

# 자유 시점 TV에서 시점 합성을 위한 시공간적 배경 정보 추정 기반 홀 채움 방식

## Hole-filling Algorithm Based on Extrapolating Spatial-Temporal Background Information for View Synthesis in Free Viewpoint Television

김 범 수\*, 응웬 띠엔 닷\*, 홍 민 철\*\*

Beomsu Kim\*, Tien-Dat Nguyen\*, Min-cheol Hong\*\*

### Abstract

This paper presents a hole-filling algorithm based on extrapolating spatial-temporal background information used in view synthesis for free-viewpoint television. A new background codebook is constructed and updated in order to extract reliable temporal background information. In addition, an estimation of spatial local background values is conducted to discriminate an adaptive boundary between the background region and the foreground region as well as to update the information about the hole region. The holes then are filled by combining the spatial background information and the temporal background information. In addition, an exemplar-based inpainting technique is used to fill the rest of holes, in which a priority function using background-depth information is defined to determine the order in which the holes are filled. The experimental results demonstrated that the proposed algorithm outperformed the other comparative methods about average 0.3-0.6 dB, and that it synthesized satisfactory views regardless of video characteristics and type of hole region.

### 요 약

본 논문에서는 자유 시점 텔레비전에서 시점 합성 영상 획득을 위해 시공간적 배경 정보 추정 기반 홀 채움 방식을 제안한다. 신뢰할 수 있는 시간적 배경 정보를 획득하기 위해 새로운 배경 코드북의 구성 및 갱신하는 과정을 수행한다. 더불어, 공간적인 국부 배경 정보 추정을 위해 홀 영역의 배경 및 전경 영역의 구별 및 갱신 과정을 수행한다. 추정된 시공간 배경 정보를 조합하여 홀 채움 과정을 수행하고, 잔여 홀 채움을 수행하기 위해 깊이 배경 정보를 이용한 우선순위 함수를 결정하여 표본 기반 인페인팅 기법을 적용한다. 실험 결과를 통해 제안 방식은 기존 방식들과 비교하여 평균 0.3~0.6dB의 성능 향상이 있음을 확인하였으며, 동영상 특성 및 홀 형태에 관계없이 제안된 방식이 새로운 시점 영상을 효과적으로 합성할 수 있음을 확인할 수 있었다.

*Key words* : hole-filling, spatial-temporal background information, background codebook, priority function, DIBR, View Synthesis

\* A School of Electronic Engineering, Soongsil University

★ Corresponding author (mhong@ssu.ac.kr / +82-2-820-0716)

※ Acknowledgment

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Plannig (2012M3C4A7032182).

Manuscript received Jan 21, 2016; revised Feb 22, 2016; accepted Feb 22, 2016

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

### 1. 서론

3-D 동영상 분야의 산업적, 학문적 발전으로 인해 입체 3-D 형태의 콘텐츠를 제공해주는 다양한 시스템 및 디스플레이에 대한 장비가 생산되고 있다[1]. 더불어, 사용자가 3-D 안경과 같은 추가 장비 없이도 깊이감을 체감할 수 있는 자동 입체 디스플레이 장비가 개발되고 있다. 입체 디스플레이는 특정 장면에 대해 다른 시점에서 획득된 다중 시점 영상을 제공함으로써 입체 효과

를 발생하게 된다[2]. 이와 같은 다중 시점 영상들은 다중 카메라들로부터 획득 및 압축되어 전송될 수 있으나, 요구되는 모든 시점 영상들을 카메라로부터 획득할 경우, 과도하게 많은 카메라 장비와 방대한 크기의 데이터를 처리하여 저장하고 전송하는데 따르는 문제점들이 발생하게 된다.

상기와 같은 문제점을 해결하기 위해 실제 시점 영상들을 사용하여 가상의 시점을 합성해내는 깊이 영상 기반 렌더링(Depth Image Based Rendering) 방식이 제안되었다[3]. 깊이 영상 기반 렌더링 기법은 새로운 시점 영상을 생성하기 위해 3-D 워핑(warping)을 사용하게 되며, 이로 인해 합성된 가상 시점 영상에는 홀(hole) 현상이 발생하게 된다. 작은 크기의 홀은 깊이 값의 추정 오류에 의해 발생하게 되며, 큰 크기의 홀은 실제 시점 영상에서는 노출된 영역이 가상 시점 영상에서 가려지는 문제에 의해 발생된다. 이와 같은 홀들의 깊이 정보 또는 텍스처(texture) 정보는 매우 제한적이므로 만족스러운 합성 텍스처의 복원 및 넓은 크기의 가려짐 영역을 제거하기 위해 많은 연구가 진행되어 왔다.

일반적으로 가상 시점은 두 개의 근접 시점으로부터 보간 되거나, 하나의 단일 시점으로부터 가상 시점을 추정하게 된다. 두 개의 근접 시점으로부터 가상 시점 영상을 재구성하는 방식은 가상 시점 영상을 추정하기 위해 많은 제약 환경이 존재하며, 이로 인해 단일 시점 영상을 이용한 가상 시점 영상의 추정 방식이 현실적인 대안으로 연구되고 있다. 단일 시점 영상으로부터 추정된 가상 시점 영상에 존재하는 홀들을 채우는 방식은 깊이 영상 기반 방식 및 텍스처 영상 기반 방식으로 분류될 수 있다. 깊이 영상 기반 홀 채움 방식에서는 배경(background) 영역과 전경(foreground) 영역의 경계를 따라 생기는 불연속면을 제거하기 위해 저역통과(low-pass) 필터가 깊이 영상의 전처리 과정으로 사용된다[4][5]. 그러나 홀 크기가 큰 영역에서 기하학적 왜곡 현상이 발생하며, 이와 같은 문제를 해결하기 위해 보간 방식과 인페인팅 방식이 활발히 연구되어 왔다.

홀 영역을 채우기 위한 보간 방식은 배경 영역과 전경 영역 사이의 경계를 따라 기하학적 왜곡과 열화(blurring) 현상이 발생하며 홀 영역의 크기가 클수록 열화 현상도 증가하게 된다[6]. 더불어, 깊이

정보를 홀 채움 과정에 포함시킨 피라미드형 구조에 기반한 계층적 홀 채움(HHF: Hierarchical Hole-Filling)이 제안되었다[7]. 상기 방식은 홀 영역이 작은 경우에는 객체 경계상의 결점들의 수를 감소시킬 수 있으나, 홀 영역의 크기가 클 경우 과도한 열화 현상을 발생하는 문제점을 갖고 있다.

위와 같은 왜곡 문제점을 해결하기 위해 인페인팅(inpainting) 기법을 홀 채움 과정에 활용하는 연구가 활발히 연구되었다[8]-[10]. 인페인팅 기법은 구조 기반 방식(structure-based method)과 표본 기반 방식(exemplar-based method)으로 분류될 수 있으며, 구조 기반 인페인팅 방식이 구조적 구성요소를 효과적으로 보존하기 때문에 보간 기반 방식보다 개선된 성능을 갖고 있다. 그러나 홀 영역이 큰 경우에는 구조 요소들의 확산으로 인해 세부 정보의 손실이 발생한다. 표본 기반 인페인팅 방식은 이웃 패치(patch)로부터 복제된 화소들을 이용하여 홀 영역을 채우며, 구조 기반 인페인팅 방식을 포함한 다른 방식들과 비교해서 향상된 결과를 나타냄이 입증되었다[11]-[14].

자유 시점 영상의 홀 채움 성능 향상을 위해 표본 기반 인페인팅 방식[11]의 다양한 변형 기법이 연구되어 왔다. 대표적인 방식으로는 연속적인 프레임들의 영상 등록(registration) 패턴이 시점 합성을 갱신하는데 사용하는 방식[12], 시간적으로 추출된 정적인 장면을 가상 시점에 병합하는 방식[13] 등이 있다. 상기 방식들은 전경 객체의 활동성이 클 때는 정적인 배경 화소들을 성공적으로 얻게 되지만, 전경 객체의 활동성이 적을수록 남아있는 전경 화소들의 수가 증가하여 기하학적 왜곡과 열화 현상을 여전히 존재한다. 또한, 홀 화소들이 보상된 배경 참조 영상을 얻기 위해 가우시안 혼성 모델(GMM: Gaussian Mixture Model)방식이 발표되었다[14]. 상기 방식은 그 성능이 초기화된 분포에 매우 민감하다는 문제점이 있다.

3-D 동영상의 시점 합성에서 홀 화소들에 대한 시간적인 정적(stationary) 정보를 다른 프레임들로부터 코드북(codebook) 과정을 통해 추정할 수 있다[15]. 상기 방식은 짧은 시간 동안 존재하는 정적 배경을 추정 시에 오류가 심각하며, 전경 객체의 경계에 위치한 노이즈에 민감한 단점이 있다. 이를 해결하기 위해 깊이 정보를 코드북 생성에 포함한 시차 함수(disparity

function) 활용 방식이 연구되었다[16].

본 논문에서는 단일 시점 영상과 해당 깊이 영상으로부터 임의의 시점 합성 영상 획득을 위해 추정된 시공간적 배경 정보를 이용한 표본 기반 홀 채움 방식을 제안한다. 배경 코드북을 기반으로 신뢰할 수 있는 시간적 배경 정보를 추정하고 갱신하는 방식을 제안한다. 더불어, 공간적 배경 정보를 추정하기 위해 깊이 맵 정보의 국부 특성을 활용하는 방식에 대해 기술한다. 상기와 같은 시공간 배경 정보를 조합하여 홀 채움 과정을 수행하고, 잔여 홀들에서 대해서는 새롭게 정의된 우선순위 함수 (priority function)를 기반으로 하는 표본 기반 인페인팅 방식이 적용된다.

본 논문은 다음과 같이 구성되어 있다. II장에서는 시공간적 추정 배경 정보 및 깊이 향을 포함한 우선순위 함수를 적용한 표본 기반 홀 채움 방식을 제안한다. III장 및 IV장에서는 실험 결과 및 결론에 대해 기술한다.

## II. 제안된 홀 채움 방식

제안된 시점 합성을 위해 홀 채움 방식은 그림 1에 도시된 바와 같이 4단계로 구성된다. 첫 번째 단계에서는  $k$ -평균 군집( $k$ -means cluster) 방식을 사용하여 기존 코드북을 배경 코드워드와 전경 코드워드들로 분류한다. 이를 통해 특정 프레임들에 존재하는 배경 화소들을 갱

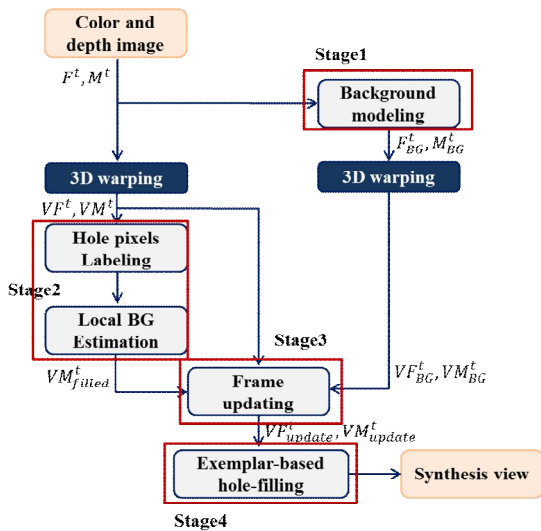


Fig. 1. Block diagram of proposed hole filling method  
그림 1. 제안된 홀 채움 방식의 블록 다이어그램

신된 배경 코드북의 배경 정보로써 표현될 수 있도록 추정한다. 고립된 영역(disocclusion)에 대한 정보는 홀 채움 순서 및 윈도우 크기를 결정하기 위해 효과적으로 사용될 수 있다. 이를 위해 두 번째 단계에서 깊이 맵 정보를 이용하여 홀 화소들을 레이블링(labeling)하고 공간적 국부 배경 정보를 예측한다. 세 번째 단계에서 시공간적 배경 정보를 이용하여 유사도가 큰 홀 화소들을 채우는 과정을 수행한다. 네 번째 단계에서 가상 시점의 홀 화소들에 대해 국부 배경 임계치를 결정하고, 레이블링 과정을 이용하여 각 홀 영역의 배경 및 전경 영역을 예측한다. 상기 과정의 결과를 이용하여 홀 채움 우선순위 함수를 새롭게 정의하여 잔여 홀에 대해 표본 기반 홀 채움 과정을 수행하게 된다. 각 단계별 상세 내용은 다음과 같다.

### 1. 시간적 정적 배경 정보 추정

코드북은 양자화 및 군집 방식(quantization and clustering method)에 기반한 배경 모델을 생성하는데 사용되어 왔다[15] [16]. 기존의 코드북 방식은 코드북 생성, 전경 추출 및 모델 유지 과정으로 구성되어 있다. 그러므로 기존 모델은 배경 영역에 많은 홀 화소들이 발생하는 문제점이 존재하며, 이와 같은 문제점을 해결하기 위해 본 논문에서는 특정 프레임에서 추출된 배경 정보를 다음 프레임의 배경 화소에 해당되는 홀 화소들을 보상하기 위해 사용한다.

홀 채움 문제에서 배경 정보를 수집하기 위해 그림 2와 같이 코드북에 기반한 배경 모델을 제안한다. 제안된 배경 모델은 코드북 생성, 배경 코드북 추출, 임시 (temporary) 배경 추출 및 배경 정보 갱신 과정으로 구성되어 있다.

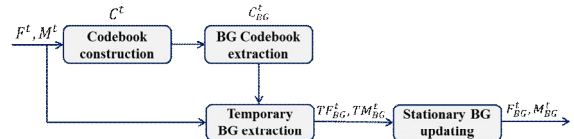


Fig. 2. Block diagram of proposed BG modeling  
그림 2. 제안된 배경 모델링의 블록 다이어그램

배경 정보를 추출하기 위해 현재 프레임( $F^t$ )과 대응되는 깊이 맵( $M^t$ )을 이용하여 적응적인 깊이 코드북 방식[16]에 의해 코드북( $C^t$ )을 생성한

다. 코드북( $C^t$ )에  $k$ -평균 군집 방식을 적용하여 배경 정보에 대한 코드북( $C_{BG}^t$ )을 구성하며,  $C_{BG}^t$ 를 이용하여 임시 배경 프레임 ( $TF_{BG}^t$ )과 임시 배경 깊이 맵 ( $TM_{BG}^t$ )를 결정하게 된다. 마지막으로, 임시 배경 정보를 배경 프레임 ( $F_{BG}^t$ )과 이에 대응되는 배경 깊이 맵 ( $M_{BG}^t$ )를 갱신하는데 사용되고, 갱신된 배경 현재 프레임의 정적 배경 정보로서 사용된다.

위의 과정의 세부 내용은 다음과 같다. 현재 프레임의 컬러 영상( $F^t$ ) 및 깊이 맵( $M^t$ )에 대해  $N$  개의 이전 프레임의 임의 화소의 훈련 집합 ( $X$ )은 다음과 같이 정의될 수 있다.

$$X = \{x_1, x_2, x_3, \dots, x_N\} = \{(p_1, d_1), \dots, (p_N, d_N)\}.$$

위에서  $p_n = \{R_n, G_n, B_n\}$ 은 RGB 벡터를 뜻하고,  $d_n$ 은 해당 화소의 깊이 값을 의미한다.  $L$  개의 코드워드로 구성된 코드북을  $C = \{c_1, c_2, c_3, \dots, c_L\}$ 로 표시한 경우,  $l$ 번째 코드워드는 벡터  $\bar{x}_l = (\bar{p}_l, \bar{d}_l) = (\bar{R}_l, \bar{G}_l, \bar{B}_l, \bar{d}_l)$ 와 5개의 요소로 구성된  $aux_l = (I_{l,\min}, I_{l,\max}, D_{l,\min}, D_{l,\max}, f_l)$ 로 이루어진다. 위의 코드워드에서  $\bar{R}_l$ ,  $\bar{G}_l$  및  $\bar{B}_l$ 는 각 컬러 채널의 평균값을 뜻하며,  $\bar{d}_l$ 는 깊이 평균값을 의미한다. 또한  $I_{l,\min}$ 은 최소 밝기,  $I_{l,\max}$ 은 최대

밝기,  $D_{l,\min}$ 은 최소 깊이 값,  $D_{l,\max}$ 은 최대 깊이 값,  $f_l$ 는 코드워드의 발생 빈도수를 의미한다. 이와 같이 구성된 훈련 집합을 기반으로 코드북은 **알고리즘 1**과 같이 생성된다.

**알고리즘 1**에서  $\leftarrow$ 는 큐(queue)의 끝에 한 개의 요소를 추가하기 위해 사용되는 연산자를 의미한다. 코드북 생성 과정에서 컬러 왜곡 정도, 밝기 유사도, 그리고 시차(disparity) 유사도가 임의의 화소에 대응하는 코드워드를 결정하는데 사용되며, 상기 함수들은 다음과 같이 정의된다[16].

$$\begin{aligned} \text{colordist}(p, \bar{p}_l) &= \sqrt{\|p\|^2 - h^2}, \\ h^2 &= \|p\|^2 \cos^2 \theta = \frac{\langle p, \bar{p}_l \rangle}{\|\bar{p}_l\|^2}. \end{aligned} \quad (1)$$

위 식에서  $\langle a, b \rangle$ 는  $a$ 와  $b$ 의 내적(inner product) 연산자를 나타내며,  $\|\cdot\|$ 은  $l_2$  norm을 의미한다. 더불어, 밝기 유사도 함수는 다음과 같이 정의된다.

$$\begin{aligned} \text{brightness}(I, aux_l) &= \begin{cases} \text{True} & \text{if } I_{low} \leq I \leq I_{hi} \\ \text{False} & \text{otherwise} \end{cases}, \\ I_{low} &= \alpha I_{l,\max}, I_{hi} = \min\left(\beta I_{l,\max}, \frac{I_{l,\min}}{\alpha}\right). \end{aligned} \quad (2)$$

식 (2)에서  $\alpha < 1$  및  $\beta > 1$ 인 상수 값을 의미한다. 더불어, 시차 유사도 함수는 다음과 같이 정의된다.

$$\begin{aligned} \text{disparity}(d, aux_l) &= \begin{cases} \text{True} & \text{if } d_{low} \leq d \leq d_{hi} \\ \text{False} & \text{otherwise} \end{cases}, \\ d_{low} &= \alpha d_{l,\max}, d_{hi} = \min\left(\beta d_{l,\max}, \frac{d_{l,\min}}{\alpha}\right). \end{aligned} \quad (3)$$

위에서 정의된 컬러 영상과 깊이 맵을 사용하는 융합 방식은 코드북 생성 측면에서 첨가 노이즈에 강인한 특성을 갖고 있다[16]. 그러나 상기 방식은 배경 코드워드들을 추출하기 위해 고정된 임계치를 사용하기 때문에 짧은 시간 동안 존재하는 배경 화소를 수집하기에 어려움이 있다. 이와 같은 문제를 해결하기 위해  $k$ -평균 군집 방식을 생성된 코드워드로부터 배경 코드워드들과 전경 코드워드들을 분류하는데 사용된다. 두 분류( $k=2$ )로 코드북을 분류하는 경우, 배경 영역에 대한 코드북은 다음과 같이 생성될 수 있다.

#### 알고리즘 1 코드북 생성

**step 1:** 초기 코드북 모델은 다음과 같이 구성된다.  $C \leftarrow \emptyset$  (empty set) at the time  $t_0$  and  $L=0$  ( $L$ : codebook length).

**step 2:** 임의 화소  $x_l = (p, d) = (R, G, B, d)$ 에 대해 밝기 값은  $I = R + G + B$ 로 정의되고, 코드북  $C = \{c_l | 1 \leq l \leq L\}$ 의 코드워드 중에서  $x_l$ 와 다음 같은 세가지 조건을 만족하는 코드워드를 결정한다.

- $\text{colordist}(p, \bar{p}_l) \leq \epsilon_1$ ,
- $\text{brightness}(I, aux_l) = \text{True}$ ,
- $\text{disparity}(d, aux_l) = \text{True}$

**step 3:** step 2의 세 가지 조건을 만족하는 코드워드가 없는 경우, 다음과 같이 새로운 코드워드 생성한다.

$$\begin{aligned} L &\leftarrow L+1 \\ u_L &\leftarrow (R, G, B, d), \\ aux_L &\leftarrow (I, I, d, 1). \end{aligned}$$

**step 2**의 세가지 조건을 만족하는 코드워드가 있는 경우, 다음과 같이 해당 코드워드를 갱신한다.

$$\begin{aligned} u_l &\leftarrow \left( \frac{f_l \bar{R}_l + R}{f_l + 1}, \frac{f_l \bar{G}_l + G}{f_l + 1}, \frac{f_l \bar{B}_l + B}{f_l + 1}, \frac{f_l \bar{d}_l + d}{f_l + 1} \right), \\ aux_l &\leftarrow (\min(I, I_{l,\min}), \max(I, I_{l,\max}), \\ &\quad \min(d, d_{l,\min}), \max(d, d_{l,\max}), f_l + 1). \end{aligned}$$

**step 4:** 다음 프레임에서 코드북  $C$ 는 **step 2**를 반복 수행해서 갱신된다.

$$C_{BG}^t = \{c_m \in BG \mid 1 \leq m \leq M\}, \quad (4)$$

$$C^t = (C_{BG}^t \cup C_{FG}^t), (C_{BG}^t \cap C_{FG}^t) = \emptyset.$$

임의 화소에 대한 코드워드들의 목록이 오직 1개의 코드워드만 포함하는 경우 해당 배경 코드워드는 군집화(clustering) 과정 없이 배경 코드워드로 간주된다.

배경 코드워드에 대한 분류 과정 후, 각 화소와 배경 코드북을 대응시킴으로서 현재 프레임에서 배경 화소들을 탐지할 수 있다. 그러나 전경 화소들은 여전히 홀 영역들이 존재한다. 전경 영역에서의 배경 정보는 이전 프레임들에 존재할 수 있으므로 배경 정보를 누적할 필요가 있다. 본 논문에서는 배경 코드북을 이용하여 현재 프레임으로부터 추출된 배경 정보를 임시 배경 정보로 정의하며, 전경 영역에서 배경 정보를 다른 프레임에 존재하는 배경 정보를 이용하여 보상하기 위해 이전 프레임의 배경 정보와 병합하는 과정을 수행한다. 임의 화소에 대한 임시 배경 정보는 **알고리즘 2**에서와 같이 각 화소와 배경 코드북을 비교함으로써 결정된다.

### 알고리즘 2 임시 배경 정보 추출

**step 1:**  $x_t = (p, d) = (R, G, B, d)$ 를 식 (4)의 배경 코드북( $C_{BG}^t$ )의 각 코드워드에 대해 다음 함수를 수행한다:

- (a)  $T_1 = \text{colordist}(p, \bar{p}_m)$ ,
- (b)  $T_2 = \text{brightness}(I, \text{aux}_m)$ ,
- (c)  $T_3 = \text{disparity}(d, \text{aux}_m)$ ,
- (d)  $T_c = (T_1 \leq \epsilon_1 \vee (\epsilon_1 < T_1 \leq \epsilon_2 \wedge T_3)) \wedge T_2 \wedge T_3$

**step 2:**  $T_c$ 가 False이면 해당 화소는 임시 배경 영상에서 홀 화소로 정의된다.

$$TF_{BG}^t(i, j) = 0, TM_{BG}^t(i, j) = 0$$

$T_c$ 가 True이면 해당 화소는 임시 배경 영상에서 홀 화소가 아닌 것으로 정의된다.

$$TF_{BG}^t(i, j) = p, TM_{BG}^t(i, j) = d$$

**알고리즘 2**에서 step 1은 깊이 맵 기반 배경 모델 방식[16]과 동일하게 정의되었다. 더불어,  $\epsilon_1$ 과  $\epsilon_2$ 는 임계치를 나타내고, 연산자  $\vee$ 와  $\wedge$ 는 부울대수 “OR”와 “AND” 연산자를 의미한다. 임시 배경 정보에 기반한 정적 배경 정보는 **알고리즘 3**과 같이 갱신된다.

그림 3에 “Street” 동영상의 99번째 프레임에서 배경 정보 추출 성능 비교 결과를 나타내었다. 상기 결과로부터 제안된 배경 모델링을 이용한

정적 배경 정보의 추출 결과가 효과적임을 확인할 수 있다.

배경 모델링 후에, 가상 정적 배경 영상 ( $VF_{BG}^t$ )과 가상 정적 배경 깊이 지도 ( $VM_{BG}^t$ )를 얻기 위해 정적 배경 영상 ( $F_{BG}^t$ )과 정적 배경 깊이 지도 ( $M_{BG}^t$ )가 워핑(warping)된다.

### 알고리즘 3 정적 배경 정보 갱신

**step 1:** 임시 배경 정보의 임의 화소가 홀 화소가 아니면 정적 배경 정보는 다음과 같이 갱신된다:

$$F_{BG}^t(i, j) = TF_{BG}^t(i, j), M_{BG}^t(i, j) = TM_{BG}^t(i, j).$$

홀 화소인 경우 이전 프레임의 정적 배경 정보의 깊이 값에 대해 다음 함수를 수행한다:

$$T_4 = \text{disparity}(M_{BG}^{t-1}(i, j), \text{aux}_m),$$

$$\text{where } c_m = (\bar{p}_m, \text{aux}_m) \in C_{BG}^t.$$

**step 2:**  $T_4$ 가 True이면 현재 프레임의 정적 배경 정보는 이전 프레임 정보에 의해 다음과 같이 갱신된다:

$$F_{BG}^t(i, j) = F_{BG}^{t-1}(i, j), M_{BG}^t(i, j) = M_{BG}^{t-1}(i, j)$$

$T_4$ 가 False면 현재 화소는 임시 배경 정보의 홀 값으로 대체된다:

$$F_{BG}^t(i, j) = 0, M_{BG}^t(i, j) = 0$$

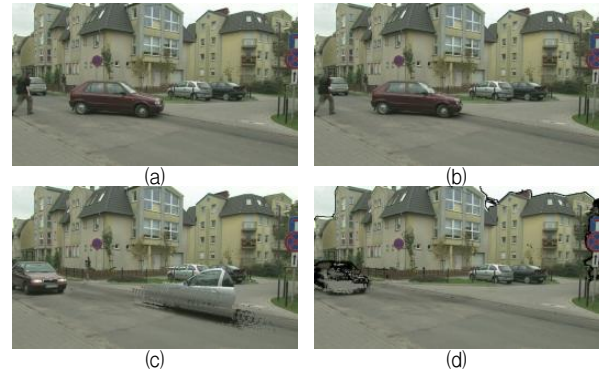


Fig. 3. Performance comparisons of BG extraction of the 99th frame of the “Street” video sequence: (a) original frame, (b) result with stationary BG extraction in [13], (c) result with BG modeling using codebook in [16]. (d) result with proposed background modeling.

그림 3. “Car Park” 동영상의 99번째 프레임 배경 추출 성능 비교: (a) 원 프레임, (b) 정적 배경 추출 결과 [13], (c) 코드북을 이용한 배경 모델링 결과 [16], (d) 제안된 배경 모델링 결과

## 2. 공간적 배경 정보 추정

가상 영상 ( $VF^t$ )과 가상 깊이 맵 ( $VM^t$ )은 많은 가려짐 영역들을 포함하고 있다. 홀 채움의 성능은 그림 4에 도시된 바와 같이 홀 채움 순서와 홀 채움의 윈도우 크기에 민감하므로 홀 화소들

에 대해 레이블링이 필요하다. 본 논문에서는 가상 깊이 맵의 가려짐 영역의 레이블링을 위해 플러드 필(flood-fill) 방식[17]을 사용하였으며, 레이블링(labeling)의 의사코드는 알고리즘 4에 기술하였다.

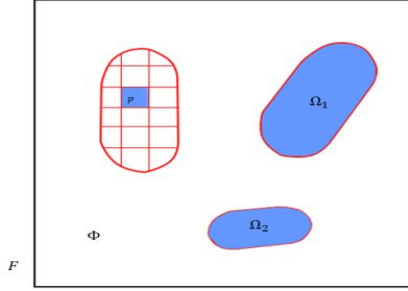


Fig. 4. Example of hole pixel labeling.  
그림 4. 홀 화소 레이블링(labeling) 예

#### 알고리즘 4 홀 화소 레이블링

**step 1:**  $S \leftarrow \emptyset$  (empty set),  $Q \leftarrow \emptyset$ ,  $VM^t(i, j) \in \Omega_k$ .  
**step 2:**  $Q \leftarrow VM^t(i, j)$   
**step 3:** While  $Q \neq \emptyset$  do  
 $Q \rightarrow n(i, j)$   
 if  $n(i, j) = 0$  then  
 $n(i, j) = 1$   
 $S \leftarrow n(i, j)$   
 If  $n(i \pm 1, j \pm 1) = 0$  then  
 $Q \leftarrow n(i \pm 1, j \pm 1)$   
 return  $S$ .

알고리즘 4에서  $S$ 와  $Q$ 는 큐(queue)들을 나타내며,  $\rightarrow$ 는 큐의 첫 번째 요소를 꺼내고 큐에서 제거하는 연산자를 의미한다. 위 레이블링 과정에 따라서  $k$ 번째 고립된 홀 영역을 포함하는 가장 작은 사각 영역은 다음과 같이 정의된다.

$$\Omega_k \subset R_k. \quad (5)$$

그림 5는 홀 화소 레이블링의 결과를 보여주고 있으며 홀 영역들을 사각형으로 표시하였다.

가상 깊이 맵의 배경 화소들과 전경 화소들은 경계영역 상의 왜곡을 방지하기 위해 구별되어야 하며, 경계 영역 구별을 위해 가상 깊이 맵의 국부 배경 정보를 제약 조건으로 활용하는 것이 효과적이다. 국부 통계 특성을 결정하기 위해 홀 영역을 포함하는 레이블링된 사각 영역을 작은 사각영역들로 나눈다.  $M \times N$  크기의  $k$  번째 레이블링된 사각영역( $R_k$ )에 대해 자녀 사각영역(child rectangle)은 다음과 같이 정의된다.

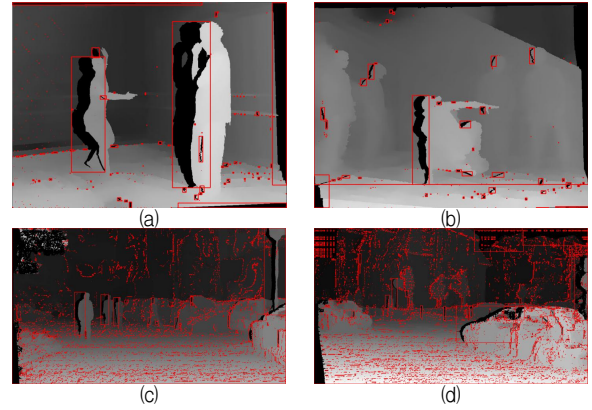


Fig. 5. Results of hole-pixel labeling: (a) "Ballet", (b) "Break Dancers", (c) "Car Park", (d) "Street" video sequences.

그림 5. 홀 화소 레이블링 결과: (a) "Ballet", (b) "Break Dancers", (c) "Car Park", (d) "Street" 동영상

$$R_k = CR_{k,1} \cup \dots \cup CR_{k,z}, \quad (6)$$

$$CR_{k,1} \cap \dots \cap CR_{k,z} = \emptyset.$$

위 식에서 각 자녀 사각 영역들의 크기는  $M \times (N/z)$ 이며, 각 자녀 사각 영역 내에 있는 홀이 아닌 화소들에  $k$ -평균 군집 알고리즘( $k=2$ )을 적용하여 다음과 같은 두 개의 군집들을 얻게 된다.

$$CR_{k,l} = CR_{k,l,low} \cup CR_{k,l,high}, \quad (7)$$

$$CR_{k,l,low} \cap CR_{k,l,high} = \emptyset$$

식 (7)에서  $CR_{k,l,low}$ 와  $CR_{k,l,high}$ 는 자녀 사각영역  $CR_{k,l}$ 의 낮은 깊이 군집과 깊은 깊이 군집을 의미한다. 더불어, 낮은 깊이 군집의 최대값은 다음과 같이 정의된다.

$$d_{k,l,max} = \max\{d_m | d_m \in CR_{k,l,low}\}. \quad (8)$$

식 (8)의 결과는 국부 배경 값으로 사용되며 가상 깊이 맵의 홀 화소는 다음 조건이 만족되는 경우 위의 결과에 의해 채워지게 된다.

$$VM_{filled}^t(i, j) = \begin{cases} d_{k,i,max} & \text{if } T_5(i, j) = True \\ VM^t(i, j) & \text{otherwise} \end{cases}, \quad (9)$$

$$T_5(i, j) = ((VM^t(i, j) \in CR_{k,l,low}) \wedge (VM^t(i, j) = 0)).$$

그림 6에 국부 정보를 이용하여 추정된 배경 정보에 의해 채워진 가상 깊이 맵 결과( $VM_{filled}^t$ )의 예를 도시하였다.

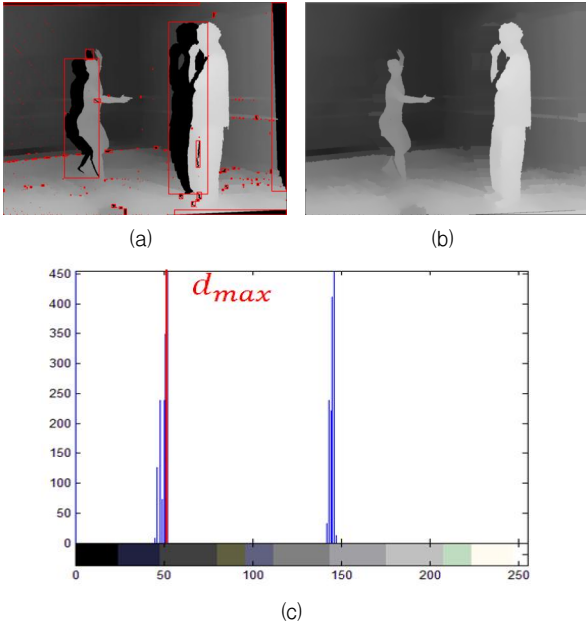


Fig. 6. Example of local BG estimation: (a) virtual depth map with grouped hole pixels, (b) result after local BG estimation, (c) histogram of child rectangle.

그림 6. 국부 배경 추정 예: (a) 그룹화된 홀 화소들의 가상 깊이 맵 (b) 국부 배경 추정 결과, (c) 자녀 시각 영역 히스토그램

### 3. 시공간적 배경 정보에 의한 프레임 갱신

본 절에서는 추정된 시간적 정적 배경 깊이 정보( $VM_{BG}^t$ )와 추정된 공간적 배경 정보( $VM_{filled}^t$ ) 간의 유사성을 이용하여 가상 영상( $VF^t$ )을 갱신하는 과정에 대해 기술한다. 이미 기술한 바와 같이  $VM_{BG}^t$ 는 프레임들 간의 상관관계를 결합시킨 배경 모델링인 가상 배경 깊이 맵을 나타내고,  $VM_{filled}^t$ 는  $VM^t$ 내의 공간적 상관관계를 이용한 가상 깊이 맵을 나타낸다. 이와 같은 특성을 이용하여  $VM_{BG}^t(i, j)$ 가  $VM_{filled}^t(i, j)$ 의 일정 범위 내에 있을 때, 홀 화소가 아닌  $VF_{BG}^t(i, j)$ 는 홀 화소인  $VF^t(i, j)$ 를 위해 유효한 배경 화소로 다음과 같이 고려될 수 있다.

$$VF_{update}^t(i, j) = \begin{cases} VF_{BG}^t(i, j) & \text{if } A \leq VM_{BG}^t \leq B \\ VF^t(i, j) & \text{otherwise} \end{cases} \quad (10)$$

식 (10)은  $VM^t$ 의 홀 화소에 대해 적용되며,  $A = (VM_{filled}^t(i, j) - T)$  및  $B = (VM_{filled}^t(i, j) + T)$ 로 정의된다. 더불어  $T$ 는 양수 임계치를 의미한다. 더불어, 가상 깊이 맵도 다음과 같이 깊이 맵의 일

관성을 유지하기 위해 갱신된다.

$$VM_{update}^t(i, j) = \begin{cases} VM_{BG}^t(i, j) & \text{if } A \leq VM_{BG}^t(i, j) \leq B \\ VM^t(i, j) & \text{otherwise} \end{cases} \quad (11)$$

위의 결과에서  $T$ 가 감소할수록  $VF_{BG}^t$ 의 대응되는 유효 화소에 의해  $VF^t$ 의 배경 화소로 갱신되지 않는 화소의 수는 증가한다. 한편으로,  $T$ 가 증가할수록  $VF^t$ 가 부정확하게 갱신되는 화소의 수는 증가하게 된다.

### 4. 신규 우선순위 함수 기반 홀 채움

정적 배경 정보를 이용하여  $VF^t$ 를 갱신한 후에도 다수의 홀 화소들이 존재한다. 상기 홀 화소들은 표본 기반 인페인팅 기법을 이용하여 채워질 수 있다. 인페인팅 기법을 적용하기 전에 미디언 예측기(median estimator)로 홀 화소들을 초기화함으로써 시점 합성의 시각적 품질이 향상될 수 있음이 입증되었다[12]. 전경 영역은 다수의 동적(non-stationary) 영역을 포함하는 관계로 미디언 예측기로 추정된 값은 배경 화소 영역에만 적용하는 것이 효과적이다. 이를 위해 홀 화소들의 배경 영역과 전경 영역 간의 경계를 구별할 필요가 있다.

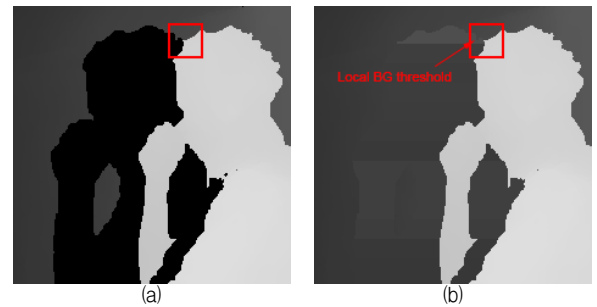


Fig. 7. Example of estimating local BG threshold: (a) updated virtual depth map( $VM_{updated}^t$ ), (b) virtual depth map filled by local BG estimation( $VM_{filled}^t$ ).

그림 7. 국부 배경 임계치 추정의 예: (a) 갱신된 가상 깊이 맵( $VM_{updated}^t$ ), (b) 국부 배경 추정에 의한 가상 깊이 맵( $VM_{filled}^t$ )

배경 영역과 전경 영역 간의 경계 구별을 위해 그림 7에 도시한 바와 같이 홀 화소인  $VM_{update}^t(m, n)$ 에 대응되는  $VM_{filled}^t(m, n)$ 를 중앙 화소로 하는  $U \times V$  크기의 윈도우( $\Psi_u$ )를 설정하여 배경 임계치를 다음과 같이 결정한다.

$$Thr(\Psi_d) = \max(VM_{filled}^t(m, n) \in \Psi_d). \quad (12)$$

식 (12)는  $VM_{update}^t(m, n) = 0$  경우에 적용한다. 위의 결과를 통해  $Thr(\Psi_d)$  값보다 작은 윈도우 내의  $VM_{filled}^t$ 의 홀 화소들은 배경 화소들로 간주한다. 이와 같은 결과를 기반으로 미디어 예측기를 결정하는 과정은  $VM_{update}^t(m, n)$ 를 중심 화소로 하는  $U \times V$  크기의 윈도우 ( $\Psi_p$ )를 고려하여 진행된다. 유용한 미디어 예측값을 산출하기 위해 식 (13)과 같이  $VF_{update}^t$ 의 윈도우( $\Psi_p$ ) 내에 있는 홀이 아닌 화소들만 미디어 필터의 요소들로 고려된다.

$$T_6 = (VM_{update}^t(m, n) \neq 0) \wedge (VM_{filled}^t(m, n) \leq Thr(\Psi_d)). \quad (13)$$

$T_6 = True$ 인 경우, 식 (13)의 조건을 이용하여  $\Psi_p$ 의 추정된 배경 영역 내에 있는 홀 화소들에 초기 배경 값이 다음과 같이 결정된다.

$$BG_{\Psi_p} = median(VF_{update}^t(m, n) \in \Psi_p). \quad (14)$$

식 (14)를 기반으로 표본 기반 인페인팅을 위한 초기 가상 영상은 다음과 같이 결정된다.

$$VF_{init}^t(m, n) = \begin{cases} BG_{\Psi_p} & \text{if } VM_{update}^t(m, n) = 0 \\ VF_{update}^t(m, n) & \text{otherwise} \end{cases}. \quad (15)$$

위 식에서  $VF_{init}^t(m, n) \in \Psi_p$ . 위의 필터링 과정은  $VF_{update}^t$ 에서 홀이 아닌 인접한 화소들로부터 유효 배경 정보를 추출하는 것으로 해석될 수 있다.

잔여 홀을 채우기 위해 사용하는 표본 기반 인페인팅 기법에서는 가려진 영역의 경계 영역에서 홀이 아닌 화소들에 대한 우선순위를 결정한다.

우선순위에 따라 에너지 함수를 이용하여 홀 화소를 갖고 있는 임의 패치와 가장 유사한 참조 패치를 탐색하게 되며, 다수의 패치들이 동일한 에너지 함수 값을 가질 경우 가장 유사한 패치를 결정하기 위해 구조적인 유사도(SSIM)를 사용할 수 있다[18]. 본 논문에서는 표본 기반 인페인팅

방식을 수행하기 전에 홀 화소들이 채워지는 순서를 결정하기 위해  $VF_{update}^t$ 의 홀 화소들을 **알고리즘 4**를 적용하여 레이블링한다. 레이블링 결과에 따라 가장 작은 홀 영역에서부터 시작하여 가장 큰 홀 영역으로 옮겨가며 표본 기반 홀 채움을 수행한다. 그림 8에 도시된 예에서 레이블링된 홀 영역 ( $F$ ) 내에서 제안하는 우선순위 함수를 적용한 표본 기반 홀 채움은 **알고리즘 5**와 같이 수행되어 그림 9와 같은 결과를 얻게 된다.

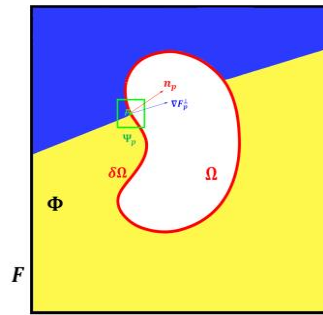


Fig. 8. Example of priority determination ( $\Omega$ : hole region,  $\Phi$ : non-hole region,  $\delta\Omega$ : boundary region).

그림 8. 우선순위 결정 예 ( $\Omega$ : 홀 영역,  $\Phi$ : 홀이 아닌 영역,  $\delta\Omega$ : 경계 영역)

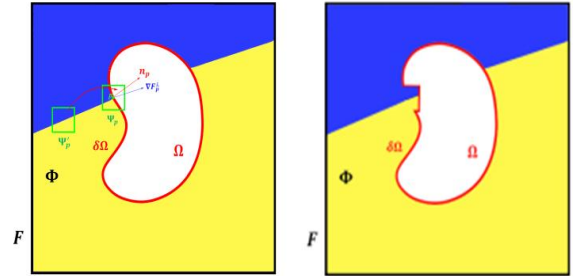


Fig. 9. Example of exemplar-based hole-filling process. 그림 9. 표본 기반 홀 채움 과정 예

**알고리즘 5**에서  $C(p)$ ,  $D(p)$  및  $Z(d)$ 는 신뢰 항(confidence term), 데이터 항(data term) 및 깊이 항(depth term)을 의미한다. 첫 두 항은 [11]에서 정의되었고, 세 번째 항은 본 논문에서 새롭게 정의하였다. 신뢰 항은 다음과 같이 윈도우 내에서 홀이 아닌 화소들의 비율로 정의된다:

$$C(p) = \frac{|\Phi|}{|\Psi_p|}. \quad (16)$$



**알고리즘 5** 우선순위 함수를 적용한 표본 기반 홀 채움

**step 1:** 그림 8에 도시된 바와 같이 경계 영역 화소  $p = VF_{update}^t(i, j)$ 를 중심으로 하는  $U \times V$  크기의 윈도우  $\Psi_p$ 에 대해 우선순위는 다음과 같이 결정한다.

$$Pri(p) = C(p)D(p)Z(d)$$

( $d$ : the corresponding depth value of  $VM_{filled}^t$ )

**step 2:** 우선순위에 따라 홀 채움 과정을 수행한다. 경계 화소  $q_k = VF_{init}^t(i, j)$ 를 중심으로 하는  $U \times V$  크기의 패치( $\Psi_q$ )에 대해 홀 영역의 집합과 비 홀 영역의 집합을 다음과 같이 정의한다:

$$\Phi_q = \{q_1, \dots, q_k\}, \Omega_q = \{q_{k+1}, \dots, q_{U \times V}\},$$

$$\Psi_q = \Phi_q \cup \Omega_q.$$

**step 3:**  $\Psi_q$ 와 가장 유사한 참조 패치는 다음 에너지 함수를 최소화하는 것으로 결정된다:

$$E = \sum_{i=1}^k \|q_i - r_i\|^2 + \omega \sum_{j=k+1}^{U \times V} \|q_j - r_j\|^2.$$

( $r_i, r_j$ : 참조 패치에서 대응되는 홀 및 비홀 컬러 화소)

**step 4:** The hole pixels of  $VF_{update}^t$ 의 홀 화소들은 그림 9에서와 같이 가장 유사한 패치에서 대응되는 화소에 의해 대체된다.

**step 5:**  $VF_{update}^t$  내의  $F_k$ 의 경계 화소들을 갱신한다.

**step 6:**  $F_k$ 의 모든 홀 화소가 채워질때까지 **step 1**을 반복한다.

위 식에서  $|\Psi_p|$ 는  $\Psi_p$ 의 영역을,  $|\Phi|$ 는 윈도우 내에서 홀이 아닌 화소들의 영역을 나타낸다. 그러므로 윈도우 내에서 홀이 아닌 화소들의 수가 증가할수록  $C(p)$ 는 증가하게 된다. 더불어, 데이터 항은 다음과 같이 정의된다:

$$D(p) = \frac{\langle \nabla F_p^\perp, n_p \rangle}{\lambda}. \quad (17)$$

식 (17)서  $\lambda$ 는 정규화 요소(normalization factor)를 의미하며, 전형적으로 한 화소가  $n$ 비트로 표현될 때,  $\lambda = (2^n - 1)$ 로 설정된다. 또한,  $n_p$ 는  $\delta\Omega$ 에 수직인 법선 단위 벡터(normal unit vector)를 나타내고  $\nabla F_p^\perp$ 는  $\Psi_p$ 의 중심과 동일한 밝기의 윤곽(isophote)을 의미하며,  $\nabla F_p^\perp$ 와 법선 단위 벡터의 방향이 동일할 때  $D(p)$ 가 가장 크게 된

다. 정적 영역 내에 있는 패치는 유사한 패치를 상대적으로 쉽게 결정할 수 있으므로 기하학적 왜곡을 최소화하기 위해 홀 채움 과정을 배경 화소들에서 시작하는 것이 효과적이다. 이와 같은 특성을 반영하여 우선순위에 깊이 항을 사용하였으며, 깊이 항에 의한 우선순위는 다음과 같이 정의된다:

$$Z(d) = \begin{cases} 1 & \text{if } d \leq Thr(\Psi_d), d \in \Psi_d \subset VM_{filled}^t. \\ 0 & \text{otherwise} \end{cases}. \quad (18)$$

**알고리즘 5**의 step 3에서  $VF_{init}^t$ 가 가장 유사한 참조 패치를 구하기 위해 사용되는데, 이는 에너지 함수를 최소화하는데 있어  $VF_{update}^t$ 보다  $VF_{init}^t$ 가 더 신뢰할 만한 정보를 지원하기 때문이다. 뿐만 아니라, 최적 대응 참조 패치를 찾는 데 있어 계산적 부담을 줄이기 위해 국부 배경 임계치 보다 더 큰 깊이 값을 갖는 참조 패치는 배제된다.

### III. 실험 결과

제안된 방식의 성능을 검증하기 위해 다양한 동영상을 이용하여 실험을 수행되었다. 본 논문에서는 “Ballet”, “Break Dancers”, “Car park”, “Street”, 및 “Book Arrival” 동영상 결과에 대해 기술한다. “Ballet”, “Break Dancers” 및 “Book Arrival” 동영상의 해상도는  $1024 \times 768$ 이며, “Car Park”와 “Street” 동영상의 해상도는  $1920 \times 1088$ 이다. 본 논문에서 제안한 방식의 성능 확인을 위해 DHHF[7], Ming 방식[13], Koppel 방식[12], Yao 방식[14]과 비교하였다. 알고리즘들의 성능을 평가하기 위한 지표로 PSNR과 SSIM를 사용하였다.

배경 모델링에 대한 코드북을 만들기 위해서

Table 1. Average PSNR and SSIM comparisons

표 1. 평균 PSNR과 SSIM 비교

Video Seq.	Camera	PSNR (dB)					SSIM				
		DHHF [7]	Ming [13]	Koppel [12]	Yao [14]	proposed method	DHHF [7]	Ming [13]	Koppel [12]	Yao [14]	proposed method
Ballet	3→4	25.943	<b>27.929</b>	27.721	26.058	27.859	<b>0.7966</b>	0.7945	0.7920	0.7793	0.7959
Break Dancers	4→5	30.838	30.584	30.782	29.968	<b>30.996</b>	<b>0.8011</b>	0.7948	0.7941	0.7921	0.7954
Car Park	4→5	<b>27.418</b>	26.167	25.886	26.474	26.749	0.8020	0.8039	0.8018	0.7975	<b>0.8066</b>
Street	4→5	23.757	<b>24.104</b>	23.862	23.838	23.984	0.6530	0.6784	0.6749	0.6682	<b>0.6797</b>
Book Arrival	7→8	<b>24.933</b>	23.539	23.972	24.727	24.796	<b>0.6257</b>	0.6209	0.6220	0.6179	0.6226
<b>Average</b>		26.558	26.465	26.445	26.213	<b>26.877</b>	0.7357	0.7385	0.7370	0.7310	<b>0.7400</b>

다양한 매개변수들이 정의되었다. [16]에서와 같이 알고리즘 1 및 2의 매개변수로써  $\epsilon_1 = 10$ ,  $\epsilon_2 = 16$ ,  $\alpha = 0.75$ ,  $\beta = 1.3$ 으로 설정하여 사용되었다. 또한, II-2절에서 기술한 바와 같이 국부 배경 값을 추정하기 위해 부모 사각영역을 자녀 사각영역들로 분할하였다. 본 실험에서는 분할된 작은 사각영역의 너비는 분할 전 부모 사각영역의 너비와 동일하며 분할된 자식 사각영역의 높이는 16으로 설정하였다.

이미 기술한 바와 같이 식 (10)에서 임계값  $T$ 가 증가함에 따라  $VF_{BG}^t$ 에 의해 부정확하게 갱신되는  $VF^t$ 의 배경 화소 수는 증가한다. 반면에  $T$ 가 감소할수록  $VF_{BG}^t$ 에 의해 갱신되는  $VF^t$ 의 유용한 배경 화소 수는 감소하게 된다. 실험을 통해  $T=5$ 일 때 안정적인 결과가 나오는 것을 확인하였다.

본 논문에서 제안된 표본 기반 인페인팅 방식에서 검색 영역 및 패치 크기는 홀 영역의 크기에 따라서 결정된다. 실험적으로 홀 화소가 50개 미만인 영역에서는  $20 \times 20$ (높이 $\times$ 너비)의 검색 영역과  $3 \times 3$ 의 패치 크기를, 50~1000개인 영역에서는  $60 \times 60$ 의 검색 영역과  $5 \times 5$ 의 패치 크기를 사용하는 것이 효과적임을 확인하였으며, 이 이외의 경우에는  $80 \times 200$  크기의 검색 영역과  $9 \times 9$ 의 패치 크기를 사용했다. 또한, 최적 대응 참조 패치를 검색하기 위해 알고리즘 5의 step 3에서 사용되는 에너지의 함수에서의 홀 영역의 가중치는  $\omega=0.2$ 로 설정하였다.

표 1에 실험에 사용한 각 동영상들의 100개 프레임들에 대한 평균 PSNR과 SSIM 비교결과를 기술하였다. “Car Park”와 “Book Arrival” 동영상에서는 깊이 DHHF 방식이 다른 방식들보다 우수한 성능을 보이며, “Ballet”과 “Street” 동영상에 대해서는 Ming 방식이 다른 방식들에 비해 우수한 성능을 보임을 확인할 수 있었다. 본 논문에서 제안된 방식은 “Break Dancers” 동영상에 대해서는 다른 방식들과 비교하여 개선된 성능을 나타냈다. 상기 결과로부터 제안된 방식과 Ming 방식이 가려짐 현상이 심해질수록 좋은 성능을 나타냈고, DHHF 방식은 가려짐 현상이 적은 경우에 좋은 성능을 나타내는 것으로 분석되

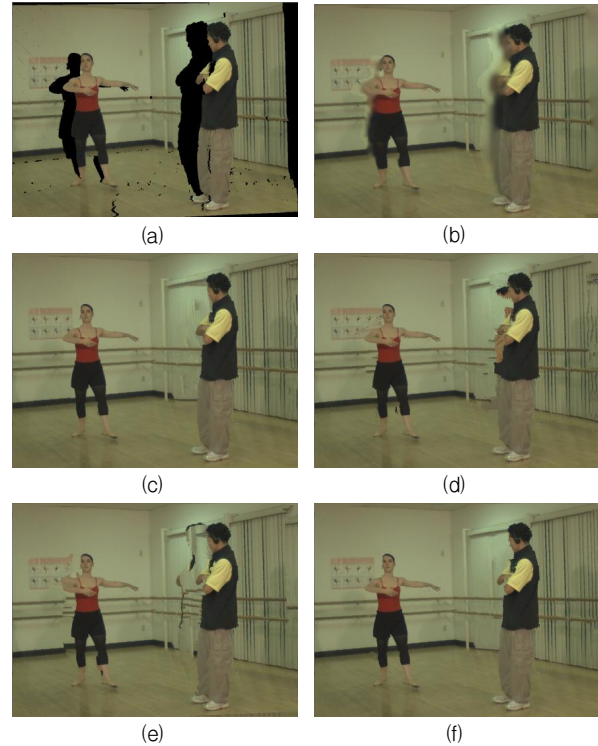


Fig. 10. Visual comparisons of the 46th frame of the “Ballet” video sequence with synthesizing from camera 3 to camera 4: (a) warped result, (b) result with the DHHF method, (c) result with Ming’s method, (d) result with Koppel’s method, (e) result with Yao’s method, (f) result with proposed method.

그림 10. 3번 카메라로부터 4번 카메라로 합성된 “Ballet” 동영상 46번째 프레임의 시각적 비교: (a) 워핑된 결과, (b) DHHF 방식 적용 결과, (c) Ming 방식 적용 결과, (d) Koppel 방식 적용 결과, (e) Yao 방식 적용 결과, (f) 제안 방식 적용

었다. 그러나, DHHF 방식은 배경 영역과 전경 영역 사이의 경계면에서 심각한 열화 현상이 발생하는 문제점이 발생했다. 또한 Ming 방식은 가려짐 현상의 유형에 따라서 성능의 편차가 큰 문제점을 갖고 있음을 알 수 있었다.

반면에 제안된 방식은 가려짐 현상의 유형에 관계없이 일정한 성능을 보장하는 것을 확인할 수 있었다. SSIM 지표 측면에서는 제안된 방식과 DHHF 방식이 다른 방식들에 비해 우월한 성능을 나타내고 있으나, 이미 언급한 바와 같이 텍스처를 합성하기 위해 오직 전경 정보만을 고려하는 DHHF 방식은 경계면 영역에서 심한 열화 현상이 발생되었다. 더불어, Ming 방식은 PSNR에 비해 SSIM이 만족스럽지 못한 결과를 보



Fig. 11. Visual comparisons of the 33rd frame of the “Car Park” video sequence with synthesizing from camera 3 to camera 4: (a) warped result, (b) result with the DHHF method, (c) result with Ming’s method, (d) result with Koppel’s method, (e) result with Yao’s method, (f) result with proposed method.

그림 11. 3번 카메라로부터 4번 카메라로 합성된 “Car Park” 영상 33번째 프레임의 시각적 비교: (a) 워핑된 결과, (b) DHHF 방식 적용 결과, (c) Ming 방식 적용 결과, (d) Koppel 방식 적용 결과, (e) Yao 방식 적용 결과, (f) 제안 방식 적용 결과

였는데, 이는 배경 영역에서의 기하학적 왜곡 현상이 심한 것으로 분석되었다.

그림 10은 “Ballet” 동영상의 46번째 프레임이 3번 카메라로부터 4번 카메라로 합성된 시점에 대한 시각적 비교 결과, 그림 11은 “Car Park” 동영상의 33번째 프레임이 4번 카메라에서 5번 카메라로 합성된 시점에 대한 시각적 비교 결과를 나타낸다. 상기 결과에서 (a)는 워핑된 프레임, (b)는 DHHF 방식 적용 결과, (c)는 Ming 방식 적용 결과, (d)는 Koppel 방식 적용 결과, (e)는 Yao 방식 적용 결과 및 (f)는 제안된 방식이 적용된 결과를 나타낸다. 시각적인 결과의 상세 비교를 위해 그림 12 및 13에 위의 결과의 부분 확대하여 발췌한 영상을 나타내었다. DHHF 방식 결과는 배경 영역과 전경 영역 사이 경계 영역에서 심각한 열화 현상이 존재함을 알 수 있었으며, 이와 같은 왜곡 현상은 계층적 구조를 생성하고 깊이

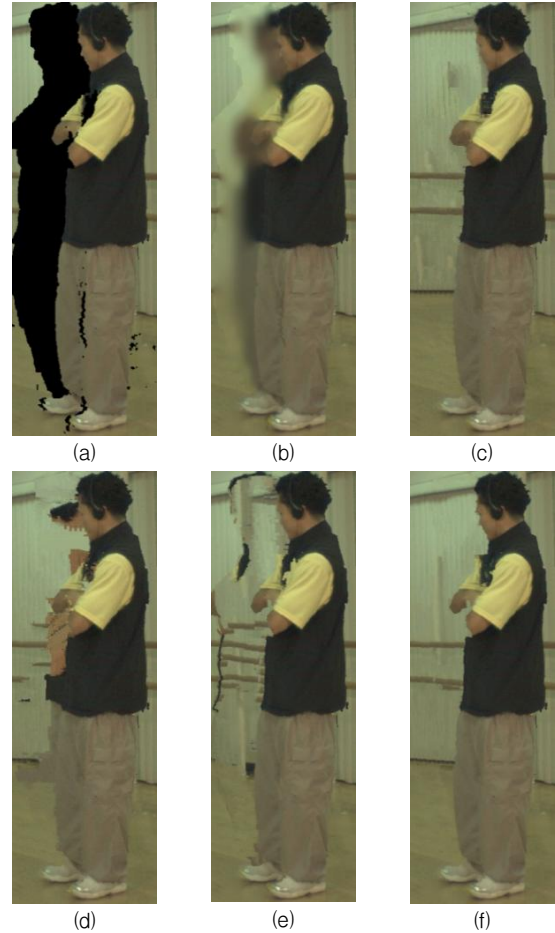


Fig. 12. Visual comparisons of zoomed and cut image of Figure 11: (a) warped result, (b) result with the DHHF method, (c) result with Ming’s method, (d) result with Koppel’s method, (e) result with Yao’s method, (f) result with proposed method.

그림 12. 그림 11의 확대 발췌된 영상의 시각적 비교: (a) 워핑 결과, (b) DHHF 방식 결과, (c) Ming 방식 결과, (d) Koppel 방식 결과, (e) Yao 방식 결과, (f) 제안 방식 결과

값에 대응시키는 함수를 토대로 가우시안 필터를 적용해서 발생된 것으로 관찰되었다. 합성된 시점 영상의 열화 정도는 워핑된 영상의 홀 크기가 증가할수록 증가함을 확인할 수 있었으며, Koppel 방식 결과 영상은 DHHF 방식과 비교하여 열화 정도가 작게 나타났다. 그러나, Koppel 방식은 배경 영역과 전경 영역의 부정확한 구분에 의해 다른 왜곡 현상을 발생하는 결과를 초래하였으며, 특히 잠시 동안 나타나는 연속적인 특징점들이 심각하게 왜곡된 것으로 관찰되었다. 또한, Yao의 방식은 배경 영역과 전경 영역의 경계면들에 전이 결점들(transition artifacts)을 발생시켰으며,

이러한 결점들은 홀 영역들을 갱신 오류에 의해 발생함을 확인할 수 있었다. 이와 같은 오류는 배경 정보를 표본 기반 인페인팅 과정에서 우선순위 함수에서 고려하지 않은 관계로 발생하였음을 확인할 수 있었다. Ming 방식은 우선순위 함수를 결정하는 과정에서 배경 영역에 대한 깊이 정보를 효과적으로 사용하여 경계 영역에서 가장 우수한 성능을 나타내었다. 그러나, 홀 화소들의 갱신 과정에서 단기간에 존재하는 배경 정보를 고려되지 않아서 텍스처 영역에서 시각적인 왜곡 현상이 존재한다. 반면에 제안된 방식은 비교 방식보다 개선된 우월한 성능을 나타내고 있다. 이와 같은 결과는 신뢰할 수 있는 배경 모델링과 국부 배경에 의한 추정 및 깊이 항을 고려한 우선순위 함수의 활용에 따른 것임을 알 수 있었다.

본 논문에서는 시공간적 배경 정보와 깊이 정보를 우선순위 함수를 정의하여 홀 채움 과정에 결합하여 효과적으로 시점 합성 영상을 획득하는 방식을 제안하였다. 신뢰할 수 있는 시간적 배경 모델링을 추출하기 위해 배경 코드북을 이용하여 정적 배경 정보를 갱신하였으며, 배경 영역과 전경 영역 사이의 경계면의 구분을 위해 공간적 배경 정보의 국부 배경 값을 추정하였다. 상기와 같이 추정된 시간적 정적 배경 정보와 공간적 배경 정보를 조합하여 홀 화소들을 효율적으로 채웠으며, 잔존하는 홀들을 표본 기반 인페인팅 기법으로 채우기 위해 깊이 항을 포함한 새로운 우선순위 함수를 제안하였다. 실험 결과를 통해 제안된 방식이 동영상의 특성이나 홀 유형에 관계없이 우수한 성능을 보이고 있음을 확인할 수 있었다.

#### IV 결론

본 논문에서는 자유 시점 TV에서 시점 합성을 위한 새로운 홀 채움 방식을 제안하였다. 시간적 정적 배경 정보와 공간적 배경 정보를 추정하기 위해 새로운 배경 코드북과 국부 배경 추정 방식을 제안하였으며, 시공간 배경 정보를 결합하여 배경 영역들에 속한 홀 화소들은 효율적으로 채울 수 있음을 확인할 수 있었다. 더불어, 잔존하는 홀들을 채우기 위해 표본 기반 인페인팅 기법 적용을 위한 우선순위 함수를 제안하였다. 실험



Fig. 13. Visual comparisons of zoomed and cut image of Figure 12: (a) warped result, (b) result with the DHHF method, (c) result with Ming's method, (d) result with Koppel's method, (e) result with Yao's method, (f) result with proposed method.

그림 13. 그림 12의 확대 발췌된 영상의 시각적 비교: (a) 워핑 결과, (b) DHHF 방식 결과, (c) Ming 방식 결과, (d) Koppel 방식 결과, (e) Yao 방식 결과, (f) 제안 방식 결과

결과를 통해 제안된 방식의 우수성을 확인할 수 있었다. 현재 카메라 추적과 결합한 신규 배경 모델링 방식을 연구 중에 있으며, 상기 접근 방식을 통해 카메라 동적 환경에서 성능 개선이 가능할 것으로 기대된다.

## References

- [1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang, "Three-dimensional video postproduction and processing," *Proc. of the IEEE*, vol. 99, no. 4, pp. 607-625, Apr. 2011.
- [2] N. A. Dodgson, "Autostereoscopic 3D displays", *Computer*, vol. 38, no. 8, pp.31-36, Aug. 2005.
- [3] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D TV," *SPIE Proc. Stereoscopic Image Proc. and Rendering*, pp. 93-104, May 2004.
- [4] P.-J. Lee, and Effendi, "Nongeometric distortion smoothing approach for depth map preprocessing," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 246-254, Apr. 2011.
- [5] S. Zinger, D. Ruijters, L. Do and P. H. N. de With, "View interpolation for medical images on autostereoscopic displays," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 21, no. 5-6, pp. 533-541, Jan. 2012.
- [6] S. Zinger, L. Do and P. H. N. de With, "Free-viewpoint depth image based rendering," *J. of Visual Commu. and Image Representation*, vol. 21, no. 5-6, pp. 533-541, Jul. 2012.
- [7] M. Solh and G. Alregib, "Hierarchical hole-filling for depth-based view synthesis in FTV and 3D video," *IEEE J. of Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 495-504, Sep. 2012
- [8] B. Marcelo, S. Guillermo, C. Vincent and B. Coloma, "Image inpainting," *The 27th Annual Conf. on Comput. Graphics and Interactive Techniques*, pp. 417-424, 2000.
- [9] K. Oh, S. Yea and Y.-S. Ho, "Hole-filling method using depth based inpainting for view synthesis in free viewpoint television and 3D video," *Picture Coding Symposium*, pp. 1-4, June 2009.
- [10] I. Daribo, H. Saito, "A novel inpainting-based layered depth video for 3D TV," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 533-541, June 2011.
- [11] A. Criminisi, P. Perez and K. Toyama, "Object removal by exemplar-based inpainting," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 721-728, June 2003.
- [12] M. Koppel, X. Wang, D. Doshkov, T. Wiegand and P. Ndjiki-Nya, "Consistent spatio-temporal filling of disocclusions in the multiview video plus depth format," *IEEE Int. Workshop on Multimedia Signal Process.*, pp. 25-30, Sept. 2012.
- [13] M. Xi, L.-H. Wang, Q.-Q. Yang, D.-X. Li and M. Zhang, "Depth image based rendering with spatial and temporal texture synthesis for 3DTV," *EURASIP J. on Image and Video Process.*, vol. 2013, no. 1, DOI:10.1186/1687-5281-2013-9, Feb. 2013.
- [14] C. Yao, T. Tillo, Y. Zhao, J. Xiao, H. Bai and C. Lin, "Depth map driven hole filling algorithm exploiting temporal correlation information," *IEEE Trans. Broadcasting*, vol. 60, no. 2, pp. 394-404, June 2014.
- [15] K. Kim, T. H. Chalidabhongse, D. Harwood and L. Davis, "Background modeling and subtraction by codebook construction," *IEEE Int. Conf. on Image Processing*, vol. 5, pp. 3061-3064, Oct. 2004.
- [16] E. J. Fernandez-Sanchez, J.r Diaz and E. Ros, "Background subtraction based on color and depth using active sensors", *Sensors*, vol. 13, no. 7, pp. 8895-8915, July 2013.
- [17] A. S. Glassner, *Graphics Gems*, Academic Press, Cambridge, MA, 1990.
- [18] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

---

**BIOGRAPHY**


---

**Beomsu Kim** (Member)

2007 : BS degree in School of Electronic Engineering, Soongsil University.

2009 : MS degree in School of Electronic

Engineering, Soongsil University.

2010~2012 : Associate Researcher, Korea Internet & Security Agency (KISA)

2012~ : PhD candidate, School of Electronic Engineering, Soongsil University.

**Tien-Dat Nguyen** (Student Member)

2012 : BS degree in Electronic and Telecommunication Engineering, Hanoi University of Science and Technology, Vietnam.

2013~2015 : Staff, FPT Corporation in Vietnam.

2015~ : MS degree candidate, School of Electronic Engineering, Soongsil University.

**Min-Cheol Hong** (Member)

1988 : BS degree in Electronic Engineering, Yonsei University.

1990 : MS degree in Electronic Engineering, Yonsei University.

1997 : PhD degree in Electrical Engineering, Northwestern University.

1990~1992 : Research Engineer, LG Electronics.

1997~1998 : Research Fellow, Northwestern University.

1998~2000 : Senior Research Engineer, LG Electronics.

2000~ : Professor, School of Electronic Engineering, Soongsil University.