

다시점 동영상의 중간시점영상 생성을 위한 화소 기반의 변이 예측 기법

*최미남 *백윤기 *김동욱 *유지상

*광운대학교

*minam@image.gwu.ac.kr

Pixel-based disparity estimation for Intermediate View Reconstruction of Multi-view images

*Choi, Mi-Nam *Baek, Yun-Ki *Kim, Dong-Wook *Yoo, Ji-Sang

Kwangwoon University

요약

본 논문은 다시점 카메라로부터 획득된 다시점 영상을 이용하여, 영상내의 모든 화소에 대한 정확한 변이 정보를 획득하는 알고리즘을 제안한다. 제안한 방법은 정합창(measurement window)을 사용하는 화소기반 정합방법을 통해 정확한 변이 추정과 비용 함수 계산시 반복적인 계산을 줄임으로써 전체적인 수행 속도를 향상시킬 수 있다. 정확한 변이 정보를 획득하는데 있어서 가려진 영역이 많은 영향을 미친다. 또한 가려진 영역의 판단 및 처리는 2대의 카메라로는 한계가 있다. 따라서 본 논문에서는 평행식으로 구성된 5대 이상의 다시점 카메라로부터 획득된 다시점 영상을 사용함으로써 변이 정보의 신뢰도를 높이고, 가려진 영역을 처리할 수 있다. 또한 비용 함수 계산의 반복성을 이용하여 계산량을 줄임으로써 전체적인 수행시간을 줄인다. 이렇게 획득된 변이 정보는 다시점 영상의 중간시점영상 생성을 위해 사용된다.

1. 서론

디지털 IT의 바람은 정보 고속도로와 무선통신의 진보에 힘입어 대화형 TV, 네트워크를 통한 정보 가진, 센서 네트워크 등으로 발전해 가고 있다. 이와 더불어 이동통신의 이용 또한 국민들 사이에 깊이 침투하여 사람들의 생활에 있어서 “당연”하고 “없어서는 안될”공기와 같은 생활의 기본요소가 되어가고 있다. 이처럼 우리 사회는 정보통신이 공기나 물처럼 어디서든 이용이 가능하여 모든 사람들의 경제, 사회, 문화 등 모든 활동의 기반이 되는 유비쿼터스 사회로 진화하여 가고 있다. 멀티미디어의 개념 또한 기존의 2D나 고해상도에서 입체감과 자연감을 중요시하는 3DAV(3D audio-visual) 형태로 발전하고 있다 [1].

현실감, 입체감을 제공할 수 있는 3DAV 콘텐츠 제작을 위해서는 기본적으로 양안식/다시점 영상 기술 개발이 필요하다. 기본적으로 다시점 영상은 여러 개의 카메라로 획득한 영상을 말한다. 현재 다시점 카메라는 스포츠 중계부분과 광고/영화의 특수효과로 사용되고 있다. 예를 들어 영화 'MATRIX'에서 공간적으로 배치된 다시점 카메라에서 획득한 영상을 시간적으로 재배치함으로써 시간이 정지된 듯한 형태의 영상물을 제작하는 것이 가능하였고 미국 CMU에서 개발한 eye-vision도 앞의 예와 동일한 영상물을 제작하여 미식축구에서 주요 장면을 재구성 하여 시청자들에게 보여줌으로써 고도의 기술적 효과를 자랑하였다. 하지만 카메라 수와 카메라 간격의 제한으로 인해 시점의 이동시 불연속성이 발생할 수 있고 카메라 수를 늘릴 경우에는 데이터양이 많아지는 단점을 가지고 있다. 이에 대하여 다시점 동영상의 중간시점영상 생성 기법을 사용함으로써 시점의 자유도와 관찰자의 위치에 따라 다른 입체감을 느낄 수 있다[2]. 따라서 본 논문의 다시점 동영상의 중간시점영상 생성을 위한 변이 예측 기법은 사용자가 원하

는 시점의 영상을 자연스럽게 감상할 수 있게 하는데 필요하다. 본 논문에서는 5대 이상의 카메라를 사용하여 보다 정확한 변이 정보를 추출하여 중간시점영상을 생성하는 기법을 제안한다.

논문의 구성은 다음과 같다. 2장에서는 마스크를 사용한 화소 기반의 변이 예측 기법을 설명하며 3장에서는 실험결과를 확인, 분석하였다. 4장에서는 본 논문의 결론과 향후 연구 방향을 기술한다.

2. 제안하는 기법

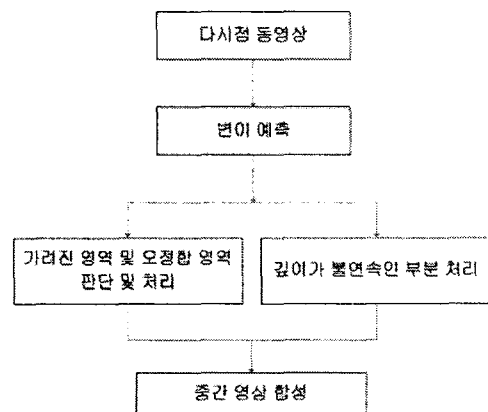


그림 1. 전체 블록도

제안하는 기법은 그림 1과 같이 다시점 동영상을 입력으로 변이 예측을 한다. 보다 정확한 변이 예측과 가려진 영역에 대한 문제를 처리할 수 있도록 5대 이상의 평행식으로 구성된 카메라로 가정하였으며, 깊이가 불연속인 부분의 변이를 예측하기 위해 기준 영상의 현재

영상과 이전 영상의 차를 사용한다. 변이맵을 사용해 중간 시점영상을 생성하여 결과를 확인하였다.

1) 변이 정보 추출

실제 영상에서 변이 정보를 추출하는 방법으로 많은 연구가 진행되어져 왔다. 정합 방법에는 크게 영역에 기반한 방법(area-based method)과 특징에 기반한 방법(feature-based method)이 있다. 본 논문에서는 영역에 기반한 정합 방법을 사용하여 변이 정보를 추출한다 [3].

기준이 되는 영상과 참조가 되는 영상의 정합을 수행하여 변이 정보를 추출한다. 하지만 카메라 두 대를 사용한 정합에서는 가려진 영역에 대한 문제를 해결할 수 없다. 본 논문에서는 5대 이상의 평행식으로 구성된 카메라로부터 획득된 다시점 영상으로 가정한다. 실험에 사용한 영상은 640 * 480 의 4:2:0 YUV Ballroom 영상이다. 그림 2는 카메라 구성과 실험에 사용한 원영상을 나타낸다.

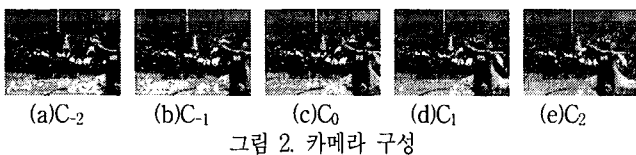


그림 2. 카메라 구성

제안하는 변이 예측 기법에서는 정합의 척도로서 일반적으로 많이 사용하는 SAD(sum of absolute difference)를 비용함수로 사용한다. 정합 과정은 기준 영상과 참조 영상간의 탐색 범위 내에서 정합창내의 화소들의 비용함수 값이 최소값을 갖는 화소를 찾는 과정이다. 기준영상의 정합창을 1화소에서 SR화소까지 이동하면서 SAD를 계산한다[4].

$$C(w) = \min \sum_{i=0}^{SR} |f_0(w) - f_1(w + i)| \quad (\text{식 1})$$

여기서 f_0 는 기준영상의 화소값, f_1 는 참조 영상의 화소값, SR은 탐색범위를 나타내며 w 는 정합창을 나타낸다.

정합창을 사용한 정합방법의 비용함수 계산시 발생하는 중복적인 계산을 제거하기 위해 각행의 첫 번째 화소의 비용함수의 값을 그림 3과 같이 저장한다.

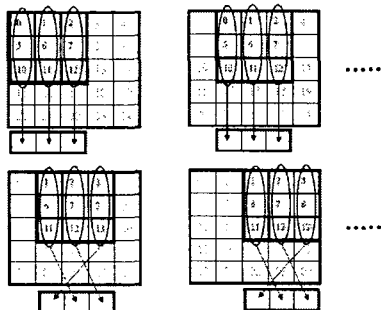


그림 3. 비용함수 결과 저장

두 번째 화소의 비용함수 계산시 정합창의 마지막 열의 화소만을 계산한다.

두 대의 카메라를 사용한 정합으로부터의 변이맵은 가려진 영역의 문제 해결에 한계가 있다. 가려진 영역과 오정합 영역을 해결하기

위해 다음과 같은 방법을 사용한다.

기준 영상 C_0 와 참조 영상 C_1 를 사용하여 정합을 통해 얻어진 변이 정보를 이용하여 참조 영상 C_2 에서의 정합된 화소의 위치를 예측할 수 있다.

(1)기준 영상과 참조 영상 C_2 의 예측된 화소의 위치에서 비용 함수를 계산하여 임계값 이하이면 정확한 정합 영역으로 판단하고, 임계값 이상이면 오정합 영역 또는 가려진 영역으로 판단한다.

$$\text{if } |f_0(w) - f_2(w + 2dv)| > Th_1 \quad (2)\text{번 과정 수행} \quad (\text{식 2})$$

f_2 는 참조 영상 C_2 의 화소이고, dv 는 기준 영상 C_0 와 참조 영상 C_1 를 사용하여 정합을 통해 얻어진 변이 정보이다.

(2)(1)번의 과정에서 오정합 또는 가려진 영역으로 판단된 영역은 기준 영상 C_0 와 참조 영상 C_1 를 이용하여 정합창을 1화소씩 이동시키면서 정합을 수행한다.

$$C(w) = \min \sum_{j=0}^{SR} |f_0(w) - f_{-1}(w + j)| \quad (\text{식 3})$$

(3)(2)번의 과정으로 얻어진 변이 정보를 이용하여 참조 영상 C_2 에서의 비용함수를 계산하고, 비용 함수의 값이 임계값 이하이면 정확한 정합 영역으로 판단하고 임계값 이상이면 오정합 영역 또는 가려진 영역으로 판단한다.

$$\text{if } |f_0(w) - f_{-2}(w + 2dv)| > Th_2 \quad (4)\text{번 과정 수행} \quad (\text{식 4})$$

여기서 f_{-2} 는 참조 영상 C_{-2} 의 화소이고, dv 는 기준 영상 C_0 와 참조 영상 C_1 를 사용하여 정합을 통해 얻어진 변이 정보이다.

(4)(3)번의 과정에서 잘못된 정합으로 판단된 영역은 4개의 참조 영상을 모두 이용하여 정합을 수행한다.

$$C_i(w) = \min \sum_{j=0}^{SR} |f_0(w) - f_i(w + j)| \quad (\text{식 5})$$

i 는 참조 영상의 번호이다.

$$C(w) = \min \sum_{i=0}^3 C_i(w) \quad (\text{식 6})$$

기준 영상과 4개의 참조 영상의 비용함수의 값을 합한다. 이 때 잘못된 정합을 피하기 위해 4개의 참조 영상에서의 비용 함수를 계산하여 비용함수 값이 가장 큰 2개의 값을 버리고 나머지 값만을 합하여 비용함수를 계산한다.

$$C'(w) = \min \sum_{i=0}^1 C_i(w) \quad (\text{식 7})$$

위의 과정을 정리하면 그림 4와 같다.

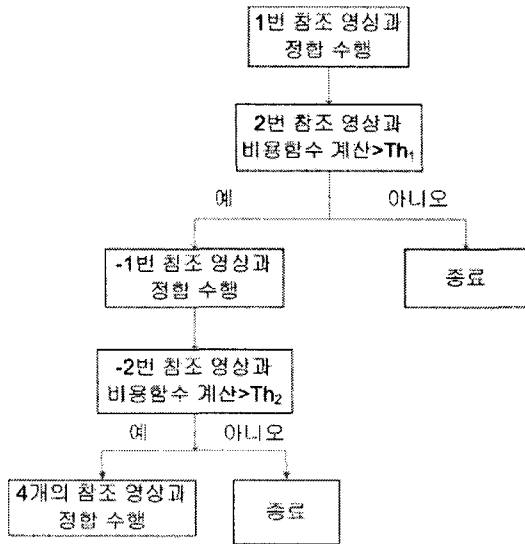


그림 4. 변이 정보 추출 기법

2) 깊이가 불연속인 부분의 변이 정보 추출

정합창을 사용한 정합방법은 정합창의 크기가 성능을 좌우한다.[5] 정합창의 크기는 정합창 내의 영역의 특징을 나타낼 수 있을 정도로 커야한다. 하지만 정합창의 크기가 커지게 되면 계산량이 많아지고, 물체의 외곽과 같은 깊이가 불연속인 영역에 있어서는 정확한 정합 정보를 기대할 수 없다[6]. 따라서 물체의 외곽의 정합을 행할 때는 정합창의 크기를 작게 하여 수행한다. 이 논문에서 물체의 외곽에 대한 정합창의 크기를 결정하기 위해 다음 방법을 제안한다.

카메라가 고정되어 있기 때문에 기준영상의 현재영상과 이전영상의 차이 영상에서 배경은 제거되고, 움직임이 있는 물체의 외곽 부분은 남게 된다. 그림 5는 기준 영상의 현재 영상과 이전 영상이다.



그림 5. 차이 영상

이러한 차이 값을 이용하여 차이 영상에서 정합창 내의 화소값의 평균 값이 임계값 이상이면 물체의 외곽으로 판단한다. 외곽 영역은 처음 설정된 정합창의 크기로 정합을 수행하여 얻어진 변이 정보를 기준으로 정합창의 크기를 작게하여 정합을 다시 수행한다.

$$\text{if } \frac{1}{M * M} \sum_{i=0}^{M * M} \text{diff}(w_i) > th \quad (\text{식 } 8)$$

여기서 M은 차이 영상의 평균값을 구하기 위한 정합창의 크기, diff는 차이 영상, w_i 는 정합창 내의 화소를 의미한다.

3) 중간시점영상 생성

스테레오 영상은 시점이 고정되어 있어서 여러 방향에서 관찰할 수 있는 시점의 자유도가 없다는 단점이 있다. 이를 해결할 수 있는 방

법이 영상들 사이의 시점을 생성하는 중간시점영상 생성 기법이다. 본 논문에서는 앞의 과정에서 생성된 변이맵을 확인하기 위해 중간시점영상을 생성하였다. 중간시점 영상을 생성하기 위해 기준이 되는 좌우 영상과 α 에 대응하는 변이맵이 필요하다. 하나의 변이에 해당하는 중간시점영상의 화소값은 기준 영상의 변이를 고려하여 보간한다. 식 9는 중간시점영상의 보간법을 나타낸다.

$$f_{\alpha}(x+d \times \alpha) = (1-\alpha) \times f_0(x) + \alpha \times f_1(x+d) \quad (\text{식 } 9)$$

여기서 α 는 좌우영상의 거리를 1이라 했을 때 좌 영상에서 생성하고자 하는 중간시점영상까지의 거리를 나타낸다.

3. 실험 결과 및 분석

본 논문에서 제안한 기법을 검증하기 위해 640 * 480 의 4:2:0 YUV 영상을 실험 영상으로(Ballroom) 사용하였다. 그림 6은 실험에 사용한 원영상의 기준 영상을 나타내었다.



그림 6. Ballroom 영상

그림 7은 기준 영상과 참조 영상을 하나를 이용하여 정합창을 이용한 화소 기반의 정합 방법을 수행하여 얻어진 변이맵이다. 정합창의 크기는 15*15로 하였다.

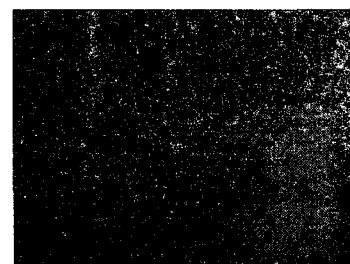


그림 7. 기준 영상 C_0 와 참조 영상 C_1 의 정합을 통한 변이맵

그림 7은 기준 영상 C_0 와 참조 영상 C_1 을 이용하여 정합을 수행하여 얻은 변이맵이다. 변이맵의 신뢰도가 떨어지고, 물체의 외곽도 정확하지 않은 것을 확인할 수 있다. 그림 8은 제안한 기법으로 첫 번째 탐색 영역을 0화소에서 60화소로 설정하고, 정합창의 크기를 15*15로 설정하여 정합을 수행하였고, 외곽으로 판단된 영역은 정합창의 크기를 3*3으로 설정하여 정합을 수행하였다. 또한 물체의 외곽 정보를 유지하면서, 잡음을 제거하기 위해 외곽영역을 제외한 영역에 대해 중간값 필터(median filter)를 사용했다



그림 8. 제안한 정합 방법의 변이맵

제안한 정합 방법에 의해 가려진 영역 및 오정합 영역에 대한 문제를 해결할 수 있었고, 향상된 변이 맵을 얻을 수 있었다.

그림 9는 차이 영상과 변이맵을 같이 나타낸 영상이다.

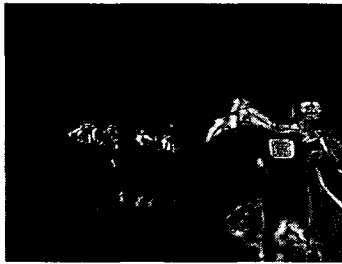


그림 9. 차이 영상과 변이맵

영상에서 물체의 외곽과 변이맵의 외곽이 거의 일치 하는 것을 확인할 수 있다.

그림 10은 제안한 방법으로 얻어진 변이맵을 사용하여 중간시점 영상을 생성한 결과이다.



그림 10. 중간시점영상

본 논문에서는 변이맵의 성능을 확인하기 위해 α 값에 따라서 생성된 두 개의 중간 영상을 사용하여 스테레오 영상을 합성하였다.



그림 11. 스테레오영상합성

그림 11은 $\alpha = 0.5, 0.7$ 일 때의 중간 영상을 이용하여 스테레오 영상을 합성한 것이다.

4. 결과

본 논문에서는 5대의 카메라로부터 획득된 다시점 영상을 이용하여 변이 정보를 추출하고, 추출된 변이맵을 이용하여 중간시점영상을 생성하였다. 기존 영상의 변이 정보를 추출하기 위해 4개의 참조 영상을 이용하였으며, 실제 물체의 외곽과 변이맵에서 물체의 외곽 일치율을 위해 기존 영상의 현재 영상과 이전 영상의 차 영상을 이용하였다. 여러 개의 참조 영상을 이용한 정합을 함으로서 발생하는 계산량과 가려진 영역에 대한 문제는 향후 더 보완되어야 할 과제이다.

<감사의 글>

본 연구는 산업자원부, 노동부의 출연금 및 보조금으로 수행한 최우수 실험실지원사업의 연구결과로 수행되었음.

<참조 문헌>

- [1] "Description of Exploration Experiments in 3D AV," ISO/IEC JTC1/SC29/WG11 N4929, July 2002.
- [2] JS McVeigh, Efficient compression of arbitrary multi-view video signal , Ph.D. dissertation, Ch.4, CMU, 1996.
- [3] R. Franich, Disparity estimation in stereoscopic digital images. PhD thesis, Technical University of Delft, 1996.
- [4] Stereo image analysis for multi-viewpoint telepresence applications. Author: Izquierdo M.E.1. Source: Signal Processing: Image Communication, Volume 11, Number 3, pp. 231-254, January 1998,
- [5] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 9, pp. 920 - 932, September, 1994.
- [6] S. D. Cochran and G. Medioni, "'3-D surface description from binocular stereo,'" IEEE Trans. Pattern Anal. Machine Intell., vol. 14, Oct. 1992.