

논문 2008-45SP-4-3

2차원/3차원 자유시점 비디오 재생을 위한 가상시점 합성시스템

(Virtual View Rendering for 2D/3D Freeview Video Generation)

민 동 보*, 손 광 훈**

(Dongbo Min and Kwanghoon Sohn)

요 약

3DTV를 위한 핵심 기술 중의 하나인 다시점 영상에서 변이를 추정하고 가상시점을 합성하는 새로운 방식을 제안한다. 다시점 영상에서 변이를 효율적이고 정확하게 추정하기 위해 준 N -시점 & N -깊이 구조를 제안한다. 이 구조는 이웃한 영상의 정보를 이용하여 변이 추정 시 발생하는 계산상의 중복을 줄인다. 제안 방식은 사용자에게 2D와 3D 자유시점을 제공하며, 사용자는 자유시점 비디오의 모드를 선택할 수 있다. 실험 결과는 제안 방식이 정확한 변이 지도를 제공하며, 합성된 영상이 사용자에게 자연스러운 자유시점 비디오를 제공한다는 것을 보여준다.

Abstract

In this paper, we propose a new approach for efficient multiview stereo matching and virtual view generation, which are key technologies for 3DTV. We propose semi N -view & N -depth framework to estimate disparity maps efficiently and correctly. This framework reduces the redundancy on disparity estimation by using the information of neighboring views. The proposed method provides a user 2D/3D freeview video, and the user can select 2D/3D modes of freeview video. Experimental results show that the proposed method yields the accurate disparity maps and the synthesized novel view is satisfactory enough to provide user seamless freeview videos.

Keywords: 가려진 영역 처리, 준 N -시점 & N -깊이 구조, 스테레오 정합, 가상시점 합성

I. 서 론

멀티미디어 처리 분야의 급속한 발전으로 인해 3차원 TV (3DTV)는 차세대 방송 시스템 시장에서 가장 주목 받는 제품 중의 하나가 될 것으로 예상된다. 3DTV는 사용자에게 3D 깊이감과 상호 작용을 제공할 수 있다. 사용자가 원하는 시점을 자유롭게 선택할 수 있고, 입체감을

제공함으로써 사용자가 마치 그 곳에 있는 듯한 효과를 줄 수 있다. 3DTV의 상용화를 위해서는 다시점 영상을 획득하여 처리하고 통신 환경에서 막대한 양의 데이터를 압축하여 전송하는 기술이 발전해야만 한다^[1].

가상시점 재생 기술은 3DTV에서 가장 중요한 기술 중의 하나이다. 컴퓨터 비전 분야에서 오랫동안 연구되어 왔으며, 영상 기반 합성 방식 (Image-based rendering)에 기반하여 많은 방식들이 제안되었다. 제한된 수의 카메라만으로 다양한 시점을 제공하기 때문에, 가상시점 재생 기술은 3DTV 시스템을 개발하는데 있어 데이터양과 비용을 줄이는데 매우 유용하다. 또한, 이를 통해 3D 영상 획득 장치와 디스플레이 장치 간의 규격에서 발생할 수 있는 불일치를 보상할 수 있다. 본 논문에서는 영상 기반 합성 방식 중에 상대적인 기하정보를 통

* 학생회원, ** 정회원, 연세대학교 전기전자공학부 (School of Electrical & Electronic Engineering, Yonsei University)

※ 본 연구는 지식경제부 및 정보통신연구진흥원의 IT 신성장동력핵심기술개발사업 (2007-F036-02, 인간 시각 특징 기반 주관적 화질 측정 방법 및 장치 기술)과 대학 IT연구센터 지원사업 (IITA-2008-C1090-0801-0011)의 연구결과로 수행되었음.

접수일자: 2007년12월4일, 수정완료일: 2008년4월18일

한 합성을 이용하여 가상시점을 제공하는 방식을 제안한다. 실 영상들은 기하학적 정보를 이용하여 원하는 시점으로 워핑되며, 워핑된 영상들을 가중치 보간함으로써 가상시점이 재생된다^[2]. 많은 방식들이 가중치 보간 방식의 성능을 향상시키기 위해 제안되었다.

Zhang은 스테레오 영상에서 중간 시점을 재생하는 방식을 제안하였다^[3]. 중간 시점은 투영된 영상의 화질에 따라 가중치 평균을 사용하여 합성되었다. Criminisi는 원격 회의를 위한 가상시점 합성 시스템을 개발하였다^[4]. 동적 프로그래밍 (dynamic programming)에 기반한 스테레오 알고리즘을 개발하고, 최소 비용 표면의 투영에 의해 효율적으로 가상시점을 재생하는 방식을 제안하였다. Kauff는 비디오와 깊이 데이터 표현 방식(video-plus-depth data representation)에 기반한 3DTV 시스템을 소개하였다^[5]. 본 논문에서는 비디오와 깊이 데이터 표현 방식을 N -시점 & N -깊이 구조라 한다. 3차원 영상 획득 장치와 디스플레이 장치 간의 호환성과 다시점 영상 조절을 고려하기 위해, 깊이 정보를 추정하고 가상시점을 합성하는 방식을 제안하였다. Zitnick는 다시점 비디오에서 고화질의 가상시점을 합성하는 방식을 제안하였다^[6]. 깊이 정보는 다시점 영상을 교정하지 않고, 교정 정보를 이용하여 바로 획득한다. 깊이 정보는 N -시점 & N -깊이 구조에서 추정되며, 컬러 분할 방식이 깊이 추정의 성능을 높이기 위해 사용되었다.

본 논문에서 우리는 가상시점을 합성하는 새로운 방식을 제안한다. 다시점 영상에서 변이를 추정할 때 발생하는 계산상의 중복을 줄이기 위해 준 N -시점 & N -깊이 구조를 제안한다. N 개의 영상에서 동일한 방식으로 모든 변이 정보를 추정하는 기존의 N -시점 & N -깊이 구조와는 달리 준 N -시점 & N -깊이 구조에서는 이웃 영상에서 추정되는 정보를 이용함으로써 중복되는 계산을 최대한 줄인다. 또한, 가려진 영역의 문제는 다시점 영상에서 계산되는 비용 함수를 사용하여 해결된다.

우리는 제안 방식에서 2D/3D 자유시점 비디오를 제공할 수 있다. 이 때, 3D 자유시점은 스테레오 영상의 형태로 제공된다. 기존의 방식들이 대부분 2D 자유시점 비디오나 고정된 시점에서 3D 비디오를 제공하는 반면, 본 논문에서는 사용자가 2D/3D 자유시점 비디오 모드를 선택하도록 함으로써 좀 더 일반화된 시스템을 제안한다.

II장에서는 시스템의 동기와 개괄을 서술하고, III, IV장에서는 본 논문에서 제안하는 다시점 스테레오 정합과 가상시점 합성 방식에 대해 설명한다. V장에서 실험 결과를 보여주고, VI장에서 결론을 맺는다.

II. 시스템 동기와 개괄

N 개의 영상이 주어졌을 때, 가상시점을 합성하기 위해서는 N 개의 깊이 지도를 획득해야만 한다. Zitnick는 다시점 영상에서 N 개의 깊이 지도를 추정하고 중간 시점을 합성하는 방식을 제안하였다^[6]. 이 때, 변이 추정은 막대한 계산량으로 인해 오프라인으로 미리 수행되었다. 변이 추정의 복잡도가 높은 이유 중의 하나는 모든 영상에 대해 동일한 방식으로 변이 추정을 수행하기 때문이다. 일반적으로 이웃한 영상들의 변이 지도는 가려진 영역을 제외하고는 서로 매우 비슷하다. 이러한 점을 고려하여 우리는 변이 추정 시 발생하는 계산상의 중복을 줄이는 준 N -시점 & N -깊이 구조를 제안한다. 그림 1은 N -시점 & N -깊이 구조와 준 N -시점 & N -깊이 구조의 차이를 보여준다. N -시점 & N -깊이 구조에서는 각 영상의 변이 지도가 독립적으로 구해지는 반면, 준 N -시점 & N -깊이 구조에서는 참조 영상의 변이 지도가 이웃한 영상들의 변이 지도를 구하는데 사용된다. 다시점 영상은 참조 영상과 대상 영상으로 구분되며, 대상 영상은 대상 영상과 준 대상 영상으로 구분된다. 제안된 비용 집합 방식은 참조 영상에서만 변이 지도를 추정하기 위해 사용되며, 대상 영상의 변이 지도는 참조 영상의 비용 함수를 워핑함으로써 계산된다. 그림 1에서 보는 바와 같이 대상 영상에서는 전방과 후방 워핑이 모두 사용되는 대칭 워핑이 수행되며, 준 대상 영상에서는 전방 또는 후방 워핑 중의 하나만이 사용되는 비대칭 워핑이 수행된다. 이 때, 좌우 끝에 있는 영상들은 항상 준 대상 영상이 된다.

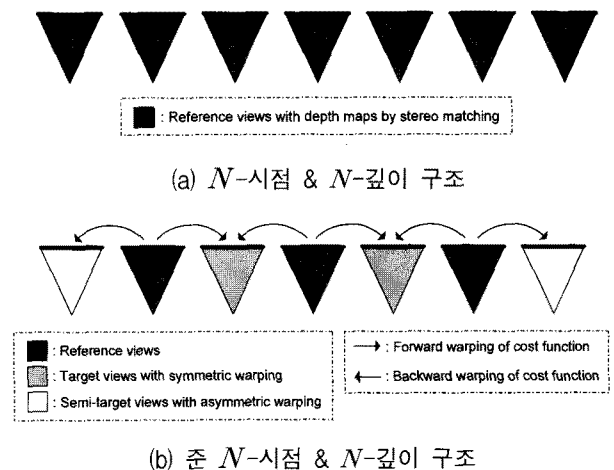


그림 1. 다시점 영상과 깊이 지도의 표현 방식
Fig. 1. Representation of multiview images and depth data.

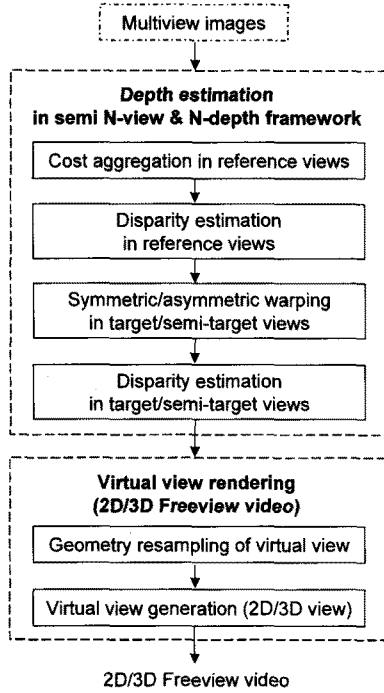


그림 2. 제안되는 시스템의 전체 구조
Fig. 2. Overall framework of the proposed method.

그림 2는 제안하는 시스템의 전체 구조를 보여준다. 준 N -시점 & N -깊이 구조에서 N 개의 깊이 지도가 추출되며, 주어진 가상시점에 따라 영상이 합성된다. 3D 자유시점은 두 장의 가상 영상을 합성함으로써 제공된다. 그림 2의 각 부분은 다음 절에서 설명된다.

III. 다시점 스테레오 정합

변이 추정과 가상시점 재생을 위해 우리는 다시점 카메라 구조를 사용하였다. 스테레오 정합에 대한 자세한 리뷰는 [7]에 나와 있다. 본 논문은 2D/3D 자유시점 비디오 재생 시스템을 개발하는 것을 목적으로 하기 때문에 평행 카메라 구조를 사용하였다. 아크 형태의 카메라 구조는 3D 자유시점 비디오를 합성할 때 많은 구멍을 발생시킬 수 있기 때문이다. 우리는 카메라 사이의 거리는 B 로 동일하다고 가정하였다.

1. 화소당 비용 함수 계산

변이 지도를 추정할 때, 두 장 또는 그 이상의 영상이 사용된다. $i-1, i, i+1$ 번째 영상을 각각 좌, 중앙, 우 영상이라 할 때, 중앙 영상의 차영상은 좌(우)영상을 오른쪽(왼쪽)으로 이동하고 차이를 구하여 계산된다. i 번째 영상의 차영상 $e_{i,j}(p,d)$ 는 i, j 번째 영상으로 다음과 같이 계산된다.

$$\begin{aligned} e_{i,i+1}(p,d) &= \min \{|I_i(x,y) - I_{i+1}(x+d,y)|, T\} \\ e_{i,i-1}(p,d) &= \min \{|I_i(x,y) - I_{i-1}(x-d,y)|, T\} \end{aligned} \quad (1)$$

p 와 d 는 화소의 2D 위치와 변이를 나타내며, I 는 RGB로 구성된 컬러를, T 는 정합 비용의 상한선을 나타낸다. 화소당 비용함수 $e_i(p,d)$ 는 중앙 영상의 점의 가시도를 고려하여 $e_{i,i+1}$ 과 $e_{i,i-1}$ 를 통해 계산된다. 본 논문에서는 중앙 영상의 모든 점들은 최소한 하나 이상의 대응되는 점들을 가지고 있다고 가정한다. 이는 몇몇 부분에서 유효하지 않지만, 가려진 영역을 해결하는데 매우 효과적이다. 가려진 영역에서 계산된 정합 비용이 일반적으로 더 크다는 가정 하에 화소당 비용함수 $e_i(p,d)$ 는 다음과 같이 계산된다.

$$e_i(p,d) = \min(e_{i,i+1}(p,d), e_{i,i-1}(p,d)) \quad (2)$$

2. 가중치 최소 자승을 이용한 비용 집합

i 번째 영상에서 최적의 $E_i(p,d)$ 를 구하기 위해 비용함수가 물체의 경계를 제외한 나머지 부분에서 매우 부드럽게 변한다는 조건을 이용한다. 이를 통해 다음과 같은 가중치 최소 자승을 이용한 에너지 모델을 정립하고, 이를 최소화함으로써 비용 함수를 추정할 수 있다.

$$\begin{aligned} \epsilon(E) &= \int_{\Omega} (E(p) - e(p))^2 dp \\ &+ \lambda \int_{\Omega} \sum_{n \in N_1} \frac{w(p, p+n)(E(p) - E(p+n))^2}{w(p, p+n) + w(p, p+n)} dp \\ N_1 &= \{(x_n, y_n) | 0 < x_n \leq M, 0 \leq y_n \leq M\} \end{aligned} \quad (3)$$

w 는 이웃한 점들 간의 가중치 함수를 의미한다. 각 변이에 대해 동일한 과정이 수행되므로 $E_i(p,d)$ 는 $E(p)$ 로 간략하게 표현한다. n 과 n^\perp 는 서로 수직인 2차원 벡터를 나타내며, M 은 이웃한 점들로 구성된 집합의 크기이다. E 에 대해 식 (3)을 1차 미분하면, 다음과 같은 식을 얻을 수 있다.

$$\begin{aligned} E(p) - e(p) \\ + \lambda \sum_{n \in N_1} \frac{w(p, p+n)(E(p) - E(p+n))}{w(p, p+n) + w(p, p+n)} \\ + \lambda \sum_{n \in N_1} \frac{w(p, p+n^\perp)(E(p) - E(p+n^\perp))}{w(p, p+n^\perp) + w(p, p+n^\perp)} = 0 \end{aligned} \quad (4)$$

위의 식을 간단하게 표현하기 위해 우리는 이웃한 점들의 집합을 재정의한다. p 가 (x,y) 일 때, 점들의 집합

은 다음과 같이 표현된다.

$$N(p) = \{(x + x_n, y + y_n) | -M \leq x_n \leq M, -M \leq y_n \leq M, x_n + y_n \neq 0\}$$

위의 식을 이용하여 식 (4)는 다음과 같이 표현된다.

$$E(p) - e(p) + \lambda \sum_{m \in N(p)} w(p, m)(E(p) - E(m)) = 0 \tag{5}$$

위의 식에서 E 를 구하기 위해 반복 계산법을 이용한다. $(k+1)^{th}$ 번째 반복 계산의 해는 다음과 같이 계산된다.

$$E^{k+1}(p) = \frac{e(p) + \lambda \sum_{m \in N(p)} w(p, m)E^k(m)}{1 + \lambda \sum_{m \in N(p)} w(p, m)} \tag{6}$$

식 (6)은 정규화된 화소당 비용 함수와 이웃한 점들의 비용을 가중치 평균한 비용 함수로 구성되어 있다. 반복 계산을 수행함으로써 비용 함수 E 는 주변의 이웃한 점들의 비용 함수에 의해 정규화된다. 반복 계산법은 반복 계산의 횟수가 1인 경우 적응적 가중치 방식과 비슷하다^[8]. 본 논문에서 우리는 CIE-Lab를 이용한 비대칭적인 가우시안 가중치 함수를 사용한다. r_c 와 r_s 는 각각 컬러와 공간상의 거리에 대한 가중치 상수로, C_i 가 i^{th} 번째 영상에서 계산되는 컬러 거리라 하면 가중치 함수는 다음과 같이 정의된다.

$$w(p, m) = \exp\left(-\left(\frac{C_i(p, m)}{2r_c^2} + \frac{S(p, m)}{2r_s^2}\right)\right)$$

$$C_i(p, m) = (L_i(p) - L_i(m))^2 + (a_i(p) - a_i(m))^2 + (b_i(p) - b_i(m))^2 \tag{7}$$

$$S(p, m) = (p - m)^2$$

3. 비용 함수 수렴의 가속화 방식

가. 가우스-사이드 방식에 의한 수렴 가속화

식 (6)에서 수렴 속도를 늦추는 원인 중의 하나는 각 점에서 이미 계산된 정보가 한 번의 반복 계산이 완료된 후에만 사용된다는 점이다. 계산된 점의 정보를 바로 이용함으로써 이러한 문제점을 해결할 수 있다. 이웃한 점들의 집합 $N(p)$ 는 인과적 부분 $N_c(p)$ 와 비인과적인 부분 $N_n(p)$ 로 구분된다. 이러한 기호를 이용하여 식 (6)은 다음과 같이 표현된다.

$$E^{k+1}(p) = \frac{e(p) + \lambda \sum_{m \in N_c(p)} w(p, m)E^{k+1}(m) + \lambda \sum_{m \in N_n(p)} w(p, m)E^k(m)}{1 + \lambda \sum_{m \in N(p)} w(p, m)} \tag{8}$$

나. 다해상도 접근 방식

신뢰도 있는 정합을 수행하기 위해서는 멀리 있는 점의 정보를 모으는 것이 필요하다. 이는 정확한 비용 함수를 추정하기 위해서는 굉장히 많은 수의 반복 계산을 수행해야 한다는 것을 의미한다. 이러한 문제를 해결하기 위해 우리는 다해상도 접근 방식을 사용한다. 본 논문에서 사용되는 다해상도 접근 방식은 비용 함수에 직접 적용된다는 점에서 기존의 다해상도 접근 방식과는 다르다. 저해상도에서 계산된 비용 함수를 고해상도의 비용 함수를 구하기 위한 초기치로 사용할 수 있다. 식 (8)에서 제안된 방식은 x, y, d 로 구성된 3D 비용 볼륨에서 동일한 변이 d 를 갖는 각 섹션에 대해 독립적으로 비용 함수를 추정한다. 그러므로 우리는 3D 비용 볼륨을 우선 계산하고, 각 2D 비용 함수에 대해 제안된 다해상도 접근 방식을 수행한다. 제안된 다해상도 방식은 최저해상도에서 $e(p, d)$ 를 초기치로 사용하여 반복 계산을 수행한다. K 번의 반복 계산 후, 계산된 비용 함수는 다음 해상도에서의 비용 함수를 초기화하기 위해 사용되며, 이러한 과정은 최고해상도까지 반복된다. 그림 3은 제안된 다해상도 방식을 보여준다. l^{th} 레벨에서의 비용 함수 $E^l(p)$ 에 대한 초기치는 선형 보간법을 이용하여 $E^{l+1}(p)$ 의 해상도를 증가시킴으로써 계산된다. 그러나, 선형 보간법은 경계 영역에서 오류를 전파하는 문제점을 가지고 있다. 이를 해결하기 위해, 우리는 가중치 최소 자승을 이용한 적응적 보간 방법을 사용한다.

$$E_l(p) = \frac{e_l(p) + \lambda_a \sum_{p_m \in N(p_i)} w(p, p_m)E_{l+1}(p_m)}{1 + \lambda_a \sum_{p_m \in N(p_i)} w(p, p_m)} \tag{9}$$

여기서 $p_i = (x_i, y_i)$ 는 저해상도 레벨에서의 점을, $(l+1)^{th}$ 레벨에서의 $N(p_i)$ 는 4개의 이웃한 점들의 집합을 의미한다. 식 (9)에서 w 는 식 (6)에서 사용된 것과 동일한 가중치 함수이다. 우리는 가중치 상수 λ_a 를 15로 정의한다. 연속된 레벨에서의 화소 값에 의한 적응적 보간은 경계에서의 불연속을 보존하는 업-샘플링을

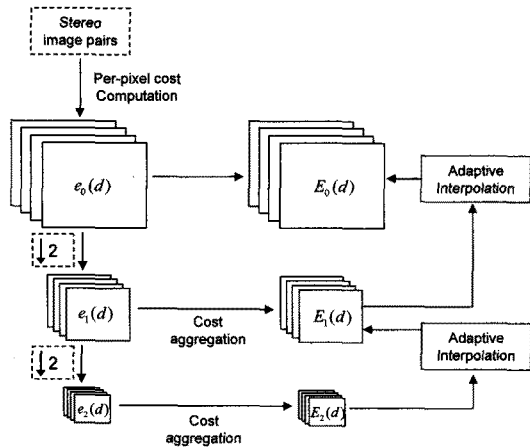


그림 3. 제안되는 비용 집합 방식의 전체 구조
Fig. 3. Overall framework of proposed cost aggregation.

가능하게 한다. 그러므로 최고해상도에서는 비용 집합을 수행할 필요가 없으며, 이는 제안한 방식의 계산량을 크게 감소시킨다.

IV. 가상시점 합성

우리는 비용 집합 방식을 사용하여 변이를 추정하고, 다시점 영상에서 계산된 화소당 비용 함수를 이용하여 가려진 영역을 해결한다. 다시점 영상에서 가상시점을 합성하기 위해서는 N 개의 깊이 영상을 획득하는 것이 필요하다. 가상시점은 각 영상을 변이 정보에 의해 워핑하여 합성된다. 제안된 시스템은 2D와 3D 자유시점을 동시에 제공하기 때문에, 사용자는 2D와 3D 모드를 선택할 수 있으며 3D 자유시점 비디오에서 여러 변수들을 조정함으로써 입체감을 조절할 수 있다.

1. 준 N -시점 & N -깊이 구조

대부분의 방식들은 N 개의 영상에서 동일한 방식으로 변이를 획득하였다. 이는 변이 추정시 막대한 계산상의 중복을 초래한다. 본 논문에서 우리는 준 N -시점 & N -깊이 구조에서 변이 추정 시 발생하는 계산상의 중복을 줄이는 새로운 방식을 제안한다. 참조 영상의 비용 함수는 제안된 비용 함수 집합 방식을 이용하여 추정된다. 대상 영상의 비용 함수는 참조 영상의 비용 함수를 워핑함으로써 계산된다. 이는 이웃한 영상의 대응되는 점들은 비슷한 비용 함수를 갖는다는 점에 기초한다. 참조 영상의 비용 함수는 변이 정보를 이용하여 대상 (준-대상) 영상으로 이동된다. 대상 영상에서는 전방과 후방 워핑이 모두 사용되므로, 워핑에 의해

발생되는 가려진 부분이 서로 보상이 되어 적은 수의 구멍만이 존재한다. 그러나 준-대상 영상에서는 전방 또는 후방 워핑 중의 하나만이 사용되므로 가려진 영역에 대한 처리가 필요하다. 예를 들어, 최 좌측 시점 0에서는 후방 워핑만이 수행되므로, 물체의 좌측 부분에 가려진 영역이 발생한다. 가려진 영역에 적절한 비용 함수를 할당하기 위해 우리는 신뢰도 있는 주변 점들을 이용한다.

일반적으로 두 영상에서 모두 보이는 점들의 비용 함수만이 이동되어야 한다. 참조 영상의 한 점의 가시도를 판별하기 위해 우리는 기하학적 및 광학적 제한 조건을 이용한다. 우선 전방 워핑 방식을 설명한다. 우리는 이웃한 점들의 변이를 평가함으로써 한 점의 가시도를 추정한다. 일반적으로 가리는 점은 가려지는 점보다 큰 변이를 가지고 있다. 가시도 함수를 정하기 전에 우리는 참조 영상의 점들로 구성된 대상 영상에 대한 함수 $S_i(j)$ 를 정의한다.

$$S_i(j) = \{i | i - d_r(i) = j, \text{ all } i \text{ with } 0 \leq i \leq W-1\}$$

i 와 j 는 참조와 대상 영상의 x 좌표를 나타내며, W 는 영상의 가로 크기를 의미한다. 대상 영상에 다중 정합 점들이 존재한다면, 즉, $\#(S_i(j)) \geq 1$ 라면, $S_i(j)$ 에서 가장 큰 변이를 갖는 점이 보이는 점이 되고 나머지는 가려진 것으로 간주된다. 그러나 이는 가리는 점, 즉, 가장 큰 변이를 갖는 점이 신뢰도가 높은 변이를 갖는 경우에만 유효하다. 그림 4는 전방 워핑의 여러 경우를 보여준다. 가려진 점의 변이가 보이는 점의 변이보다 작을 경우, 가려진 영역은 정확하게 추정된다. 그러나 가려진 점이 다른 보이는 점들을 가리는 경우에는 문제가 발생한다. 이를 막기 위해 우리는 가리는 점들의 신뢰도를 평가하는 광학 제한 조건을 사용한다. 이러한 조건을 이용하여 대상 영상에서 가려진 영역 대신 가려진 영역의 후보군을 추정한다. 가려진 점의 비용 함수는 일반적으로 보이는 점들의 비용 함수보다 크다. 기하학적 조건에 의해 가리는 점으로 추정된 점의 비용 함수가 나머지 가려진 점들의 비용 함수보다 작지 않으면, 우리는 가리는 점의 신뢰도를 보장할 수 없다. 그러므로 $S_i(j)$ 에 속한 모든 점들은 그림 4에서 보는 바와 같이 가려진 영역의 후보로 사용되며, $\#(S_i(j))$ 는 0으로 재설정된다. 대상 영상에서의 가시도 함수 $O_i(j)$ 는 $\#(S_i(j)) = 0$ 일 때 0, 아니면 1로 설정된다. 대상 영상에서의 가시도 함수 $O_i(j)$ 를 이용하여 참조 영상의 비

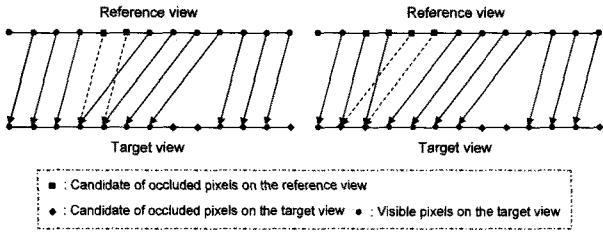


그림 4. 전방 워핑의 여러 가지 경우
Fig. 4. Several cases of forward warping.

용 함수는 다음과 같이 워핑된다.

$$E_t(i - d_r, d_r) = E_r(i, d_r), \quad \text{if } O_t(i - d_r) = 1 \quad (10)$$

위 식에서 동일한 과정이 각 라인마다 반복되기 때문에, y 좌표는 생략되었다. 후방 워핑도 전방 워핑과 비슷한 방식으로 수행된다. 후방 워핑에서 $S_t(j)$ 는 다음과 같이 정의된다.

$$S_t(j) = \{i | i + d_r(i) = j, \text{ all } i \text{ with } 0 \leq i \leq W-1\}$$

대상 영상의 가시도 함수 $O_t(j)$ 를 이용하여 다음과 같이 후방 워핑이 수행된다.

$$E_t(i + d_r, d_r) = E_r(i, d_r), \quad \text{if } O_t(i + d_r) = 1 \quad (11)$$

참조 영상의 변이가 아닌 비용 함수가 대상 영상으로 이동된다는 것을 주의해야 한다. 전방과 후방 워핑이 모두 수행되는 대상 영상의 점들에서는 해당하는 변이에 대한 비용 함수가 작은 워핑을 선택한다. 준-대상 영상에서는 전방이나 후방 워핑 중의 하나만 수행되기 때문에 가려진 영역이 처리되어야만 한다. 대상 또는 준-대상 영상에서의 가시도 함수 O_t 를 이용하여 우리는 반복 계산식을 다음과 같이 재정의한다.

$$E^{k+1}(p) = \frac{O_t(p)e(p) + \lambda \sum_{m \in N_r(p)} O_t(m)w(p, m)E^{k+1}(m) + \lambda \sum_{m \in N_t(p)} O_t(m)w(p, m)E^k(m)}{O_t(p) + \lambda \sum_{m \in N(p)} O_t(m)w(p, m)} \quad (12)$$

그림 5는 대칭/비대칭 워핑의 과정을 보여준다. 비대칭 워핑에서는 가려진 영역이 처리되어야 하기 때문에, 일반적으로 비대칭 워핑이 대칭 워핑보다는 느리다. 본 논문에서 사용된 가려진 영역 처리 방법은 널리 사용되고 있는 변이 외삽 방식과는 다르다. 외삽 방식이 단순히 보이는 점들의 변이를 사용하여 채우는 반면, 제안한 방식은 보이는 점의 정보를 가려진 영역에 전달해준다. 이러한 과정은 경계 보존 비선형 확산 방식과 비슷

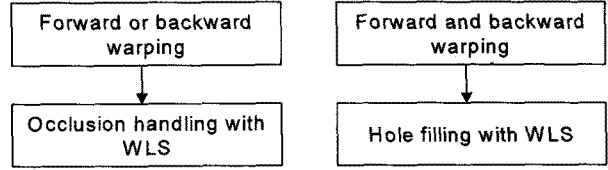


그림 5. 대칭/비대칭 워핑
Fig. 5. Symmetric and asymmetric warping.

한 개념을 이용하고 있다. 본 논문에서 우리는 최적화 방식으로 WTA (Winner-Takes-All) 방식을 사용하지만, graph cut이나 belief propagation과 같은 다른 방식이 사용될 수 있다^[9-10].

2. 2D 가상시점 재생

주어진 N 개의 영상과 변이 지도를 이용하여 각 영상을 워핑하고, 워핑된 영상을 내삽하여 가상시점이 합성된다. 가상시점에 해당하는 영상을 합성하는 방식은 다음과 같다.

1. 변이 정보를 이용하여 실제 영상에 있는 모든 점들을 3D 공간상으로 역투영한다.
2. 실제 영상의 좌표를 가상 영상의 좌표로 변환한다.
3. 가상 영상의 2D 평면으로 3D 좌표를 투영한다.

위의 과정을 통해 실제 영상의 텍스처 정보가 가상 영상으로 맵핑되며, 이 과정은 모든 실제 영상에 대해 수행된다. 가상 카메라의 회전은 합성된 영상에 많은 구멍을 발생시킬 수 있으며, 이는 3DTV에 적합하지 않기 때문에 우리는 평행 이동만을 고려한다. 주어진 가상시점에 대해 가장 가까운 두 영상 (i^{th} 와 $(i+1)^{th}$)이 선택되고, 가상시점에 투영된다. i^{th} 영상에서 변이 d_i 를 갖는 점 $m_i(x_i, y_i)$ 는 3D 공간상의 점 M_i 로 다음과 같이 변환된다.

$$\left(\frac{(x_i - x_0)B}{d_i}, \frac{(y_i - y_0)B}{d_i}, \frac{fB}{d_i} \right) \quad (13)$$

(x_0, y_0) 은 영상의 기저점을 의미한다. M_i 을 평행 이동 (T_x, T_y, T_z)를 이용하여 가상 카메라의 좌표로 변환하고, 가상 카메라의 영상 표면으로 투영하면, 실제와 가상 영상에서 대응하는 점들의 관계를 계산할 수 있다. 가상시점의 점 $m_i^v(x_i^v, y_i^v)$ 는 다음과 같은 식으로 계산된다.

$$\begin{aligned} x_i^v - x_0 &= f \frac{(x_i - x_0)B/d_i + T_x}{fB/d_i + T_z} = \frac{x_i - x_0 + d_i \alpha_x}{1 + d_i \alpha_z/f} \\ y_i^v - y_0 &= f \frac{(y_i - y_0)B/d_i + T_y}{fB/d_i + T_z} = \frac{y_i - y_0 + d_i \alpha_y}{1 + d_i \alpha_z/f} \end{aligned} \quad (14)$$

수식을 간단하게 하기 위해 정규화된 좌표 $(\alpha_x, \alpha_y, \alpha_z)$ $= (T_x, T_y, T_z)/B$ 를 사용하며, 카메라 간의 거리는 1로 정의한다. I_i 와 I_i^v 를 실제와 투영된 영상이라 할 때, 식 $I_i^v(x_i^v, y_i^v) = I_i(x_i, y_i)$ 이 성립한다.

실제 영상에서 가상 영상으로 텍스처 정보를 전달하는 전방 맵핑이 수행될 때, 몇가지 문제점이 발생한다. 식 (14)에서 일대일 맵핑이 되지 않고, 가상시점에 다중 투영이나 구멍이 발생한다. 가상시점에 다중 투영이 발생하는데는 두 가지 이유가 있다: 깊이 불연속과 영상 리샘플링. 깊이 불연속 지점에 존재하는 점들은 그들이 다른 변이를 가지고 있음에도 불구하고, 가상시점에서 동일한 점에 투영된다. 이 때, 카메라로부터 가까운 거리에 있는 점이 나머지 점들을 감싸기 때문에, 투영된 점들 중 가장 큰 변이를 갖는 점만을 보존한다. 다른 문제는 영상의 리샘플링에 의해 발생한다. 가상시점 재생 시 물체의 크기가 변할 때, 같은 변이를 갖는 점들에 대해서도 다중 투영과 구멍이 발생할 수 있다. 더욱이, 가상시점의 점 (x_i^v, y_i^v) 는 정수 단위가 아닐 수도 있다. 이러한 문제점들을 해결하기 위해, 우리는 가상시점 합성 시 가상 영상에서 실사 영상으로 투영하는 역방향 맵핑과 선형 보간법을 이용한다. 역방향 맵핑은 영상 리샘플링에서 발생할 수 있는 화질의 저하를 막아준다. 그림 6에서 보는 바와 같이 가상 카메라는 좌, 우, 전방, 후방을 포함하여 x 와 z 축으로 움직인다. 가상시점에 구멍을 발생시킬 수 있기 때문에, y 축으로의 움직임은 제한된다. $(\alpha_x^g, \alpha_y^g, \alpha_z^g)$ 는 가상 카메라의 실제 위치를 의미한다. $V(p)$ 를 가상시점의 점이 실제 영상에서 보이는 지 여부를 판별하는 가시도 함수라 할 때, 가상시점은

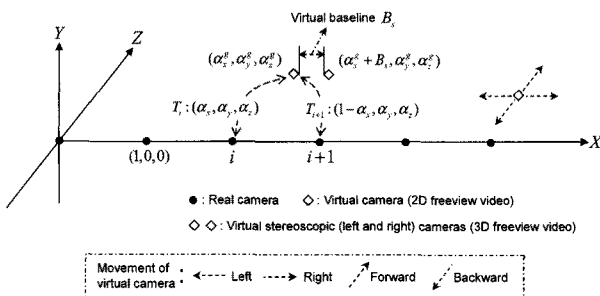


그림 6. 가상 카메라의 움직임

Fig. 6. Movement of virtual camera.

투영된 영상의 내삽을 통해 다음과 같이 계산된다.

$$I_v(p) = V^i(p)(1 - \alpha_x)I_v^i(p) + V^{i+1}(p)\alpha_x I_v^{i+1}(p) \quad (15)$$

3. 3D 가상시점 재생

3D 가상시점은 두 장의 시점을 합성하여 제작할 수 있다. 두 가상시점 사이의 거리를 B_s 라 할 때, 제로 시차 설정 (zero parallax setting, ZPS)을 하기 위해 가상 카메라의 CCD 센서가 렌즈의 위치에 대해 h 만큼 평행 이동된다. 이는 3D 장면이 디스플레이 장치에서 재연될 때, 수렴 거리 Z_c 를 선택할 수 있도록 해준다. 센서의 이동은 가상 카메라의 기저점을 이동시킴으로써 간단히 계산된다. 기저점의 수평 이동을 h 라 정의할 때, 가상시점의 점은 식 (16)에 의해 계산된다. $\pm h$ 는 이동된 우, 좌 영상을 의미한다. 좀 더 자세한 설명은 [11]을 참고하기 바란다.

$$\begin{aligned} x_i^v - (x_0 \pm h) &= \frac{x_i - x_0 + d_i \alpha_x}{1 + d_i \alpha_z/f} \\ y_i^v - (y_0 \pm h) &= \frac{y_i - y_0 + d_i \alpha_y}{1 + d_i \alpha_z/f} \end{aligned} \quad (16)$$

V. 실험 결과

준 N -시점 & N -깊이 구조의 성능을 분석하기 위해 Middlebury 테스트 영상들에 대해 실험을 수행하였다 [12]. 우리는 'Tsukuba', 'Venus', 'Teddy', 'Cone' 영상을 이용하였다. 'Tsukuba' 영상은 5개 (시점 0-4), 'Venus', 'Teddy', 'Cone' 영상은 8개 (시점 0-8)로 구성되어 있다. 우리는 'Venus', 'Teddy', 'Cone' 영상에 대해서는 짝수 시점만을 사용하였다. 즉, 실험은 $N=5$ 인 경우에 대해 수행되었다. 실험은 모든 테스트 영상들에 대해 동일한 변수를 가지고 수행되었다. 가중치 함수의 두 변수 r_c 와 r_s 는 8.0으로, 가중치 변수 λ 는 1.0으로 설정되었다. 4 레벨에서 다해상도 접근 방식이 사용되었으며, 반복 계산의 횟수는 (3, 2, 2, ×)였다. 최고해상도 레벨에서는 비용 함수 집합이 수행되지 않았으며, 이웃한 점들의 집합의 크기는 $5 \times 5, 7 \times 7, 9 \times 9, 9 \times 9$ 이다.

그림 7은 제안된 비용 집합 방식을 이용한 변이 추정 결과를 보여준다. Middlebury 테스트 베드와 동일한 추정 거리를 이용하여 변이 추정을 수행하기 위해 'Tsukuba' 영상에서는 시점 1, 2, 3을 이용하여 시점 2의 변이 지도가 추정되었고, 'Venus', 'Teddy', 'Cone' 영상

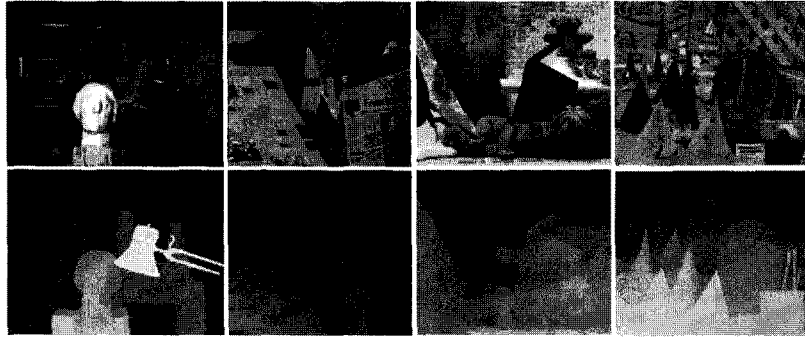


그림 7. 'Tsukuba', 'Venus', 'Teddy' and 'Cone' 영상에 대한 결과
 Fig. 7. Results for the 'Tsukuba', 'Venus', 'Teddy' and 'Cone' image pairs.

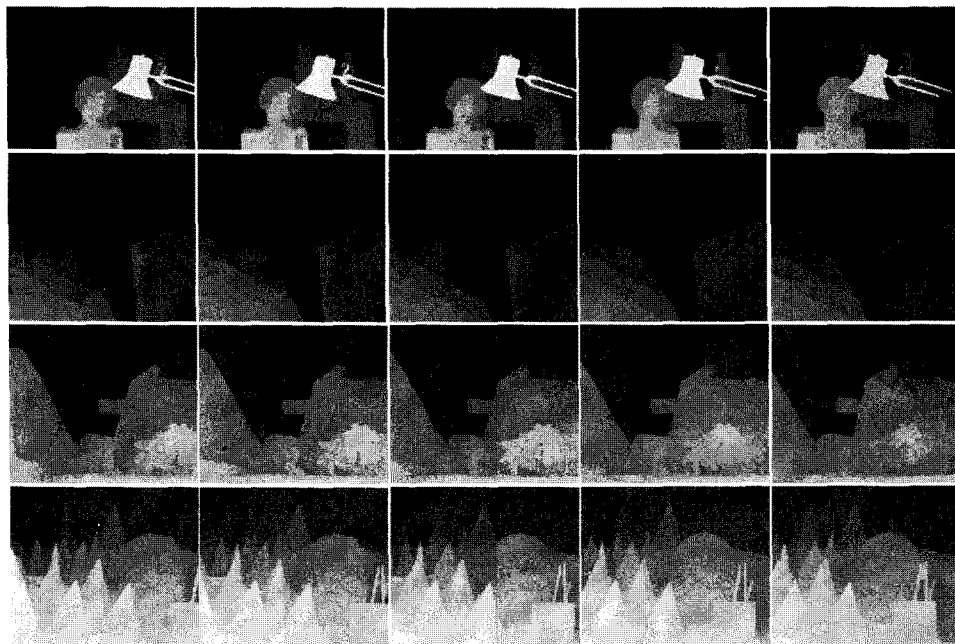


그림 8. 준 N -시점 & N -깊이 구조에서 다시점 영상에 대한 변이 추정 결과 ($N=5$)
 Fig. 8. Results for the multiview image pairs in the semi N -view & N -depth framework ($N=5$)

표 1. N -시점 & N -깊이 구조와 준 N -시점 & N -깊이 구조에서의 처리 시간 비교 ($N=5$)
 Table 1. Processing time for N -view & N -depth and semi N -view & N -depth frameworks ($N=5$).

시점	<i>Tsukuba</i> (s)		<i>Venus</i> (s)		<i>Teddy</i> (s)		<i>Cone</i> (s)	
	<i>N.N.</i>	<i>Semi N.N.</i>	<i>N.N.</i>	<i>Semi N.N.</i>	<i>N.N.</i>	<i>Semi N.N.</i>	<i>N.N.</i>	<i>Semi N.N.</i>
시점 0	2.43	0.59	3.03	0.48	6.58	1.97	6.32	2.09
시점 1	2.39	2.42	2.92	3.05	6.42	6.62	6.27	6.20
시점 2	2.36	0.11	2.89	0.13	6.38	0.28	6.24	0.34
시점 3	2.37	2.39	2.93	3.02	6.33	6.55	6.22	6.22
시점 4	2.41	0.53	3.08	0.47	6.52	1.91	6.37	2.05
합계	11.96	6.05	14.84	7.14	32.23	17.36	31.41	16.91

에서는 시점 0, 4, 8을 이용하여 시점 4의 변이 지도가 추정되었다. 즉, 변이 지도는 N 이 3일 때, 준 N -시점 & N -깊이 구조에서 추정되었다. Ground truth maps은 테스트 영상의 시점 2에 대해서만 제공되기 때문에, 주관적인 평가만이 수행되었다. 제안한 방식이 불연속 지점,

가려진 영역, 텍스처가 부족한 영역에서 정밀한 결과를 도출하는 것을 알 수 있다. 또한, 다시점 영상의 비용 함수를 비교하는 간단한 방식으로 가려진 영역이 정확하게 처리되었다.

다시점 영상($N=5$)에 대해 추정된 변이 지도는 그림



그림 9. 2D 자유시점 합성의 결과 ($N=5$)
 Fig. 9. Results for 2D freeview generation ($N=5$).

표 2. N -시점 & N -깊이 구조와 준 N -시점 & N -깊이 구조에서 합성된 영상의 PSNR 결과
 Table 2. PSNR results of reconstructed views in N -view & N -depth and semi N -view & N -depth frameworks.

$N=5$	<i>Tsukuba</i> (dB)	<i>Venus</i> (dB)	<i>Teddy</i> (dB)	<i>Cone</i> (dB)
<i>N.N.</i>	<i>N. A.</i>	36.99	33.66	31.57
<i>Semi N.N.</i>	<i>N. A.</i>	36.98	33.57	31.67

8과 같다. 그림 8 (b)와 (d)는 참조 영상에 대해 추정된 변이 지도를 보여준다. 그림 8 (c)는 대칭 워핑을 이용하여 계산된 대상 영상의 변이 지도이며, 그림 8 (a)와 (e)는 비대칭 워핑을 통해 계산된 준-대상 영상의 변이 지도이다. 후방(전방) 워핑만이 수행되었기 때문에 제안된 가려진 영역 처리 방식을 이용하여 가려진 영역에 정밀한 비용 함수가 할당되었다. 워핑 방식에 의해 비용 함수가 추정된 대상과 준-대상 영상에서 정확한 경계를 갖는 정밀한 변이 지도가 추정된 것을 알 수 있다.

표 1은 N -시점 & N -깊이 구조와 준 N -시점 & N -깊이 구조에 의한 변이 추정 시 발생하는 복잡도를 보여준다. 준 N -시점 & N -깊이 구조의 처리 시간이 거의 반 정도에 불과한 것을 알 수 있다. 그림 9은 제안된 방식에 의해 합성된 가상시점을 보여준다. 물체 경계나 가려진

영역에서 매우 자연스러운 영상이 합성되었으며, 사용자에게 3DTV를 위한 자유시점 비디오를 제공하기에 충분히 만족스러운 화질을 보여주었다. 'Teddy'와 'Cone' 영상에서는 시점 0과 8에서 영상의 경계에 정보가 없기 때문에, $1.0 \leq \alpha_x^g \leq 3.0$ 에서만 가상시점이 합성되었다. 객관적인 평가를 위해 표 2에서 N -시점 & N -깊이 구조와 준 N -시점 & N -깊이 구조에서 합성된 영상의 PSNR을 비교하였다. 합성된 영상의 PSNR이 거의 비슷하다는 것을 알 수 있다. 3D 가상시점 합성에 대한 결과는 두 장의 영상을 합성함으로써 제공된다. 자세한 결과는 [13]에서 제공된다.

VI. 결 론

본 논문에서 우리는 다시점 영상에서 2D/3D 자유시점을 제공하는 새로운 방식을 제안하였다. 이웃한 영상의 추정된 비용 함수를 이용하여 다시점 영상의 변이 추정 시 발생하는 계산상의 중복을 줄였다. 가려진 영역의 문제는 다시점 영상의 화소당 비용 함수 계산 과정에서 해결되었다. 가상시점은 사용자의 선택에 따라 2D나 3D로 제공될 수 있었다. 실험 결과를 통해 준 N -시점 & N -깊이 구조에서 변이 추정 시 발생하는 복잡도가 약 50%정도 감소하였으며, 가상시점 합성 성능도 유지된다는 것을

알 수 있었다. 향후 연구로 우리는 다양한 카메라 구조에서 가상시점을 제공하는 시스템을 개발하고, 변이 추정의 오류에 강인한 가상시점 합성 방식을 개발할 예정이다.

참 고 문 헌

[1] <https://www.3dtv-research.org>.

[2] E. Chen and L. Williams, "View interpolation for image synthesis," SIGGRAPH, pp. 279-288, 1993.

[3] L. Zhang, D. Wang, and A. Vincent, "Adaptive Reconstruction of Intermediate Views From Stereoscopic Images," IEEE Trans. Circuits and Systems for Video Technology, vol. 16, no. 1, pp. 102-113, Jan. 2006.

[4] A. Criminisi, A. Blake and C. Rother, "Efficient Dense Stereo with Occlusions for New View-Synthesis by Four-State Dynamic Programming," International Journal of Computer Vision, vol. 71, no. 1, pp. 89-110, 2007.

[5] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," Signal Processing: Image Communication, vol. 22, pp. 217-234, 2007.

[6] L. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," SIGGRAPH, pp. 598-606, 2004.

[7] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision, vol. 47, no. 1-3, pp. 7-42, Apr. 2002.

[8] K. Yoon and I. Kweon, "Adaptive support-weight approach for correspondence search," IEEE Trans. PAMI, vol. 28, no. 4, pp. 650-656, Apr. 2006.

[9] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," Proc. IEEE International Conf. Computer Vision, pp. 508-515, 2001.

[10] J. Sun, N-N. Zheng, and H-Y. Shum, "Stereo Matching Using Belief Propagation," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 25, no. 7, pp. 787-800, 2003.

[11] C. Fehn, R. Barre, and S. Pastoor, "Interactive 3-DTV-Concepts and Key Technologies," Proceedings of the IEEE, vol. 94, no. 3, pp. 524-538, Mar. 2006.

[12] <http://vision.middlebury.edu/stereo>.

[13] <http://diml.yonsei.ac.kr/~forevertin>.

저 자 소 개



민 동 보(학생회원)
 2003년 연세대학교 기계전자
 공학부 학사 졸업(공학사)
 2005년 연세대학교 전기전자
 공학과 석사 졸업
 (공학석사)
 2007년 현재 연세대학교 전기전자
 공학과 박사 과정

<주관심분야 : 컴퓨터 비전, 3차원 영상 신호 처리>



손 광 훈(정회원)
 1983년 연세대학교 전자공학과
 학사 졸업(공학사).
 1985년 Univ. of Minnessota 석사
 졸업 (공학석사).
 1991년 North Carolina State
 Univ. 졸업(공학박사)

2007년 현재 연세대학교 전기전자공학과 교수
 1988년 9월~1992년 2월 North Carolina A&T
 State Univ. Research Associate.
 1992년 4월~1993년 10월 한국전자통신연구소
 선임연구원.
 1994년 1월~1994년 12월 Georgetown Univ.
 Post-Doc. Fellow

<주관심분야 : 얼굴인식, 컴퓨터 비전, 3차원 영상 신호처리>