

# 영상 인페인팅을 이용한 다중 시점 영상 생성시의 가려짐 영역 처리

\*김용진 \*\*박종일

한양대학교 가상현실연구소

\*ladinjin@mr.hanyang.ac.kr, \*\*jipark@hanyang.ac.kr

## Occlusion Handling in Generating Multi-view Images Using Image Inpainting

\*Kim, Yong-Jin \*\*Park, Jong-Il

Virtual Reality Lab., Hanyang Univ., Seoul, Korea.

### 요약

본 논문에서는 한 장의 기준 영상과 그에 상응하는 참 깊이 맵을 이용하여 가상의 다중 시점 영상 생성 시 발생하는 가려짐 영역 보간 방법을 제안한다. 이 방법은 영상 인페인팅 기술과 각각의 깊이 정도에 따른 층별 보간 기술을 이용한다. 우선, 기준 영상을 깊이 정보에 따라 여러 개의 층으로 분할한다. 각각의 층에 대해 가려짐 영역 내의 화소들은 영상 인페인팅 기술을 이용하여 보간한다. 마지막 단계에서 개별적으로 보간된 층 영상들은 하나로 합성되어 가상 시점의 영상을 이룬다. 영상을 깊이 정보에 따라 분할함으로써, 각 깊이 정도에 대한 텍스처의 연관성을 보존하며 보간 할 수 있으므로 기존의 방법에 비하여 보다 정확하고 세밀한 가려짐 영역 보간이 가능하다. 본 논문에서는 여러 가지 실험 결과를 통하여 제안한 방법의 효율성을 입증하였다.

### 1. 서론

기술이 진보함에 따라 디스플레이에 있어서 사용자들의 몰입감과 임장감에 대한 요구는 날이 증대되고 있다. 현재 이러한 시장의 기대에 따라 날이 3D 미디어 콘텐츠가 2D 미디어 시장을 점유해 가고 있다.

오랜 시간 동안 3D 영상처리에 관련된 연구가 관련 분야에서 활발히 수행되어 왔다[1,2,6]. 최근, 3D디스플레이 기술은 전자공학과 컴퓨터 기술의 발전에 힘입어 보다 현실감 있는 콘텐츠를 사용자에게 제공할 수 있게 되었으며 이에 따라 3D 디스플레이의 표준화 작업이 컴퓨터 비전 분야에서 큰 이슈로 떠오르고 있다.

일반적으로 3D 영상 디스플레이는 하나의 스테레오 영상을 기반으로 행해진다. 그러나 스테레오 영상은 오직 두 시점만의 뷰(view)를 가지므로 다시점 3D 디스플레이에는 적합하지 않다. 이러한 문제를 해결하기 위하여 모든 시점에서의 영상을 획득하여 전송하는 방법을 생각해 볼 수 있다. 그러나 그러한 방법은 매우 많은 데이터량으로 인하여 전송 선로의 대역폭(bandwidth) 점유의 측면에서 문제점이 있으며 영상의 획득 측면에서도 금전적, 시간적 비용이 매우 크다. 따라서 실용적인 미래형 3D디스플레이 시스템은 한 장의 기준 영상과 그에 상응하는 깊이 맵(depth map)을 이용한 실시간 합성 시스템이 될 것이다 [1].

지금까지 한 장의 영상과 깊이 맵을 이용하여 다중 시점 영상을 생성해 내는 연구가 활발히 진행되어 왔다. 또한, Z-cam이나 레이저 스캐너로 획득한 깊이 맵이 아닌 한 장의 영상과 영상 내부에 포함된 정보를 이용한 깊이 맵 생성 방법이 연구되고 있다[3,5]. 그러나 단일 영상 기반의 다시점 영상 합성 방법은 깊이에 따른 각기 다른 화소(pixel)들의 이동으로 인하여 가려짐 영역(occlusion)이 발생하고, 정보가 없는 이러한 가려짐 영역으로 인하여 정확한 합성 영상을 얻기

힘들다. 그러므로 이러한 가려짐 영역에 대한 효과적인 보간(interpolation)방법에 대한 연구가 필요하다.

본 논문에서는 한 장의 기준 영상과 그에 상응하는 참 깊이 맵(ground-truth depth map)을 이용하여 여러 곳의 주변 시점에서의 영상 합성 방법을 제안하였다. 본 방법은 영상 인페인팅(image inpainting)과 계층적 보간(layered interpolation)을 이용하여 가려짐 영역에 대한 보간을 수행한다. 우선, 기준 영상을 그 깊이 정보에 따라 다수의 영역으로 분할한다. 각각의 분할된 영상에서 가려짐 영역 상의 화소들은 영상 인페인팅 알고리즘을 통하여 보간 한다. 최종적으로, 보간된 각 영역 영상들은 하나의 영상으로 합성되어 가상 시점 영상을 이룬다. 가려짐 영역을 깊이 정보를 고려하여 보간 함으로써 기존의 방법에 비하여 보다 깨끗하고 세밀한 결과 영상을 획득할 수 있다.

이후의 장에서는 관련연구에 대하여 간략히 살펴본다. 또한 본 논문에서 제안하는 시스템에 관하여 설명하고 그 결과를 나타내었으며, 마지막에서는 결과에 대하여 고찰하였다.

### 2. 관련 연구

#### 2.1 가상 다시점 영상 합성 시스템

지금까지 다양한 가상 시점 합성 시스템이 제안되었다. 일반적인 영상 합성 시스템은 다시점 영상을 기반으로 하고 있다[12,13]. 그러나 Fhen은 단일 기준 영상과 그에 상응하는 깊이 맵을 기반으로 영상 획득에서부터 압축, 전송, 합성 과정을 포괄하는 전체적 프레임워크(framework)를 제안하였다[1]. Fhen은 가려짐 영역에 대하여 가우시안(gaussian) 보간을 적용하였다. 하지만 이와 같은 방법은 가상 시점 영상 합성 시 기하학적 왜곡을 발생시켜 세밀한 보간이 어렵기 때문에 실제 입체 디스플레이에 적용하였을 경우 사용자로 하여금 시각적인

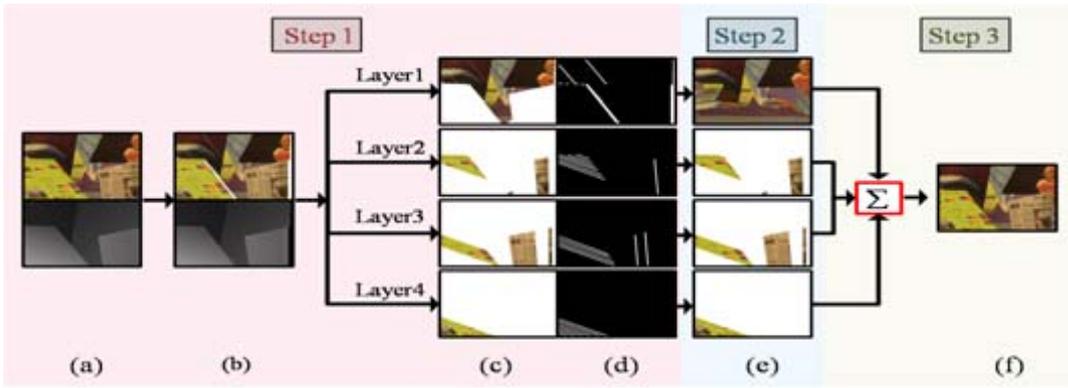


그림 1. 제안한 시스템의 처리 구조(n=4) (a) 기준 영상  $I_s$ 와 상응하는 깊이 맵  $D_s$  (b) 시점 이동 된 기준영상  $I_{mov}$ 와 깊이영상  $D_{mov}$  (c) 계층화 된 영상  $I_1, I_2, \dots, I_n$  (d) 계층화 된 가려짐 영역  $Occ_1, Occ_2, \dots, Occ_n$  (e) 영상 인페인팅을 통해 가려짐 영역이 보간 된 각 계층 영상 (f) 합성 된 영상  $I_{syn}$

거슬림을 느끼게 하는 요인이 된다.

## 2.2 영상 인페인팅

영상 인페인팅 기술은 최근 영상 신호 처리 분야에서 주목 받고 있는 분야로써 다양한 방법들이 제안되고 있다. 과거에 영상 인페인팅을 위하여 사용된 방법은 대부분 보간법에[7]에 의한 것이었으나 보간법을 통한 방법은 에러가 전파(propagation)되는 특성이 있기 때문에 넓은 영역의 보간을 위해서는 적합하지 않다. 이러한 이유로 더 나은 보간 방법을 찾고자 활발히 연구가 진행되었다.

Efros등은 텍스처 합성 기술을 소개하였다. 텍스처 합성 기술은 제한된 텍스처를 합성하여 새로운 텍스처를 생성하는 기술이다[8,9]. 또한 그들은 웹(web) 상의 방대한 영상 자료를 이용하여 이미지 정보의 화소가 없는 부분을 보간하는 획기적인 알고리즘을 제안하였다[10]. Sun 등은 영상 보간 시 구조체가 왜곡 되는 현상을 보완하기 위해 사용자로 하여금 직접 구조체의 모양을 선택 할 수 있도록 하는 방법을 제안하였다. Criminisi등은 하나의 영상 내에는 수많은 화소 정보가 담겨 있다는 전제하에 Efros의 텍스처 합성 방법을 토대로 보간 할 영역에 대하여 가장 적절한 텍스처를 화소 정보가 존재하는 다른 영역에서 찾아 차용하는 방법을 제안하였다[4]. 이른바 모형 기반 방법이라 불리는 이 방법은 그 이후 Drori와 Sun등에 의해 발전되고 보완되었다[11,14].

본 논문에서 사용한 방법은 Criminisi등이 제안한 방법을 시스템에 알맞게 수정하여 사용하였다. 본 방법은 정보가 없는 영역에 대하여 외곽 쪽의 경계(contour)상의 화소부터 한 화소씩 차례로 우선순위(priority)를 정하여 보간하는 방법으로 차용할 화소로는 정보가 있는 영역 상의 가장 유사도가 높은 화소가 선택된다.

## 3. 가상 다시점 영상 합성

제안한 시스템은 다음과 같이 3단계로 나누어진다.

- 단계 1: 기준 영상의 계층적 분리
- 단계 2: 각 계층 영상에서의 영상 인페인팅 보간 수행
- 단계 3: 보간 된 계층 영상들의 합성

그림1 에는 단계1-3의 전반적인 처리 구조를 블록 다이어그램으로 나타내었다. 단계1에서 단계3까지의 처리과정은 원하는 시점 영상이 합성 될 때까지 반복적으로 수행된다. 이 장에서는 각 단계의 세부

적인 처리 과정에 대해 설명하였다.

## 3.1 기준 영상으로 부터의 계층 영상 생성

본 논문에서는 정렬된(aligned) 단일 기준 영상  $I_s$ 와 상응하는 참 깊이 맵  $D_s$ 가 존재함을 가정한다. 가상 시점 영상 합성을 위하여 기준 영상  $I_s$ 와 참 깊이 맵  $D_s$ 의 모든 화소들은 깊이 값에 따라 수평방향으로 이동하게 된다. 시점이동 된  $I_s$ 와  $D_s$ 는  $I_{mov}$ 와  $D_{mov}$ 가 되며 이 가상 시점 영상들은 각각 가려짐 영역을 가진다. 그 후,  $I_{mov}$ 를  $D_{mov}$ 의 값의 분포에 따라 k-means 알고리즘을 이용하여 여러 개의 계층 영상( $I_1, I_2, \dots, I_n$ )으로 나눈다. 나누어진 각 계층 영상의 텍스처들은 각각 유사한 깊이 값을 가진다. 따라서 영상 인페인팅 시, 유사 깊이 값을 갖는 텍스처만을 참고하게 되므로 보다 올바른 가려짐 영역 처리를 할 수 있다. 또한 보간해야 할 가려짐 영역들을 시점 이동 방향의 반대편 방향에 인접한 화소의 깊이 값을 참조하여 구분함으로써 ( $Occ_1, Occ_2, \dots, Occ_n$ ) 영상 인페인팅 시 참조 할 수 있도록 한다.

## 3.2 각 계층 영상에 대한 영상 인페인팅

각 계층 영상  $I_n$ 상의 가려짐 영역들은 모형 기반(exemplar-based) 영상 인페인팅 방법에 의해 각각 보간된다[4]. 이 방법은 각 계층상의 가려짐 영역이 모두 보간 될 때까지 수행되며 보간의 단계는 다음과 같다.

- A. 패치(patch)의 우선순위 계산
- B. 최우선순위(highest priority) 패치에 대한 유사 텍스처 차용
- C. 우선순위 재계산

과정 A에서는 보간 할 영역의 경계 상의 화소에 대해서 보간의 우선순위를 계산한다. 우선순위는 올바른 텍스처와 구조체 보간을 위하여 매우 중요하다. 패치의 우선순위는 경계 상의 화소를 중심으로 하는 패치의 중심점이 갖는 신뢰도(confidence)와 패치 내의 구조(structure)와 관련된 값의 곱으로 결정된다.

우선순위 계산식은 다음과 같으며, 그림 2에 수식의 이해를 돕기 위한 다이어그램을 도시하였다.

$$P(p) = \frac{\sum_{q \in W_p \cap (I-O)} C(q)}{|W_p|} \cdot \frac{|\nabla I_p^\perp \cdot n_p|}{\alpha} \quad (1)$$

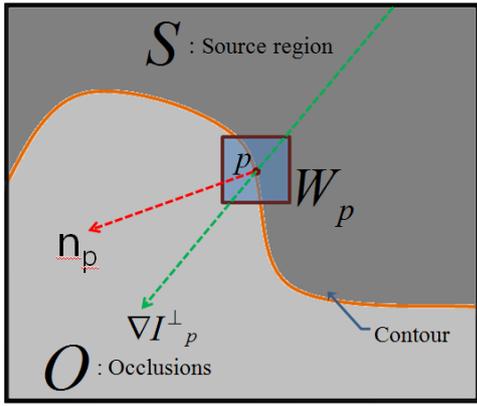


그림 2. 우선 순위 계산 방법에 관한 도식.

(1) 식에서,  $P(p)$ 는 보간 하고자 하는 화소 p의 우선순위를 의미 하고,  $C(q)$ 는 패치내의 화소에서의 신뢰도 값,  $O$ 는 보간해야 할 영역 을 나타낸다.  $W_p$ 는 패치를 의미하고  $|W_p|$ 는 패치의 크기,  $\nabla I_p^\perp$ 는 영상 내 구조체의 방향 단위 벡터,  $n_p$ 는 화소 p에서의 경계(contour)에 대 한 법선 단위 벡터를 나타낸다.  $\alpha$ 는 정규화 상수(normalization constant)이고, 화소의 신뢰도는 소스영역(S)에 포함되는 화소에 대해 서는 1, 그외에는 0으로 부여하였다. 정리하면, 영상 내 구조체의 방향 과 일치하는 방향의 화소에게 더 높은 우선순위를 부여함으로써 구조 체를 더 잘 복원 할 수 있도록 되어있으며, 주변에 신뢰도가 높은 화소 가 많이 분포하는 화소부터 보간 함으로써 보다 정확한 보간이 가능하 다.

최 우선순위가 결정되면 그 화소에 대해 표본 정합(template matching)을 통해 일정 범위 내의 패치와의 컬러 강도(color intensity)비교를 통해 유사도가 최대가 되는 영역을 차용하여 목표 화 소상의 패치영역과 혼합(blending)하게 된다. 이 과정은 앞에서 언급한 과정 B에 해당한다. 본 논문에서는 템플릿 정합의 방법으로 RGB 컬러 공간에서의 SAD(sum of absolute difference)를 사용하였다. 마지막 으로 과정 C에서는 경계 상의 화소중의 최 우선순위를 재계산하며 목 표 화소를 모두 보간 할 때까지 A에서 C의 과정을 반복한다.

### 3.3 각 계층 영상에 대한 영상 인페인팅

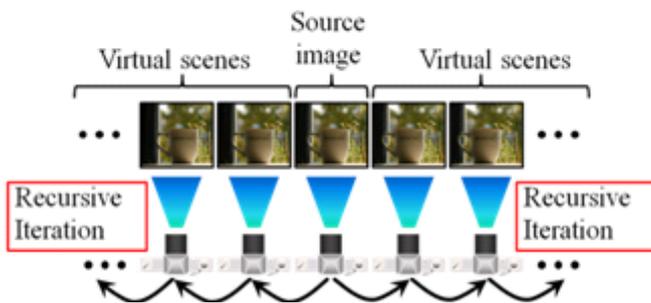


그림 3. 재귀적 반복 구조. 단계 1-3의 반복마다 단위 거리만큼 시점 이동 된 가상 시점 영상을 획득 할 수 있다. 원하는 시점의 개수가 획득 될 때까지 반복을 수행한 다.

각각 개별적으로 가려짐 영역 보간 된 계층 영상  $I_1, I_2, \dots, I_n$ 은 최 종적으로 하나의 영상으로 합성된다.

$$I_{syn}(x, y) = I_{mov}(x, y) \prod_{k=1}^n I_k(x, y) \quad (2)$$

(2) 식에서  $I_k(x,y)$ 는 각 계층에서의 가려짐 영역상의 보간 된 텍

스처 화소를 의미하며,  $I_{syn}$ 은 이동 된 시점에서의 합성 완료된 영상을 나타낸다.  $I_{syn}$ 은 그 다음의 반복과정에서는 기준 영상으로 입력되게 되며 마찬가지로  $D_{mov}$  또한 다음의 과정에서 기준 참 깊이 영상으로 입 력된다.

시스템은 3절의 초반부에서 언급한 단계1-3을 재귀적으로 반복함 으로써 다시점 합성 영상을 획득한다. (그림 3)

## 4. 결과

4장의 기준 영상에 대한 결과를 그림 4부터 그림6 까지 나타내었 다. 본 논문의 시스템에서는 영상의 상태에 따라 계층의 개수 n을 영상 에 따라 수동으로 분할하였으며, 각 패치의 크기는 가로, 세로 9로 설 정하였다.

그림 4에서 사용된 기준 영상은 미들베리(middlebury) 스테레오 비전 홈페이지[15]에서 제공되는 Venus영상이며 간단한 깊이 구조와 복잡한 텍스처 패턴을 가진다. 그림4에서 가려짐 영역 부분에 대한 텍 스처가 적절히 보간 되었음을 확인 할 수 있다. 그림 5에서는 Tripot 영상에 대한 결과들을 나열하였다. 각 영상들은 스텝1-3의 각 반복마 다 얻어진 영상 시퀀스이다. 각각의 영상에서 배경 부분의 체스보드 패 턴의 구조가 왜곡 없이 올바르게 생성되었음을 확인 할 수 있다. 그림 6에서는 본 논문에서 제안한 계층적 보간의 효율성을 나타내었다. 그 림 6의 (a)는 계층적 보간을 수행하지 않은 경우로, 다른 깊이를 갖는 텍스처의 고려로 인하여 그림6의 (b)에 비하여 적절히 보간되지 않았 음을 확인할 수 있다. 하지만 영상 내의 붉은 캔의 일부가 사라지는 문 제가 발생하였다. 이는 가려짐 영역을 다수의 계층으로 나눌 때 각 가 려짐 영역을 가상 카메라의 이동 방향에 대해 인접한 반대 방향 화소 의 깊이 정보를 이용하여 계층화 하는데, 이때 부적절한 깊이 값을 참 조하여 계층화합에서 발생한 문제점이다. 가려져 있던 부분에 대해서 는 전혀 깊이 정보를 알 수 없으므로, 이는 한 장의 기준 영상과 깊이 맵만으로는 해결하기 어려운 문제라 생각 되며, 따라서 향후 연구에서 다룰 예정이다.

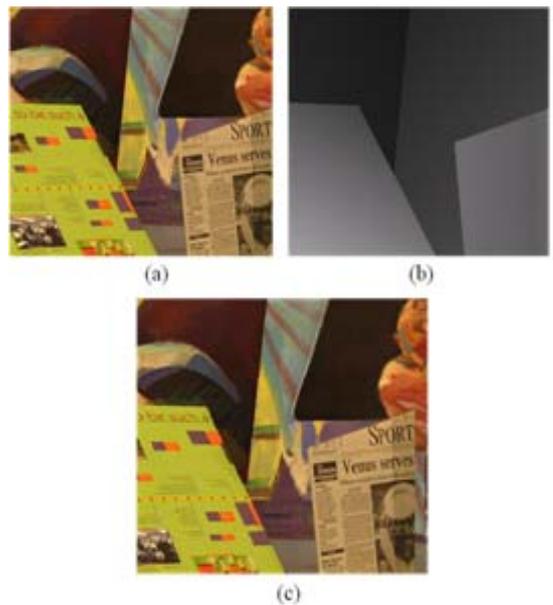


그림 4. Venus영상에 대한 결과 영상. (a) 기준 영상 (b) 참 깊이 맵 (c) 합성 된 영상



그림 5. 합성 된 영상 시퀀스. 가장 좌측의 영상은 기준 영상이며 그 우측 영상들은 계산 된 가상 시점의 영상이다. 가장 우측의 영상은 기준 영상에 대해 가장 큰 변위(disparity)를 갖는다. 각 영상에서, 배경 부분의 체스보드 패턴의 텍스처가 올바르게 보간 되었음을 확인 할 수 있다.

## 5. 결론 및 향후 연구

본 논문에서는 한 장의 기준 영상과 상응하는 깊이 맵을 이용하여 가상의 다시점 영상을 생성해 내는 알고리즘을 제안하였다. 이는 계층별 보간과 영상 인페인팅 방법을 결합함으로써 이루어진다. 영상 인페인팅 방법을 통하여 가려짐 영역에 대해 가장 유사한 텍스처 패턴을 영상 내의 다른 부분에서 차용함으로써 보다 세밀하고 정확한 보간이 가능하며, 계층별 보간을 통해 실 공간상에 존재하는 깊이-텍스처 연계를 살려 보간 시의 텍스처 오류를 감소시킬 수 있었다.

그러나 본 논문에서 제안한 시스템은 몇 가지의 한계점을 가지고 있으며 그것들은 향후 과제로 남겨두었다. 우선, 한 장의 영상으로는 가려짐 영역 부분의 깊이 값을 알 수 없음에서 발생하는 잘못된 텍스처 차용의 경우가 발생하였다. 또한 제안한 방법은 깊이 맵의 품질에 의존적인 경향이 있기 때문에 올바른 가상 시점 영상 생성을 위해서는 정확한 깊이 맵을 사용하여야 하는 한계점이 있다.

현재 이러한 단점을 극복한 보다 견고한 가상 시점 합성 알고리즘을 개발하고 있다. 또한 향후 GPU(Graphic Processing Unit)의 병렬 처리 프로세스 구조를 사용한다면 시스템의 속도를 더욱 상승시킬 수 있을 것이라 생각된다.

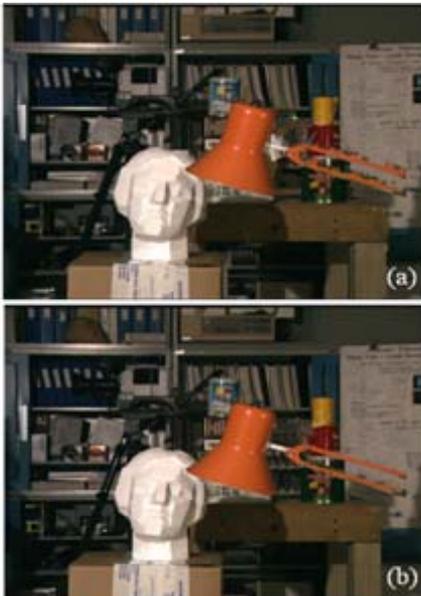


그림 6. 계층별 보간의 효율성. (a) 계층별 보간을 수행하지 않은 경우 (b) 계층별 보간을 수행한 경우

## 6. 참고문헌

- [1] C.Fehn: Depth-image-based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV, Proc. of SPIE, Stereoscopic Displays and Virtual Reality Systems XI, pp. 93-104, 2004.
- [2] J.R.Ohm: Stereo/Multi-view Video Encoding Using the MPEG Family of Standards, Invited Paper, Proc. of Electronic Imaging, 1999.
- [3] A.Saxena, S.H.Chung, and A.Y.Ng: 3-D Depth Reconstruction from a Single Still Image, Proc. of International Journal of Computer Vision (IJCV), 2007.
- [4] A.Criminisi, P.Perez and K.Toyama: Region Filling and Object Removal by Exemplar-based Image Inpainting, Trans. Image Processing, vol.13, no.9, pp. 1200-1212, 2004.
- [5] S.Battiato, S.Curtib, M.L.Casciuc, M.Tortorac, and E. Scordato: Depth Map Generation by Image Classification, Proc. of SPIE Electronic Imaging, pp. 95-104, 2004.
- [6] C.Fehn and P.Kauff: Interactive Virtual View Video (IVVV) "The Bridge between 3D-TV and Immersive TV, Proc. of SPIE Three-Dimensional TV, Video and Display, pp. 14-25, 2002.
- [7] M. Manuel et al.: Fast Digital Image Inpainting, Proc. of the International Conference on Visualization Imaging and Image Processing, pp. 261-266, 2001.
- [8] A. Efros and T. Leung: Texture Synthesis by Non-parametric Sampling, Proc. of IEEE ICCV, pp. 1033-1038, 1999.
- [9] A. Efros and W.T. Freeman: Image Quilting for Texture Synthesis and Transfer, Proc. of ACM SIGGRAPH, pp. 341-346, 2001.
- [10] J.-F. Lalonde and A. Efros: Using Color Compatibility for Assessing Image Realism, Proc. of IEEE ICCV, 2007.
- [11] I. Diori et al.: Fragment-Based Image Completion, Proc. of ACM SIGGRAPH, Vol. 22, pp. 303-312, 2003.
- [12] S.Avidan and A.Shashua: Novel View Synthesis by Cascading Trilinear Tensors, Trans. Visualization and Computer Graphics, vol.4 no.4, pp.293-306, 1998.
- [13] D.Scharstein: View Synthesis Using Stereo Vision, Springer Press, 1999.
- [14] J. Sun et al.: Image Completion with Structure Propagation, Proc. of ACM SIGGRAPH, Vol.24, pp. 861-868, 2005.
- [15] <http://vision.middlebury.edu/stereo>