다시점 영상의 가상시점 합성을 위한 깊이 및 수퍼픽셀 기반 가중치 블렌딩 방법

윤준영 *신홍창 박종일†

한양대학교 *한국전자통신연구원

jyyun@mr.hanyang.ac.kr hcshin@etri.re.kr jipark@hanyang.ac.kr

Depth and Superpixel Based Weighted Blending for Virtual View Synthesis

Junyoung Yun *Hong-Chang Shin Jong-Il Park

Hanyang University *Electronics and Telecommunications Research Institute

요약

본 논문에서는 다시점 영상과 스테레오 매칭을 수행하여 얻은 깊이 정보를 통해 가상시점을 합성할 때 적용되는 블렌딩 기법을 제안한다. 다시점 영상에서 스테레오 매칭으로 얻을 수 있는 깊이정보는 물체와 배경의 경계부분에서 큰 오차를 갖는다. 이러한 이유로 원본 시점들의 영상을 깊이정보를 통해 특정 가상시점으로 워핑할 때 물체의 경계 부분에 배경의 화소가 일부 포함되는 문제가 발생한다. 이 문제는 워핑된 영상들을 블렌딩하여 하나로 합성 시 영상의 품질에 영향을 주는 요인이 된다. 본 논문에서 제안하는 방법은 깊이정보와 함께 영상의 수퍼픽셀 분할에서 얻은 정보를 이용하여 블렌딩 대상이 되는 영상에 화소 단위로 가중치를 부여하여 해당 문제의 영향을 완화한다.

1. 서론

최근 스마트폰과 머리에 착용하는 형태의 디스플레이 장치인 HMD(Head Mounted Display)의 보급을 통해 360도 영상과 같은 몰입형 비디오에 대한 접근성이 높아지고 그에 대한 수요가 증가하고 있다. 기존의 360도 영상은 사용자의 머리 위치가 고정되어 주변을 둘러볼 수 있는 전방향 비디오로 고개를 앞으로 빼거나 걷는 등의 사용자의 움직임에 의한 시차는 제공되지 않는다. 이러한 한계를 넘어 임의의 위치에서 보는 것과 같은 3차원 영상을 구현하는 방법 중 하나가 다시점 영상을 통해 가상시점을 합성하는 기법이다.

특정 시점에서 촬영된 영상과 그 영상을 구성하는 환경의 3차원 기하정보가 주어지면 해당 환경을 다른 시점에서 촬영할 때 얻을 수 있는 영상을 가상으로 생성할 수 있다. 하지만 하나의 시점 영상만으로는 보이지 않는 영역을 다른 시점에서 생성하기 어렵고, 3차원 기하정보를 얻어내기 위해 카메라 이외의 추가적인 장비가 필요한 문제가 있다. 이것을 극복하기 위해 서로 다른 평행한 시점에서 촬영된 다수의 영상을 이용해 가상시점의 영상을 생성하는 방법이 제시되었다. 해당 방법에서는 인접한 두 시점 영상의 스테레오 매칭을 통해 촬영한 환경의 3차원 깊이정보를 복원할 수 있다. 더불어 하나의 시점에서는 보이지 않는 영역을 다른 시점의 영상을 통해 알 수 있다. 이러한 장점 덕분에 다수의 시점을 합성하고자 하는 가상시점에서 재구성하고 그 영상을 블렌딩하여 영상정보가 없는 영역의 크기를 최소화 할 수 있다.

스테레오 매칭을 통해 얻어지는 깊이정보에는 두 영상 사이의 매칭 오류에 의해 발생하는 노이즈가 포함된다. 이러한 노이즈는 한쪽 시점에서는 보이지만, 다른 시점에서는 가려져 보이지 않아 매칭 대상이 사라지는 물체와 배경의 경계 부분에서 자주 발생한다. 이 원인에 의해

† 교신저자

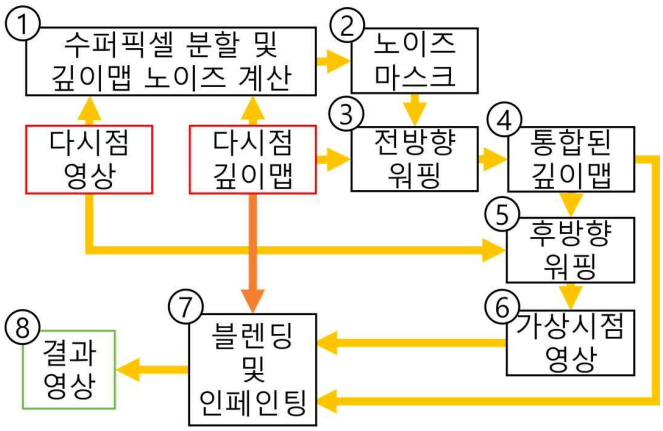

그림 1 제안하는 방법의 흐름도

영상을 가상의 시점으로 재구성하는 워핑을 수행하면 물체의 경계에 배경 영역이 일부분 함께 이동되어 영상의 블렌딩 시 배경과 물체의 화소가 섞이게 된다.

본 논문에서는 주어진 다시점 영상에 수퍼픽셀 분할을 수행하여 영상의 색상 정보를 통해 물체와 배경의 영역을 지역적인 수준에서 분할한다. 분할된 각 영역 안에서 노이즈에 해당되는 깊이의 마스크 정보를 생성한다. 이어서 생성된 노이즈 마스크와 다시점의 깊이맵들에 대해 전방향 워핑이 수행되어 가상시점에서 통합된 깊이맵이 생성된다. 통합된 깊이맵을 통해 역방향 워핑을 수행하여 원본 시점들에 대한 가상시점의 영상을 얻는다. 마지막으로 이전 단계에서 얻은 가상시점 영상들은 깊이정보와 원본 시점의 위치에 따른 가중치에 의해 블렌딩되어 하나의 영상으로 합성된다.

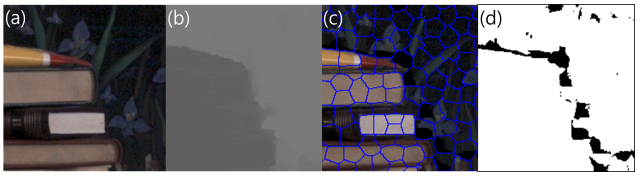

그림 2 (a)시점영상, (b)시점깊이맵, (c)수퍼픽셀 분할 결과, (d)노이즈 마스크

2. 수퍼픽셀 영상분할을 통한 깊이맵 정제

일반적으로 동일한 물체의 내부 영역의 깊이 값은 연속성을 갖는다. 이와 반대로 물체와 배경의 경계 영역에서 잘못 추정된 깊이는 실제 깊이와 상이한 값을 가지므로 인접한 영역에 대해 그 값이 불연속적이며, 물체의 영역 안에서 작은 면적을 차지한다고 가정할 수 있다. 본 연구에서는 물체영역과 배경영역을 분할하는데 수퍼픽셀 영상 분할 기법 중 하나인 SLIC[2]을 사용하였다. 하나의 수퍼픽셀 영역에 있는 깊이 값을 표 1.의 알고리즘에 의해 노이즈에 해당되는지 판단한다.

수퍼픽셀 영상분할은 화소의 색상정보와 이미지 공간상에서 화소간 거리를 토대로 유사한 화소들을 무리지어 더 큰 단위인 수퍼픽셀 단위로 영상을 분할하는 기법이다. 영상 속에서 동일한 물체는 지역적으로 유사한 색상을 갖는다고 가정할 수 있으므로, 수퍼픽셀 분할 결과를 통해 각 물체의 영역을 지역적인 수준에서 분할할 수 있다. 그러나, 분할된 수퍼픽셀들은 물체의 경계를 정확히 나누지 못하는 경우가 있을 수 있다. 이런 경우를 처리하기 위해 표 1.의 과정을 여러번 반복하여 하나의 화소에 대해 노이즈로 판단되는 횟수가 반수를 초과하면 해당 화소의 깊이값을 노이즈라고 판단하도록 한다.

각 시점에서 깊이 값이 노이즈가 아닌 것으로 분류된 영역에 대한 깊이맵을 가상시점으로 전방향 워핑을 수행한다. 이때 가상시점에서 각 픽셀의 깊이 값을 해당 픽셀에 워핑된 깊이 값들의 중간 값으로 선택하여 하나로 통합된 깊이맵을 생성한다. 그 후 통합된 깊이맵을 사용하여 후방향 워핑을 통해 원본시점 영상들을 가상시점으로 워핑한다.

표 1. 수퍼픽셀 내부 깊이 값의 노이즈 판별 알고리즘

Variables: The cluster of each pixel **cluster**; the mask of pixels with outlier depth value **noise_mask**; the number of pixels of each cluster **pixel_count**; the number of clusters **cluster_num**; the largest cluster **largest_cluster**; the number of pixels in largest cluster **max_pixel_count**; the depth value of each pixel **d**; the continuity threshold ϵ; initialize **cluster, noise_mask, pixel_count** to be all 0.

```
cluster_num ← 1
max_pixel_count ← 0
largest_cluster ← 0
function join(i, j, c):
    cluster(i, j) ← c
    pixel_count(c)++
    if pixel_count(c) > max_pixel_count:
        max_pixel_count ← pixel_count(c)
        largest_cluster ← c
    for (ii, jj) in [(i-1, j), (i, j-1), (i+1, j), (i, j+1)]:
        if cluster(ii, jj) == 0 and |d(i, j) - d(ii, jj)| < ϵ:
            join(ii, jj, c)
for (i, j) in superpixel:
    if cluster(i, j) == 0:
        join(i, j, cluster_num)
        cluster_num++
for (i, j) in superpixel:
    if cluster(i, j) != largest_cluster:
        noise_mask(i, j) ← 1
```

3. 워핑된 가상시점의 블렌딩

원본시점의 영상 I_i를 가상시점으로 워핑해 생성한 영상 I_i'를 블렌딩하여 합성한 영상을 I_{blend}라 할 때, 영상좌표 p에 위치하는 화소의 값 $I_{blend}(p)$는 가상시점과 원본시점 사이의 거리 d_i, 가상시점으로 워핑된 원본시점의 깊이맵 D_i'와 가상시점에서 통합된 깊이맵 D^*에 의해 정의된 α_i에 의해 다음과 같이 가중치가 결정되어 블렌딩된다.

$$\alpha_i(p) = \left| D_i'(p) - D^*(p) \right| \tag{1}$$

$$w_i(p) = e^{-c_d d_i^2 - c_\alpha \alpha_i(p)} \tag{2}$$

$$I_{blend}(p) = \frac{\sum_i w_i(p) I_i'(p)}{\sum_i w_i(p)} \tag{3}$$

가상시점과 거리가 먼 시점의 영상일수록 워핑된 영상에 대해 깊이맵에 전체적으로 포함된 노이즈가 미치는 영향이 커지는 것을 고려하여 해당 영상의 전체 화소의 블렌딩 가중치가 낮아지도록 하였다. 깊이맵 D^*는 다수 시점의 깊이 값이 일치하는 신뢰도 높은 깊이맵이라 가정할 수 있다. 이때 워핑된 I_i'영상 안에서 D_i'과 D^*의 값의 차이가 큰 화소는 그에 대응되는 깊이 값의 신뢰도가 떨어진다고 판단할 수 있으므로 블렌딩 가중치가 낮아지도록 하였다.

가상시점 영상에서 블렌징 과정을 진행한 이후에도 색상이 채워지지 않은 영역에 대해서 이미지 인페인팅을 수행한다. 인페인팅은 [3]의 방법을 사용하여 결과 영상을 얻어내었다.

4. 실험 결과

그림 3 (a) (b)원본영상, (c)제시된 방법에 의해 생성된 영상, (d)가상시점으로 워핑된 영상들의 평균영상

본 실험에서는 MPEG-I Visual에서 테스트 목적으로 제작하여 사용되고 있는 TechnicolorPainter[1]가 사용되었다. 해당 영상은 수직 및 수평 방향으로 일정한 간격을 두고 평행하게 위치한 카메라 배열을

통해 얻어졌다. 영상의 깊이정보는 OpenCV extra modules 라이브러리의 cv::StereoSGBM 클래스에 구현된 스테레오 매칭 알고리즘을 사용하여 인접한 두 영상의 스테레오 매칭을 통해 계산하였다. 스테레오 매칭을 통해 일차적으로 계산되는 깊이정보는 weighted lest squares filter를 통해 필터링한 후, 깊이맵으로 변환하여 사용하였다. 결과영상 간의 품질 비교를 위해 비교 대상이 되는 시점영상은 입력에서 제외한 뒤 비교대상 시점의 영상을 합성하였다. 실험결과는 그림 3과 같다.

(c)는 제시된 방법에 의해 블렌딩되어 생성된 영상이며, (d)는 각 입력 시점을 원본 깊이맵을 통해 가상시점으로 워핑한 뒤, 그 평균을 블렌딩 결과로 사용한 영상이다. (c)에서는 물체와 배경의 경계부분에서 (d)보다 원본시점에 가까운 결과를 확인할 수 있으며, 구조적 유사지수(Structural Similarity: SSIM)의 측면에서도 품질이 향상됨을 확인하였다.

5. 결론

본 논문에서는 각 시점 깊이맵에 수퍼픽셀 기법을 적용하여 영상 속 객체 간 경계 부분에 존재하는 깊이 정보의 노이즈를 제거하였다. 이를 통해 노이즈가 제거된 가상시점의 깊이맵과 워핑된 영상을 생성하였고, 노이즈가 제거된 깊이맵을 통해 워핑된 영상에 가중치를 부여하여 가상시점의 영상을 합성한 결과 영상의 품질이 향상됨을 확인하였다.

감사의 글

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 '범부처 Giga KOREA 사업'의 지원을 받아 수행된 연구임 (GK19C0200, 모바일 완전입체 단말 및 콘텐츠 기술 개발)

참고 문헌

[1] ISO/IEC JTC1/SC29/WG11, "EE_DEPTH: New version of the pseudo-rectified TechnicolorPainter content", M43366, (2018).

[2] Achanta, Radhakrishna, et al. "SLIC superpixels compared to state-of-the-art superpixel methods." IEEE transactions on pattern analysis and machine intelligence 34.11 (2012): 2274-2282.

[3] Bertalmio, Marcelo, Andrea L. Bertozzi, and Guillermo Sapiro. "Navier-stokes, fluid dynamics, and image and video inpainting." Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Vol. 1. IEEE, 2001.