

텍스처 추출시 제한된 수의 참여 영상을 이용한 Multi-View 영상 개선 알고리즘

*김도현, **양영일
경상대학교 전자재료공학과
경남 진주시 가좌동 900

e-mail : *dhkim@nongae.gsnu.ac.kr , **yyang@nongae.gsnu.ac.kr

An Algorithm for the Multi-View Image Improvement with The Restricted Number of Images in Texture Extraction

Do-Hyun Kim , Yeong-Yil Yang

Members of Research Center for Aircraft Parts Technology
Dept. of Electronic Materials Eng. GyeongSang Nat'l. Univ.
Gazwa-Dong 900, Chinju, KyungNam, 660-701, Korea

Abstract

In this paper, we propose an efficient multi-view images coding algorithm which finds the optimal texture from the restricted number of multi-view images. The X - Y plane of the normalized object space is divided into triangular patches. The depth value of the node is determined by applying the block based disparity compensation method and then the texture of the each patch is extracted by applying the affine transformation based disparity compensation method to the multi-view images.^[3,4] We restricted the number of images contributed to determining the texture compared to traditional methods^[1,3,4] which use all the multi-view images in the texture extraction. Experimental results show that the SNR of images encoded by the proposed algorithm is better than that of images encoded by the traditional method by the amount about 0.2dB for the test sets of multi-view images called *dragon*, *kid*, *city* and *santa*. The recovered images from the encoded data by the proposed method show the better visual images than the recovered images from the encoded data by the traditional methods.

Key words : Multi-view images, disparity, depth, texture

I. 서 론

지금은 음성, 영상, 데이터가 하나의 미디어로 처리되고 전송되는 멀티미디어의 시대다. 멀티미디어 관련 기술의 계속적인 발전은 1.2Gbps의 전송을 요구하는 고선명 텔레비전 방송을 가능하게 하였다. 이런 발전추세를

보면 고선명 텔레비전보다 더 많은 데이터량의 전송처리가 요구되는 3차원 텔레비전은 고도의 인간 감각에 좀더 다가갈 수 있는 영상을 보내어 새로운 문화를 창조해 가는 다음 세대의 미디어로 부각될 것이다.

3차원 영상코딩 기술은 스테레오 영상처리^[6]와 더불어 입체의 위치에서 보았을 때 영상을 재구성하는 것이 용이한 multi-view 영상처리에 관한 연구가 진행되었다.^[1] Multi-view 영상처리는 여러 장의 영상으로부터 깊이와 텍스처로 추출한 후 저장 또는 전송한 후 복원하는 영상 압축 기술이다. Fujii 등은^[1] 정규화 좌표계를 사용하여 물체를 표현한 후 블록에 기반한 변이 보상 방법을 사용하여 깊이와 텍스처를 동시에 구하거나, 아원 변환에 기반한 변이 보상 모델을 사용하여 깊이와 텍스처를 구하였다. 그리고 참고문헌 [3]과 [4]에서는 깊이와 텍스처를 추출하는 단계를 분리함으로써, 깊이와 텍스처 모두 지역적인 값에 의해 결정되고 artifact가 없는 복원영상을 얻는 알고리즘을 제안하였다. 이들 방법에서는 깊이를 구한 후, 주어진 깊이에서 텍스처를 추출할 때 해당되는 깊이에 관여한 모든 영상들의 평균값으로부터 결정하였다. 많은 수의 영상들이 텍스처 추출에 기여하므로 흐려진 영상을 얻게된다. 보다 선명한 상을 얻기 위하여, 본 논문에서는 텍스처를 깊이 결정에 관여한 영상 중 일부만을 사용하여 추출하는 알고리즘을 제안한다. 제안된 방법은 첫 채 정규화 공간에서 X - Y 면을 삼각패치로 나눈 후 깊이 추출 단계에서 각 노드를 둘러싸는 사각형 영역에 대하여 블록에 기반한 변이 보상 방법을 적용함으로써 노드의 깊이를 결정한다. 마지막 단계로 삼각패치의 텍스처는 아원 변환에 기반한 변이 보상 모델을 이용하여 깊이 결정에 관여한 영상 중 우선 순위가 높은 일부 영상으로부터 구한다.

제 2 장에서는 multi-view 영상과 multi-view 영상 압축 알고리즘을 기술하였고, 제 3 장에서는 본 논문에서 제안한 알고리즘에 대해 기술하였다. 실험 결과를 제 4 장에 설명하였고, 제 5 장에서 결론을 내렸다.

II. Multi-view 영상과 물체공간의 기하학적 관계

여러 대의 카메라를 사용하여 물체를 촬영하거나, 한 대의 카메라를 이동시키면서 물체를 촬영하였을 때, 동일한 물체에 대하여 여러 장의 영상이 얻어진다. 이 때 얻어진 여러 장의 영상을 multi-view 영상 세트라 한다. 그림 1은 dragon 물체를 촬영하여 얻어진 multi-view 영상 세트를 보여준다. 카메라는 +z축을 향하여 $(-2D, 0, z_0)$, $(-D, 0, z_0)$, $(0, 0, z_0)$, $(D, 0, z_0)$ 과 $(2D, 0, z_0)$ 에 놓여있다. 뷰포인트에서 +z 방향으로 f만큼 떨어진 곳에 상이 맺히게 되며, 상이 맺히는 x-y 평면을 영상 평면 x_i-y_i 라 한다.

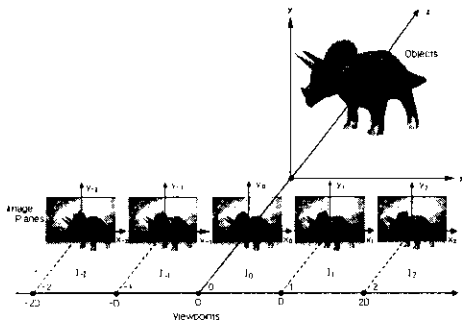


그림 1. Multi-view 영상 세트
Fig. 1. The set of the multi-view images.

그림 2는 multi-view 영상의 기하학적 관계를 보여준다. 그림 2에서 왼쪽 그림은 물체공간에서 multi-view 영상을 획득하는 과정이다. 카메라는 $y=0, z_0=-F$ 인 위치에 x축으로 D만큼씩 떨어져 있다. 그러므로 뷰포인트는 $V(nD, 0, -F)$ 로 표현되고, n은 정수이며 $n=0$ 이면 z 축상의 뷰포인트를 나타낸다. n번째 영상은 뷰포인트 $V(nD, 0, -F)$ 에서 얻어진 영상을 말한다.

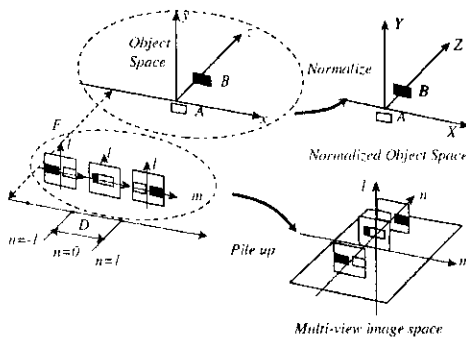


그림 2. 정규화 물체 공간과 multi-view 영상
Fig. 2. Normalized object space and multi-view images.

물체공간 좌표로 표현된 물체가 그림 2의 오른쪽 위 부분에 보인 것처럼 정규화 공간 좌표계로 변환되어 표

현되고, 정규화 공간상의 점 (X, Y, Z) 는 식 (1)과 식 (2)에 의하여 n번째 영상 $I_n(m, l)$ 에 투영된다.^[1,3] 식(1)은 multi-view 영상간의 좌표를 Z 픽셀(pixel) 만큼의 변이(disparity)로 정규화하는 것이다.

$$m = X + nZ \quad (1)$$

$$l = Y \quad (2)$$

그림 2의 오른쪽 아래 부분에 보인 것처럼, $I_n(m, l)$ 를 n 방향에 따라 배열하여 형성되는 n-m-l 공간을 multi-view 영상 공간(multi-view image space)이라 한다. Multi-view 영상 공간의 특징을 이용해 multi-view 영상의 압축과 복원을 수행한다.

그림 3은 multi-view 영상으로부터 깊이와 텍스처를 추출한 후 영상을 복원하는 과정을 보여준다.

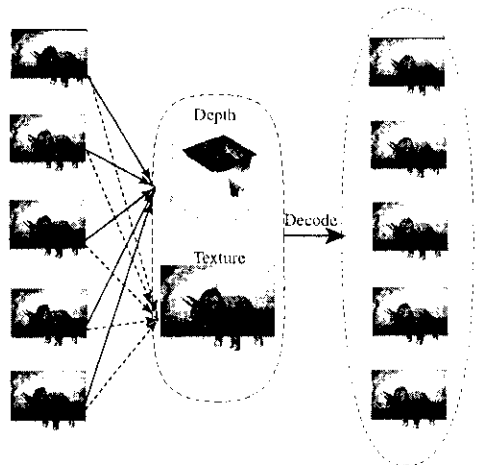


그림 3. 깊이와 텍스처 추출을 이용한 multi-view 영상의 압축과 복원
Fig. 3. Compression and Decompression of multi-view images by extracting the depth and the texture

III. Multi-view 영상에서 제한된 수의 참여 영상을 이용한 텍스처 추출 알고리즘

이 장에서는 multi-view 영상의 압축과 복원을 수행하기 위해 깊이와 텍스처를 추출하는 방법을 설명한다.

알고리즘 1은 제안한 multi-view 영상 압축 알고리즘의 pseudo code 이다. 첫 단계는 정규화 공간에서 $Z=0$ 인 X-Y 평면을 삼각패치로 나눈다. 둘째 단계는 깊이 추출 단계로 그림 4처럼 노드 d 깊이를 결정하기 위해서, 노드 d를 둘러싼은 사각형 영역 W_d 에 대하여 블록에 기반한 변이 보상 방법을 적용하여 가능한 Z 값에 대하여 식 (3)과 식 (4)에 의하여 휘도 분산값 $I_{var}(X, Y, Z)$ 을 구한다. 구하여진 휘도 분산값 중 가장 적은 휘도 분산값을 갖는 위치를 노드의 깊이로 결정한다. 각 삼각패치 노드의 깊이가 결정되었으므로 각 삼각패치의 텍스처는

새 번째 단계에서 결정된다. 정규화된 공간상의 각 삼각패치에 대하여 아핀변환을 이용하여 삼각패치에 대응되는 multi-view 각각의 영상에서의 삼각패치 p_1, p_2, \dots, p_n 을 구한다. 식 (3)을 이용하여 평균값 $I_{avg}(X, Y, Z)$ 을 구한다. 삼각패치 p_1, p_2, \dots, p_n 에 대하여 삼각패치의 휘도값과 평균값과의 차의 절대값 $Diff(p_i)$ 를 구한다. 절대값 $Diff(p_i)$ 가 작은 순서대로 삼각패치를 나열한 후, 정해진 수만큼의 삼각패치를 선택한다. 선택된 삼각패치에 대하여 식 (3)의 방법으로 평균값을 구한 후 정규화 공간의 삼각패치에 배정한다.

$$I_{avg}(X, Y, Z) = \frac{1}{(2N+1)} \sum_{n=-N}^N I_n(X+nZ, Y) \quad (3)$$

$$I_{var}(X, Y, Z) = \frac{1}{2N+1} \sum_{n=-N}^N (I_n(X+nZ, Y) - I_{avg}(X, Y, Z))^2 \quad (4)$$

Input : Multi-view 영상 세트

Output : 깊이와 텍스처

STEP1 : 정규화 공간에서 $Z=0$ 인 X - Y 평면을 삼각패치로 나눈다. (그림 4 참조) 삼각패치 노드의 $Z=0$ 이다.

STEP2 : 각 노드에서 노드를 둘러싼 사각형 영역(블록)에 대하여 노드의 Z 값을 변화시키면서 블록에 기반한 변이 모델을 이용하여 텍스처 분산값이 가장 적은 노드의 깊이, Z 값을 찾는다.

STEP3 : Multi-view 영상에서 깊이 결정에 관련한 영상 중에서 일부만 아핀 변환을 이용하여 각 삼각패치에 대응되는 텍스처들의 평균값을 구한 후, 삼각패치에 할당한다.

알고리즘 1 : Multi-view 영상처리 알고리즘의 슈도 코드.

Algorithm 1 : The pseudo-code of multi-view images coding algorithm.

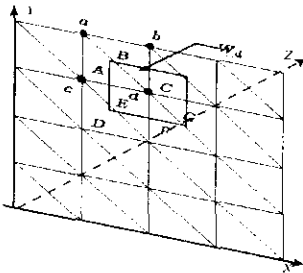


그림 4. Multi-view 영상처리에서 깊이 결정
Fig. 4. Decision of depth in multi-view images coding

IV. 실험 결과

본 논문에서 제안한 알고리즘을 실험하기 위하여 컴퓨터로 생성한 dragon 영상과 일본 Tsukuba 대학에서 제공한 실 영상 kid, city, santa 영상을 사용하였다. 각 영상은 9개의 256×256 그레이 영상으로 이루어져 있다.

본 논문에서는 Fujii의 방법과 9개의 참여 영상과 5개

의 참여 영상과 3개의 참여 영상으로부터 텍스처를 추출하는 실험을 하여 영상의 신호 대 잡음비(SNR)를 비교하였다.

$$SNR = 20 \log_{10} \frac{255}{\sqrt{MSE}} \quad (5)$$

여기서 MSE 는 복원 영상과 원래 영상의 에리 제곱의 평균을 나타낸다.

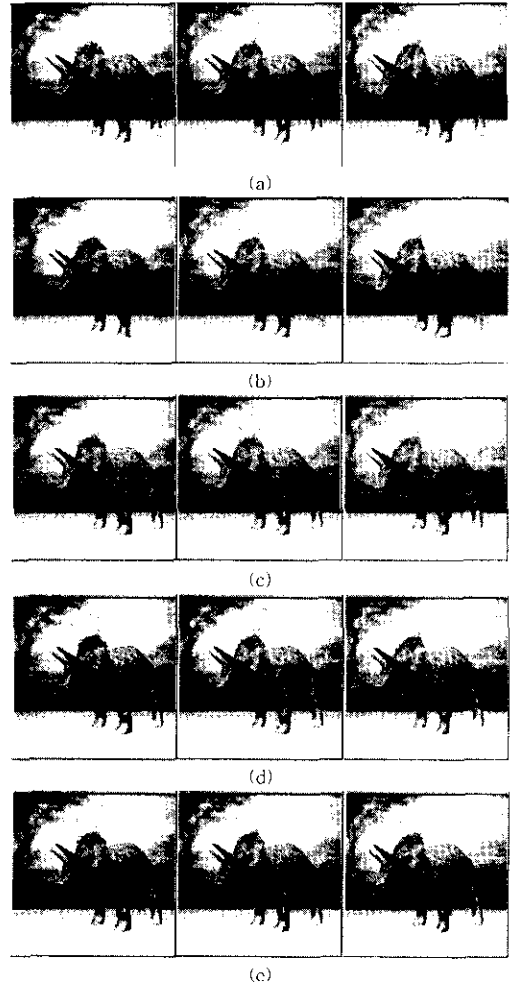


그림 5. Dragon multi-view 영상 세트 (a) 인코딩하기 전의 원 영상($n=-3, 0, 3$) (b) Fujii 방법에 의해 복원된 영상 (c) 9개 영상을 텍스처 추출에 사용하였을 때 복원된 영상 (d) 5개 영상을 텍스처 추출에 사용하였을 때 복원된 영상 (e) 3개 영상을 텍스처 추출에 사용하였을 때 복원된 영상
Fig. 5. Sets of the multi-view images called dragon, (a) the original images before encoding($n=-3, 0, 3$) (b) is the decoded images by Fujii method, and (c), (d), (e) is the decoded images according to the number of participated images(9, 5, 3) to obtain the optimal texture

그림 5는 dragon 영상에 대하여 Fujii 알고리즘과 제안한 알고리즘을 사용하여 압축한 후 복원한 multi-view 영상 중 $n=-3, 0, 3$ 번째 영상을 보여 주고 있다. 그림 5(b)는 Fujii가 제안한 알고리즘에 의해 복원된 영상을 보여 주고 있다. 그리고 그림 5(c), 그림 5(d), 그림 5(e)는 각각 참여 영상이 9개, 5개, 3개로 제한하여 텍스처를 추출한 후 복원된 dragon 영상을 보여준다. Table 1은 dragon 영상에 대하여 Fujii가 제안한 알고리즘과 텍스처 추출시 참여 영상의 수를 제한하여 압축 후 복원하였을 때 신호 대 잡음비를 나타낸다. 깊이와 텍스처 분리 단계의 알고리즘은 Fujii가 제안한 모델보다 dragon 영상에서 평균 신호 대 잡음비는 약 3dB 향상됨을 알 수 있다. 제안한 알고리즘에서 참여 영상이 5일 때 최적의 텍스처를 얻을 수 있었다. 참여 영상이 5일 때의 신호 대 잡음비는 참여 영상이 9일 경우 비해 0.2dB 정도 증가하였다. 텍스처 추출시 영상을 제한하여 얻어진 정보로부터 복원된 영상은 모든 영상으로부터 추출된 텍스처로부터 복원된 영상보다 영상의 화질이 개선됨을 관찰 할 수 있었다.

Table 1. Comparison of SNR(Signal-to-Noise Ratio) for the multi-view images called dragon according to the number of participated images to obtain the optimal texture

Number of image	SNR(dB)			
	Fujii 알고리즘	제안한 알고리즘		
		9개 영상참여	5개 영상 참여	3개 영상참여
-4	25	26	26	26
-3	26	28	28	28
-2	27	30	30	30
-1	29	33	34	34
0	30	34	36	37
1	30	33	34	34
2	28	31	31	30
3	26	29	28	28
4	24	27	26	26

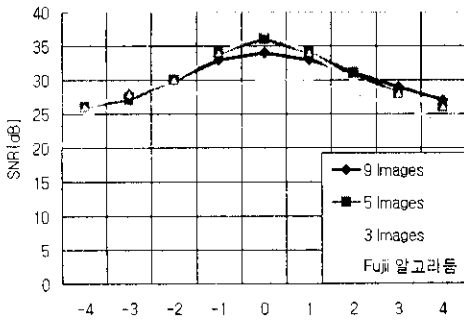


그림 6. 참여 영상의 수에 따른 dragon 영상의 신호 대 잡음비의 비교

Fig. 6. Comparison of SNR(Signal-to-Noise Ratio) for the set of the multi-view images called dragon according to the number of participated images to extract the texture

V. 결론

본 논문에서는 텍스처 추출시 제한된 수의 참여 영상을 이용한 개선된 multi-view 영상코딩 알고리즘을 제안하였다. 제안한 알고리즘은 Windows NT에서 Visual C++를 사용하여 구현하였고, 9개의 영상, 5개 영상 3개의 영상 참여로 이루어진 multi-view 영상 셋에 대하여 Pentium II를 사용하여 수행하였을 때 인코딩 할 때 약 12~15초 소요되고, 디코딩시 약 2초가 소요되었다. 그리고 참여 영상이 5일 경우 가장 좋은 화질을 얻을 수 있었다. 제안된 알고리즘에 의해 복원된 영상은 기존의 알고리즘으로 복원된 영상 비해 신호 대 잡음비가 약 0.2 dB 정도 향상되었다. 결론적으로 텍스처 추출시 영상을 제한하여 얻어진 정보로부터 복원된 영상은 모든 영상으로부터 추출된 텍스처로부터 복원된 영상보다 영상의 화질이 개선됨을 관찰 할 수 있었다

참고 문헌

- [1] T. Fujii and H. Harashima, "Data Compression and Interpolation of Multi-View Image Set," *IEICE Trans. Inf. & Syst.*, E77-D, 9, pp. 987-995, 1994.
- [2] M. E. Lukas, "Predictive Cording of Multi-Viewpoint Image Sets," *IEEE ICASSP'86*, pp. 521-524, 1986.
- [3] 김도현, "Multi-view 영상표현을 위한 개선된 알고리즘," *경상대학교 석사학위논문*, 1998.
- [4] 김도현, 양영일, "Multi-view 영상표현을 위한 개선된 알고리즘," *제 10회 영상처리 및 이해에 관한 워크숍*, pp. 3-8, 1998.
- [5] H. Aydinoglu and M. H. Hayes III, "Compression of Multi-View Images," *IEEE ICIP'94*, pp. 385-389, 1994.
- [6] U. R. Dhond and J. K. Aggarwal, "Structure from Stereo - A Review," *IEEE Trans. Systems, Man, Cybernetics*, vol. 19, no. 6, pp. 1489-1510, 1989.
- [7] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 333-340, July, 1980.
- [8] R. C. Bolles, H. H. Baker and D. H. Marimont, "Epipolar-Plane Image Analysis : An Approach to Determining Structure from Motion," *International Journal of Computer Vision*, 1, pp. 7-55, 1987.
- [9] Y. Ohta and T. Kanade, "Stereo by intra-and inter-scanline search using dynamic programming," *IEEE Trans. PAMI*, vol. 7, no. 2, pp. 139-154, 1985.