가 빠르게 움직이는 경우가 많으며, 원거리 샷의 경우 축구 경기장이 영상의 대부분을 차지하여, 영상 내의 공간적 변화 정도가 작기 때문에, 스테레오 변환에 활용할 수 있는 정확한 모션 정보를 추출하기가 어렵다. 따라서, 우리는 축구 비디오의 분석을 통해서 일정한 수로 샷의 종류를 분류하고, 샷의 종류에 따른 깊이지도 생성 방법을 제안한다. 그림 2는 제안한 알고리즘의 전체 처리과정을 나타내는 다이어그램이며, 크게 축구장의 축구 운동장 분석 과정, 샷 분류 과정, 깊이지도 생성 과정으로 나누어질 수 있다.

본 논문의 구성은 다음과 같다. II 장에서는 축구 운동장 분석과 샷 종류 판별에 대해 다루고, III 장에서는 분류된 각 샷에 따른 깊이지도 생성 방법을 제안한다. IV장에서는 생성한 깊이 지도를 이용하여 좌, 우 영상을 생성하는 방법을 소개하고, V 장에서는 제안한 방법을 통해 생성한

그림 2. 축구 비디오의 깊이지도 생성 과정

Fig 2. Overall procedure of depth map generation for soccer videos

깊이지도 영상과 생성된 스테레오스코픽 영상을 실험 결과로 제공하고, 생성한 결과 영상에 대한 주관적 깊이감 품질 조사를 결과를 제공한다. 마지막으로 VI장에서 결론을 맺는다.

II. 축구 운동장 분석 및 추출과 샷 분류 처리 과정

축구 비디오의 분석을 위하여 축구 운동장 영역을 추출하는 운동장 추출 과정과 샷 분류 처리는 필수적인 과정이다. 우리는 이전의 연구^[9]에서 운동장 영역을 추출하고, 이 운동장 발견 알고리즘을 바탕으로 샷 분류 과정을 수행하였다. 우리는 [9]에서 제시한 방법을 통하여, 현재 샷 종류의 판별과 함께, 운동장 영역을 추출하여 축구 운동장 이진화 지도 G_{binary} 는 다음과 같이 생성한다. (그림 3-(b))

$$G_{binary}(x, y) = \begin{cases} 1, & \text{if } (x, y) \text{ is ground pixel} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

여기서 (x, y) 는 $0 \leq x < \text{framewidth}$, $0 \leq y < \text{frameheight}$ 의 범위를 갖는다.

또한 운동장 영역 내에 존재하는 선수들과 구획선 등을 포함하는 운동장 영역을 정의하기 위하여, 영상을 16×16 크기의 블록 지도 B_{ij} 로 나누고, 다음 수식을 통하여 운동장 영역 블록 지도 $GB(i, j)$ 를 (2)와 같이 생성한다. (그림 3-(c))

$$GB(i, j) = \begin{cases} 1, & \text{if } \frac{\sum_{(x, y) \in B_{ij}} G_{binary}(x, y)}{16 \times 16} \geq 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

여기서 (i, j) 는 $0 \leq i < \text{framewidth}/16$, $0 \leq j < \text{frameheight}/16$ 의 범위를 갖는다.

그림 3. (a) 입력 영상 (b)추출된 축구 운동장 영역, (c) 초기 운동장 블록 지도 (d)개선된 운동장 블록 지도

Fig 3. (a)Input Video (b) Extracted ground region (c) Initial ground block map (d) Refined ground block map

그림 4. (a) 추출된 운동장 이진화 지도 G_{binary} (b) 개선된 운동장 이진화 지도 $G_{refined}$

Fig 4. (a) Extracted ground binary map G_{binary} (b) Refined ground binary map $G_{refined}$

$framewidth/16$ 의 범위를 갖는다. 우리는 샷 경계 검출 (shot boundary detection)과 샷 분류(shot classification), 깊이지도(depth map) 생성을 위하여 관중석 영역과 축구 운동장 영역을 분리하였다. 앞에서 구한 $GB(i,j)$ 의 주변 블록과의 연계성을 측정하여, 운동장 내에 생성된 작은 비 운동장 블록을 매워 보다 개선된 운동장 블록 지도 $GB_{refined}$ 를 얻을 수 있었다. 그림 3-(d)는 개선된 운동장 블록 지도 $GB_{refined}$ 를 나타낸다. 자세한 샷 경계 검출과정과 tit 분류 과정에 대한 설명은 [9]에서 참조할 수 있으며, [9]의 방법을 통하여, 축구 비디오를 원거리 샷(Long shot)과 비원거리 샷(Non-long shot)으로 분류할 수 있다.

$$G_{refined}(x,y) = \begin{cases} G_{binary}(x,y), & \text{if } G_{binary}(x,y)=1 \text{ and} \\ & GB_{refined}(x/16,y/16)=1 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

우리는 픽셀 단위의 축구 운동장 영역과 관중석 영역 분리를 위하여, 수식 (3)을 통하여 개선된 운동장 이진화 지도 $G_{refined}$ 를 추출하였다.

그림 4(a)는 처음 생성한 운동장 이진화 지도 G_{binary} 이며, 그림 4(b)는 관중석 영역을 정확히 분리하여 얻은 개선된 운동장 이진화 지도 $G_{refined}$ 이다. 그림 4에서 확인 할 수 있듯이 영상 상부의 관중석 영역의 하얀 점을 보이는 운동장 발견 오류들이 제거 된 것을 확인할 수 있다. 이것은 운동장 블록 지도를 이용하여 작은 크기의 노이즈 픽셀들을 제거 할 수 있기 때문이다.

III. 각 분류된 샷에 대한 깊이 지도 생성 방법

우리는 수 많은 축구 비디오의 관찰을 통해서, 일반적인 축구 비디오의 경우, 영상의 아래 영역은 촬영하는 카메라와 가까운 영역이고, 영상의 위 영역은 카메라와 멀리 떨어진 영역임을 알 수 있었다. 따라서 우리는 영상의 아래 영역이 시청자에게 보다 가까운 영역이라는 기본 가정으로 하여 깊이 지도를 생성하였으며, 더욱 효과적인 깊이 지도의 생성을 위하여 II장에서 분류된 샷의 종류를 기준으로, 각 샷에 따른 깊이지도 생성 방법을 다음과 같이 제안한다.

1. 원거리 샷에서의 깊이지도 생성 방법

원거리 샷에서는 우리가 대부분의 경우가 우리가 앞에서 설정한 가정, 영상의 아래 영역이 시청자와 가까운 영역이라는 점에 해당된다. 따라서 우리는 앞에서 얻은 개선된 운동장 이진화 지도 $G_{refined}$ 에 대하여 깊이지도 (Depth map)

그림 5. 원거리 샷 영상과 해당 영상의 깊이지도

Fig 5. Long shot frames and corresponding depth maps

를 생성하였다. 깊이 지도를 생성하기 위하여, 우리는 다음과 같이 개선된 운동장에서 추출한 운동장 픽셀들 중에서 운동장 영역의 최상위 지점과 최하위 지점의 세로축 좌표를 각각 grd_upper_y 와 grd_lower_y 라 정의한다. 즉,

$$grd_upper_y = \min(y), \text{ where } (x, y) \in G_{refined} \quad (x, y) = 1 \quad (4)$$

그리고,

$$grd_lower_y = \max(y), \text{ where } (x, y) \in G_{refined} \quad (x, y) = 1 \quad (5)$$

또, 운동장 영역의 시작 지점에서 운동장 영역의 종료 지점 사이의 거리, 즉 운동장 영역의 세로 길이는 다음과 같이 정의한다. (그림 4(b) 참고)

$$grd_height = grd_lower_y - grd_upper_y \quad (6)$$

우리는 앞에서 정의한 변수들을 이용하여 다음과 같이 영상의 각 픽셀의 깊이 값을 결정하였다.

$$depth_step = \frac{Max_Depth}{grd_height} \quad (7)$$

$$Depth(x, y) = \begin{cases} \min(255, (y - grd_upper_y) \times depth_step), \\ \text{if } G_{refined}(x, y) = 1 \text{ AND } y \geq grd_upper_y \\ \min(255, (y - grd_upper_y) \times depth_step + \alpha), \\ \text{if } G_{refined}(x, y) = 0 \text{ AND } y \geq grd_upper_y \\ 0, \text{ otherwise} \end{cases} \quad (8)$$

여기서 $depth_step$ 은 단계별로 증가되는 깊이 값의 단위 증가 크기 값을 나타내며, Max_depth 는 운동장 픽셀이 갖는 최대 깊이 값을 나타낸다. (본 논문에서는 200로 설정하였다.) 그리고 $Depth(x, y)$ 는 (x, y) 에서의 할당된 깊이 값을 나타낸다. 추출된 운동장 위의 오브젝트의 경우,

선수, 심판, 공인 경우이므로, 보다 사용자에게 입체감을 느낄 수 있도록 하기 위하여 주변 운동장 영역의 픽셀보다 α 값만큼의 높은 깊이 값을 갖도록 하였다. 그림 5는 입력된 축구비디오의 원거리 샷과 해당 영상의 생성된 깊이 지도이다.

2. 비 원거리 샷 분류와 깊이지도 생성 방법

우리는 앞의 축구 비디오 분석 과정을 통하여 원거리 샷과 비 원거리 샷을 분류할 수 있었다. 비 원거리 샷의 깊이 지도를 생성하기 위하여 분류된 비 원거리 샷을 다시 표 1과 같이 3가지 종류로 분류 하였으며, 그림 6은 분류된 3가지 종류의 비 원거리 샷의 예제를 나타낸다.

(a)

(b)

(c)

그림 6. 비 원거리 프레임의 스크린 샷 (a) NL1, (b) NL2, (c) NL3

Fig 6. Screen shots of non-long shot frames (a) NL 1 (b) NL2 (c) NL3

표 1. 비 원거리 샷의 분류

Table 1. Three types of non-long shot frames

Types	비 원거리 샷 분류 설명
NL1	오브젝트가 운동장 영역으로 둘러싸인 경우
NL2	오브젝트가 운동장 영역에 둘러싸이지 않으며, 스킨 블록이 존재하는 경우
NL3	오브젝트가 운동장 영역에 둘러싸이지 않으며, 스킨 블록이 존재하지 않는 경우

비 원거리 샷으로 분류된 프레임을 $frame_{NL}(k)$ 라 정의

할 때, 우리는 다음과 같이 프레임 내의 전체 운동장 블록 수 N_{GB_Total} 와 프레임의 상위 1/2 영역의 운동장 블록 개수 N_{GB_Upper} 를 비교하여 Type 1 과 Type 2, 3을 분류하였다.

$$Type(frame_{NL}(k)) = \begin{cases} NL1, & \text{if } N_{GB_Total} > NB_{Total} / 2 \text{ and,} \\ & N_{GB_Upper} > NB_{Total} / 3 \\ NL2 \text{ or } NL3, & \text{otherwise} \end{cases} \quad (9)$$

여기서 NB_{Total} 은 블록 지도의 전체 블록 수를 나타낸다. 우리는 실험을 통하여, 전체 영상에 대한 운동장 블록의 수가 전체 블록의 1/2이상이며, 영상의 프레임의 상위 1/2 영역의 운동장 블록수가 블록 지도의 전체 블록 수의 1/3 이상인 샷들을 *NL1 Type*의 샷으로 결정하였다. *NL1 Type*으로 분류된 비 원거리 샷은 운동장 발견 알고리즘을 활용하여 전경 영역을 추출할 수 있으므로, 비교적 간단히 깊이 지도를 생성할 수 있다. 깊이지도 생성 방법은 2.1에서 제시한다.

Type NL2 와 *NL3* 의 비 원거리 샷의 경우, 깊이지도 생성을 위하여 전경 영역과 배경 영역을 정확히 추출하여 분리해 내는 것은 어렵기 때문에 우리는 영상 내의 특징을 추출하여 대략적인 전경 위치를 결정하고, 결정된 전경 영역을 중심으로 깊이 지도를 생성하였다. 대략적인 전경 영역을 추출하여, 그 영역을 중심으로 깊이 지도를 생성하였다. 영상 내의 전경 영역을 찾기 위하여 우리는 관찰을 통해, 일반적으로 축구 비디오에서의 근거리 샷은 선수나 감독 등의 인물을 중심으로 구성되어 있음을 알 수 있었다. 따라서, 우리는 운동장이 아닌 영역에 대하여 피부색 발견 알고리즘을 적용하여 인물의 대략적인 위치를 파악하고, 그 위치를 기준으로 한 영상의 대략적인 전경영역을 예측하는 방법을 제안하고자 한다.

먼저, 우리는 효과적인 피부색 발견 알고리즘 적용을 위하여 HSI 컬러 공간^[10]을 사용하였으며, 적용된 피부색 발견 알고리즘은 다음과 같다.

$$skin(x, y) = \begin{cases} 1, & \text{if } 20 < H(x, y) < 60 \text{ and,} \\ & 0 < I(x, y) < 200 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

여기서 $H(x, y)$ 는 (x, y) 에서의 색상(Hue) 값을 나타내며, $I(x, y)$ 는 (x, y) 에서의 명도(Intensity) 값을 나타낸다. 우리는 이 피부색 발견 알고리즘을 활용하여 영상을 16X16 크기의 블록 맵 B_{ij} 로 나누었을 때, 스킨 블록 지도 $SB(i, j)$ 는 (11)과 같이 생성할 수 있다.

$$SB(i, j) = \begin{cases} 1, & \text{if } \frac{\sum_{(x,y) \in B_{ij}} Skin(x, y)}{16 \times 16} \geq 2/3 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

여기서 (i, j) 는 각각 $0 \leq i < framewidth/16$, $0 \leq j < framewidth/16$ 의 범위를 갖는다. 그리고 우리는 생성된 스킨 블록 지도를 주변 블록과의 관계성을 이용하여 노이즈로 판단되는 블록을 제거하고, 영상 내의 스킨 블록 중에 가장 위에 위치한 블록을 영상 내의 선수 등의 얼굴이 위치하는 영역으로 가정하여, 그림 7과 같이 최종 스킨 블록 지도 $SB_{refined}$ 를 얻을 수 있었다.

우리는 최종 스킨 블록 지도 $SB_{refined}$ 의 블록의 개수, N_{SB} 를 비교하여 비 원거리 샷을 다음과 같이 다시 분류하였다. k 번째 비 원거리 프레임을 $frame_{NL}(k)$ 라 할 때,

그림 7. 입력 영상(첫 번째 행)과 해당 영상의 스킨 블록 지도(두 번째 행)
Fig 7. Input frames(First row) and corresponding skin block map (second row)

$$Type(frame_{NL}(k)) = \begin{cases} NL2, & \text{if } N_{SB} \neq 0 \\ NL3, & \text{if } N_{SB} = 0 \end{cases} \quad (12)$$

즉, 영상 내의 스킨 블록이 한 개라도 존재하는 경우, 해

당 비원거리 샷을 *NL2 Type*으로 분류하였고, 스킨 블록이 존재하지 않는 경우, *NL3 Type*으로 분류하였다.

2.1 오브젝트가 운동장으로 둘러싸인 비원거리 샷 (*NL1*)

그림 8. *NL1 Type*의 비 원거리 샷의 깊이 지도 생성
Fig 8. Depth maps for *NL1 type* non-long shot frames

*NL1 Type*의 비 원거리 샷의 경우, 축구 운동장 발견 알고리즘을 통하여 운동장 영역과 오브젝트 영역을 간단히 분리해 낼 수 있으므로, 원거리 샷의 경우와 유사한 방법으로 깊이 지도를 생성할 수 있다. 우리는 축구 운동장 발견 알고리즘을 통하여 추출된 운동장의 경우, 배경 영역으로, 그리고 나머지 오브젝트 영역을 전경 영역으로 설정하였다. 그리고 원거리 샷과 비슷한 방법을 통하여, 다음과 같이 깊이 지도를 생성할 수 있다.

$$depth_step = \frac{Max_Depth}{obj_height} \quad (13)$$

그리고,

$$Depth(x, y) = \begin{cases} y \times depth_step, & \text{if } G_{refined}(x, y) = 0 \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

원거리 샷과 같이 여기서 $depth_step$ 은 단계별로 증가되는 깊이 값의 단위 증가 크기를 나타내며, Max_Depth 는 깊이 지도가 갖는 최대 깊이 값을 나타낸다. 그리고 $Depth(x, y)$ 는 (x, y) 에서의 깊이 값을 나타낸다. 그림 8은 *NL1 Type*의 비 원거리 샷에 대하여 생성된 깊이 지도를 나

타낸다.

2.2 스킨 블록이 존재하는 비 원거리 샷의 깊이지도 생성 방법 (*NL2*)

그림 9. *NL2 Type*의 비 원거리 샷의 전경 영역 설정 과정 (a) *NL2 Type* 비원거리 샷, (b) 최종 스킨 블록, (c) 전경 영역 확장, (d) 예측된 전경 영역.
Fig 9. Foreground determination for *NL2 type* shots (a) *NL2 Type* non-long shot (b) Skin block map (c) Foreground extention (d) Estimated foreground region

영상 내에 스킨 블록이 존재하는 경우(*NL2*), 찾은 스킨 블록들 중에서 가장 높은 위치에 있는 블록을 영상내의 사람의 얼굴의 시작 위치로 가정하였다. 그리고 대략적인 전경 영역을 설정하기 위하여 얼굴의 위치로 가정한 최종 스킨 블록의 한 칸 위 블록의 왼쪽, 오른쪽 끝 블록의 위치를 시작으로 하여, 다음 행의 블록에 대해서는 좌우로 한 블록씩 증가시키는 방법으로 그림 9-(c)와 같이 피라미드 형태로 블록 영역을 확장하였다. 그리고 그림 9-(d)와 같이 좌우로 피라미드 블록 영역을 포함하는 두 개의 선을 그려서 두 선의 안쪽 영역을 대략적인 전경 영역으로 설정하였다.

예측된 전경 영역을 기준으로 깊이 값을 할당하기 위하여 다음의 과정을 수행한다. 영상의 아래 영역이 시청자와 가까운 영역이라는 가정을 설정 하에, 스테레오스코픽 영

그림 10. 가우시안 분포 확률을 구하기 위한 인자값 설정
Fig. 10. Parameter setting for gaussian function

상 생성 시, 깊이 값의 불연속성에 의한 영상의 왜곡을 최소화하기 위한 깊이 값 할당 방법을 다음과 같이 제시한다. 우리는 예측된 전경 영역을 기준으로 깊이 지도를 생성하기 위하여, 영상의 각 행의 깊이 값을 가우시안 분포 형태로 모델링하여 깊이 값을 할당하였다. 각 행의 깊이 값을 결정하기 위한 첫 번째 단계로, 영상의 각 행 위의 각 픽셀에 대한 확률을 다음과 같이 계산하였다.

$$P_y(x) = \frac{1}{\sqrt{2\sigma_y^2}} \exp\left(-\frac{(x-m_y)^2}{2\sigma_y^2}\right) \quad (15)$$

여기서 $P_y(x)$ 는 y 번째 행의 x 번째 픽셀의 확률 값을 나타내며, $0 < x < \text{framewidth}$, $0 < y < \text{frameheight}$ 의 범위를 갖는다. 그리고 (15)에서 각 행의 σ_y 값은 앞의 전경 영역을 설정하기 위하여 그린 좌우 두 라인 위의 두 점 사이의 거리로, m_y 값은 그 두 점 사이의 중심의 x 값으로 설정하였다. (그림 10 참조)

앞에서 계산한 확률 $P_y(x)$ 을 이용하여, y 번째 행의 깊이 값 $\text{Depth}_y(x)$ 은 다음과 같은 수식을 통해서 구할 수 있다.

$$\text{Depth}_y(x) = y \times \omega_{\text{incre}} \times \frac{P_y(x)}{\max_x \{P_y(x)\}} \quad (16)$$

여기서 x 는 y 번째 행에 대한 $0 < x < \text{framewidth}$ 범위

의 값이며, w_{incre} 는 행에 따른 깊이 값 증가량을 조절하기 위한 가중치이다. 그리고 $\text{MAX}_x(P_y(x))$ 는 y 번째 행의 확률 $P_y(x)$ 중에서 가장 높은 확률 값을 나타낸다. 그림 11은 입력 영상과 예측된 전경 영역 그리고 가우시안 분포 모델링을 통해 얻어진 깊이 지도이다.

(a)

(b)

(c)

그림 11. (a) 입력 영상, (b) 추출된 전경영역 (c) 깊이지도
Fig. 11 (a) Input frame (b) Estimated foreground (c) Depth map

2.3 스킨 블록이 존재하지 않는 비 원거리 샷의 깊이지도 생성 방법 (NL3)

(a)

(b)

그림 12. (a) NL3 Type의 비 원거리 샷과 (b)해당 영상의 깊이지도
Fig. 12 (a) NL3 Type non-long shots (b) Depth map for NL3 type non-long shots

영상 내에 스킨 블록이 존재하지 않는 경우(NL3) 영상 (그림 12-(a))내에서 정확히 전경 영역을 예측하기는 어렵

다. 따라서, 영상의 아래 영역이 시청자와 가까이 위치한다는 가정을 활용하여 그림 12-(b)와 같이 영상의 위에서 아래로 갈수록 점점 밝아지는 그레디언트 지도를 깊이 지도로 설정하였다.

IV. 스테레오스코픽 영상 생성

우리는 생성한 깊이 지도를 이용하여 좌, 우 영상을 생성하기 위한 시차(parallax)를 구하기 위하여 [1]에서 제안한 방법을 활용하였다.

$$Parallax(x, y) = Max_Parallax \times \left(1 - \frac{Depth(x, y)}{Max_Depth} \right) \quad (17)$$

입력 영상과 깊이 지도를 활용하여 계산한 시차 값을 이용하여, 좌우 영상을 생성할 수 있다. 입력 영상을 왼쪽 시점 영상으로 설정하고, 입력 영상의 각 픽셀을 해당 시차 값만큼 오른쪽으로 이동 시켜준 영상을 오른쪽 시점 영상으로 설정하여, 스테레오스코픽 영상을 생성하였다.

V. 실험 및 결과

1. 인지된 깊이감 품질(Perceived Depth Quality)에 대한 주관적 평가

우리는 앞에서 분류한 총 4가지(원거리샷, 비 원거리샷: *NL1*, *NL2*, *NL3*)의 축구비디오의 샷의 깊이 지도를 통하여 생성한 스테레오스코픽 영상에서 실제로 시청자가 어느 정도 깊이감을 느끼는지 평가하기 위하여 다음과 같이 실험을 수행하였다. 우리는 앞에서 분류한 4가지 샷에 대하여 각 각의 샷으로만 구성된 스테레오스코픽 동영상 3개씩을 준비하였으며, 각 각의 동영상은 200 프레임으로 구성되어 있고, MPEG-4, Xvid 코덱을 통하여 압축되었으며, 320X240의 해상도를 갖는다. 우리는 제안한 알고리즘의 결과 영

그림 13. 깊이감 품질 평가를 위해 사용된 12개의 동영상에서 추출한 프레임들
Fig. 13 Screen shots of 12 test videos for the subjective assessment

상의 깊이감 평가를 위하여, 일반 영상과 제안한 알고리즘을 통해 얻어진 스테레오 영상, 그리고 [7]에서 제안한 움직임 벡터를 이용한 *Modified Time Difference (MTD)* 알고리즘을 통해 얻어진 스테레오 영상을 준비하였다.

(a)

(b)

그림 14. 인지된 깊이감 품질 평가를 위한 (a)평가표와 (b)평가 결과
Fig 14. (a) Rating scale for the subjective assessment (b) Ratings of the perceived depth quality

실험자는 총12개의 동영상에 대하여 원본 2차원 동영상, 제안된 알고리즘을 통한 스테레오 영상, *MTD* 알고리즘을 통한 스테레오 영상을 순서대로 두 번 시청하게 되며, 각 각의 영상에 대한 깊이감을 그림 14(a)의 평가표에 0~100

의 값 사이에서 연속적으로 평가하도록 하였다. 인지된 깊이감 품질 평가는 총 20명에 대하여 수행되었으며, 품질 평가의 결과는 그림 14(b)에서 확인 할 수 있다. 그림 14(b)는 제안한 알고리즘에서 분류한 샷들에 따른 평가된 인지된 깊이감 품질의 평균을 나타낸다. 총 4가지 타입의 샷, 원거리샷, 비 원거리샷 $NL1$, $NL2$, $NL3$ 에 대하여, 일반 2D 동영상에 대해서는 각 각 22.3, 27.3, 15.7, 24.5의 평가 값을 보였으며, 제안한 방법을 통해 얻어진 영상에 대해서는 각각

그림 15. 제안한 스테레오 변환 결과 영상(a),(c) 와 [7]의 스테레오 변환 결과 영상(b),(d)의 비교

Fig.15 Comparison of stereoscopic images generated by proposed algorithm and by [7].

58.8, 63.4, 68.7, 55의 인지된 깊이감 평가 값을 보였다. 그리고 [7]의 알고리즘을 이용한 스테레오 영상의 깊이감 평가 값은 46.2, 40, 51.2, 52 와 같았다. 그림 18-(b)에서 확인할 수 있듯이, 제안된 알고리즘을 통한 스테레오 영상은 2차원 영상에 비하여 깊이감이 상승한 것을 확인할 수 있었다.

또, 제안한 알고리즘의 결과 영상은 [7]의 방법보다 높은 깊이감을 갖는 것으로 평가되었다. [7]의 방법은 과거의 프레임들을 스테레오 변환에 활용한다는 면에서 빠른 처리 속도를 보였으나, 프레임 내의 선수가 복잡한 움직임을 보이는 경우(그림 15-(b) 참조), 또는 카메라가 급격히 빠르게 움직이는 경우(그림 15-(d) 참조), Delay 된 프레임과 현재 프레임을 좌, 우 영상으로 활용하여 스테레오 영상을 생성했을 때 시청자에게 불편함을 주어, 제대로 된 깊이감을 제공하기 어려운 경우가 발생하였다. 따라서, 복잡하고 빠른 움직임이 존재하고, 장면 전환이 빠른 축구비디오의 경우, 현재 프레임의 종류에 따라 깊이 지도를 생성하고 좌, 우 영상을 생성하는 제안한 방법에 의한 스테레오 영상이 시청자에게 보다 높은 깊이감을 제공할 수 있다고 판단할 수 있다.

2. 각 Type의 샷에 따른 깊이 지도와 스테레오스코픽 영상 결과

그림 16-19는 논문에서 제안한 알고리즘을 이용하여, 각 샷의 종류 별로 추출된 깊이지도와 깊이 정보를 이용하여 생성된 스테레오스코픽 영상들을 보여준다.

그림 16. 원거리 샷의 깊이 지도와 스테레오스코픽 영상

Fig. 16 Depth map and stereoscopic images for long shot frames

그림 17. NL1 Type의 비 원거리 샷의 깊이 지도와 스테레오스코픽 영상
Fig. 17 Depth map and stereoscopic images for NL1 type non-long shot frames

그림 18. NL2 Type의 비 원거리 샷의 깊이 지도와 스테레오스코픽 영상
Fig. 18 Depth map and stereoscopic images for NL2 type non-long shot frames

그림 19. NL3 Type의 비 원거리 샷의 깊이 지도와 스테레오스코픽 영상
Fig. 19 Depth map and stereoscopic images for NL3 type non-long shot frames

VI. 결 론

본 논문에서는 단일 시점의 축구 비디오를 스테레오스코픽 동영상으로 변환하는 새로운 스테레오 변환 방법을 제안하였다. 우리는 축구 비디오 분석 과정을 통하여 축구 비디오를 일정한 종류의 샷으로 분류하고, 분류된 샷 종류에 따라 다른 깊이 정보 예측 알고리즘을 적용하여 깊이 지도를 생성하였다. 우리는 관찰을 통하여 발견한, 대부분의 축구비디오의 경우 영상 프레임의 아래쪽이 시청자와 가까운 영역이라는 점을 기본 가정으로 하고, 원거리 샷의 경우, 운동장 발견 알고리즘을 통하여, 관중석 영역과 축구 운동장 영역을 분리한 후, 운동장 영역에 대하여 깊이 값을 할당하여 깊이 지도를 생성하였다. 그리고 비 원거리 샷의 경우, 피부색 발견 알고리즘을 사용하여 비 원거리 샷에서의 사람의 위치를 예측해 맴으로써 전경영역을 예측하고, 예측된 전경영역을 기준으로 하여 가우시안 분포를 모델링하여 깊이값을 할당하였다. 그리고 그 이외의 경우는 영상 위에서 아래로 점점 깊이가 가까워지도록 깊이 지도를 생성함으로써, 피부색을 찾지 못한 예외적 영상에 대한 깊이 지도를 생성하였다. 우리는 각 종류의 샷에 따라 생성된 깊이 지도를 이용하여, 영상 내의 각 픽셀의 시차 값을 계산하고, 이 시차값을 이용하여 새로운 시점 영상을 생성할 수 있었다. 원본 영상과 생성된 새로운 시점의 영상을 이용하여 스테레오스코픽 영상을 생성할 수 있었다. 그리고 인지된 깊이감 품질에 대한 평가 실험을 통하여, 원본 영상에 비하여 시청자가 영상을 시청할 때 향상된 깊이감을 느낄 수 있다는 것을 확인 할 수 있었으며, 축구비디오의 경우, 기존의 움직임 정보를 이용한 스테레오 변환 방법에 비해 높은 깊이감을 제공한 다는 것을 확인 할 수 있었다.

제안한 알고리즘을 통하여 우리는 빠른 샷 전환으로 인해 모션 정보를 정확하게 얻기 힘든 축구 비디오에 대하여, 모션 이외의 정보를 사용하여, 깊이 지도를 생성하고, 스테레오스코픽 영상을 생성할 수 있었다. 프레임 단위로 깊이 지도를 생성하기 때문에, 영상 내의 오브젝트의 수직 움직임에 대해서도 자연스러운 스테레오스코픽 영상을 생성할

수 있다는데 장점이 있다. 하지만 프레임 단위로 깊이 지도를 생성하면서, 노이즈나 알고리즘 오류로 인하여 발생하는 깊이 지도의 시간적 불연속성에 의하여 시청자가 불편함을 느낄 수 있다. 우리는 이 연구를 시작으로 하여, 축구 비디오의 프레임 사이의 관계성을 분석하여, 프레임 사이에 깊이 지도가 자연스럽게 전환되도록 보완할 것이며, 더 나아가, 축구비디오가 아닌 뉴스, 영화, 공연 영상 등의 다양한 일반 영상들에 대해서도 적용 가능한 스테레오 변환 기법을 연구를 진행할 것이다.

참 고 문 헌

- [1] Battiato, S.; Capra, A.; Curti, S.; and La Cascia, M., "3D Stereoscopic Image Pairs by Depth-Map Generation", in Proc. of 2nd International Symposium on 3D Data Processing Visualization and Transmission, 3DPVT(2004)
- [2] S.A. Valencia, R.M. Rodriguez-Dagnino, "Synthesizing Stereo 3D View from Focus Cues in Monoscopic 2D Images", in Proc. Of SPIE, Vol. 5006, pp.377-388 (2003)
- [3] W.J Tam, F. Speranza, L.Zhang, R. Renaud, J. Chan, and C. Vazquez, "Depth image based rendering for multiview stereoscopic displays: Role of information at object boundaries," in Proc. of the SPIE, Vol. 6016, pp. 75-85 (2005)
- [4] Jaeseung Ko, Manbae Kim, and Changick Kim, "2D-to-3D stereoscopic conversion: depth-map estimation in a 2D single-view image," in Proc. Of SPIE, Vol. 6696, (2007)
- [5] B.J. Garcia "Approaches to stereoscopic video based on spatio-temporal interpolation," in Proc. Of SPIE, Vol. 2653, pp. 85-95 (1996)
- [6] J. Kim, Y. Kim, J. Park, Jin-mo Kang, and B. Lee, "Stereoscopic conversion of two-dimensional movie encoded in MPEG-2," in Proc. Of SPIE, Vol. 6311 (2006)
- [7] T. Okino, H. Murata, K. Taima, T. Iinuma, and K. Oketani, "New television with 2D/3D image conversion technologies," in Proc. Of SPIE, Vol. 2653, pp. 96-103 (1996)
- [8] Ilkwon Park, Manbae Kim, and Hyeran Byun, "Robust Zooming Motion Detection for Motion-based stereoscopic conversion," International Workshop on Advanced Image echnology 2007 (IWAIT 2007), pp.370-374, 2007
- [9] Keewon Seo, Jaeseung Ko, Ilkoo Ahn and Changick Kim, "An Intelligent Display Scheme of Soccer Video for Mobile Devices," IEEE Transactions on Circuits and Systems for Video Technology (CSVT), vol.17, no.10, pp.1395-1401, October 2007.
- [10] R. C. Gonzalez, R.E. Woods, Digital Image Processing Second Edition, Prentice Hall, 2002, pp. 295- 302.

저 자 소 개

고 재 승

- 2006년 2월 : 서울시립대학교 전자전기컴퓨터 공학부 (학사)
- 2008년 2월 : 한국정보통신대학교 공학부 (석사)
- 2008년 2월 ~ 현재 : 한국정보통신대학교 공학부 (박사)
- 주관심분야 : 3D Video Coding, ROI based Video Coding, 2D to 3D Conversion

김 영 우

- 2007년 2월 : 광운대학교 컴퓨터 공학전공 (학사)
- 2007년 2월 ~ 현재 : 한국정보통신대학교 공학부 (석사)
- 주관심분야 : 2D-to-3D Conversion, Machine Learning, 3D Image Acquisition

정 용 주

- 1999년 2월 : 홍익대학교 컴퓨터공학과 (학사)
- 2000년 8월 : 한국정보통신대학교 공학부 (석사)
- 2001년 : (주)아이브이시스템 연구원
- 2005년 8월 : 한국정보통신대학교 공학부 (박사)
- 2005년 9월 ~ 현재 : 삼성전자/종합기술원 전문연구원
- 주관심분야 : 이미지/비디오 처리 및 분석, 이미지/비디오 인덱싱, Content adaptation, 3D display processing

김 창 익

- 2000년 12월 : 워싱턴주립대학교 전기공학과 (박사)
- 2000년 12월 ~ 2005년 1월 : Senior Member of Technical Staff, Epson Palo Alto Laboratory, Epson R&D Inc.
- 2005년 2월 ~ 현재 : 한국정보통신대학교(ICU) 공학부 조교수
- 주관심분야 : Intelligent Video, 3D Video, Next Generation Video Communication Systems, Video Analysis