일반논문-08-13-6-13

다시점 동영상의 중간시점영상 생성을 위한 변이 예측 기법

최 미 남^{a)}, 윤 정 환^{a)}, 유 지 상^{a)‡}

Disparity Estimation for Intermediate View Reconstruction of Multi-view Video

Mi-Nam Choi^{a)}, Jung-Hwan Yun^{b)}, Ji-Sang Yoo^{a)‡}

요 약

본 논문은 다시점 카메라로부터 획득된 영상을 이용하여 영상내의 모든 화소에 대한 정확한 변이 정보를 구하는 알고리즘을 제안한다. 제안한 방법은 객체의 경계 정보를 고려하여 초기 변이를 예측한 후 획득된 변이 정보를 이용하여 탐색 범위를 줄임으로써 효율적으로 변이를 예측한다. 또한 가변 블록을 사용하여 텍스처 정보가 부족한 영역과 경계부분에서 발생하는 오정합 문제를 줄일 수 있다. 획득된 변이 맵 정보를 이용하여 중간시점영상을 생성한 결과 기존의 블록기반 변이 추정방식과 화소기반의 변이 예측방식에 비해 0.1dB ~ 1.2dB의 PSNR(Peak signal to noise ratio)이 향상되는 것을 확인하였다.

Abstract

In this paper, we propose an algorithm for pixel-based disparity estimation with reliability in the multi-view image. The proposed method estimates an initial disparity map using edge information of an image, and the initial disparity map is used for reducing the search range to estimate the disparity efficiently. Furthermore, disparity-mismatch on object boundaries and textureless-regions get reduced by adaptive block size. We generated intermediate-view images to evaluate the estimated disparity. Test results show that the proposed algorithm obtained 0.1 ~ 1.2 dB enhanced PSNR(peak signal to noise ratio) compared to conventional block-based and pixel-based disparity estimation methods.

Keyword : 다시점 비디오, 변이 예측, 경계 정보, 정합창

1. 서 론

디지털 IT의 바람은 정보 고속도로와 무선통신의 진보에 힘입어 대화형 TV, 네트워크를 통한 정보 가전, 센서 네트워크 등으로 발전해 가고 있다. 이와 더불어 이동통신의 이

용 또한 국민들 사이에 깊이 침투하여 사람들의 생활에 있어서 “당연”하고 “없어서는 안 될” 공기와 같은 생활의 기본 요소가 되어가고 있다. 이처럼 우리 사회는 정보통신이 공거나 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].

3DAV는 세계적 비디오 표준화 단체인 MPEG(Moving

a) 광운대학교 전자공학과

Dept. of Electronics Engineering, Kwangwoon Univ.

‡ 교신저자 : 유지상(jsyoo@kw.ac.kr)

※ "이 논문은 2007년도 광운대학교 교내 학술연구비와 한국과학재단 특정기초연구(R01-2006-000-10199-0(2008))로 수행되었음."

Picture Expert Group)에서 최근 부각되고 있는 다시점 비디오 및 3차원 오디오/비디오 기술과 관련하여 2002년 12월 58차 회의부터 명칭한 이름이다. 3DAV 그룹의 비디오 부분 활동과 관련해서는 3가지 기능 제공을 표준화 활동의 목표로 하고 있다. 첫째로 사용자에게 영상을 관찰함에 있어서 시점의 자유를 제공하고, 둘째로는 객체 기반 상호작용성 제공을 포함한다. 셋째로는 영상기반 렌더링을 이용한 3D 객체 생성기능을 제공하는 것이다.

이러한 현실감, 입체감을 제공할 수 있는 3DAV 콘텐츠 제작을 위해서는 기본적으로 양안식/다시점 영상 기술 개발이 필요하다. 양안식/다시점 영상이란 두 대 혹은 그 이상의 카메라로부터 획득된 영상을 말한다. 한 대의 카메라로 촬영하여 사용자에게 제한된 시점을 제공하는 기존의 방법에서 벗어나 여러 대의 카메라로부터 영상을 획득하여 사용자가 원하는 시점의 영상을 보여준다. 현재 다시점 카메라는 스포츠 중계부분과 광고/영화의 특수효과로 사용되고 있다. 예를 들어 영화 'MATRIX'에서 공간적으로 배치된 다시점 카메라에서 획득한 영상을 시간적으로 재배치하여 시간이 정지된 듯한 형태의 영상물을 제작하는 것이 가능하였다. 미국 CMU에서 개발한 eye-vision도 앞의 예와 동일한 영상물을 제작하여 미식축구에서 주요 장면을 재구성하여 시청자들에게 보여줌으로써 고도의 기술적 효과를 자랑하였다. 또한 일본의 ponycanyon사는 쇼·오락분야에 관련된 영상을 12에서 17대의 카메라로부터 획득하고 사용자에게 제공함으로써 더욱 현실감 있는 서비스를 제공하였다.

그러나 현실감과 입체감을 제공하는 양안식/다시점 영상은 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 기법은 시점의 자유도와 관찰자의 위치에 따라 다른 입체감을 제공한다^[2]. 또한 적은 수의 카메라로 영상을 획득하였다 하더라도 훨씬 많은 수의 시점의 영상을 생성할 수 있기 때문에 데이터 량 측면에서 매우 효율적이다. 이러한 중간시점영상 생성기술은 변이 예측, 가려진 영역 처리, 실제 영상생성으로 나눌 수 있다. 이 중 가장 중요한 과정은 각각의 입력받은 영상으로부터 서로 대응되는 점을 찾는 변이 예측 과정이다. 변이 예측이 얼마

만큼 잘 이루어졌는가에 따라 생성되는 영상의 질이 결정된다. 본 논문에서는 5대 이상의 카메라를 사용하여 보다 정확한 변이 정보를 추출하여 중간시점영상을 생성하는 기법을 제안한다. 정합창(measurement window)을 사용하는 화소기반 정합방법을 통해 변이를 추정하며, 영상에서 객체의 경계정보를 이용하여 변이를 구함으로써 경계부분에서의 오류를 줄일 수 있다. 또한 객체의 경계정보를 이용하여 얻어진 초기 변이를 통해 탐색 범위를 줄임으로써 변이 추정 속도를 향상 시킨다. 최종적으로 획득된 변이 정보를 이용하여 다시점 영상의 중간 시점 영상을 생성한다.

논문의 구성은 다음과 같다. 2장에서는 기존의 연구와 관련된 기술에 대해 설명하고, 3장에서는 제안하는 변이 예측 기법에 대해 설명한다. 4장에서는 실험결과를 확인 및 분석하고, 5장에서는 본 논문의 결론과 향후 연구 방향을 기술한다.

II. 다시점 동영상 서비스를 위한 입체감 및 관련 기술

다시점 동영상 서비스는 실사 영상에서 전경과 배경의 깊이감을 느낄 수 있으며, 관찰자의 움직임에 따라 해당되는 시점의 영상을 제공할 수 있어야 한다. 이를 위해서 인간 시각 시스템(human visual system : HVS)에 기인한 입체감의 요소를 분석하고, 이에 적합한 효과를 제공해 주어야 한다. 입체감을 느낄 수 있는 요소는 크게 단안 렌즈의 초점 조절에 의한 입체감, 양안에 의한 입체감, 경험적 요인에 의한 입체감으로 구분될 수 있다. 현존하는 기술 중에 이러한 입체감의 요인을 만족시킬 수 있는 기법으로는 홀로그램(hologram)이 있지만, 아직까지 데이터의 획득이나 전송, 표현 등과 같은 문제점을 해결하여야 한다. 일반적인 입체 디스플레이 장치에서는 양안 시차(binocular disparity)를 이용한 방법이 가장 많이 활용되고 있다. 양안시차를 제공하기 위해서는 카메라 두 대를 설치하여 영상을 획득하고, 최종적으로 스테레오 모니터를 사용하여 입체 영상을 감상할 수 있다^[3].

그림 1. 평행식 카메라 모델
Fig. 1. Parallel camera model

1. 평행식 스테레오 카메라 모델

일반적으로 양안 시차의 원리가 입체 디스플레이 장치에서 가장 많이 활용되고 있다. 양안시차를 제공하기 위해서는 인간의 눈과 유사한 구조로 카메라 두 대를 설치하여 영상을 획득한다. 그림 1은 평행식으로 구성된 스테레오 카메라 모델을 나타낸다. 그림에서 L은 좌 영상, R은 우 영상을 나타내고 b는 기선(base line)을 f는 초점거리(focal length)를 나타낸다. Z는 카메라의 상에 투영되는 공간상의 한 점과 카메라 사이의 거리 즉 깊이를 나타낸다. 평행식 카메라 모델에서 공간상의 한 점은 다음 식 (1)의 관계에 의해 좌, 우 카메라에 투영된다.

$$x_l = f \frac{x'_l}{Z_l}, \quad x_r = f \frac{x'_r}{Z_r} \quad (1)$$

여기서 (x'_l, y'_l, Z_l) 는 좌측 카메라 렌즈의 중심축이 기준인 공간상의 한점의 좌표를 나타내고, (x'_r, y'_r, Z_r) 는 우측 카메라 렌즈의 중심축이 기준인 공간상의 한 점의 좌표를 각각 나타낸다. 평행식 카메라 모델인 경우 좌, 우 카메라의 좌표계의 차이는 횡축 좌표에만 존재한다. 최종적으로 깊이 정보를 나타내는 Z는 식 (2)를 통해 구할 수 있다

$$Z = \frac{bf}{d} \quad (2)$$

d는 $x'_l - x'_r$ 로서 변이가 되고 변이와 z는 반비례 관계에 있는 것을 알 수 있다. 변이를 알게 되면 깊이 정보를 알 수 있고, 이를 이용해 다른 시점의 영상에서의 좌표를 찾아 낼 수 있다. 양안 시차를 이용한 스테레오 영상을 인간의 두 눈에 각각 보여줌으로써 입체감을 느낄 수 있다.

2. 변이 예측 방식

실제 영상에서 깊이 정보를 얻는 방법에 대하여 많은 연구가 진행되어 왔다. 그 중에서 스테레오 정합 방법은 공간상의 같은 한 점이 스테레오 영상에서 서로 다른 좌표에 맷히는 이른바 좌우 영상간의 화소의 위치 차이를 이용하여 깊이 정보를 얻어내는 방법이다. 3차원 영상처리에 있어서 가장 중요한 문제 중 하나는 동일한 물체를 서로 다른 위치의 카메라로부터 획득된 두 장의 영상 내에서 정합점을 찾는 변이 예측(disparity estimation)이다. 변이란 깊이감을 나타내는 시차를 나타내는 것으로서, 실제 영상 좌표상에서 기준이 되는 좌우 영상 사이에 정합점들의 좌표 차를 나타내며, 변이 예측 과정은 중간 영상 합성 방식의 결과에 가장 중요한 영향을 줄 수 있는 부분이다. 스테레오 정합 과정은 먼저 두 영상 중에서 한 영상을 기준영상으로 정하고 다른 영상을 대상영상으로 정한 뒤 기준영상의 임의의 화소를 대상영상에서 찾는다. 기준영상에서의 화소와 대상영상에서 정합된 화소간의 좌표차이가 변이가 된다. 스테

레오 정합 방법을 확장하여 다시점 영상에 적용하면 다시점 영상에서의 변이를 예측할 수 있다. 변이 예측 방법으로는 영상의 특징을 찾고 그 특징간의 정합을 하는 특징 기반 방식, 영상을 일정 크기의 블록으로 나누어 블록 단위로 변이를 예측하는 블록 기반의 방식, 영상을 여러 다각형 조각으로 나누어 다각형의 꼭지점을 중심으로 변이를 예측하는 메쉬(mesh) 기반 방식 등이 있다.

2.1 블록 기반 정합

블록 기반의 정합 방식은 영상을 고정 크기의 블록으로 나누어 변이를 예측하는 방식이다. 이는 MPEG-1/2, H.26x와 같은 동영상 압축 표준에서 움직임 예측 시 사용되는 기법과 동일하다. 기준이 되는 영상을 블록 단위로 나누어 다른 한 영상에서 비용함수(cost function)를 만족하는 위치를 찾아 이동량을 그 블록의 변이 값으로 결정한다. 비용함수로는 일반적으로 평균절대오차(mean absolute error : MAE), 평균제곱오차(mean square error : MSE), 상관도(cross correlation)등이 사용된다^[4]. 하지만 평균제곱오차의 경우에는 하드웨어적으로 제곱연산 구현의 복잡성 때문에 잘 활용되지 않고 있다. 식 (3)은 블록 정합에 사용하는 평균절대오차를 나타낸 것이다.

$$MAE = \frac{1}{N \times M} \sum_{i=1}^N \sum_{j=1}^M |I_L[i, j] - I_R[i + d_k, i + d_l]| \quad (3)$$

여기서 IR 는 예측하고자 하는 우 영상, IL 은 좌영상이고, N 과 M 은 블록의 가로, 세로 크기를 말한다. 블록 기반의 정합 방식은 블록의 크기와 탐색 범위에 따라 변이 예측 결과에 많은 영향을 미친다. 블록의 크기가 커지면 다른 변이를 가지는 화소가 동시에 블록 내에 존재 할 수 있으므로, 변이가 부정확해지고 블록화 현상이 커진다. 반대로 블록의 크기가 작으면 세밀한 시차벡터를 얻을 수 있는 반면, 신뢰도가 감소할 뿐 아니라 변이의 정보가 증가한다.

2.2 화소 기반 정합

화소 기반의 정합 방식의 과정은 블록 기반의 정합 방식과 비슷한 방법으로 화소 단위로 주변 영역의 화소 값들을

고려하여 변이를 찾는 방식이다. 주변 영역을 가리켜 정합창(measurement window)이라 하고, 정합창 내의 화소 값들이 비용함수를 만족하는 화소를 찾아 변이를 예측한다. 화소 기반 정합 방식은 세밀한 정합은 가능하지만 수행시간이 길다는 단점이 있다. 블록 기반의 정합 방식과 화소 기반의 정합 방식은 화소들의 밝기 값을 이용하여 변이를 예측한다. 화소의 밝기 값만으로 변이를 예측하게 되므로 영상의 광도 변화나 물체의 경계 부분에서 정확한 변이를 예측하기 어렵다.

2.3 특징 기반 정합

특징 기반의 정합 방식은 밝기 값 대신 영상의 특징을 사용하여 변이를 예측하는 방식이다. 이때 사용되는 영상의 특징으로는 영점 교차(zero crossing)의 부호, 경도의 첨두(gradient peak), 영역, 선, 경계 등이 있다. 특징 기반 방식은 영상의 종류에 따라 민감하게 반응하게 되고, 영상의 전체 영역에 대한 변이 예측은 불가능하며, 전체 영상에 대한 조밀한 변이를 구하기 위해서는 내삽과정(interpolation)을 거쳐야 하는 단점이 있다^[5].

2.4 메쉬 기반 정합

메쉬 기반의 정합 방식은 시차가 있는 물체를 블록이 아닌 임의의 형태로 분할하여 각 분할 영역에 대해 평행 이동만으로 시차를 보상하는 대신 회전, 확대, 축소, 변형 등과 같은 공간 변환을 하여 변이를 예측한다. 즉, 영상을 다각형 조각으로 나누어 다각형 조각의 꼭짓점들을 중심으로 변이를 예측하는 방식을 메쉬 기반의 정합 방식이라 한다^[6]. 메쉬 기반의 정합 방식은 메쉬의 형태를 결정하는 특정한 과정이 필요하며, 절점 정보만으로 불규칙 삼각망을 생성하는 델로니 삼각망(delaunay triangulation)형성 방식이 대표적이다. 블록화 현상이 없이 효율적인 부호화가 가능하지만 잘못된 절점의 추출로 잘못된 변이 예측이 발생할 수 있다.

2.5 영상 분할 기반 정합

영상 분할 기반의 정합 방식은 영상이 독립적인 객체의 조합으로 이루어져 있다는 가정하에 영상을 객체 단위로

분할하여 변이를 예측하는 방식이다^[7]. 영상을 유사한 특성을 갖는 영역으로 분할(segmentation)하여, 분할된 영역에 대하여 변이를 예측하고, 이와 함께 윤곽선을 전송한다. 영상의 분할된 영역 단위로 변이를 예측할 경우 블록화 현상이 발생하지 않는 장점이 있으나, 분할된 영역에 대한 정보로 인해 데이터 양이 커지는 단점이 있다.

2.6 변이 공간 기반

변이 공간 영상에서의 정합 방식은 변이 공간 영상(disparity space image)에서 최소의 비용 경로를 찾아 정합하는 방법이다. 변이 공간 영상이란 좌영상과 우영상의 밝기값 차로 표현되는 영상을 말한다. 최소 비용의 경로는 동적 계획법(dynamic programming method)에 의해 구하며, 구해진 정합의 경로를 변이로 변환한다. 이 방식은 비용 경로가 연속적인 특성으로 인해 기존 방법들보다 정확하고 안정된 결과를 나타낸다. 그러나 계산 시간이 오래 걸리는 단점이 있다. 따라서 정확한 변이 예측 및 빠른 처리 속도를 갖는 방법의 개발이 필요하다^[8].

3. 다시점 동영상의 변이 예측 기법

스테레오 영상의 경우에는 시점이 고정되어 있고, 다시점 영상의 경우 시점의 수는 증가하지만 시점의 수에 비례하여 카메라 수가 증가하고 관찰자의 시점 이동시 시점의 불연속성이 발생하는 단점이 있다. 이러한 문제를 해결하기 위해 다시점 영상을 이용하여 변이 예측을 하고, 중간 시점 영상 합성 방법을 사용하여 다시점 스테레오 영상을 생성하여 시점의 자유도와 입체감을 줄 수 있다.

III. 제안하는 변이 예측 기법

제안하는 변이 예측 기법의 전체 구조는 그림 2와 같다. 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

경계 정보를 사용한다. 변이 예측의 연산량을 줄이기 위해 초기 변이 정보를 이용하여 변이 예측의 탐색 시작점을 결정한다. 변이 정보를 사용해 중간 시점영상을 생성하여 결과를 확인하였다.

그림 2. 제안하는 변이 예측 기법

Fig. 2. Proposed disparity estimation method

1. 초기 변이 예측

그림 3은 초기 변이 예측 기법의 흐름도를 나타낸 것이다. 좌, 우 입력 영상으로부터 경계 정보를 추출하고 경계 정보의 유무에 따라 라인별로 가변적인 블록을 생성한다. 경계 정보에 따라 가변적인 정합블록을 사용하게 되면 텍스처(texture)가 없는 영역의 경우 잘못된 정합이 줄게 되며, 경계 영역에 대한 정확한 변이를 예측할 수 있다. 이후에 변이 예측에 있어서 탐색 시작점을 예측하는 중요한 정보로 사용된다. 정합이 이루어진 후에는 정합된 두 영역의 MAD(Mean Absolute difference)값을 분석하여 잘못된 정합에 의해 결정된 불확실한 변이 여부인지를 판단하고 불확실한 변이로 판단될 경우 동일 영역 내에서의 변이가 유사한 성질을 이용하여 주변 영역의 변이 벡터를 사용하여

그림 3. 초기 변이 예측

Fig. 3. Initial disparity estimation

그림 4. 경계 정보의 추출 : (a) 소벨 마스크 (b) Teddy 영상 (c) 경계 정보 추출

Fig. 4. Edge extraction (a) Sobel mask (b) Teddy image (c) Edge-extracted image

그림 5. 라인별 가변블록의 생성 과정

Fig. 5. Decision of adaptive block-size

보간한다.

적응적인 정합 블록을 생성하기 위해 가장 먼저 영상의 경계 정보를 추출한다. 경계 정보를 추출하는 데는 일반적으로 널리 사용되는 Sobel 마스크를 사용하였다. Sobel 마스크는 에지 추출의 가장 대표적인 1차 미분 연산자로써 모든 방향의 에지를 추출한다. 정합 블록을 결정하기 위해 수평 라인만 고려하기 때문에 Sobel 마스크 중에서 수직성분의 마스크를 사용한다. 또한 가변 블록을 결정하는 데에 있어서 경계 정보의 잡음 성분에 의한 영향을 줄이기 위해 3X3 블러링(blurring) 마스크를 적용하였다. 그림 4는 사용된 Sobel 마스크와 영상의 경계 정보 추출 결과이다.

그림 5는 영상으로부터 추출된 경계 정보의 유무에 따라서 라인별 가변 블록의 생성 과정을 나타낸 그림이다. 이때 불필요한 블록의 생성을 방지하기 위해서 작은 블록의 경우 주변 블록으로 합치게 된다.

가변블록을 생성한 후 한쪽 영상에서 다른쪽 영상으로의 정합점을 찾는 정합블록으로 사용한다. 초기 변이를 예측하기 위한 비용함수는 다음의 평균절대오차(Mean Absolute Error)를 사용한다.

$$\frac{1}{N} \sum_{k=1}^N |I_r(i, j) - I_r(i + k, j)| \quad (4)$$

그림 6. 초기 변이 예측 : (a) Teddy image (b) 가변블록 (c) 초기 변이

Fig. 6. Initial disparity estimation : (a) Teddy image (b) Adaptive block-size (c) Initial disparity map

I 은 좌영상 I_r 은 우영상을 의미하며 k 는 변이 벡터, N 은 블록의 크기를 나타낸다. 식 (4)에 의해 최소 비용 값을 갖는 k 를 초기 변이로 결정한다.

그림 6은 경계 정보를 이용한 가변 블록의 생성 및 초기 변이 결과를 나타낸다.

초기 변이 예측은 영상의 경계 정보를 이용하여 객체별로 변이를 예측하여 경계 부분에서의 오류를 줄일 수 있고, 텍스처(texture)가 없는 부분에 대해 강인한 결과를 얻을 수 있다. 그러나 이러한 초기 변이 예측 기법은 영상의 경계 정보에 따라서 결과에 큰 영향이 미칠 수 있고, 주변의 상, 하 영역을 고려하지 않음으로써 잘못된 정합이 가로로 전파될 수 있다. 하지만 정합창을 이용한 정합 기법을 사용하여 잘못된 정합 및 가려진 영역에 대한 문제를 해결 할 수 있다.

2. 변이 예측

현재 화소의 탐색 시작점을 적응적으로 결정하기 위해 초기 변이를 이용한다. 이 때 정확한 탐색 시작점을 결정하기 위해 현재 화소의 초기 변이와 이웃 화소의 변이를 이용하여 값이 가장 작은 변이를 탐색 시작점으로 결정한다. 정합 과정은 기준 영상과 참조 영상간의 탐색 범위 내에서 정합창 내의 화소들의 비용함수 값이 최소값을 갖는 화소를 찾는 과정이다.

$$C(w) = \min \sum_{i=0}^{SR} |f_0(w) - f_1(w+i)| \quad (5)$$

여기서 f_0 는 기준영상의 화소값, f_1 는 참조 영상의 화소값, SR 은 탐색범위를 나타내며 w 는 정합창을 나타낸다. 정합창을 사용한 정합방법의 비용함수를 계산할 때 발생하는 중복적인 계산을 제거하기 위해 각행의 첫 번째 화소의 비용함수의 값을 그림 7과 같이 저장한다.

그림 7. 비용함수의 중복계산을 위한 처리

Fig. 7. Algorithm with low complexity for cost-function calculation

두 번째 화소의 비용함수를 계산할 때 정합창의 마지막 열의 화소만을 계산하고 나머지 열은 저장한 결과를 사용하게 됨으로써 연산량을 줄일 수 있다.

가려진 영역과 잘못된 정합 영역에 대한 변이 예측을 해결하기 위해 다음과 같이 (1)~(4)의 과정을 수행한다.

(1) 기준 영상 C_0 와 참조 영상 C_1 를 사용하여 정합을 통

해 얻어진 변이 정보를 이용하여 참조 영상 C2에서의 정합된 화소의 위치를 예측 할 수 있다. 기준 영상과 참조 영상 C2의 예측된 화소의 위치에서 비용 함수를 계산하여 임계값 이하이면 정확한 정합 영역으로 판단하고, 임계값 이상이면 잘못된 정합 영역 또는 가려진 영역으로 판단한다.

$$\text{if } |f_0(w) - f_2(w + 2dv)| > Th_1 \text{ (2)번 과정 수행 (6)}$$

f_2 는 참조 영상 C2의 화소이고, dv 는 기준 영상 C0와 참조 영상 C1을 사용하여 정합을 통해 얻어진 변이 정보이다.

- (2) (1)번의 과정에서 잘못된 정합 또는 가려진 영역으로 판단된 영역은 식 (7)을 이용하여 기준 영상 C0와 참조 영상 C-1을 이용하여 정합 과정을 수행한다.

$$C(w) = \min \sum_{j=0}^{SR} |f_0(w) - f_{-1}(w + j)| \quad (7)$$

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

$$\text{if } |f_0(w) - f_{-2}(w + 2dv)| > Th_2 \text{ (4)번 과정 수행 (8)}$$

여기서 f_{-2} 는 참조 영상 C-2의 화소이고, dv 는 기준 영상 C0와 참조 영상 C-1을 사용하여 정합을 통해 얻어진 변이 정보이다.

- (4) (3)번의 과정에서 잘못된 정합으로 판단된 영역은 식 (9)를 이용하여 4개의 참조 영상을 모두 이용하여 정합을 수행한다.

$$C_i(w) = \min \sum_{j=0}^{SR} |f_0(w) - f_i(w + j)| \quad (9)$$

i 는 참조 영상의 번호이다.

$$C(w) = \min \sum_{i=0}^3 C_i(w) \quad (10)$$

식 (10)과 같이 기준 영상과 4개의 참조 영상에서의 비용 함수의 값을 합하여 최소가 되는 값을 변이로 결정한다. 값이 연속적으로 변하는 영역에서는 이 방법에 의해 정확한 결과가 얻어진다. 그러나 값이 불연속적인 영역에서는 잘못된 정합이 발생한다. 따라서 올바른 정합을 방해하는 데이터는 제거하여야 한다. 잘못된 정합을 방지하기 위해 4개의 참조 영상에서의 비용 함수를 계산하여 비용함수 값이 가장 큰 2개의 값을 버리고 나머지 2개의 값을 합하여 비용함수를 계산한다.

$$C'(w) = \min \sum_{i=0}^1 C_i(w) \quad (11)$$

위의 과정을 정리하면 그림 8과 같다.

그림 8. 다시점 동영상을 이용한 변이 예측 기법

Fig. 8. Disparity estimation method for multi-view video

3. 깊이가 불연속인 영역의 변이 예측

정합창을 사용한 정합방법은 정합창의 크기가 성능을 좌우한다. 정합창의 크기는 정합창 내의 영역의 특징을 나타낼 수 있을 정도로 커야한다. 하지만 정합창의 크기가 커지게 되면 연산량이 많아지고, 물체의 경계와 같은 깊이가 불연속인 영역에 있어서는 정확한 정합 정보를 기대할 수 없다^[9]. 따라서 물체의 경계 영역의 정합을 행할 때는 정합창의 크기를 작게 하여 수행한다. 물체의 경계에 대한 정합창의 크기를 결정하기 위해 영상의 경계 정보를 사용한다. 현재 블록의 경계 영역 유무를 식 (12)를 이용하여 판단한다.

$$\text{if } \sum_{i=0}^{BS} \sum_{j=0}^{BS} I(x+i, y+j) > 0 \quad \text{edge region} \quad (12)$$

BS는 블록 크기이며, I는 경계 영상, x, y는 현재 블록의 위치이다. 경계 영역으로 판단된 블록은 작은 크기의 정합창을 사용하여 변이 추정을 수행한다.

그림 9. 경계 영역에서의 변이 추정 : (a) Teddy image (b) 경계 영역 (c) 경계 영상 (d)~(e) 경계 영역을 고려한 변이 추정 전후
Fig. 9. Disparity estimation on object boundary : (a) Teddy image (b) Object boundary (c) Edge-extracted image (d)~(e) Disparity estimation with object boundary

4. 중간 시점 영상 생성

다시점 영상을 이용하여 각 시점 사이의 중간 시점 영상을 생성하기 위해서는 그림 10과 같이 기준 영상의 변이 정보와 생성하고자 하는 위치의 가중치를 이용한다.

그림 10에서 α 는 좌영상과 우영상 사이의 거리를 1로 했을 때 중간 시점 영상의 거리를 표현한 것이다. 생성하고자 하는 시점의 영상은 α 의 값이 1에 가까우면 우영상에 가까운 시점의 영상을 0에 가까우면 좌영상에 가까운 시점의 영상을 나타낸다. 중간 시점 영상 생성에는 일반적으로 기준 영상의 변이를 고려한 선형 양방향 보간법을 사용한다. 좌영상에서의 좌표와 우영상에서의 좌표, 중간 시점 영상에서의 좌표, 변이에 대한 관계는 식 (13)과 같이 표현할 수 있다^[10].

$$I_{\alpha}(x, y) = (1-\alpha)I_{left}(x - \alpha d(x, y), y) + \alpha I_{right}(x + (1-\alpha)d(x, y), y) \quad (13)$$

I_{α} 는 중간 시점 영상에서의 합성하고자 하는 화소이고 d 는 변이를 나타낸다. I_{left} 와 I_{right} 는 각각 좌영상과 우영상에 해당하는 화소이다.

IV. 실험 결과 및 분석

1. 정지영상에서의 변이 추정 결과

변이 추정 알고리즘 성능을 평가하기 위해서는 변이 맵의 객관적인 판단 기준이 필요하다. 변이맵 평가를 위한 일

그림 10. 좌, 우영상으로 부터의 중간영상 생성방법
Fig. 10. Intermediate view reconstruction from stereo image

그림 11. 실험 영상 : (a) cone (b) aloe (c) art (d) venus (e) teddy (f) baby1

Fig. 11. Test images : (a) cone (b) aloe (c) art (d) venus (e) teddy (f) baby1

반적인 2가지 방법은 얻어진 변이 맵을 ground truth 데이터와 비교하여 통계적 오류를 구하는 방법과 변이 정보를 이용하여 새로운 시점의 영상을 생성 후 실제 중간시점에 해당하는 영상과 비교하는 방법이다^[11]. 전자의 경우 텍스트 정보 부속한 영역, 가려진 영역을 제외한 모든 영역, 깊이 불연속 영역으로 영상을 분리한 후 각각의 영역에 대한 통계적 오류를 구하고 오류가 작을수록 성능이 우수한 것으로 판단한다^[12]. 후자의 경우 생성된 중간시점 영상과 본래 중간시점에 해당하는 카메라로부터 획득된 영상의 PSNR(peak signal to noise ratio)을 통하여 성능을 판단한다. 본 논문에서는 중간시점 영상 생성을 통한 알고리즘의 성능평가를 실시하였으며 middlebury 대학 컴퓨터 비전 연구실에서 제공하는 영상을 이용하였다^[12]. 그림11은 실험에 사용된 영상을 나타낸다.

각 영상은 7대 혹은 9대로부터 획득된 다시점 영상으로 1번과 5번 카메라로부터 획득된 영상을 이용하여 3번 시점의 영상을 생성 후 실제 카메라로부터 획득된 영상과의 PSNR을 측정하였다. 비교 알고리즘으로는 화소기반의 정합방식과 블록기반의 정합방식을 사용하였다.

화소기반의 정합방식은 정합창 내의 화소 값들을 비교하여 변이를 찾는 방식으로 모든 화소에 대하여 세밀한 정합을 수행한다. 본 논문에서는 5x5 크기의 정합창을 사용하여 변이 추정을 수행하였다.

블록기반 정합방식은 대표적인 영역기반 정합방법으로 영상을 일정 블록 단위로 나눈 뒤 다른 영상에서 정합되는 블록을 찾아 변이를 추정하는 방식이다. 이는 동영상 압축 코덱에서 움직임 예측 시 사용되는 기법과 동일하며 각종 스테레오 비전 시스템에서 가장 많이 쓰이는 알고리즘이다. 본 논문에서는 절대오차합(sum of absolute difference : SAD) 비용함수를 이용하여 변이 추정을 하였으며 8x8,

16x16 크기의 블록을 사용하였다.

전역적 방법인 그래프 컷(graph cut)이나 신뢰확산(belief propagation), 다이나믹 프로그래밍(dynamic programming) 알고리즘의 경우 정합창을 이용한 화소기반의 변이 추정 방식에 비해 좋은 결과를 얻을 수 있다. 그러나 이러한 전역적 방법은 수십 초에서 수분의 수행시간이 걸리기 때문에 1초 내외의 변이 추정을 하는 제안된 알고리즘과의 성능 비교에서 제외하였다.

표1 PSNR 비교(dB)

Table 1. Result of PSNR

구분		cone	aloe	art	venus	teddy	baby1
블록기반	8x8	24.9	26.4	23.2	30.2	27.1	31.2
	16x16	24.8	25.5	23.3	30.5	27.2	31.2
화소기반		25	25.7	23.3	30.3	27	31.7
제안한 방식		25.5	26.6	24.1	31.4	27.4	31.8

PSNR측정 결과 제안한 방식이 cone영상의 경우 블록기반 방식에 비해 0.6dB~0.7dB, 화소기반방식과 비교하여 0.5dB정도 향상되는 것을 확인할 수 있었다. 이는 단순 블록이나 정합창 사용으로 인해 텍스트 정보가 부족한 영역이나 깊이 불연속지점에서의 정합 오류가 줄어든 결과이며 cone 영상 외에 나머지 영상에 대해서도 약 0.1~1.2dB 향상되는 것을 확인하였다. 특히 venus 영상과 같이 텍스트 정보가 부족한 영역이 많이 존재하는 경우 다른 일반 영상들에 비해 더욱 좋은 성능을 나타내었다.

2. 동영상에서의 변이 추정결과

영상 시퀀스의 경우 MERL에서 제공하는 Ballroom영상 중에서 3-7번 시점의 영상을 사용하였고 영상의 특성은 표

표 2. 실험 영상의 특성

Table 2. Test data sets

Data Set	Sequence	Image Property	Camera Arrangement
MERL	Ballroom	640x480	8 cameras with 20cm spacing, 1D/parallel

1과 같다. 표에서 보는 바와 같이 평행식 카메라 모델이고 영상에 대한 보정이(rectify) 이루어져 있어 라인단위의 초기 변이 추정을 하는 본 논문의 알고리즘에 적합하였다. 그러나 MERL에서 제공하는 영상 시퀀스의 경우 카메라 간 간격이 매우 크기 때문에 연속된 시점의 카메라가 아닌 1시

점 떨어진 카메라를 이용하여 좌·우 영상을 획득할 경우 가려진 영역이 매우 크기 때문에 중간시점 영상을 생성하기 어렵다. 따라서 동영상의 경우 PSNR측정은 따로 수행하지 않았다.

그림 12은 실험에 사용한 카메라의 구성과 Ballroom 영상 3~7번 시점의 첫 번째 프레임에 해당하는 영상을 나타낸다.

그림 13과 그림 14는 제안하는 기법을 통해 기준 영상의 경계정보를 이용하여 라인별로 가변 블록을 결정하고, 초기 변이를 구한 결과이다. 초기 변이 결과에서 일부 잘못된 변이가 발생하지만 전체 영상에 대해서 충분히 정확한 결

그림 12. 실험 영상 : (a)~(e) Ballroom 3~7번 시점의 영상
Fig. 12. Test images : (a)~(e) Ballroom image ; view 3 ~ view 7

그림 13. 첫 번째 프레임 초기 변이 결과
Fig. 13. Initial disparity map of first frame (a) Fundamental-view image (b) Edge-extracted image (c) Adaptive block-size (d) Initial disparity map

그림 14. 열 번째 프레임 초기 변이 결과

Fig. 14. Initial disparity map of tenth frame (a) Fundamental-view image (b) Edge-extracted image (c) Adaptive block-size (d) Initial disparity map

그림 15. 제안한 변이 예측 방법의 변이 결과

Fig. 15. Test result from proposed disparity estimation method

과를 얻을 수 있고, 기존의 방법에 비해 텍스처가 없는 영역에서 정확한 변이 예측 결과를 볼 수 있다.

그림 15는 제안한 방법으로부터 획득된 변이 결과이다. 정합창의 크기는 15×15 를 사용하였고, 경계 영역에서의 정합창은 3×3 을 사용하였다. 초기 변이와 주변 벡터를 이용하여 탐색 시작점을 결정함으로써 변이 예측의 연산량을 줄일 수 있고, 잘못된 변이 예측을 피할 수 있는 것을 확인

할 수 있다. 제안한 4개 시점의 참조 영상을 사용하는 변이 예측을 통해 가려진 영역에 대한 문제를 해결 할 수 있고, 정확한 변이 예측 결과를 확인 할 수 있다. 또한 영상의 경계 영역에서 작은 정합창을 사용하여 정확한 변이를 구할 수 있다.

그림 16는 변이 정보의 신뢰도를 확인하기 위해 제안한 방법으로 얻어진 변이 정보를 사용하여 중간시점영상을 생

(a) 첫 번째 프레임

(b) 열 번째 프레임

그림 16. 중간시점영상

Fig. 16. Intermediate-view image (a) First frame (b) Tenth frame

(a) 첫 번째 프레임

(b) 열 번째 프레임

그림 17. 스테레오영상합성

Fig. 17. Stereoscopic image synthesis (a) First frame (b) Tenth frame

성한 결과이다. 최종적으로 스테레오 모니터를 이용하여 입체 영상을 보기 위해 기준영상의 우영상에 해당하는 시점의 영상을 생성하여 스테레오 영상을 합성한 것이다.

그림 17은 α 가 0.5, 0.7일 때의 중간 영상을 이용하여 스테레오 영상을 합성한 것이다. 그림에서 보는 바와 같이 상대적으로 뒤에 있는 배경보다 앞에 있는 물체에 더 큰 시차가 적용된 것을 확인할 수 있다.

V. 결 론

현재 3차원 TV의 구현 방법으로서 양안 시차를 이용한 여러 가지 방법들이 제안되고 있다. 양안에 해당하는 두 대

의 카메라를 이용하여 획득되는 스테레오 영상의 경우 입체감은 느낄 수 있지만 고정된 시점의 영상만을 봄으로써 시점의 자유도가 없다. 다시점 영상 획득 시스템의 경우 시점의 수는 증가하나 시점 이동시 불연속성이 발생한다. 이에 따라 한정된 시점의 영상으로 자연스럽게 연속적인 영상을 제공하고 동시에 전송 데이터량도 감소시킬 수 있는 기술로 중간 시점 영상을 합성하는 방식에 대한 연구가 필요하다. 본 논문에서는 5대의 카메라로부터 획득된 다시점 영상을 이용하여 변이 정보를 추출하고, 추출된 변이를 이용하여 중간시점영상을 생성하였다. 기준 영상의 변이 정보를 추출하기 위해 4개의 참조 영상을 이용하였으며, 실제 물체의 외곽과 변이 맵에서 물체의 외곽 일치율을 위해 기준 영상의 경계 정보를 이용하였다. 또한 초기 변이를 이용하여 탐색 시작점

을 결정함으로써 연산량을 줄일 수 있었다. 제안한 변이 예측 결과 가려진 영역 및 오정합 영역에 대한 문제를 해결할 수 있었고, 향상된 변이 맵을 얻을 수 있었다. 여러 개의 참조 영상을 이용한 정합을 함으로서 발생하는 연산량과 가려진 영역에 대한 문제는 향후 더 보완되어야 할 과제이다. 또한 보정이 되지 않은 영상에 대해서도 좋은 결과를 도출할 수 있는 강건한 방법이 연구되어야 할 것이다.

참 고 문 헌

- [1] A Smolic, H Kimata, "Description of Exploration Experiments in 3DAV," ISO/IEC JTC1/SC29/WG11 N4929, July 2002.
- [2] J. S. McVeigh, "Efficient Compression of Arbitrary Multi-view Video Signal," Ph.D. dissertation, CMU, June 1996.
- [3] Stephen T. Barnard, and William B. Thompson, "Disparity Analysis of Image," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 2, pp. 333-340, July 1980.
- [4] J. S. McVeigh, M. Siegel, and A. Jordan, "Intermediate View Synthesis Considering Occluded and Ambiguously Referenced Image Regions," Signal Processing Image Communication, vol. 9, pp. 21-28, Sep 1996.
- [5] J. Y. Goulermas and P. Liatsis, "Hybrid Symbiotic Genetic Optimization for Robust Edge-based Stereo Correspondence," Pattern Recognition, vol. 34, pp. 2477-2496, Dec 2001.
- [6] Yao Wang, and Ouseb Lee, "Use of 2-D Deformable Mesh Structures for Video Coding," IEEE Trans. Circuits and systems for video technology, vol. 6, pp. 636-646, Dec 1996.
- [7] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based Coding of Stereo Image Sequences Using Joint 3-D Motion Disparity Compensation," IEEE Trans. Circuits and Systems for Video Technology, vol. 7, pp. 312-327, Apr 1997.
- [8] Y. Ohta and T. Kanede, "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 7, pp. 139-154, Mar 1985.
- [9] S. D. Cochran and G. Medioni, "3-D Surface Description from Binocular Stereo," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 14, pp. 981-994, Oct 1992.
- [10] R. Franich, "Disparity Estimation in Stereoscopic Digital Images," PhD thesis, Technical University of Delft, 1996.
- [11] R. Szeliski, "Prediction Error as a Quality Metric for Motion and Stereo," IEEE International Conference Computer Vision, Vol. 2, pp. 781-788, Sep 1999.
- [12] <http://vision.middlebury.edu/stereo>

저 자 소 개

최 미 남

- 2006년 2월 : 광운대학교 전자공학과 졸업
- 2008년 2월 현재 : 광운대학교 전자공학과 졸업
- 주관심분야 : 동영상 codec, 3차원 영상 신호처리, MVC

윤 정 환

- 2007년 2월 : 광운대학교 전자공학과 졸업
- 2007년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- 주관심분야 : Stereo Matching, 동영상 코덱, 3차원 영상처리

저 자 소 개

유 지 상

- 1985년 2월 : 서울대학교 전자공학과 졸업(공학사)
- 1987년 2월 : 서울대학교 대학원 전자공학과 졸업(공학석사)
- 1993년 5월 : Purdue 대학교 전기공학과 졸업(Ph.D)
- 1993년 9월 ~ 1994년 8월 : 현대전자산업(주) 산전연구소 선임연구원
- 1994년 9월 ~ 1997년 8월 : 한림대학교 전자공학과 조교수
- 1997년 9월 ~ 2001년 8월 : 광운대학교 전자공학과 조교수
- 2001년 9월 ~ 현재 : 광운대학교 전자공학과 정교수
- 주관심분야 : 웨이블릿 기반 영상처리, 영상압축, 영상인식, 비선형 신호처리