영상의 영역 분할을 이용한 다시점 영상 생성기법

*백윤기 *최미남 **허남호, **김진웅, *유지상

*광운대학교 전자공학과, **한국전자통신연구원

*byk100@image.gwu.ac.kr

The Conversion of 2 Dimension Images by Segmentation into Multi-View Images

*Baek, Yun-Ki *Choi, Mi-Nam **Hur Nam-Ho **Kim Jin-woon *Ji-Sang

*Kwangwoon University, **Electronics and Telecommunications Research Institute

요약

영상의 영역 분할을 통한 다시점 영상의 생성기법을 제안한다. 본 논문에서는 다시점 영상을 생성하기 위해 공간적으로 색상 정보와 시간적으로 움직임 정보를 이용하여 객체를 추출하고 이로부터 다시점 영상을 생성하게 된다. 색상 정보는 움직임 정보로는 정확한 외곽을 추출하지 못하기 때문에 정확한 외곽을 추출하기 위해 사용되어진다. 색상의 동질영역을 구분하기 위하여 휘도와 색차를 이용하며, 정합창을 사용하여 화소기반의 움직임 예측을 수행한다.

다음 단계로 색상 정보와 움직임 정보를 통해 독립적으로 얻어진 결과를 결합하게 된다. 움직임 예측을 통해 얻은 움직임의 세기값을 색상정보를 이용해 얻은 영역에 할당하고 이를 깊이값으로 변환하게 된다. 2차원의 입력 영상과 변환되어진 깊이값을 통해 회전변환의 과정을 거쳐서 최종의 다시점 영상이 생성된다. 실험을 통해서 제안된 알고리즘이 효과적으로 다시점 영상을 생성함을 확인할 수 있었다.

1. 서론

수요 변화와 사용자들의 욕구 증대로 인해 CRT를 이용한 TV, 평면 브라운관, LCD, PDP등의 소자들과 여러 영상 압축 기술들이 개발되었다. 그로인해 현실감이 증대되었고 보다 나은 영상 표시가 가능하게 되었다. 하지만 인간의 끊임없는 욕구는 보다 사실적이고 실감 있는 영상을 추구하게 되었다. 이러한 욕구는 자연스럽게 3차원 영상으로의 관심의 증대로 이어졌다. 3차원 영상에 대한 관심이 증대하면서 유럽에서는 3DTV에 대한 연구로 COST230(1992~1994) 프로젝트, PANORAMA(1995~1998), ATTEST(2002~2004)를 수행하였으며 3DTV(2004~2009) 프로젝트가 수행되고있다[1][2]. 또한 국내에서는 한국전자통신연구원이 2002 FIFA 한일 월드컵을 스테레오 동영상 방송 시스템을 구축하여 실험 방송을 하였다. 일본에서는 NHK의 3차원 HDTV 프로젝트 등 3차원 관련 분야에 대한 다양한 연구가 진행 중이며 MPEG(Moving Picture Experts Group)의 3DAV 그룹에서도 다시점 동영상 부호화 표준화 작업이 진행 중이다[3].

다시점 영상은 여러 시점에서 영상을 획득하기 때문에 시점이 한 곳에 고정되는 스테레오(stereo) 영상의 단점을 극복할 수 있다. 다시점 영상의 시점의 제약을 극복할 수 있는 특징은 3차원 영상의 디스플레이(display)에서 가장 중요한 기능이다. 하지만 다시점 영상은 시점의 증가에 따라 데이터양이 방대하게 커지는 단점이 있다.

이러한 문제점들을 해결하면서 효율적으로 스테레오 영상이나 다시점 영상을 얻기 위한 방법으로 depth 카메라(Camera)를 이용한 방법이 있다.[4]. depth 카메라를 이용하면 RGB 텍스쳐 데이터와 8 비트의 깊이 정보를 얻을 수 있다. depth 정보를 이용한 스테레오 영상 합성 기법에 대한 연구가 진행되었다[5]. 스테레오 영상이나 다시점 영상을 얻기 위해서는 depth 카메라에서 획득된 깊이 정보를 시차 정보로 변환하고 주어진 시점에 따른 적절한 스테레오 영상을 합성할 수 있는 알고리즘이 필요하다. 또한 depth 카메라의 보급은 아직 미흡한 편이고 그에 따른 영상 컨텐츠(Contents)도 부족하다.

본 논문에서는 2차원 영상을 영역 분할(Segmentation)을 통해 다시점 영상으로 변환하는 방법을 제안한다. 영역 분할을 위해 공간적으로 색상정보를 이용하고 시간적으로 움직임 정보를 이용하여 깊이 정보를 생성한다. 다음 단계로 회전(Rotation)의 기하학적 변환을 통해서 다른 시점의 영상을 생성하게 된다.

본 논문의 2장에서는 다시점 영상을 생성하기 위한 기법을 설명하며 3장에서는 실험결과를 확인, 분석하며 4장에서 결론을 내린다.

2. 제안하는 기법

제안하는 기법은 그림 1과 같이 영상의 영역 분할, 깊이 정보 생성, 다시점 영상 생성의 3단계로 나누어진다. 입력 영상을 색상 정보를 이용해 영역 분할을 한다. 다음 단계로 현재 영상과 이전 영상과의 움직임 예측을 통해 움직임의 세기를 구하게 된다. 마지막 단계로 영역 분할 된 영상과 깊이 정보를 이용하여 움직임의 세기값을 깊이 정보로 변환하고 결과 깊이 정보를 이용하여 입력 2차원영상의 회전변환을 통해 다시점 영상을 생성한다. 세부과정에 대해서는 다음 절을 통해서 설명하게 된다.

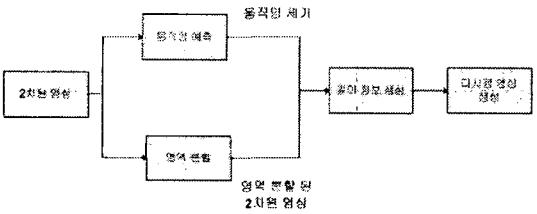


그림 1. 제안하는 기법의 블록도

가. 영역 분할

본 논문에서는 영역을 분할하기 위하여 색상 정보를 이용하였다. 색상 정보는 동질의 영역에 대한 정보를 포함하고 있는 정보로서 색상의 유사도는 영상을 분할하기 위한 척도로 사용되어진다[6].

색상 정보를 이용한 영역 분할의 목적은 객체의 외곽을 따라가기 위함이다. 정합창을 사용한 움직임 예측 시에 객체의 외곽보다 큰 움직임 영역이 구해지기 때문에 색상정보를 이용한 영역 분할 결과가 이를 보정해 주는 역할을 한다.

영역 분할은 영역 초기화단계, 영역 융합(region merging), 최소 화소이하의 영역처리 단계로 나누어진다. 초기화 단계에서는 2차원 영상의 화소 각각이 하나의 영역으로 초기화되어 진다.

그리고 영역 융합 단계에서는 유사한 색상을 가진 영역을 하나의 영역으로 융합하는 과정이 이루어진다. 유사한 색상을 구분하기 위한 척도로 본 논문에서는 휘도 값과 채도 값을 사용하였다. Y,U,V의 각각의 비용함수(Cost function)로 식 1을 사용한다.

$$C_Y(R_i,R_j)=\{Y(R_i)-Y(R_j)\}$$
$$C_U(R_i,R_j)=\{U(R_i)-U(R_j)\}$$
$$C_V(R_i,R_j)=\{V(R_i)-V(R_j)\} \quad \text{(식 1)}$$

$Y(R_i)$:영역 R_i의 Y성분의 평균값
$U(R_i)$:영역 R_i의 U성분의 평균값
$V(R_i)$:영역 R_i의 V성분의 평균값
$C_Y(R_i,R_j)$:영역 R_i와 R_j 사이의 Y성분의 비용값
$C_U(R_i,R_j)$:영역 R_i와 R_j 사이의 U성분의 비용값
$C_V(R_i,R_j)$:영역 R_i와 R_j 사이의 V성분의 비용값

식 1과 같이 계산된 모든 비용값(Cost value)이 설정되어진 오차값(error)보다 작다면 두 영역을 융합하는 과정이 수행되어진다. 식 2는 이 과정을 설명해준다.

$$C_Y(R_i,R_j)<error \cap C_U(R_i,R_j)<error \cap C_V(R_i,R_j)<error \quad \text{(식 2)}$$

2단계 과정을 거치고 나면 영역이 과도하게 분할된 결과를 얻을 수 있다. 하지만 과분할(Oversegmentation) 영역의 결과는 차후 과정인 영상에 깊이값을 부여하는 과정에서 오차를 발생하게 되는 요인으로 작용한다.

마지막 3단계에서는 영역의 화소의 수가 설정되어진 최소 화소보다 적은 수의 화소를 보유하고 있다면 수행되어진다. 먼저 영역에 이웃하고 있는 모든 영역과의 유사도를 비교하기 위하여 다음 식 3과 같이 비용값을 계산한다.

$$C(R_i,R_j)=C_Y(R_i,R_j)+C_U(R_i,R_j)+C_V(R_i,R_j) \quad \text{(식 3)}$$

그리고 각각의 영역은 이웃한 영역들과의 색상의 유사도를 비용값을 통해 비교하여 가장 작은 비용값을 가진 영역과 융합하는 과정을 정해진 최소의 화소 이상이 될 때까지 반복한다. 지금까지의 과정을 정리하면 그림 2와 같다.

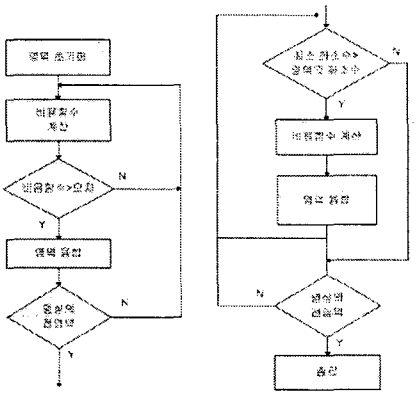


그림 2. 영역 분할 과정

나. 움직임 예측 및 깊이 정보 생성

본 논문에서는 영상에서 움직임을 예측하기 위하여 화소 기반의 움직임 예측을 한다. 화소 기반의 움직임 추정은 모든 화소가 변이를 갖는 방식으로 블록 기반의 움직임 예측보다 신뢰도 있는 세밀한 움직임 예측을 할 수 있다. 화소 기반의 움직임 예측을 하기 위해 찾고자 하는 화소의 주변영역의 값을 정합창(measurement window)으로 사용하였다. 움직임 예측 과정은 탐색 범위 내에서 정합창 내의 화소값의 비용값을 비교함으로써 이루어진다. 탐색 범위내의 가장 작은 비용값을 가지는 화소값을 시간상의 정합되는 점으로 간주한다. 화소기반의 움직임 예측을 정확히 하기 위해서는 정합창의 크기와 탐색 범위의 설정이 중요하다. 정합창의 크기는 정합되는 점의 신뢰도를 높일 수 있으나 정합창이 커지게 되면 계산량도 많아지고, 영상내의 객체의 외곽을 정확히 따라가기가 어렵게 된다. 또한 탐색 범위의 설정도 결과에 중요한 영향을 미친다. 너무 작거나 큰 탐색 범위의 설정은 잘못된 정합점을 찾을 가능성을 크게 만든다. 또한 수평의 움직임이 있는 영상에서는 수평방향으로의 탐색범위가 정합점을 찾을 가능성을 높여주고 수직 움직임이 있는 영상에서는 수직 방향으로의 탐색 범위가 정합점을 찾을 가능성을 높여준다. 본 논문에서는 영상의 움직임의 유형에 따라 수평방향과 수직방향, 정방향의 탐색범위를 사용하였다.

지금까지의 과정은 움직임 추정을 하기위한 과정이었다. 움직임 값을 깊이정보로 변환하기 위하여 식4의 과정을 거친다.

$$D(x,y)=Motion_x+Motion_y \quad \text{(식 4)}$$

여기서 $D(x,y)$는 화소 x, y의 깊이값이다. 그리고 $Motion_x$는 x방향으로의 움직임의 세기(intensity)이고, $Motion_y$는 y방향으로의 움직임의 세기이다.

$D(x,y)$는 많은 오차값들을 가지고 있고 한 영역 내에서 화소들 간에 편차가 크기 때문에 다른 처리가 필요하다. 본 논문에서는 구한

$D(x,y)$ 값들을 한 영역 내에서는 하나의 값을 가지도록 영역내의 화소값의 평균값을 구해서 한 영역 내에서는 같은 값을 가지도록 한다. 이 과정은 식 5이다.

$$D(R_i) = \frac{1}{m \times n} \sum_{k=0}^{m \times n} D(k) \qquad \text{(식 5)}$$

식 5에서 $D(R_i)$ 는 $m \times n$ 화소를 가진 I번째 영역에서의 깊이값을 의미하며 $D(k)$ 는 영역의 k번째 화소의 깊이값이다.

영상에서 물체나 카메라의 움직임은 일정하지가 않다. 그렇기 때문에 위 과정의 결과를 깊이 정보로 사용하게 되면 다른 시점의 영상을 생성하였을 경우 영상이 끊기는 느낌을 받게 되거나 영상이 깨지는 것 같이 느끼게 된다. 위와 같은 현상을 방지하기 위해 논문에서는 과거 영상과 미래 영상 그리고 현재 영상의 평균값을 취하게 된다. 지금까지의 과정을 정리하면 그림 3과 같다.

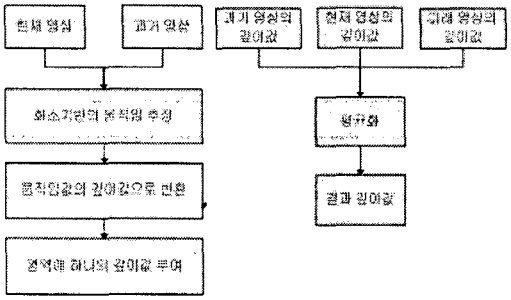


그림 3. 움직임 추정 및 깊이 정보 생성

다. 다시점 영상 생성

그림 4는 디스플레이 장치를 바라보는 관찰자의 위치를 나타낸다.

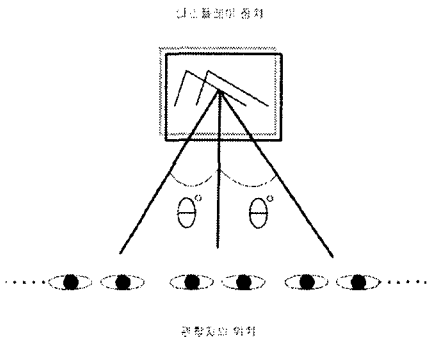


그림 4. 디스플레이 장치를 바라보는 관찰자의 위치

관찰자의 위치 이동에 따른 각을 Θ도라 하면 새로운 위치의 카메라에 투영되는 영상은 Θ도 회전된다. Θ도 회전된 영상은 다음 식 6에의해서 생성된다. $(x_\theta, y_\theta, z_\theta)$ 는 Θ도 회전된 영상의 좌표, (x, y, z) 는 원 영상의 좌표를 의미한다.

$$\begin{bmatrix} x_\theta \\ y_\theta \\ z_\theta \end{bmatrix} = R(\theta) \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad \text{(식 6)}$$

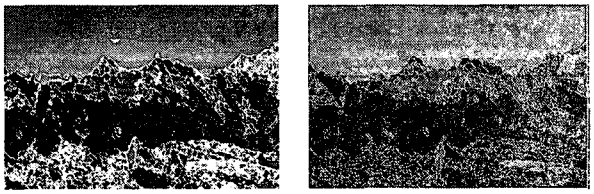
$R(\theta)$ 는 식 7의 회전 변환 행렬이다.

$$R(\theta) = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \qquad \text{(식 7)}$$

3. 실험결과 및 분석

실험으로 영역분할, 움직임 예측 및 깊이 정보 생성의 각 과정과 생성되어진 깊이 정보를 이용한 다시점 영상 생성 결과를 분석한다.

그림 5는 영역분할 결과 영상이다. 오차값은 3으로 설정 되었으며 영역의 최소화소 수는 150화소이다. 그림 5 (a)는 최소화소 적용이전의 결과 영상으로 비용함수만을 계산한 영역 융합의 결과이다. 영상에서 위쪽 하늘영역은 거의 하나의 영역으로 융합되었고 밑쪽의 복잡도가 높은 영역은 많은 미소 영역(small region)이 존재한다. 그림 5 (a)와 같은 결과는 미소 영역이 존재하기 때문에 영역 분할 결과로는 부적합하다. 그림 5 (b)는 최소화소의 설정을 통해서 미소 영역을 일정 수준 이상으로 융합한 결과이다. 그림 5를 통해서 영상에서 유사한 색상의 영역이 하나의 영역으로 융합 된 것을 확인 할 수 있다.



(a) 최소화소 적용이전 결과 (b) 최소화소 적용이후 결과

그림 5. 영역분할 결과 영상

다음의 과정은 현재 영상과 과거 영상의 화소 기반 움직임 예측이다. 그림 6은 움직임 예측 결과이다. 본 논문에서는 10×10의 정합창과 10×2의 탐색범위를 설정하였다.

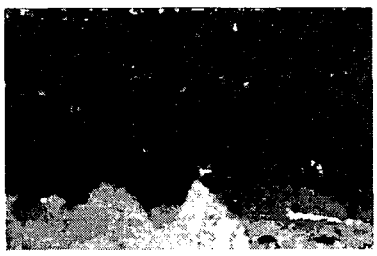
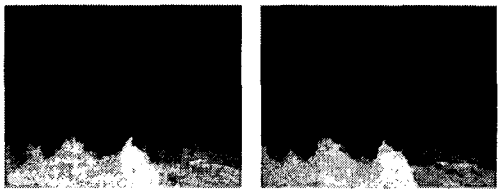


그림 6 . 움직임 추정 결과

움직임 예측 후 결과를 깊이값으로 변환한다. 그림 5 (a), 6으로 그림 7의 영상을 얻는다. 그림 7은 하나의 영역에 그 영역의 깊이값의 평균값이 부여된 결과이다.



(a) 1번 프레임 (b) 2번 프레임

(c) 3번 프레임 (d) 4번 프레임

그림 7. 깊이값을 부여한 결과

그림 7은 잡음이 있고 그림 7 (c)와 같이 카메라나 물체의 움직임의 세기의 변화, 움직임 예측의 오차로 인해 불연속적인 결과를 얻게 된다.

본 논문에서는 그림 7과 같이 연속된 영상에서 변화가 큰 프레임의 문제를 처리하기 위해서 현재 영상과 과거, 미래 영상과의 평균값을 사용하였다. 그림 8은 그림 7의 3,4번 프레임의 결과 영상이다.

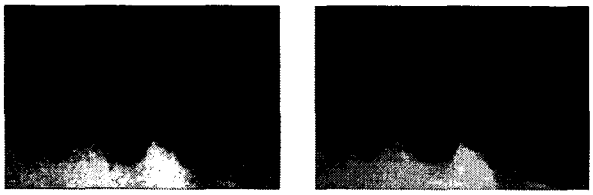


그림 8. 평균값이 부여된 결과

그림 9는 그림 8의 결과 깊이 영상과 입력 영상을 회전변환하여 얻은 9시점의 영상이다. 그림 10은 9시점의 영상 중 -4도 영상과 +4도 영상의 차영상으로 앞으로 나와 있는 부분이 시차가 크다는 것을 확인할 수 있다.

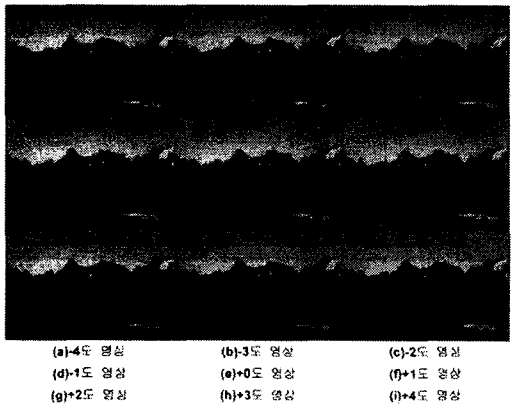


그림 9. 9시점의 결과 영상

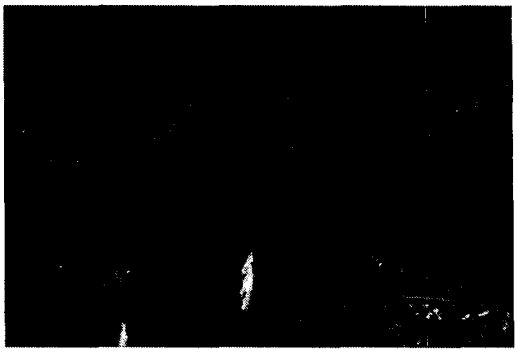


그림 10. -4도 +4도 영상의 차영상

4. 결과

본 논문에서는 2차원의 영상에서 다시점 영상의 생성기법을 제안하였다. 다시점 영상 생성을 위해서 영역분할과 움직임 추정을 하였다. 움직임 추정의 결과를 깊이정보로 사용하기 위해서 변환하였으며 변환되어진 깊이정보와 입력 2차원영상을 회전변환을 하여 다시점의 영상을 생성하였다. 영역을 분할하기 위하여 색상성분을 이용하여 원하는 최소한의 화소이상의 영역이 되게 설정하였으며 움직임 추정을 위해 정합창을 이용한 화소단위의 움직임 예측을 통해서 보다 정확한 움직임 예측을 하였다. 또한 분할 된 영역에 깊이 정보를 부여하고 단일 시점의 영상에서 얻은 움직임 값을 사용함으로써 생길 수 있는 변화가 심한 현상을 과거, 미래 영상을 사용하여 방지하였다. 실험을 통해서 각 과정의 결과를 확인하였고 최종 결과를 다시점 모니터 (SynthaGram 222)를 통해 확인함으로써 다시점 영상이 생성 된 것을 확인하였다.

<감사의 글>

본 연구는 정보통신부의 연구사업인 "지능형 통합정보방송 (SmarTV) 기술 개발" 사업 지원으로 수행되었습니다

<참조 문헌>

[1] R. Franich, R. Lagendijk and R. Horst, Reference model for hardware demonstrator implementation, RACE DISTIMA deliverable 45/TUD/IT/DS/B/003/b1, October, 1992

[2] S. Malassiotis, M. G. Strintzis, Coding of video-conference stereo image sequences using 3D models, Signal Processing: Image Communications vol. 9, no. 1, pp 125-135, January, 1997

[3] MPEG/ISO/IEC JTC1/SC29/WG11 "Subjective test results for the CfP on Multi-view Video Coding," N7779, Bangkok, Thailand, Jan. 2006.

[4] ATTEST : Advanced Three-Dimensional Television System Technologies, http://eigg.res.cse.dmu.ac.uk/attest/ 2002 - 2004

[5] C. Fehn,"Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV", In Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI, pages 93-104, San Jose, CA, USA, January 2004

[6] S. Cooray, N. O'Connor, S. Marlow, N. Murphy, T. Curran, "SEMI-AUTOMATIC VIDEO OBJECT SEGMENTATION USING RECURSIVE SHORTEST SPANNING TREE AND BINARY PARTITION TREE" WIAMIS 2001 - Workshop on Image Analysis for Multimedia Interactive Services, Tampere, Finland, 16-17 May 2001