논문 2017-54-8-8

추정된 시공간 배경 정보를 이용한 홀채움 방식

(Hole-Filling Method Using Extrapolated Spatio-temporal Background Information)

김 범 수*, 응웬띠엔닷*, 홍 민 철**

(Beomsu Kim, Tien Dat Nguyen, and Min-Cheol Hong[©])

요 의

본 논문에서는 추정된 시공간 배경 정보를 이용하여 자유 시점에서 합성영상을 생성하기 위한 홀채움 방식을 제안한다. 시간적 배경 정보를 추정하기 위해 비겹침 패치 기반의 배경 코드북을 이용한 새로운 시간적 배경 모델을 소개한다. 더불어, 공간적 배경 후보의 하한 및 상한 값의 제약 조건을 설정하는 깊이영상 기반 공간적 국부 배경 예측 방식에 대해 제안한다. 추정된 시간적 배경 정보와 공간적 배경 정보의 유사도를 비교하여 가려짐 배경 영역의 홀채움 과정을 수행한다. 또한 3-D 워핑후 발생하는 컬러영상과 깊이영상간의 불일치 문제를 해결하기 위해 깊이영상 기반의 고스트 제거 필터를 기술한다. 마지막으로 잔여 홀을 채우기 위해 새로운 깊이 항을 포함한 우선순위 함수를 이용하여 인페인팅 방식이 적용된다. 실험 결과를 통해기존의 홀채움 방식들과 비교하여 제안하는 방식의 객관적, 주관적 성능의 우수성을 확인할 수 있었다.

Abstract

This paper presents a hole-filling method using extrapolated spatio-temporal background information to obtain a synthesized view. A new temporal background model using non-overlapped patch based background codebook is introduced to extrapolate temporal background information In addition, a depth-map driven spatial local background estimation is addressed to define spatial background constraints that represent the lower and upper bounds of a background candidate. Background holes are filled by comparing the similarities between the temporal background information and the spatial background constraints. Additionally, a depth map-based ghost removal filter is described to solve the problem of the non-fit between a color image and the corresponding depth map of a virtual view after 3-D warping. Finally, an inpainting is applied to fill in the remaining holes with the priority function that includes a new depth term. The experimental results demonstrated that the proposed method led to results that promised subjective and objective improvement over the state-of-the-art methods.

Keywords: Hole-filling, Temporal background modeling, Local background estimation, Ghost removal, Inpainting

I. 서 론

자유 시점 렌더링 (Free-viewpoint Rendering)은 실 제 시점 영상으로부터 가상 시점 영상을 생성하여 영상 의 입체감을 향상시키기 위해 연구되어 왔다. 특히, 단

(School of Electronic Engineering, Soongsil University)

Received; April 25, 2017 Revised; May 19, 2017

Accepted; August 1, 2017

일 컬러 및 깊이영상을 이용하여 가상 시점에서의 합성 영상을 생성하는 깊이영상 기반 렌더링(Depth Image-Based Rendering)에 대한 연구가 활발히 진행되어 왔다^[1]. 그러나 DIBR은 가상 시점 영상을 생성하기 위해 사용되는 3-D 워핑(warping) 과정에서 부정확한 깊이 정보 및 가려짐 현상에 의해 홀 현상을 발생하는 문제점이 존재한다. 부정확한 깊이 정보에 의해 발생하는 홀은 작은 영역으로 나타나며, 이웃 정보를 이용하여효과적인 홀채움(hole-filling)이 가능하다. 반면에 가려짐 현상은 참조 시점에서는 전경(foreground) 객체들에의해 가려진 영역이 가상 시점에서는 보이는 현상으로서, 가상 시점에서 영상의 화질은 가려짐 영역을 효과

 ^{**} Corresponding Author (E-mail: mhong@ssu.ac.kr)
 ** 이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2017R1A2B4002205)

적으로 채움에 의해 결정된다[2~3].

가상 시점에서 합성 영상의 화질을 개선하기 위해 다양한 홀채움 방식들이 발표되었다. 깊이영상 기반 전처리 방식^[1, 4] 및 보간 방식^[5~6] 등은 연산량 측면에서 효율적이지만 홀 영역이 클수록 전경과 배경(background)의 경계 영역에서 기하학적 왜곡이나 과도한 열화 현상이 발생하는 문제점이 존재한다.

인페인팅(inpainting) 기법은 영상 내에서 알려지지 않은 영역과 이웃한 알려진 영역이 유사한 통계적 속성 또는 기하학적 구조를 공유한다는 특성을 이용하여 홀채움에 활용되어 왔다^[7~9]. 인페이팅 기법은 배경 영역과 전경 영역을 구분할 수 있는 깊이 정보와 결합하여 효과적으로 홀채움이 가능함이 입증되었으나^[10~13] 가려짐 영역에서 배경 영역과 전경 영역에 대한 정보가 제한적인 경우 홀채움 성능에 한계가 존재한다^[14~16]. 그러므로 가상 시점에서 만족스러운 합성영상을 생성하기위해서는 가려짐 영역에서 전경과 배경 영역 분리의 정확성이 매우 중요하다.

가려짐 영역에서의 전경과 배경 영역 분리를 위해 시 간적(temporal) 정보를 이용하여 홀채움 과정을 수행하 는 다양한 연구가 진행되었다. Mdiiki-Nva 등^[17]은 깊 이영상의 배경 스프라이트(sprite)를 이용하여 전역 임 계치 결정하고, 전역 임계치를 이용하여 배경과 전경 영역을 분리하는 방식을 제안하였다. 상기 방식은 분리 된 전경 및 배경 영역에 대한 홀채움을 수동적으로 선 택하여 인페인팅을 적용하기 때문에 인페인팅 순서에 따라 홀채움 성능의 편차가 큰 문제점을 갖고 있다. Koppel 등^[18]은 GOP(Group of Pictures)에서 연속된 프 레임(frame)들 간의 전역 움직임을 추정하고, 갱신된 시 간적 정보를 이용하여 인페인팅 순서를 결정하는 방식 을 소개하였다. 상기 방식은 디스플레이 순서와 홀채움 순서가 다르기 때문에 프레임 지연이 발생하고, 프레임 내에 존재하는 각 객체의 움직임이 상이한 경우 심각한 기하학적 왜곡이 발생하게 된다.

최근에는 배경 및 전역 영역 분리의 정확성을 향상시키기 위해 일관성 있는 시간적 배경 정보를 추정하여 홀채움 과정에 적용하는 방식들이 제안되었다. Xi 등^[19]은 깊이영상 기반 구조적 유사성(Structural Similarity)을 이용하여 연속된 가상 시점 영상들 간의 배경 영역을 추정하고, 배경 정보를 인페인팅 과정에 활용하는 방식을 제안하였다. 더불어, Yao 등^[20]은 깊이영상에서 배경 스프라이트를 추정하기 위해 가우시안 혼성 모델을 제안하였다. 상기 방식들은 이전 영상들에 존재하는

배경 영역들을 추정하는데 한계가 있었으며, 이로 인해 인페이팅을 이용하여 만족스러운 영상 생성에 한계가 존재한다.

배경 정보의 정확도를 향상시키기 위해서는 시공간 (spatio-temporal) 정보를 동시에 고려할 필요가 있다. 시간적 배경 정보를 추정하기 위해 텍스처(texture), 깊이영상 또는 프레임들 간의 유사도 등이 이용되어 왔다. 코드북(codebook)[21-22]은 텍스처 및 깊이 정보에 대응되는 코드워드(codeword)를 검출하여 시간적 유사정보를 추정하기 위해 활용되었다. 그러나 상기 코드북들은 배경 정보에 해당되는 코드워드를 선택하기 위해고정된 임계치를 사용하기 때문에 짧은 기간 동안 이전영상에 존재하는 배경 화소들을 효과적으로 추정하기에는 한계가 있다.

본 논문에서는 가상 시점에서 합성영상을 생성하기 위해 추정된 시공간 배경 정보를 이용하는 홀채움 방식 을 제안한다. 시간적 배경 정보를 추정하기 위해 비겹 침 패치(non-overlapped patch) 기반의 배경 코드북을 소개한다. 또한, 공간적 배경 정보를 추정하기 위해 깊 이영상의 국부 통계 특성을 이용한 배경 후보들의 제약 조건을 설정하는 방식에 대해 기술한다. 추정된 시간적 배경 정보와 공간적 배경 정보의 유사도를 비교하여 홀 영역에서의 전경 및 배경 영역을 구분하고 배경 영역의 홀들을 채우게 과정을 설명한다. 또한, 깊이 카메라의 저해상도로 인해 3-D 워핑 후에 발생하는 고스트 (ghost artifact) 현상을 제거하기 위해 깊이영상 기반 적응 필터 방식을 제안한다. 마지막으로 잔여 홀들을 채우기 위해 깊이 정보 항을 포함한 새로운 우선순위 함수를 정의하고, 인페이팅 과정에 적용하는 방식에 대 해 기술한다.

본 논문의 구성은 다음과 같다. Ⅱ절에서는 제안하는 시공간 배경 정보를 이용하는 홀채움 방식을 소개하고, Ⅲ절 및 IV절에서는 실험결과 및 결론을 기술한다.

Ⅱ. 제안하는 홀채움 방식

제안하는 홀채움 방식은 그림 1에 나타낸 바와 같이다섯 단계들로 구성된다: (1) 시간적 배경 정보를 추론하기 위한 시간 배경 모델, (2) 3-D 워핑 및 고스트 제거, (3) 깊이 정보의 국부 통계 특성을 이용한 공간적배경 정보 추론, (4) 시간적 배경 정보 및 공간적 배경 정보를 결합한 배경 영역의 홀채움, (5) 잔여 홀을 채우기 위한 깊이영상의 배경 정보를 포함한 우선순위 함수

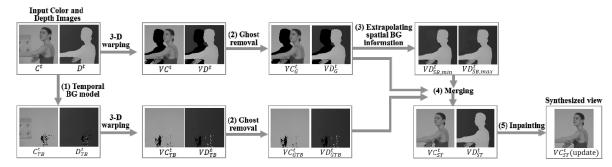


그림 1. 제안하는 홀채움 방식의 프레임워크

Fig. 1. Framework of the proposed hole-filling method.

기반 인페인팅 과정으로 구성된다. 각 단계별 상세 내용은 다음과 같다.

1. 시간적 배경 모델

연속된 프레임 간에는 시간적 상관관계가 존재하므로 현재 가상 시점 영상 내의 가려짐 영역 일부가 이전가상 시점 영상들에서 존재할 수 있다. 이와 같은 특성을 이용하여 전경 모델을 생성하기 위한 화소 단위의코드북 방식이 발표되었다^[22]. 그러나 홀 영역에는 많은배경 영역이 존재하므로 이전 프레임들로부터 배경 정보를 추정하고, 추정된 배경 정보를 현재 프레임의 배경 영역 채움에 이용할 필요가 있다. 기존 화소 단위의코드북은 많은 연산량을 필요로 하며, 전경 영역 추정을 목표로 하므로 배경 영역 추정에는 적합하지 않다.본 논문에서는 이와 같은 문제점을 해결하기 위해 비결침 패치 기반 배경 정보 코드북을 이용하는 시간적 배경 모델을 제안한다.

제안하는 시간적 배경 모델은 그림 2에 도시한 바와 같이 세 단계로 구성된다: (1) 비겹침 패치 기반 코드북 생성, (2) 패치 기반 배경 코드북 추정, (3) 시간적 배경 정보의 추정으로 구성된다.

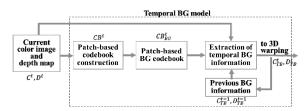


그림 2. 제안하는 시간적 배경 모델의 프레임워크 Fig. 2. Framework of the proposed temporal background model.

 $X \times Y$ 크기의 컬러 및 깊이영상으로 구성된 프레임을 $U \times U$ 크기의 비겹침 패치로 나누었을 때 i번째

비겹침 패치 벡터는 $P_i = [p_{i1}, p_{i2}, \, \cdots, p_{iU^2}]^T$ 로 표현되 $\mathbf{P}_i \mathbf{P}_i \mathbf{P}$

패치 P_i 에 대해 L개의 코드워드로 구성된 코드북을 $CB_i^t = \left\{c_i^l, c_i^l, \cdots, c_i^L\right\}$ 으로 표현할 때, 각 코드워드는 RGBD 성분을 포함한 $\overline{P}_i^l = \left[\overline{p}_{i1}^l, \overline{p}_{i2}^l, \cdots, \overline{p}_{iU^2}^l\right]^T$ 와 5개의 요소를 포함한 $AUX_i^l = \left[I_{i,\min}^l, I_{i,\max}^l, d_{i,\min}^l, d_{i,\max}^l, f_i^l\right]^T$ 로 구성된다. 코드북의 l번째 코드워드에서 \overline{P}_i^l 의 j번째 원소 $\overline{p}_{ij}^l = \left[\overline{r}_{ij}^l, \overline{g}_{ij}^l, \overline{b}_{ij}^l, \overline{d}_{ij}^l\right]^T$ 는 컬러 성분 및 깊이의 평균값으로 구성된 RGBD 벡터를 의미하며, AUX_i^l 에서 $I_{i,\min}^l$ 은 최소 휘도 벡터, $I_{i,\max}^l$ 는 최대 휘도 벡터, $d_{i,\min}^l$ 은 최소 깊이 벡터, f_i^l 는 코드워드의 발생 빈도를 나타낸다. AUX_i^l 의 각 벡터는 다음과 같이 표현된다.

$$\begin{split} I_{i,\,\text{min}}^{l} &= \left[I_{i1,\,\text{min}}^{l}\,,\,I_{i2,\,\text{min}}^{l}\,,\,\cdots,\,I_{iU^{2},\,\text{min}}^{l}\,\right]^{T}\!,\\ I_{i,\,\text{max}}^{l} &= \left[I_{i1,\,\text{max}}^{l}\,,\,I_{i2,\,\text{max}}^{l}\,,\,\cdots,\,I_{iU^{2},\,\text{max}}^{l}\,\right]^{T}\!,\\ d_{i,\,\text{min}}^{l} &= \left[d_{i1,\,\text{min}}^{l}\,,\,d_{i2,\,\text{min}}^{l}\,,\,\cdots,\,d_{iU^{2},\,\text{min}}^{l}\,\right]^{T}\!,\\ d_{i,\,\text{max}}^{l} &= \left[d_{i1,\,\text{max}}^{l}\,,\,d_{i2,\,\text{max}}^{l}\,,\,\cdots,\,d_{iU^{2},\,\text{max}}^{l}\,\right]^{T}\!. \end{split}$$

위에서 정의된 코드북 구성에 따라 패치 P_i 에 대한 코드북은 알고리즘 1과 같이 생성된다.

알고리즘 1의 step 2에서 컬러 왜곡 함수, 휘도 유사도 함수 및 깊이 유사도 함수를 이용하여 P_i 와 코드워드 c_i^l 와의 대응 여부를 결정하게 된다. 컬러 왜곡 함수는 P_i 와 c_i^l 를 구성하는 각 컬러 성분들의 유사도를 비교하기 위해 사용되며, 화소 단위의 코드북 생성 방식[21]과 유사하게 다음과 같이 정의된다.

알고리즘 1 코드북 생성

Step 1: P_i 의 코드북은 시간 t=0에서 $\mathit{CB}_i^t=\varnothing$ 와 L=0으로 초기화된다.

Step 2: 코드북 $C\!B_i^t = \{c_i^t|1 \le L\}$ 에서 P_i 에 대응되는 코드워드 c_i^t 를 다음과 같이 찾는다.

- (a) $T_{ij}^l = color distortion(p_{ij}, \overline{p}_{ij}^l),$
- (b) $T2_{ij}^{l} = brightness(I_{ij}, AUX_{i}^{l}),$
- (c) $T3_{ij}^l = disparity(d_{ij}, AUX_i^l),$

$$\text{(d)} \ \ c_i^l = \begin{cases} matched & \text{if } \left(T_i^l = \prod_{j=1}^{U \times U} T_{ij}^l = 1\right), \\ mismatched & otherwise \end{cases}$$

Step 3: 대응되는 코드워드가 없을 경우, $L\!\!\leftarrow\!\!L\!\!+\!1$ 그리고 새로운 코드워드 $c^{\!L}$ 가 다음과 같이 설정되어 생성된다.

$$\overline{P}_i^L \leftarrow P_i,$$

$$AUX_i^L \leftarrow \left[I_1, I_i, d_i, d_i, 1\right]^T.$$

$$I_i = \left[I_{i1}, \, \cdots, I_{i\,U^2}\right]^T \text{ and } d_i = \left[d_{i1}, \, \cdots, d_{i\,U^2}\right]^T$$
 반면, 대응되는 코드워드가 있는 경우, 대응되는 코드워드 c_i 는 다음과 강에 개시되다.

$$\begin{split} & \overline{P}_i^L \leftarrow & \left(\frac{f_i^l \overline{P}_i^l + P_i}{f_i^l + 1} \right), \\ & AUX_i^L \leftarrow & \left[\min_v(I_i, I_{i, \min}^l), \max_v(I_i, I_{i, \max}^l), \min_v(d_i, d_{i, \min}^l), \right. \\ & \left. \max_v(d_i, d_{i, \max}^l), f_i^l + 1 \right]^T. \end{split}$$

Step 4: 다음 프레임에서 step 2를 반복한다.

$$color distortion(p_{ij}, \overline{p}_{ij}^{l}) = \sqrt{\parallel p_{ij} \parallel^{2} - h^{2}},$$

$$h^{2} = \parallel p_{ij} \parallel^{2} \cos^{2}\theta = \frac{\langle p_{ij}, \overline{p}_{ij}^{l} \rangle}{\parallel \overline{p}_{ij}^{l} \parallel^{2}}.$$
(1)

식 (1)에서 $1 \leq j \leq U^2$ 이며, $\|\cdot\| \vdash l_2$ 놈(norm)을, $\left\langle p_{ij}, \overline{p}_{ij}^l \right\rangle$ 는 p_{ij} 와 \overline{p}_{ij}^l 의 내적(inner product)을 나타낸다. 프레임간의 휘도 변화를 반영하기 위해 P_i 와 c_i^l 간의휘도 유사도 함수를 다음과 같이 정의한다.

$$brightness(I_{ij}, AUX_{i}^{l}) = \begin{cases} 1 \text{ if } I_{low} \leq I_{ij} \leq I_{hi}, \\ 0 \text{ otherwise} \end{cases},$$

$$I_{low} = \alpha I_{ij, \max}^{l}, I_{hi} = \min \left(\beta \times I_{ij, \max}^{l}, \frac{I_{ij, \min}^{l}}{\alpha} \right).$$

$$(2)$$

위 식에서 $0<\alpha<1$ 이며 $\beta>1$ 인 상수 값을 의미한다. 프레임간의 휘도 변화는 깊이영상에도 영향을 미치므로 유사하게 깊이 유사도 함수를 정의한다.

$$disparity \left(d_{ij}, A U X_i^l\right) = \begin{cases} 1 & \text{if } d_{low} \leq d_{ij} \leq d_{hi}, \\ 0 & \text{otherwise} \end{cases}$$

$$d_{low} = \alpha d_{ij, \max}^l, d_{hi} = \min \left(\beta \times d_{ij, \max}^l, \frac{d_{ij, \min}^l}{\alpha}\right).$$

$$(3)$$

위와 같이 각 유사도 함수들을 사용하여 P_i 의 j번째 화소에 대응되는 코드워드 c_i^l 의 j번째 성분을 찾는 기준은 다음과 같이 정의된다.

$$T_{ij}^{l} = \begin{cases} 1 \text{ if } \left(\left(T \mathbf{1}_{ij}^{l} \le \epsilon_{1} \right) \wedge T \mathbf{2}_{ij}^{l} \wedge T \mathbf{3}_{ij}^{l} \right) = 1 \\ 0 \text{ otherwise} \end{cases}$$
 (4)

위 식에서 \wedge 는 부울 방식(Boolean)의 AND 연산을 의미하며 ϵ_1 은 양의 임계치를 나타낸다.

Step 3에서 \leftarrow 는 큐(queue)의 끝에 코드워드를 추가하는(push) 연산자를 나타내며, $\min_v(a,b)$ 와 $\max_v(a,b)$ 는 벡터 a와 b의 원소 단위에서의 최소 및 최대 연산자를 의미한다. 예를 들어, U^2 크기의 벡터 a와 b의 $\min_v(a,b)$ 는 다음과 같이 정의된다.

$$\min_{v}(a, b) = \{\min(a_1, b_1), \dots, \min(a_{U^2}, b_{U^2})\}.$$
 (5)

알고리즘 1을 통해 현재 프레임의 각 패치에 대한 코드북을 생성한다. 배경 영역의 깊이 값은 전경 영역의 깊이 값보다 작은 값을 갖는 특성을 이용하여 각 패치의 코드북을 구성하는 코드워드의 평균 깊이 값에 k-평균 군집(k-means clustering) 방식(k=2)을 적용하여다음과 같이 배경 코드북을 분리한다. 패치 P_i 의 코드북 CB_i^t 의 l번째 코드워드의 평균 깊이 값은 다음과 같이 결정될 수 있다.

$$\overline{d}_{i}^{l} = U^{-2} \sum_{j=1}^{U^{2}} \overline{d}_{ij}^{l} \text{ for } 1 \le l \le L.$$
 (6)

군집화 과정을 통해 분리된 두 개의 코드북 $CB_{i1}^t=\left\{c_i^m|1\leq m\leq M\right\}$ 과 $CB_{i2}^t=\left\{c_i^m|M+1\leq m\leq L\right\}$ 에 대해 배경 코드북은 다음과 같이 결정된다.

$$CB_{i,BG}^{t} = \begin{cases} CB_{i1}^{t} & \text{if } \max_{1 \leq m \leq M} \overline{d}_{i}^{m} \leq \max_{M+1 \leq m \leq L} \overline{d}_{i}^{m} \\ CB_{i2}^{t} & \text{otherwise.} \end{cases}$$

$$(7)$$

식 (4)에서 ϵ_1 을 작게 설정할수록 CB_i^t 의 코드워드 검색 조건이 엄격하게 되어 휘도 성분에 미세한 변화가 있는 패치는 전경 패치로 결정될 수 있다. 또한, 이전 프레임들에서 짧은 기간 동안 존재하는 정적 패치들이 배경 코드북에서 존재하지 않는 경우가 발생할 수 있다. 그러므로 완화된 임계치를 적용하여 패치가 배경

영역의 후보가 될 수 있는지 재검색할 필요가 있으며, 이전 프레임의 단기간에 존재하는 정적 전경 영역을 결 정할 필요가 있다. 위 특성을 반영한 현재 프레임의 시 간적 배경 정보는 알고리즘 2와 같이 결정된다.

알고리즘 2의 step 1은 패치 P_i 와 완화된 휘도 유사도를 갖는 배경 코드북 내에서 대응되는 코드워드를 찾는 과정이며, 대응되는 코드워드가 존재하는 경우, step2와 같이 P_i 영역에 대한 현재 프레임의 시간적 배경 정보를 결정한다. 배경 코드북에 대응되는 코드워드가 존재하지 않는 경우, 이전 프레임의 시간적 배경 깊이 정보와의 유사도를 계산하여 $T4_i^m=1$ 인 경우 이전 프레임의 배경 정보에 해당되는 것으로 간주하여 현재 프레임의 시간적 배경 정보를 이전 프레임의 시간적 배경 정보로 대체한다. $T4_i^m \neq 1$ 인 경우 P_i 는 전경 영역으로 간주한다.

알고리즘 2 시간적 배경 정보 추정

Step 1: P_i 에 대해, 다음을 만족하는 배경 코드북 $(C\!\!P_{i,BG}^t)$ 내에서 최적으로 대응되는 코드워드 c_i^n 를 찾는다.

- (a) $T_{ij}^m = color distortion(p_{ij}, p_{ij}^m),$
- (b) $T2_{ij}^m = brightness(I_{ij}, AUX_i^m),$
- (c) $T_{ij}^m = disparity(d_{ij}, AUX_i^m),$

$$\begin{array}{ll} \text{(d)} & T_{ij}^m = \begin{cases} 1 & \text{if} \left(\left(T_{ij}^m \leq \epsilon_2 \right) \wedge T\!\!2_{ij}^m \wedge T\!\!3_{ij}^m \right) \!\! = \! 1, \\ 0 & otherwise \\ & where & \epsilon_1 < \epsilon_1 \end{cases}$$

(e)
$$T_i^m = \prod_{j=1}^{U^2} T_{ij}^m$$

Step 2: $T_i^m = 1$ 일 경우, P_i 는 배경 패치로 간주하여 시간적 배경 정보는 다음과 같이 결정된다.

$$C_{TR}^{t}(i) = C_{i}, D_{TR}^{t}(i) = D_{i}$$

 C_i 와 D_i 는 각각 P_i 의 컬러와 깊이 요소들을 나타낸다.

 $T_i^n \neq 1$ 경우, 다음 과정을 수행한다.

(a) $T4_{ij}^{m} = disparity \left(d_{ij}^{t-1}, AUX_{i}^{m}\right)$,

(b)
$$T4_i^m = \prod_{j=1}^{U^2} T4_{ij}^m$$
, where $d_{ij}^{t-1} \in D_{TB}^{t-1}$.

Step 3: $T_i^m=1$ 인 경우 P_i 영역의 현재 프레임의 시간적 배경 정보는 이전 프레임의 시간적 배경 정보로 대체된다.

$$C_{TB}^{t}(i) = C_{TB}^{t-1}(i), \ D_{TB}^{t}(i) = D_{TB}^{t-1}(i).$$

$$C_{TB}^{t}(i) = [0, \dots, 0]^{T}, D_{TB}^{t}(i) = [0, \dots, 0]^{T}$$

그림 3에 추정된 시간적 배경 정보의 성능 비교를 나타내었다. 기존 코드북 방식은 그림 3(a)에 나타난 바와같이 이전 프레임에서 단기간 동안 존재하는 정적 전경 영역을 코드북에 수용하지 못함으로서 배경 정보 추정에 한계가 존재한다. 반면에 제안하는 배경 모델은 그



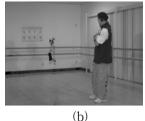


그림 3. 시간적 배경 정보 추정의 성능 비교: (a) [22]의 결과 영상, (b) 제안하는 시간적 배경 모델의 결 과(검은색 영역은 전경 영역을 나타냄)

Fig. 3. Performance comparison of extrapolating temporal background Information: (a) result in [22], (b) result of the proposed temporal background model (foreground marked in black).

림 3(b)의 결과와 같이 배경 영역 및 정적 전경 영역을 배경 코드북에 반영함으로서 효과적으로 시간적 배경 정보를 추정할 수 있음을 알 수 있다. 그림 3(b)에서 검은색으로 표현된 부분은 시간적 전경 영역으로 판별된 화소들을 나타낸다. 상기 결과를 통해 텍스처-깊이에 적응적인 패치 기반 배경 코드북 방식을 이용하여 배경 및 정적 전경 정보를 효과적인 수집 할 수 있음을 확인할 수 있다. 시간적 배경 컬러영상(C^t_{TB})과 깊이영상(D^t_{TB})은 3-D 워핑 과정을 통해 시간적 가상 배경 컬러 영상(VC^t_{TB})와 가상의 시간적 배경 깊이영상(VD^t_{TB})을 생성한다.

2. 깊이영상 기반 고스트 제거

깊이영상은 컬러영상보다 상대적으로 저해상도인 관계로 3-D 워핑 영상에서 깊이영상과 컬러 영상의 화소들 간에 불일치가 발생할 수 있으며, 이로 인해 전경 영역의 화소들이 배경 영역에서 나타날 수 있다. 이와 같은 고스트 현상은 시각적으로 불편함을 초래하며, 합성된 가상 시점 영상들의 시간적 일관성을 유지하지 못하는 원인이 된다. 고스트 현상을 제거하기 위한 기존 방식들은 홀 영역 경계의 두 화소들에 완화(smoothing) 필터를 적용하였으나, 전경 및 배경 경계 영역 화소들이 동시에 제거되는 문제점이 존재한다^[12]. 본 절에서는 전경 화소들을 기하학적 구조를 유지하면서 배경 영역에 존재하는 고스트 현상만을 제거하는 방식에 대해 소개한다.

가려짐 영역의 배경 및 전경 경계를 분리하기 위해 그림 4(a)의 가상 시점에서 배경 컬러 또는 깊이영상에 대해 다음과 같이 이진 영상을 정의한다.







그림 4. (a) 가상 시점에서 컬러 영상, (b) 이진 영상, (c) 경계 마스크

Fig. 4. (a) Virtual color image, (b) binary image, (c) boundary mask.

$$BI(\widehat{m}) = \begin{cases} 0 & \text{if } VD^{t}(\widehat{m}) = 'hole' \\ 1 & otherwise \end{cases}$$
 (8)

식 (8)에서 \hat{m} 은 2차원 영상 획득 시스템에서 화소의 위치를 나타낸다. 그림 4(b)는 그림 4(a)의 이진 영상을 나타내며, 이진 영상에 2차원 라플라시안(Laplacian) 연산자를 적용하여 그림 4(c)와 같이 경계 마스크(BM)를 생성한다.

알고리즘 3 분할(segmentation)과 레이블링

Step 1: $X \times Y$ 크기의 입력 영상 IM에 대해, 레이블 순번을 k=0으로, IM의 화소 위치를 $z_c=0$ 으로 설정한다.

Step 2: $k = k+1, z_c = z_c+1.$

Step 3: &번째로 레이블링된 집합의 화소 위치가 저장된 큐(queue) 와 임시 큐를 다음과 같이 초기화된다.

 $\phi_k = \varnothing$, $Z = \varnothing$

Step 4: z_c 가 IM의 마지막 위치인 경우 step 9로 이동하고, $IM(z_c)=0$ 인 경우 $z_c=z_c+1$ 로 변경하고 step 4를 반복한다.

Step 5: $Z \leftarrow z_c$ 로 z_c 를 임시 큐에 저장한다.

Step 6: $Z \rightarrow z_l$ 과 같이, 임시 큐의 마지막 원소를 인출한다. $\phi_k \leftarrow z_l$, $IM(z_l) = 0$.

Step 7: 다음과 같이, $IM\!(z_l)$ 의 근접한 8개의 화소들을 확인한다.

 $IM(z_l+1)\neq 0, Z\leftarrow(z_l+1).$

 $IM(z_l-1)\neq 0, Z\leftarrow(z_l-1).$

 $IM(z_l + Y) \neq 0, Z \leftarrow (z_l + Y).$

 $IM(z_l - Y) \neq 0, Z \leftarrow (z_l - Y).$

 $IM(z_l + Y+1) \neq 0, Z \leftarrow (z_l + Y+1).$

 $IM(z_l + Y - 1) \neq 0, Z \leftarrow (z_l + Y - 1).$

 $IM(z_1 - Y + 1) \neq 0, Z \leftarrow (z_1 - Y + 1).$

 $IM(z_1-Y-1)\neq 0, Z\leftarrow(z_1-Y-1).$

 $Z=\emptyset$ 가 될 때까지 **step 6**을 반복한다.

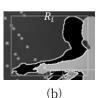
Step 8: $\phi \leftarrow \phi_k$, IM의 모든 화소들이 확인될 때까지 step 2를 반복한다. Step 9: ϕ 를 반환한다.

본 논문에서는 경계 마스크에 알고리

본 논문에서는 경계 마스크에 알고리즘 3에 기술된 flood-fill 방식^[28]을 적용하여 각 가려짐 영역을 레이블링(labeling)하고 전경 및 배경 영역의 경계 화소들에 대한 위치 정보를 다음과 같이 얻게 된다.

$$\phi = 알고리즘 3(BM). \tag{9}$$





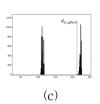


그림 5. (a) 알고리즘 3을 이용한 가려짐 영역 레이블링 (노란색: 가려짐 경계 영역, 빨간색: 경계 영역을 포함한 최소 사각 영역), (b) 최소 사각 영역 예, (c) 자녀 사각 영역의 히스토그램 예

Fig. 5. (a) Labeling disocclusions using Algorithm 3 (yellow: boundary of disocclusion, red: smallest rectangle containing boundary), (b) example of smallest rectangle, (c) example of histogram of child rectangle.

경계 화소들의 위치 정보를 깊이영상에 적용하여 다음과 같이 전경 및 배경 경계 화소들로 분리하게 된다, 그림 5(a)와 같이 레이블링된(노란색으로 표기) 가려짐 경계 영역을 포함한 최소 사각 영역 (빨간색으로 표기)을 정의한다. 그림 5(b)에 나타낸 바와 같이 i번째 레이블링 영역 ϕ_i 를 포함하는 $M \times N$ 크기의 사각영역(R_i)을 비겹침 자녀 사각 영역들(Child Rectangle)로 다음과 같이 분리한다.

$$R_i = \bigcup_{j=1}^{\lceil N/Z \rceil} CR_{ij}. \tag{10}$$

위 식에서 $\lceil N/Z \rceil$ 는 N/Z보다 크거나 같은 가장 작은 정수 값을 의미하며, $M \times Z$ 크기의 CR_{ij} 는 R_i 의 j번째 자녀 사각 영역을 나타낸다. 각 자녀 사각 영역에 k-평균 군집 방식(k=2)을 적용하여 자녀 사각 영역 내의 깊이 정보를 다음과 같이 두 개의 군집 영역으로 분리하다.

$$CR_{ij, low} \cup CR_{ij, hi} = CR_{ij},$$

 $CR_{ij, low} \cap CR_{ij, hi} = \varnothing.$ (11)

식 (11)에서 $CR_{ij,low}$ 와 $CR_{ij,hi}$ 는 CR_{ij} 의 낮은 깊이 군집(low-depth cluster)과 높은 깊이 군집(high-depth cluster)을 나타낸다. 전경 영역 및 배경 영역의 분리를 위한 임계치는 그림 5(c)와 같이 높은 깊이 군집의 최소 깊이 값으로 설정된다.

$$d_{ij,gh} = \min_{\hat{n} \in CR_{ij,k}} d(\hat{n}). \tag{12}$$

 $d_{ij,gh}$ 는 CR_{ij} 내의 배경 경계 영역에서 고스트 현상을 제거하기 위한 임계치로 사용되며, 고스트 현상을 제거

한 컬러 및 깊이영상은 다음과 같이 결정된다.

$$VD_{G}^{t}(\widehat{m}) = \begin{cases} VD^{t} & \text{if } VD^{t}(\widehat{m}) \geq d_{ij,gh} \\ 0 & otherwise, \end{cases}$$

$$VC_{G}^{t}(\widehat{m}) = \begin{cases} VC^{t} & \text{if } VD^{t}(\widehat{m}) \geq d_{ij,gh} \\ 0 & otherwise. \end{cases}$$

$$(13)$$

위 식에서 $\hat{m} \in CR_{ij,hi}$ 인 화소의 2차원 위치를 나타내며, 컬러 영상과 깊이영상간의 일관성을 유지하기 위해동일한 기준을 적용한다.

그림 6에 [12]의 팽창(dilation) 방식과 제안하는 고스트 제거 방식의 성능 비교를 나타내었다. 팽창 방식은 가려짐 영역에서 배경 및 전경의 화소들을 동시에 제거하므로 객체 경계면 성분의 손실을 발생한다. 반면에 제안하는 고스트 제거 필터는 객체 경계를 유지하면서 배경 영역의 고스트 현상을 효과적으로 제거하였음을 확인할 수 있다.

 Π -1절에서 기술한 시간적 배경 컬러 영상 (VC^t_{TB}) 및 깊이영상 (VD^t_{TB}) 도 동일한 방식으로 고스트 현상을 제거한다.







그림 6. 고스트 현상 제거 성능 비교: (a) [12] 방식을 이용한 합성영상, (b) 제안 방식을 이용한 합성영상, (c) 제안 방식을 이용한 고스트 현상 제거전후 비교 (녹색: 제거된 배경 화소, 청색: 보존된 전경 화소)

Fig. 6. Performance comparisons of ghost removal: (a) synthesized view with [12], (b) synthesized view with proposed ghost removal, (c) difference between before and after removal of ghost artifacts (green: removed background pixels, blue: preserved foreground pixels).

3. 공간적 배경 정보 추정

본 절에서는 각 가려짐 영역으로부터 공간적 배경 정보를 추정하기 위한 깊이영상 기반 국부 배경 추정 방식을 기술한다. 상기 과정을 통해 가려짐 영역에서 공간적 배경 후보 화소들의 깊이 정보의 하한 및 상한 제약 조건을 설정한다.

고스트 현상을 제거한 가상 시점 깊이영상(VD_G^t)의 각 가려짐 영역으로부터 공간적 배경 영역들을 구분해

내기 위해 다음과 같이 이진 마스크를 정의한다.

$$BM_{G}(\widehat{m}) = \begin{cases} 1 & \text{if } VD_{G}^{t}(\widehat{m}) = 'hole' \\ 0 & otherwise. \end{cases}$$
 (14)

이진 마스크 BM_G 에 알고리즘 3을 적용하여 분할 및 레이블링된 각 가려짐 영역에 대한 위치 정보를 다음과 같이 결정한다.

$$\Omega = Algorithm3(BM_G). \tag{15}$$

알고리즘 3을 통해 그림 7(a)와 같이 각 가려짐 영역에 대한 레이블링 및 위치 정보를 결정하고, 레이블링된 각 가려짐 영역을 포함한 최소 크기의 사각 영역(빨간색으로 표기)들을 정의한다. 국부적인 배경 깊이 값을 추정하기 위해 각 최소 사각 영역을 자녀 사각 영역으로 세분화한다. i번째로 레이블링된 사각 영역 (R_i) 의j번째 자녀 사각 영역인 CR_{ij} 에 k-평균 군집 방식을 적용하여 그림 7(b)와 같이 낮은 깊이 군집 $(CR_{ij,low})$ 및 높은 깊이 군집 $(CR_{ij,hi})$ 영역으로 분리한다. k-평균 군집 방식에서 자녀 사각 영역 내의 국부 정보를 활용하기 위해 자녀 사각 영역내의 홀이 아닌 화소들의 깊이 정보를 적용한다. 분리된 낮은 깊이 군집의 깊이 정보의 최대 및 최소값은 가려짐 영역내의 배경 영역을

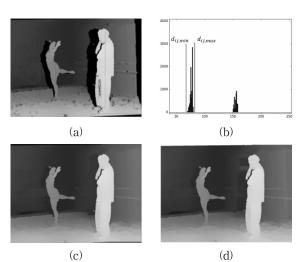


그림 7. 국부 배경 추정의 사례: (a) 고립 영역을 레이블 링한 가상의 깊이영상, (b) 자녀 사각 영역의 히 스토그램, (c) 최소값 경계에 의해 채워진 가상 의 깊이영상, (d) 최대값 경계에 의해 채워진 가 상의 깊이영상

Fig. 7. Example of local background estimation: (a) virtual depth map with labeled disocclusion, (b) histogram of a child rectangle, (c) virtual depth map filled by the lower bound, (d) virtual depth map filled by the upper bound.

결정하기 위한 제약 조건으로 활용된다. 낮은 깊이 군집 의 최대 및 최소값은 다음과 같이 결정된다.

$$d_{ij,\max} = \min_{\hat{n} \in CR_{ij,low}} d(\hat{n}),$$

$$d_{ij,\min} = \min_{\hat{n} \in CR_{ij,low}} d(\hat{n}).$$
(16)

식 (16)의 제약 조건들에 의해 추정되는 가상 시점에서의 공간적 배경 깊이영상은 다음과 같이 결정된다.

$$\begin{split} VD_{SB,\,\text{max}}^t\left(\widehat{m}\right) &= \begin{cases} d_{ij,\,\text{max}} & \text{if } BM_G(\widehat{m}) = 1 \\ VD_G^t\left(\widehat{m}\right) & otherwise, \end{cases} \\ VC_{SB,\,\text{min}}^t\left(\widehat{m}\right) &= \begin{cases} d_{ij,\,\text{min}} & \text{if } BM_G(\widehat{m}) = 1 \\ VD_G^t\left(\widehat{m}\right) & otherwise. \end{cases} \end{split} \tag{17}$$

위 식에서 \hat{m} \in $CR_{mn,\,low}$ 은 2차원 화소의 위치를 나타 낸다.

그림 7(c) 및 (d)는 깊이 정보의 국부 통계 특성을 이용한 공간적 배경 깊이 정보의 하한 및 상한 값을 나타내는 결과 영상을 나타낸다.

4. 시공간 배경 정보를 이용한 홀채움

시공간 배경 정보는 앞 절들에서 기술한 시간적 배경 깊이영상(VD_{GTB}^t) 및 공간적 배경 깊이영상들 ($VD_{SB,\,\mathrm{max}}^t$ 및 $VD_{SB,\,\mathrm{min}}^t$)의 병합 과정을 통해 얻게된다. 시간 및 공간적 배경 정보의 병합은 시간 및 공간적 배경 깊이 정보의 유사도를 비교하여 다음과 같이결정한다.

$$VD_{ST}^{t}(\widehat{m}) = \begin{cases} VD_{GTB}^{t}(\widehat{m}) & \text{if } A \leq VD_{GTB}^{t}(\widehat{m}) \leq B \\ VD_{G}^{t}(\widehat{m}) & \text{otherwise.} \end{cases}$$
(18)

위 식에서 $A = VD_{SB,\,\mathrm{min}}^t\left(\hat{m}\right)$ 및 $B = VD_{SB,\,\mathrm{max}}^t\left(\hat{m}\right)$ 로 정의되며, 깊이영상과 컬러 영상간의 일관성을 유지하기 위해 시공간 배경 컬러 영상은 결정된다.

$$VC_{ST}^{t}(\widehat{m}) = \begin{cases} VC_{GTB}^{t}(\widehat{m}) & \text{if } A \leq VD_{GTB}^{t}(\widehat{m}) \leq B \\ VC_{G}^{t}(\widehat{m}) & otherwise. \end{cases}$$
(19)

5. 새로운 우선순위 함수 기반 인페인팅

시공간 배경 정보를 이용하여 합성된 가상 시점 영상의 잔여 홀들은 인페인팅 방식을 이용하여 채우게 된다. 인페인팅 기반 홀채움의 결과는 홀채움 우선순위에 매우 민감한 특성을 갖고 있다. 홀채움 순위를 결정하

기 위해 그림 $10(\mathbf{a})$ 에 나타낸 바와 같이 VC_{ST}^t 의 각 가려짐 경계 영역 화소를 중심으로 $M\times N$ 크기의 패치 (Ψ_n) 를 설정하여 다음과 같이 우선순위를 결정한다.

$$Pri (p) = C(p)D(p)Z(d).$$
(20)

위 식에서 $p = VC_{ST}^t(\hat{m})$ 는 가려짐 경계 영역의 컬러 화소를 의미하며, d는 p에 대응되는 $VD_{SB,\max}^t$ 값을 의미한다. 식 (20)의 우선순위 함수에서 C(p)는 신뢰도 항, D(p)는 데이터 항, Z(p)는 깊이 항을 나타낸다. 신뢰도 항과 데이터 항은 기존 인페인팅 기반 홀채움 방식들에서 정의된 항들이며, 본 논문에서는 깊이 항을 추가하여 우선순위를 결정한다. 신뢰도 항은 다음과 같이 패치 내의 홀이 아닌 영역의 비율로 정의된다.

$$C(p) = \frac{|\phi_p|}{|\Psi_p|}.$$
 (21)

위 식에서 $|\phi_p|$ 와 $|\Psi_p|$ 는 ϕ_p 와 Ψ_p 의 면적을 나타낸다. 신뢰도 항은 가려짐이 있는 패치 내에서 홀이 아닌 영역의 비율이 높을수록 우선순위가 크게 된다. 더불어, 데이터 항은 다음과 같이 정의된다.

$$D(p) = \frac{\left\langle \nabla F_p^{\perp}, n_p \right\rangle}{\lambda}.$$
 (22)

식 (22)에서 n_p 는 $\delta\Omega$ 의 법선 단위 벡터를 의미하고, ∇F_p^\perp 는 Ψ_p 의 중심과 동일한 등광도선(isophote)을 의미한다. 그러므로 ∇F_p^\perp 와 법선 벡터 n_p 가 같은 방향일 때 D(p)가 가장 큰 값을 갖게 된다. 더불어, λ 는 정규화 인자(normalization factor)를 나타내며, 일반적으로 n-bit 화소에 대해 $\lambda=(2^n-1)$ 로 설정된다. 정적 영역 내에 존재하는 가려짐 패치는 일정 검색 영역 내에 유사한 패치가 존재할 가능성이 높으므로 홀채움을 배경 패치로부터 시작하는 것이 효과적이다. 이와 같은 특성을 우선순위 함수에 반영하기 위해 깊이 항을 다음과 같이 정의한다.

$$Z(d) = \exp\left(-\left|\Psi_{p, \max} - d\right|^{2}\right),$$

$$\Psi_{p, \max} = \max_{\hat{n} \in \Psi_{d}} VD_{SB, \max}^{t}(\hat{n}),$$

$$subject \ to \ VD_{ST}^{t}(\hat{n}) = 'hole'.$$
(23)

위 식에서 Ψ_d 는 Ψ_p 에 대응되는 깊이 정보 d를 중심으로 하는 패치를 의미한다.

상기 우선순위에 따라 가려짐 경계 영역 화소에 대한

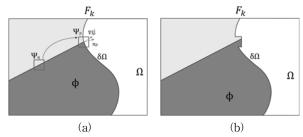


그림 8. 인페인팅 적용 예: (a) 레이블링된 가려짐 영역 (ϕ) : 홀이 아닌 영역, Ω : 홀 영역, $\delta\Omega$: 고립 영역의 경계면), (b) 인페인팅 방식 적용 결과

Fig. 8. Example of inpainting: (a) labeled disocclustion $(\phi \colon \text{non-holes}, \ \varOmega \colon \text{holes}, \ \delta \varOmega \colon \text{boundary of disocclution})$, (b) result after inpainting method.

인페인팅 과정을 수행한다. $p \in VC_{ST}^t$ 가 가장 높은 우선순위를 갖는 경우, 그림 $8(\mathbf{a})$ 와 같이 p를 중심으로 하는 패치 Ψ_p 를 비홀 영역 $(\phi_p = [p_1, \cdots, p_k]^T)$ 과 홀 영역 $(\Omega_p = [p_{k+1}, \cdots, p_{MN}]^T)$ 으로 분리하고, 검색 영역 (S_p) 내의 화소 q'를 중심으로 하는 패치 $\Psi_{q'}$ 를 ϕ_p 및 Ω_p 에 대응되도록 $\Phi_{q'} = [q_1, \cdots, q_k]^T$ 및 $\Omega_{q'} = [q_{k+1}, \cdots, q_{MN}]^T$ 를 설정한다. 검색 영역 내에서 ϕ_p 와 가장 유사한 영역을 다음과 같이 결정한다.

$$\phi_q = \arg\min_{\phi_{q'} \in S_n} \| \phi_p - \phi_{q'} \|^2. \tag{24}$$

식 (24)의 결과에 따라 그림 8(b)와 같이 \varPsi_p 의 홀 영역 Ω_p 를 ϕ_q 에 대응되는 홀 영역 Ω_q 로 채우게 된다. 상기와 같이 우선순위 결정과 인페인팅 과정을 이용하여 모든 잔여 홀들의 채움 과정을 순차적으로 수행한다.

Ⅲ. 실험 결과

제안하는 시공간 배경 정보를 이용한 홀채움 방식을 다양한 3-D 동영상을 이용하여 실험하였다. 본 논문에서는 "Ballet", "Break Dancers", "Book Arrival", "Lovebird1", "Street" 및 "Car Park"에 대해 기술한다. 제안하는 홀채움 방식은 시공간 정보를 이용한 Koppel^[18], Xi^[19], Yao^[20] 방식 및 공간적 정보를 이용한 Muddala^[16] 방식과 성능 비교를 하였다. 모든 홀채움 방식에서 가상 시점 합성영상을 획득하기 위해 [10]의 3-D 워핑 방식을 사용하였다. 성능 비교를 위해 홀 영역에 속한 화소들에 대한 PSNR (Peak Signal to Noise Ratio) 및 SSIM(Structural Similarity)^[24]을 이용하였다. 연산량 비교를 위해 인텔 Core i7-4790K CPU 4.0 GHz

및 16 GB 메모리를 사용하였으며, 각 방식들은 리눅스 환경에서 Open Multi Processing (OpenMP)으로 시뮬 레이션 하였다.

제안하는 시간적 배경 모델은 여러 매개변수를 사용 하였다. 비겹침 패치 기반 배경 코드북의 연산량은 패 치 크기가 커질수록 연산량은 작아지게 된다. 그러나 패치 크기가 커질수록 시간적 배경 영역에 대한 추정의 정확도는 낮아지게 된다. 실험을 통해 패치 크기가 3×3 또는 5×5 인 경우에 만족스러운 결과를 얻을 수 있음을 확인하였으며, 본 논문에서는 3×3 패치 크기 에 대한 결과를 기술한다. 또한 알고리즘 1 및 2의 밝기 유사도 함수에서 사용한 ϵ_1 및 ϵ_2 의 값이 작아질수록 이전 영상에 존재하는 시간적 배경 정보를 수집할 가능 성이 낮아지게 된다. 반면에 상기 임계치들이 커질수록 전경 영역에 속한 패치들이 배경 영역으로 추정되어 합 성영상의 기하학적 왜곡 현상을 증가시키게 된다. 실험 을 통해 $10 \le \epsilon_1 \le 20$ 및 $16 \le \epsilon_2 \le 26$ 에서 만족스 러운 결과를 얻을 수 있었으며, 본 논문에서는 $\epsilon_1 = 10$ 및 $\epsilon_2 = 16$ 를 사용하였다. 또한 밝기 유사도 함수에서 정의된 상수 α 및 β 는 $0.7 \le \alpha \le 0.8$ 및 $1.2 \le \beta \le 1.5$ 에서 만족스러운 성능을 나타내었으며, 본 실험에서는 $\alpha = 0.75$ 및 $\beta = 1.3$ 을 사용하였다.

II-2절 및 II-4절에서 가려짐 영역의 경계 영역 및 국부 배경 제약 조건을 설정하기 위해 가려짐 영역을 포함한 최소 부모 사각형을 자녀 사각형으로 나누었다. 본 실험에서는 자녀 사각형의 높이는 16을 사용하였으며, 폭은 부모 사각형과 동일하게 설정하였다.

인페이팅 기법을 이용하는 합성영상의 성능은 패치의 크기 및 유사 패치를 찾기 위한 검색 영역 크기에 의존적이다. 본 실험에서는 홀 영역의 크기가 50개 화소 이하인 경우 패치 크기 및 검색 영역을 3×3 및 20×20 을 사용하였으며, 50개에서 100개의 화소로 구성된 홀 영역은 5×5 크기의 패치 및 60×60 검색 영역을 이용하였다. 100개 이상의 화소로 구성된 홀 영역에 대해서는 9×9 크기의 패치 및 80×200 의 검색 영역을 적용하였다.

실험에 사용한 각 3-D 동영상 100 프레임에 대한 평균 PSNR 및 SSIM 성능 비교를 표 1에 나타내었다. Koppel 및 Yao 방식은 전경 영역에 대해 인페인팅 우선순위를 높게 설정하는 관계로 성능의 한계가 존재한다. Xi 방식은 배경 영역에 존재하는 홀 영역이 클수록 Koppel 및 Yao 방식보다 성능 향상이 있었으나, 가려

표 1. PSNR 및 SSIM 성능 비교

Table 1. PSNR and SSIM comparisons.

Table 1. 1 ON 1 and Comparisons.											
Video	Camera	PSNR (dB)					SSIM				
Seq.		Koppel	Xi	Yao	Muddala	proposed	Koppel	Xi	Yao	Muddala	Proposed
Ballet	$3 \rightarrow 4$	21.94	24.49	21.30	23.83	25.43	0.7625	0.7978	0.7423	0.7853	0.8176
	$4 \rightarrow 5$	21.57	23.42	20.45	22.98	23.99	0.7476	0.7746	0.7277	0.7449	0.7723
Break	$3 \rightarrow 4$	28.60	28.47	27.89	28.39	28.89	0.7791	0.7799	0.7740	0.7772	0.7824
Dancers	$4 \rightarrow 5$	30.97	30.67	30.11	31.34	31.58	0.7977	0.7974	0.7941	0.7986	0.8010
Book	$7 \rightarrow 8$	18.46	15.77	18.96	19.38	21.84	0.8042	0.7604	0.7515	0.7417	0.8133
Arrival	$9 \rightarrow 8$	19.50	19.68	20.03	20.38	21.01	0.8263	0.8245	0.8336	0.8378	0.8540
Lovebird	$4 \rightarrow 6$	17.25	17.27	17.18	17.63	18.02	0.8845	0.8876	0.8873	0.8905	0.9039
1	$6 \rightarrow 8$	17.11	16.94	17.12	15.53	19.83	0.8537	0.8333	0.8402	0.8022	0.8830
Car	$3 \rightarrow 4$	22.21	22.39	21.94	23.71	24.14	0.9045	0.9065	0.8969	0.9122	0.9216
Park	$4 \rightarrow 5$	21.46	21.75	21.24	23.19	23.53	0.9010	0.9014	0.8935	0.9087	0.9169
Street	$3 \rightarrow 4$	25.25	25.53	24.20	25.89	26.27	0.9649	0.9660	0.9565	0.9671	0.9707
	$4 \rightarrow 5$	25.03	25.37	24.53	25.76	26.26	0.9648	0.9662	0.9579	0.9651	0.9699
Average		22.45	22.65	22.08	23.17	24.23	0.8477	0.8476	0.8380	0.8443	0.8672

표 2. 프레임당 연산량 비교 (단위: 초)

Table2. Processing time comparisons per frame. (unit: sec.)

Spatial Resolution	Video Seq.	Camera	Koppel	Xi	Yao	Muddala	Proposed
	Dellet	$3 \rightarrow 4$	12.79	38.84	10.16	20.75	7.12
	Ballet	4 → 5	10.90	36.37	12.13	17.95	7.54
1024×768	Break Dancers	$3 \rightarrow 4$	9.72	29.76	5.74	9.59	4.37
	Dreak Dancers	$4 \rightarrow 5$	10.06	30.64	6.88	10.61	6.73
	Book Arrival	7 → 8	9.84	28.35	7.01	8.30	5.43
	DOOK AITIVAL	9 → 8	11.49	29.29	4.68	9.37	5.54
	Lovebirdl	$4 \rightarrow 6$	13.32	23.82	3.59	7.46	5.11
	Lovebirdi	6 → 8	9.34	21.84	2.64	4.15	2.78
	10.93	21.84	6.60	11.02	5.58		
1090×1000	Car Park	$3 \rightarrow 4$	34.46	70.52	17.20	60.42	9.07
	Cai Faik	$4 \rightarrow 5$	32.58	69.55	27.88	64.02	9.16
1920×1088	Ctroat	$3 \rightarrow 4$	40.98	71.86	13.31	50.74	6.32
	Street	4 → 5	37.68	71.52	12.98	47.19	6.25
	36.43	70.86	17.84	55.59	7.70		

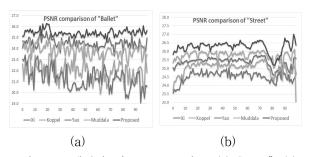


그림 9. 프레임에 따른 PSNR 비교: (a) "Ballet", (b) "Street"

Fig. 9. PSNR comparisons as s function of frame number: (a) "Ballet", (b) "Street".

집 영역의 형태에 따라 성능의 편차가 큰 단점이 있다. Muddala 방식은 Xi 방식과 비교하여 합성영상 결과가 가려짐 형태에 덜 민감하지만 1차 인페인팅에서 부정확하게 채워진 영역이 2차 인페인팅 과정에서 다른 홀 영역에 부정적인 영향을 미치게 된다. 반면에 제안하는 홀채움 방식은 가려짐 형태에 무관하고 일관성 있게 기존 방식보다 개선된 성능을 나타내고 있음을 확인할 수 있었다.

그림 9에 "Ballet" 및 "Street" 영상에 대해 프레임에 따른 PSNR 성능 비교를 나타내었다. 기존 방식들

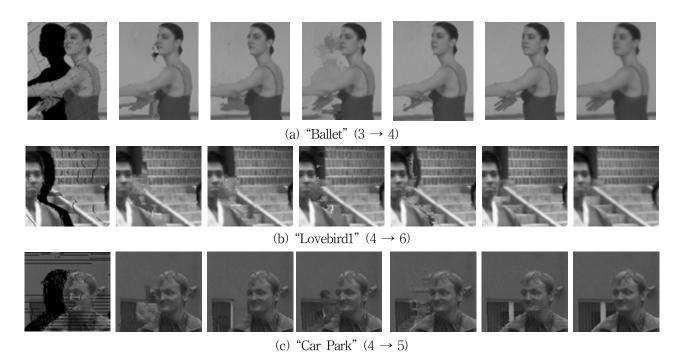


그림 10. 확대된 가상 시점 합성영상 화질 비교 (왼쪽에서 오른쪽으로): 워핑 영상, Koppel 방식, Xi 방식, Yao 방식, Muddala 방식, 제안 방식 및 원 영상

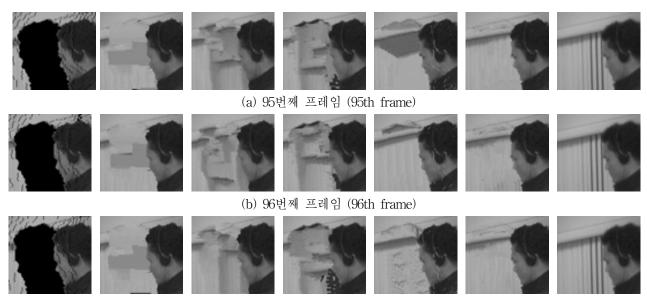
Fig. 10. Visual comparisons of zoomed synthesized image (from left to right): warping image, Koppel's method, Xi's method, Yao's method, Muddala's method, proposed method, and ground truth.

은 가려짐 형태에 따라 동일한 가려짐 영역에서의 홀채움 결과가 상이하므로 PSNR 편차가 크며, 이로 인해 합성영상의 시간적 일관성을 유지하기 어려운 경우가 존재하여, 깜박임 (flickering) 현상이 나타나게된다. 반면에 제안하는 홀채움 방식은 시간적 배경정보의 활용을 통해 기존 방식과 비교하여 프레임간 PSNR 편차가 작으며, 시간적으로 일관성 있게 합성영상을 생성할 수 있음을 확인할 수 있었다.

각 방식들에 대한 연산량 비교를 표 2에 나타내었다. Xi 방식은 가려짐 영역에서 배경 영역을 구분하기 위해 화소 단위의 SSIM을 이용하므로 가장 많은 연산량을 필요로 한다. Koppel 방식의 연산량은 주로 깊이 정보 채움 및 인페이팅 과정에 의해 결정되며, Yao 방식의 연산량은 주로 인페인팅에 의해 결정되며, 이로 인해 가려짐 영역의 크기에 따라 연산량의 편차가 크다. 더불어 Koppel 및 Yao 방식은 유사 패치 검색 영역이 커짐에 따라 연산량이 급격히 증가함에도 성능 향상이 미비함을 확인할 수 있었다. Muddala 방식에서 사용되는 two-level 인페인팅 과정은 연산량을 결정하는 중요한 요인이 되며, 연산량은 가려짐 영역의 크기에 매우 의존적이다. 제안하는 방식은 기존 홀채움 방식들과 비교하여 상대적으로 적은 연산량을 필요로 하며, 연산량의 편차가 작음을

알 수 있다. 제안하는 방식은 비겹침 패치 기반 시간 적 배경 모델은 가려짐 형태에 관계없이 일정량의 연 산량을 필요로 하며, 시공간 배경 정보를 이용한 홀 채움 과정을 통해 가려짐 영역에 존재하는 많은 배경 정보를 효과적으로 채우고, 이로 인해 인페인팅 과정 의 연산량을 줄일 수 있었다.

각 방식의 가상 시점 합성 확대 영상을 그림 10에 나타내었다. Koppel 방식은 가려짐 영역의 전경 및 배경 영역 분리를 위해 사용하는 전역 임계치로 인해 인페인팅 과정에서 부정확한 패치를 가려짐 영역으 로 복사하게 되어 기하학적 왜곡 현상이 발생하게 된 다. Xi 방식에서 사용하는 화소 단위의 SSIM은 배경 영역과 인접된 가려짐 영역에서는 효과적으로 배경 영역을 분리할 수 있으나, 고립된 가려짐 영역에서는 배경 영역을 분리하기에는 한계가 존재한다. 이로 인 해 고립된 가려짐 영역이 많이 존재할수록 기하학적 왜곡 현상이 커지게 된다. Yao 방식은 인페인팅 우선 순위 결정 과정에서 배경 및 전경 영역을 동시에 고 려하여 전경 영역이 배경 영역에 복사되는 경우가 발 생한다. Muddala 방식은 계층적(layered) 깊이영상을 이용하여 가려짐 배경 영역의 구조를 효과적으로 복 원하였으나, 전경 영역의 균열(crack) 및 수평 경계 영역의 고스트 현상이 존재하는 문제점이 있다. 더불



(c) 97번째 프레임 (97th frame)

그림 11. "Ballet" 합성영상 (3 → 4) 시간적 일관성 성능 비교: (왼쪽에서 오른쪽으로) 워핑 영상, Koppel 방식, Xi 방식, Yao 방식, Muddala 방식, 제안 방식, 원 영상

Fig. 11. Comparisons of temporal consistency of synthesized "Ballet" video $(3 \rightarrow 4)$: (from left to right) warping image, Koppel's method, Xi's method, Yao's method, Muddala's method, proposed method, ground truth.

어, 1차 인페인팅 과정에서 홀채움이 부정확하게 이루어지는 경우 부정확하게 채워진 영역이 2차 인페이팅 과정에 사용되어 고립된 가려짐 영역 및 복잡한가려짐 영역에서 기하학적 왜곡 현상이 심각하게 된다. 반면에 제안하는 홀채움 방식은 전경 및 배경 영역 분리의 정확성, 시공간 정보를 이용한 배경 영역의 홀채움 및 배경 정보를 포함한 우선순위 함수를이용한 인페이팅을 통해 화질 측면에서도 만족스러운 결과를 얻을 수 있었다.

가상 시점에서 합성된 영상의 시간적 일관성에 대한 성능 비교를 그림 11에 나타내었다. 실험 결과를 통해 Koppel, Xi 및 Yao 방식은 가려짐 영역이 큰 경우 홀채움 결과가 만족스럽지 못하고, 가려짐 형태에따라 동일한 영역의 합성 결과가 상이하게 표현됨을확인할 수 있었다. Muddala 방식은 공간적 배경 정보만을 활용한 관계로 고립된 가려짐 영역 및 시간적으로 배경 영역에서 에지 영역의 구조를 일관성 있게유지함에 한계가 존재한다. 반면에 제안하는 방식은시간적 배경 정보의 활용을 통해 합성영상의 시간적 균일성을 유지할 수 있었으며, 고스트 제거 필터를통해 전경 영역에 존재하는 에지 구조를 보존할 수 있었다.

상기와 같은 실험 결과를 통해 제안하는 홀채움 방식을 이용한 가상 시점에서 합성영상은 정량적, 정성

적인 측면에서 만족스러운 결과를 나타냄을 확인할 수 있었다.

Ⅳ. 결 론

본 논문은 가상 시점에서의 합성영상을 생성하기 위 해 추정된 시공간 배경 정보를 이용한 홀채움 방식을 제안하였다. 시간적 배경 정보를 추정하기 위해 새로운 패치 기반의 배경 코드북을 소개하였다. 또한 깊이 정 보 기반의 공간 국부 배경 예측 방식을 제안하여 고립 영역의 공간적 배경 후보들의 범위를 나타내는 제약 조 건들을 설정하였다. 시간적 배경 정보와 공간적 배경 정보를 결합하여 배경 영역의 홀들을 효과적으로 채울 수 있었다. 더불어 제안한 깊이영상 기반 고스트 필터 를 통해 전경 영역의 에지 영역의 구조를 유지하면서 합성영상의 배경 영역에 존재하는 고스트 현상을 효과 적으로 제거할 수 있음을 확인하였다. 깊이영상의 배경 정보를 포함한 새로운 우선순위 함수를 이용한 인페인 팅을 통해 잔여 홀들을 효과적으로 채울 수 있었다. 기 존 방식들과 비교하여 PSNR 측면에서 평균 1dB 이상의 홀채움 성능이 개선되었음을 확인할 수 있 었으며, 연산량 측면에서도 30% 이상의 절감 효과 가 있음을 확인할 수 있었다.

실험 결과를 통해 정량적, 정성적으로 제안한 시공간

배경 정보를 이용한 홀채움 방식의 성능의 우수성을 확 인할 수 있었다.

REFERENCES

- C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D TV," in SPIE Int. Conf. Stereoscopic Display and Virtual Reality Syst. XI, pp. 93-104, Jan. 2004,
- [2] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," Signal Process.: Image Commu., vol. 22, no. 2, pp. 217–234, Feb. 2007.
- [3] C. Zhu and S. Li, "Depth image based view synthesis: new insights and perspectives on hole generation and filling," IEEE Trans. Broadcast., vol. 62, no. 1, pp. 82–93, Mar. 2016.
- [4] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3D TV," IEEE Trans. Broadcast., vol. 51, no. 2, pp. 191–199, June 2005.
- [5] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," ACM Trans. Graphics, vol. 23, no. 3, pp. 600-608, Aug. 2004.
- [6] M. Solh and G. AlRegib, "Hierarchical hole-filling for depth-based view synthesis in FTV and 3D video," IEEE. J. Selected Topic in Signal Process., vol.6, no.5, pp. 495–504, Sep. 2012.
- [7] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in Proc. ACM the 27th Int. Conf. Computer Graphics and Interactive Tech. (SIGGRAPH), pp. 417–424, July 2000.
- [8] A. Criminisi, P. Perez, and K. Toyama, "Object removal by exemplarbased inpainting," in Proc. IEEE Int. Conf. Computer Vision and Pattern Recog. (CVPR), doi:10.1109/CVPR.2003.1211538, July 2003.
- [9] C. Guillemot and O. Le Meur, "Image inpainting: Overview and recent advances," IEEE Signal Process. Mag., vol. 31, no. 1, pp. 127-144, Jan. 2014.
- [10] Y. Mori, N. Fukushima, T. Fujii, and M. Tanimoto, "View generation with 3D warping using depth information for FTV," Signal Process.: Image Commu., vol. 24, no. 1–2, pp. 65–72, Jan. 2009.

- [11] K. Oh, S. Yea, and Y.-S. Ho, "Hole filling method using depth based inpainting for view synthesis in free viewpoint television and 3-D video," in Proc. Picture Coding Symp., doi:10.1109/PCS.2009.5167450, July 2009.
- [12] S. Zinger, L. Do, and P. H. N. de With, "Free-viewpoint depth image based rendering," J. Vis. Comm. Image Represent., vol. 21, no. 5-6, pp. 533-541, July 2010.
- [13] I. Daribo and H. Saito, "A novel inpainting-based layered depth video for 3DTV," IEEE Trans. Broadcast., vol. 57, no. 2, pp. 533-541, June 2011.
- [14] I. Ahn and C. Kim, "A depth-based disocclusion filling method for virtual viewpoint image synthesis," Journal of the Institute of Electronics Engineers of Korea SP, vol. 48, no. 6, pp. 48-60, Nov. 2011.
- [15] S. Choi, B. Ham, and K. Sohn, "Space-time hole filling with random walks in view extrapolation for 3D video," IEEE. Trans. Image Process., vol. 22, no. 6, pp. 2429–2441, June 2013.
- [16] S. M. Muddala, M. Sjostrom, R. Olsson, "Virtual view synthesis using layered depth image generation and depth-based inpainting for filling disocclusions and translucent disocclusion," J. Vis. Commu. Image Represent., vol. 38, pp. 351-366, July 2016.
- [17] P. Mdjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3-D Video," IEEE Trans. Multimedia, vol. 13, no. 3, pp. 453-465, June 2011.
- [18] M. Koppel, X. Wang, D. Doshkov, T. Wiegand, and P. Ndjiki-Nya, "Consistent spatio-temporal filling of disocclusions in the multiview-videoplus-depth format," in Proc. IEEE Int. Workshop Multimedia Signal Process., pp. 25–30, Sep. 2012.
- [19] M. Xi, L.-H. Wang, Q.-Q. Yang, D.-X. Li, and M. Zhang, "Depthimage-based rendering with spatial and temporal texture synthesis for 3DTV," EURASIP J. Image and Video Process., doi:10.1186/1687-5281-2013-9, Dec. 2013.
- [20] C. Yao, T. Tillo, X. Jimin, "Depth map driven hole filling algorithm exploiting temporal correlation information," IEEE Trans. Broadcast., vol. 60, no. 2, pp. 394-404, June 2014.
- [21] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in IEEE Int. Conf. Image Process., pp. 3061–3064, Oct. 2004.

- [22] E. J. Fernandez-Sanchez, D. Javier, and R. Eduardo, "Background subtraction based on color and depth using active sensors," Sensors, vol. 13, no. 7, pp. 8895–8915, July 2013.
- [23] A. S. Glassner, Ed., Graphics Gems I, Academic Press, Cambridge, MA, 1990.
- [24] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

— 저 자 소 개 ·



김 범 수(학생회원)

2007년 숭실대학교 정보통신전자 공학부 학사 졸업.

2009년 숭실대학교 정보통신공학 과 석사 졸업.

2012~현재 숭실대학교 정보통신 공학과 박사과정.

2010~2012년 한국인터넷진흥원 연구원. <주관심분야: 초해상도 영상처리, 3D 영상 처리>



응웬띠엔닷(학생회원)

2012년 하노이 공과대학교 정보 통신공학과 학사 졸업 2015~현재 숭실대학교 정보통신 공학과 석사과정.

<주관심분야: 3D Video Processing>



홍 민 철(정회원)-교신저자

1988년 연세대학교 전자공학과 학사 졸업.

1990년 연세대학교 전자공학과 석사 졸업.

1997년 Northwestern University Electrical & Computer Engineering 박사 졸업.

1997~1998년 Northwestern University 박사후 연구원.

1998~2000년 LG전자 선임연구원.

2000~현재 숭실대학교 전자정보공학부 교수.

<주관심분야: Image/Video Restoration, Image/ Video Enhancement, Nonlinear Signal Processing, 3D Video Reconstruction, Video Coding>