

하이브리드 카메라를 이용한 고해상도 비디오 합성

김중원^o 경민호

아주대학교 미디어학부
{allonia, kyung}@ajou.ac.kr

High Resolution Video Synthesis with a Hybrid Camera

Jong-Won Kim^o Min-Ho Kyung

Division of Digital Media, Ajou University

요약

최근들어 디지털 영상 기술의 발달로 많은 영화들이 디지털화되어 제작되고 있고, 이러한 디지털 영화를 직접 상영할 수 있는 디지털 시네마로의 전환이 빠르게 진행되고 있다. 하지만 아직까지 디지털시네마에서 요구하는 2K 이상의 고해상도를 지원하는 촬영 장비는 매우 고가이기 때문에 디지털 영화의 제작에 걸림돌이 되고 있다. 본 논문에서는 이러한 문제를 저가의 일반 비디오 카메라와 디지털 스틸 카메라의 조합을 통해 해결할 수 있는 방법을 제시한다. 즉, 비디오 카메라로 촬영한 저해상도 비디오 영상과 동시에 일정한 간격의 키프레임마다 함께 촬영된 고해상도 이미지들을 이용하여 새로운 고해상도 비디오 영상을 합성하는 것이다. 비디오 합성은 프레임 단위로 이루어지는데, 먼저 현재 프레임의 픽셀마다 대응되는 점의 위치를 앞 뒤 키프레임들에서 찾는다. 대응점은 현재 프레임과 키프레임 간의 광흐름(optical flow)을 계산하여 찾게 된다. 대응점이 구해지면 키프레임의 고해상도 이미지에서 대응점 위치의 이미지 블록을 가져와 현재 프레임의 해당 픽셀 위치에 복사한다. 만일 키프레임에서의 대응점을 찾을 수 없는 경우에는 현재 저해상도 비디오 프레임의 확대 이미지를 빈곳 채움에 이용하여 이미지를 완성한다. 본 연구의 결과는 고해상도 비디오 합성 외에 비 사실적 비디오 렌더링 등과 같은 영상 효과의 구현에도 적용할 수 있다.

Abstract

With the advent of digital cinema, more and more movies are digitally produced, distributed via digital medium such as hard drives and network, and finally projected using a digital projector. However, digital cameras capable of shooting at 2K or higher resolution for digital cinema are still very expensive and bulky, which impedes rapid transition to digital production. As a low-cost solution for acquiring high resolution digital videos, we propose a hybrid camera consisting of a low-resolution CCD for capturing videos and a high-resolution CCD for capturing still images at regular intervals. From the output of the hybrid camera, we can synthesize high-resolution videos by software as follows: for each frame, 1. find pixel correspondences from the current frame to the previous and subsequent keyframes associated with high resolution still images, 2. synthesize a high-resolution image for the current frame by copying the image blocks associated with the corresponding pixels from the high-resolution keyframe images, and 3. complete the synthesis by filling holes in the synthesized image. This framework can be extended to making NPR video effects and capturing HDR videos.

키워드: 고해상도 비디오, 광흐름, 하이브리드 카메라

Keywords: high resolution video, optical flow, hybrid camera

1. 서론

본격적인 HD방송과 대중적인 HDTV의 보급, 그리고 전통적인 영화 배급 방식에서 디지털 시네마로의 전환 등은 고해상도 디지털 비디오 영상에 대한 수요를 급격하게 증가시키고 있다. 최근들어 HD급(1920x1080) 보급형 비디오 카메라의 등장, 고해상도 비디오 처리에 필요한 컴퓨터 성능의 향상, 대용량 하드디스크 등은 고해상도 비디오 제작 비용을 많이 낮추어 이러한 수요를 더욱 촉진시키고 있다. 하지만 이러한 IT기술의 발전에도 불구하고 여전히 4K(4096x2160)이상의 해상도

를 요구하는 고품질의 디지털 시네마 영상의 제작은 매우 높은 비용을 요구한다. 이러한 비용의 상당 부분은 고가의 고해상도 디지털 촬영 장비로 인한 것으로 촬영 당시 실시간에 생성되는 대량의 영상 데이터를 처리하고 저장할 수 있는 높은 성능의 하드웨어를 사용해야 하기 때문이다. 따라서, 제작비가 충분한 블록버스터 상업 영화 제작을 제외하고는 독립 영화 등과 같은 저예산 제작에서는 아직도 이러한 고해상도 디지털 영상으로 제작하는 것은 매우 어려운 현실이다.

본 논문에서는 비교적 저렴한 비용으로 구현할 수 있는 하이브리드 카메라를 이용해 고해상도 비디오 영상을 소프트웨어

적으로 합성해 내는 방법을 제안하려고 한다. 하이브리드 카메라는 한 개씩의 고해상도 CCD와 저해상도 CCD를 사용한다. 고해상도 CCD는 일정한 시간 간격으로 정지 영상을 촬영하는데 사용하고, 저해상도 CCD는 비디오 영상을 촬영하는데 사용한다. 개념적으로는 높은 해상도의 디지털 스틸 카메라와 상대적으로 낮은 해상도의 디지털 비디오 카메라를 결합해 놓은 형태로 볼 수 있다. 따라서, 고해상도 비디오 카메라에 비해 상대적으로 낮은 비용으로 제작할 수 있다.

이 하이브리드 카메라에서 얻어진 정지 영상들과 비디오 영상으로부터 광흐름(optical flow) 분석과 이미지 복원 기술을 이용하여 정지 영상 해상도와 같은 고해상도 비디오 영상을 합성하게 된다. 합성 알고리즘을 간단히 설명하면 다음과 같다. 비디오 합성은 프레임 단위로 이루어지는데, 먼저 현재 프레임의 픽셀마다 대응되는 점의 위치를 앞뒤 정지 영상이 촬영된 키프레임들에서 찾는다. 대응점은 현재 프레임과 키프레임 간의 광흐름(optical flow)을 분석하여 구하게 된다. 대응점이 구해지면 키프레임의 고해상도 이미지에서 대응점 위치의 이미지 블럭을 가져와 현재 프레임의 해당 픽셀 위치에 복사한다. 만일 앞뒤 키프레임들 모두에서 대응점을 찾을 수 없는 경우에는 현재 프레임의 비디오 이미지를 확대하여 빠진 부분을 채워 넣는다. 이렇게 채워진 부분은 화면의 다른 부분에 비해 선명하지 않지만, 일반적으로 빠르게 움직이는 영역이기 때문에 시각적으로 두드러져 보이지 않는다.

본 연구의 결과는 고해상도 비디오 합성 외에도 비사실적 비디오 렌더링 등과 같은 다양한 영상 효과의 구현에도 적용할 수 있다. 예를 들어 키프레임에서의 정지 영상을 고해상도 영상 대신에 비사실적 렌더링(NPR) 효과를 적용한 이미지를 주게 되면 이 효과가 반영된 비디오 영상이 만들어지게 된다. 정지 영상에 대한 비사실적 렌더링 효과는 PhotoShop 등의 이미지 편집 도구를 이용하여 자동적으로 만들거나 페이팅 도구와 타블렛을 이용해 직접 손으로 그려서 줄 수 있다. 만일 하이브리드 카메라의 스틸 카메라 부분에 HDR(high dynamic range) 카메라를 사용한다면 HDR 비디오 영상을 합성할 수도 있다. 다중 노출을 통해 쉽게 HDR 촬영이 가능한 정지 영상에 비해 비디오 영상의 HDR 촬영이 훨씬 어려운 점을 고려할 때 하이브리드 카메라는 HDR 비디오 촬영을 위한 한 가지 유용한 해법이 될 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 관련 연구에 관해 살펴보고, 3장에서는 광흐름에 관한 기본적인 정의와 계산 방법에 관해 설명한다. 그리고, 4장과 5장에서는 고해상도 비디오 합성에 관한 구체적인 방법을 설명하고, 6장에서 실험 결과를 보여준다. 마지막으로 7장에서 연구 내용의 요약과 향후 연구 방향을 제시한다.

2. 관련 연구

본 연구의 기반 기술로 사용되는 광흐름 계산에 관해서는 컴퓨터 비전 분야에서 폭넓게 연구되어져 왔다. Lucas와 Kanade[1]는 인접한 픽셀들이 동일한 속도를 가지고 이동한다고 가정하고 미분적 방법으로 광흐름을 계산하는 효율적인 방법을 제안하였다. Horn과 Schunck[2]는 광흐름의 초기 예측값을 점진적으로 개선하여 최적의 결과를 구하는 방법을 제안하였다. 그런데, 일반적으로 광흐름을 넓은 영역을 탐색해서 찾게 되면 계산 시간이 길어질 뿐만 아니라 멀리 떨어져 있는 잘못된 픽셀을 대응 픽셀로 판단하는 경우가 종종 발생한다. 이러한 문제를 해결하기 위해 이미지를 계층적인 피라미드 형태의 모델로 만들고 top-down 방식으로 광흐름을 구해내려 오는 접근법이 많이 사용되고 있다. 이미지 피라미드

를 만드는 방법에는 여러가지가 있는데, Burt와 Adelson[3]은 가우시안 커널을 이용하여 가우시안 피라미드와 라플라시안 피라미드 구성 방법을 제안하였고 이들을 이용하여 이미지를 컴팩트하게 인코딩하는데 사용하였다. Bergen 등[4]은 Lucas와 Kanade[1] 방법을 top-down 순서로 가우시안 피라미드 모델의 레벨에 차례로 적용하여 점진적으로 광흐름의 해를 구해나가는 방법을 제안하였다. 이와 달리 Seliski와 Coughlan [5]은 움직임장(motion field)를 계층적 스플라인 모델로 표현하고 스플라인 제어점들을 점진적으로 구해나가 정확한 광흐름 해를 구하는 접근 방법을 사용하였다. 본 연구에서는 구현이 쉽고 비교적 정확한 결과를 구할 수 있는 Bergen 등의 방법을 사용하여 인접 프레임간의 광흐름을 구하였다.

원래 영상의 해상도를 높여려는 연구는 여러 방향으로 시도되어 왔다. 가장 간단한 방법으로는 이중 선형 보간(bilinear interpolation)을 이용하여 재샘플링을 하는 것으로 구현이 쉽지만 이미지의 블러링이 심하다. Unser[6]는 스플라인 합수 보간법을 이용하여 재샘플링을 하는 방법을 제안하였고, Kondo 등[7]은 최소평균자승(least mean square)을 척도로 하는 비선형적 방법을 제안하였다. 비디오 해상도 업스케일링에 관한 기존 연구들은 [8]에 잘 정리되어 있다. 하지만 재샘플링을 이용한 영상 해상도 향상은 영상의 세밀한 부분까지 복원해 내지는 못하는 한계가 있다. 초분해능(superresolution) 분야에서는 원래 영상의 수학적 분석을 통해 세밀한 부분까지 복원해 더욱 정밀한 고해상도 영상을 만들어 내는 방법에 관한 연구를 활발히 해 왔다. 초분해능 연구는 Tsai[9] 등에 의해 처음 제안된 이후에 많은 연구 결과들 [10, 11]이 나오고 있다.

서로 다른 두 대의 카메라를 이용하여 고해상도 비디오 영상을 만들어 내는 방법에 관한 연구로 Sawhney 등[12]이 입체 영상을 촬영하는데 사용한 하이브리드 카메라가 있다. 이들이 제안한 하이브리드 카메라는 스테레오 카메라의 왼쪽은 고해상도 카메라를 사용하고 오른쪽에는 저해상도 카메라를 사용하여 일반적인 고해상도 스테레오 카메라에 비해 크기와 비용을 줄일 수 있다. 이 카메라에서 얻어진 (고해상도, 저해상도) 비디오 쌍으로부터 프레임별로 변이도맵(disparity map)을 구하고, 이 변이도맵을 이용하여 (고해상도, 고해상도) 비디오 쌍을 합성하였다. 본 논문에서 제안하는 하이브리드 카메라는 이와 달리 저해상도 비디오 카메라와 고해상도 스틸 카메라의 조합으로부터 고해상도 비디오 영상을 합성하는 것이다.

3. 계층적 광흐름 계산

계층적으로 광흐름을 구하는 과정은 두가지로 구분될 수 있다. 첫째가 이미지 피라미드를 구성하는 단계이고, 둘째는 각 레벨에서 점진적으로 광흐름을 구해나가는 단계이다.

3.1 계층 생성

일반적으로 피라미드를 구성하는 데는 Burt와 Adelson[3]이 제안한 방법을 주로 사용한다. 이 방법은 많은 연구 결과들에서 안정적이고 정확한 결과를 보여 주었다. 여기에서는 이 방법에 대해서 간략히 설명하겠다.

이미지 피라미드에서 먼저 레벨 i 의 이미지를 g_i 라고 하고, 원본 이미지는 가장 낮은 레벨의 g_0 라고 하자. g_0 에 로우패스 필터를 적용하여 가로 세로 절반 크기의 레벨 1 이미지 g_1 를 구한다. 다음 g_2, g_3, \dots, g_n 까지 차례로 구해 나간다. 따라서, 이렇게 구해진 이미지 피라미드에서 g_{i+1} 에서 로우패스 필터의 역필터를 적용하면 g_i 가 구해진다. Bert와 Adelson은 로우패스

필터로 가우시안 필터로 사용하였기 때문에 이렇게 만들어진 이미지 피라미드를 가우시안 피라미드라고 부른다.

3.2 광흐름 계산

먼저 우리가 구하고자 하는 광흐름을 수식으로 나타내면 다음과 같다:

$$I(\mathbf{x}, t) = I(\mathbf{x} + \mathbf{u}_t^{t-1}(\mathbf{x}), t - 1). \quad (1)$$

여기서 \mathbf{x} 는 시간 t 에서의 픽셀 좌표 (x, y) 를 나타내고, 함수 $\mathbf{u}_t^{t-1}(\mathbf{x})$ 는 \mathbf{x} 이 대응되는 시간 $t-1$ 에서의 픽셀까지 상대적인 변위 $(u(x, y), v(x, y))$ 를 나타낸다. $I(\mathbf{x}, t)$ 는 시간 t 에서 픽셀 \mathbf{x} 에서의 명도를 나타낸다. 이때 광흐름 문제는 변위 함수 \mathbf{u} 를 찾는 문제가 된다.

일반적으로 $\mathbf{u}_t^{t-1}(\mathbf{x})$ 의 결정은 해당 픽셀을 포함하는 일정한 크기의 영역 R 을 비교하여 SSD(sum of squared distance)가 최소가 되는 대응 픽셀을 구하는 것으로 이루어진다. 이를 수식으로 표현하면 다음과 같다:

$$\mathbf{u}_t^{t-1}(\mathbf{x}) = \arg \min_{\mathbf{h}} \sum_{\mathbf{p} \in R(\mathbf{x})} (I(\mathbf{p}, t) - I(\mathbf{p} + \mathbf{h}, t - 1))^2. \quad (2)$$

광흐름을 구하는 것은 결국 위 식을 푸는 것이라고 할 수 있다. 이를 푸는 것은 시간이 많이 걸리는 전역탐색(exhaustive search) 대신 좀 더 효율적으로 푸는 많은 방법들이 연구되어 왔다. 우리는 그 중에서 일반적으로 많이 사용하는 Lucas-Kanade[1] 방법을 이용하였다.

계층적 광흐름 계산은 가우시안 피라미드를 top-down 순서로 내려오면서 광흐름을 점진적으로 계산하는 Bergen 등[4]의 방법을 사용하였다 (그림 1 참조). 먼저 가우시안 피라미드의 최상위 레벨 n 에서 Lucas-Kanade 방법을 이용하여 광흐름을 구한다. 다음 이중 선형 보간법을 사용하여 광흐름의 해상도를 2배로 높이고 이 결과를 레벨 $n-1$ 에서의 계산을 위한 초기값으로 사용한다. 이 초기값은 이미 레벨 n 에서의 광흐름을 계산한 것이기 때문에 구하려는 해에 상당히 근접해 있다. 따라서 광흐름 계산이 매우 빠르고 정확하게 해에 수렴해 간다. 이러한 과정을 레벨 0까지 진행하면 최종적으로 원래 이미지의 광흐름이 구해진다.

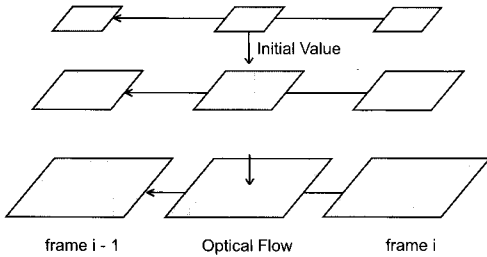


그림 1: 프레임 $i-1$ 과 프레임 i 간의 계층적 광흐름 계산.

4. 고해상도 이미지 합성

현재 프레임과 이전 프레임 간의 광흐름 계산이 끝나면 고해상도 이미지 합성을 시작할 수 있다. 먼저 현재 프레임의 각 픽셀들과 대응되는 점을 가장 가까운 키프레임 이미지에서 찾는다. 키프레임에서 대응되는 점은 현재 프레임에서 시작하여 광흐름을 역으로 거슬러 올라가면 찾을 수 있다(그림 2 참조).

즉, 현재 프레임 t 에서의 한 점 \mathbf{x} 가 대응되는 키프레임 k 에서의 점은

$$\begin{aligned} \mathbf{x}_{t-1} &= \mathbf{x} + \mathbf{u}_t^{t-1}(\mathbf{x}) \\ \mathbf{x}_{t-2} &= \mathbf{x}_{t-1} + \mathbf{u}_{t-1}^{t-2}(\mathbf{x}_{t-1}) \\ &\dots \\ \mathbf{x}' &= \mathbf{x}_{k+1} + \mathbf{u}_{k+1}^k(\mathbf{x}_{k+1}) \end{aligned} \quad (3)$$

만일 \mathbf{x} 에 대해서 중간에 변위 함수 \mathbf{u}_i^{i-1} 의 함수값이 정의되지 않는다면 \mathbf{x}' 는 *unknown*으로 표시한다. 변위 함수의 함수값이 정의되지 않는 경우는 프레임 i 에서 처음 등장하여 프레임 $i-1$ 에 대응되는 점이 없는 경우이다. \mathbf{x} 의 대응점이 *unknown*인 경우는 후처리 과정에서 주변 픽셀들을 이용하여 채워 주게 된다. 식 3을 계산하는 과정에서 변위 함수 \mathbf{u}_i^{i-1} 에 들어가는 인자는 처음을 제외하고 일반적으로 실수 좌표가 된다. 하지만 광흐름은 픽셀에 대해서만 계산하기 때문에 여기서 구해진 변위 함수 \mathbf{u}_i^{i-1} 는 정수 좌표에 대해서만 정의되어 있다. 따라서 실수 좌표에 대해서는 가장 가까운 픽셀의 변위값을 이용해 이중 선형 보간으로 구하는 방법을 사용한다. \mathbf{x}' 를 직접 광흐름 알고리즘을 이용해 계산하지 않고 식 3과 같이 단계적으로 구한 것은 프레임 k 와 프레임 t 사이가 멀어짐에 따라 광흐름 계산이 매우 부정확해지기 때문이다.

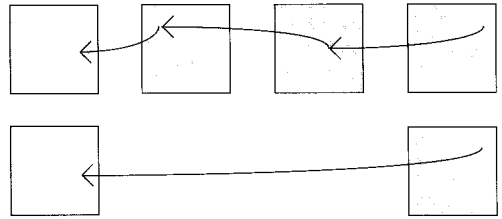


그림 2: 현재 프레임에서 키프레임까지의 광흐름 계산

현재 프레임과 키프레임 간의 대응 관계가 구해지고 나면 다음으로 현재 프레임에 대한 고해상도 이미지를 합성하게 된다. 만일 키프레임 이미지의 한 픽셀당 고해상도 이미지 $m \times m$ 크기 블록이 대응된다면, 현재 프레임의 한 픽셀을 $m \times m$ 크기의 서브픽셀 블록으로 나눈다. 각 서브픽셀에 대응되는 키프레임 이미지의 점 \mathbf{x}' 를 식 3을 이용하여 구한다. 다음 \mathbf{x}' 에 대응되는 고해상도 이미지의 점 \mathbf{x}'_H 에서의 색상을 이중 선형 보간을 이용해 구한 후 합성되는 이미지의 해당 위치에 찍어 준다. 모든 서브픽셀에 대한 처리가 끝나면 다음 픽셀로 넘어가서 합성을 계속 진행한다. 그림 3은 고해상도로 합성된 이미지의 결과를 보여주고 있다. 왼쪽의 저해상도 비디오의 확대 이미지와 비교했을 때 세밀한 부분이 잘 나타남을 알 수 있다.

키프레임에서의 멀어짐에 따라 광흐름 계산이 부정확하게 되어 *unknown* 픽셀이 많이 나타나게 된다. 만일 다음에 오는 키프레임 이미지의 정보도 동시에 사용한다면 이러한 문제를 많이 완화할 수 있을 것이다. 왜냐하면 키프레임 k 에서 멀어짐에 따라 다음 키프레임 l 에는 오히려 가까와져서 대응 점을 찾을 확률이 커지게 되기 때문이다. 이때의 대응점은 식 3의 첨자를 바꾸어 주면 동일하게 구할 수 있다:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{x} + \mathbf{u}_t^{t+1}(\mathbf{x}) \\ \mathbf{x}_{t+2} &= \mathbf{x}_{t+1} + \mathbf{u}_{t+1}^{t+2}(\mathbf{x}_{t+1}) \\ &\dots \end{aligned}$$

$$\mathbf{x}'' = \mathbf{x}_{l-1} + \mathbf{u}_{l-1}'(\mathbf{x}_{l-1}) \quad (4)$$

합성 과정에서는 \mathbf{x}' 와 \mathbf{x}'' 중에 *unknown*이 아닌 대응 점을 선택하게 되고, 만일 모두 *unknown*이 아니라면 두 점에서의 색상값을 블렌딩하여 사용한다. 그림 4은 역방향 광흐름만을 이용하여 합성한 이미지와 양방향흐름 모두 고려하여 합성한 이미지를 보여주고 있다. 후자가 훨씬 깨끗하고 정확하게 합성되었음을 알 수 있다.

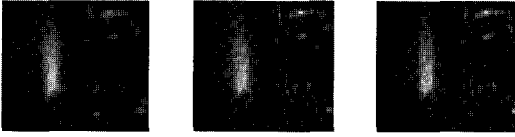


그림 3: 키프레임으로부터 10번째 이미지. 왼쪽부터 저해상도 비디오의 확대 이미지, 합성된 고해상도 이미지, 그리고 원래 고해상도 비디오 이미지.

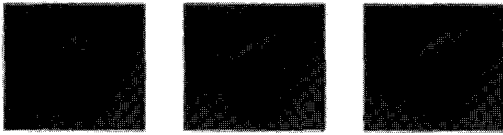


그림 4: 왼쪽부터 역방향 광흐름만으로 합성된 이미지, 순방향 광흐름까지 사용하여 합성된 이미지, 그리고 원래 고해상도 비디오 이미지.

5. 이미지 후처리

Unknown 픽셀인 경우 합성 과정에서 채워지지 않은 채로 남아 있게 되므로, 합성이 끝난 후에 이러한 빈 곳을 인위적으로 채워서 결함이 없는 이미지로 완성을 해야 한다. 주변의 픽셀을 이용하여 이미지의 빈 곳을 시각적으로 자연스럽게 채우는 것을 이미지 인페인팅(image inpainting)이라고 하는데 이미지 처리 분야와 컴퓨터 그래픽스 분야에서 여러 가지 다양한 방향으로 연구가 되어 왔다. 하지만 대부분의 *unknown* 픽셀 영역이 크지 않고 움직임이 빠른 부분에서 주로 발생하기 때문에 복잡하고 시간이 많이 걸리는 이미지 인페인팅 방법까지는 필요가 없다. 대신 간단히 원래 프레임 이미지에서 해당되는 부분을 확대하여 채워 넣는 것으로도 충분히 효과적으로 빈곳 채움을 할 수 있다.

빈곳 채움을 할 때 확대해서 채운 영역과 고해상도 이미지에서 가져온 영역간의 경계에서는 색상의 불연속이 발생할 수 있다. 따라서 이러한 불연속을 없애기 위한 색상 보정 과정이 필요하다. 본 논문에서는 Sawhney 등 [12]이 제안한 방법을 사용하였다. 경계에 놓여 있는 *unknown* 픽셀 \mathbf{x}_u 을 중심으로 5×5 크기의 윈도우를 설정한다. 색상 보정의 가정은 합성된 이미지의 픽셀 색상과 확대한 이미지의 픽셀 색상이 이 윈도우 안에서 선형적인 관계를 갖는다는 것이다. 즉, c_h 와 c_l 은 각각 합성된 고해상도 이미지와 확대한 이미지를 나타내고, R 은 *unknown* 픽셀을 중심으로 하는 윈도우 영역을 나타낼 때,

$$c_h(\mathbf{x}) = M \cdot c_l(\mathbf{x}) \text{ for } \mathbf{x} \in R.$$

이러한 관계를 만족한 행렬 M 을 R 의 *unknown*이 아닌 픽셀들에 대해서 최소자승법(*least-squares method*)를 이용해 푼다. 다음 \mathbf{x}_u 의 색상에 행렬 M 을 곱해 보정해 준다.

6. 실험 결과

본 논문에서 제안한 방법을 실험하기 위해서는 고해상도 디지털 스틸 카메라와 저해상도 디지털 비디오 카메라가 조합된 하이브리드 카메라가 필요하지만, 현재 이러한 형태의 촬영 장비가 개발되어 있지 않기 때문에 720×480 의 크기로 촬영된 일반 비디오 영상을 360×240 크기로 축소하여 실험하였다. 키프레임은 30프레임 간격으로 설정하였고, 키프레임에서의 고해상도 이미지는 원래 비디오 영상에서 해당 프레임 이미지를 캡처하여 사용하였다. 실험에는 1GB ram이 장착된 펜티엄4 3.0GHz를 사용하였고, 광흐름 계산에는 OpenCV에서 구현된 Lucas-Kanade 방법을 사용하였다.

그림 8은 실험 결과에서 5프레임 간격으로 뽑은 고해상도 이미지들이다. 한 프레임을 합성하는데 걸리는 시간을 보면 그림 5에서 보듯이 13~20초 정도의 시간이 걸렸다. 하지만 이것은 최적화를 시키지 않은 코드를 사용한 경우이고 광흐름 계산 부분을 최적화할 경우 합성을 10초 이내로 끝낼 수 있을 것으로 예상된다.

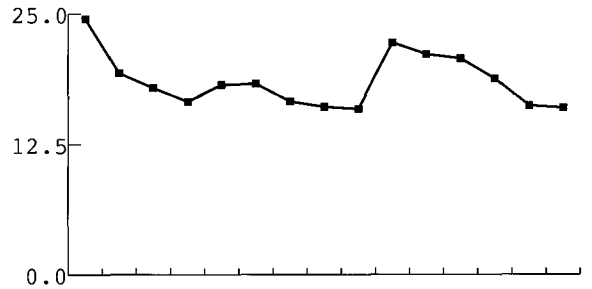


그림 5: 한 프레임을 합성하는데 걸리는 시간. 가로축은 프레임 번호, 세로축은 시간(초)를 나타냄.

그림 6은 320×240 크기의 동영상에서 발생한 *unknown* 픽셀 수를 보여준다. 그래프에서 가로축의 시작은 첫번째 키프레임, 마지막은 두번째 키프레임을 나타낸다. 순방향과 역방향 광흐름을 모두 사용하기 때문에 양 끝으로 갈 수록 *unknown* 픽셀 수가 줄어드는 것을 볼 수 있다. 그래프에서 흰색 사각형선은 움직임이 많은 비디오 영상에 대한 것이고 검은색 사각형선은 움직임이 적은 비디오 영상에 대한 것이다. 전자에서 *unknown* 픽셀 수가 많은 이유는 움직임이 많은 영상에서 광흐름 계산이 부정확해지기 때문으로, 테스트 영상에서 최대 2만 여개의 *unknown* 픽셀(전체의 26%)이 발생하였다. 이것은 매우 부정적인 결과이기는 하지만, 이러한 *unknown* 픽셀 영역에서의 영상 변화가 매우 크므로 빈곳 채움으로 인해 생긴 이미지 품질의 저하가 실제 동영상 재생 과정에서는 크게 두드러지게 나타나지 않는다.

그림 7은 키프레임에 고해상도 이미지 대신 유희 스타일로 변형한 이미지를 주어 유희 스타일의 비디오 영상을 합성한 예를 보여 주고 있다. 여기서는 고해상도 이미지 합성에서와 같이 단순히 원래 저해상도 프레임 이미지를 확대해서 빈곳 채우기를 할 수 없기 때문에 빈곳 주위의 픽셀을 복사하여 빈곳 채우기를 하였다.

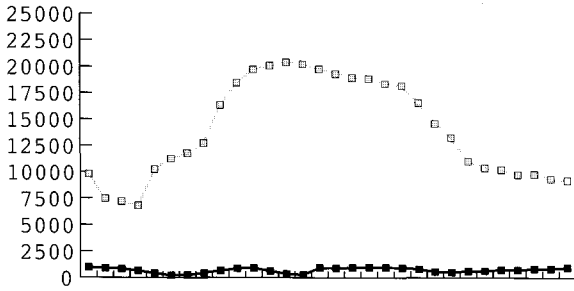


그림 6: 320x240 크기의 움직임이 적은 테스트 동영상(검은색 사각형 선)과 움직임이 큰 테스트 동영상(흰색 사각형 선)에서 발생한 Unknown 픽셀의 수. 가로축은 프레임 번호, 세로축은 픽셀 수를 나타냄.



그림 7: 비사실적 렌더링 실험.

7. 결론

본 논문에서는 하이브리드 카메라를 이용한 고해상도 비디오 합성에 관한 새로운 방법을 제시하였다. 다시 한번 정리하면 저해상도 비디오 카메라와 고해상도 스틸 카메라의 조합으로 만들어진 하이브리드 카메라에서 촬영된 영상은 저해상도 비디오 영상과 일정한 시간 간격을 두고 동시에 촬영된 고해상도 스틸 이미지로 구성된다. 이들을 이용하여 고해상도 비디오를 합성하는 것은 먼저 비디오 영상의 광흐름을 계산하여 현재 프레임의 픽셀들이 대응되는 키프레임 픽셀을 찾는다. 이 대응관계를 이용하여 현재 프레임의 각 픽셀들에 대응되는 키프레임의 고해상도 이미지 블럭을 찾고, 이 고해상도 이미지 블럭들을 차례로 배열하여 현재 프레임에서의 고해상도 이미지를 완성한다. 대응 픽셀을 찾지 못하는 경우는 비디오 프레임 이미지를 확대하여 해당 영역을 채우게 된다.

향후 연구 방향으로서는 합성된 고해상도 비디오의 영상 품질을 향상시킬 수 있는 여러 가지 방법들을 생각해 볼 수 있다. 영상 품질에 결정적인 영향을 미치는 것은 첫째 unknown 픽셀의 갯수, 그리고 둘째는 unknown 영역을 채우는 알고리즘의 성능이다. Unknown 픽셀의 갯수를 줄이는 것은 키프레임 간의 간격을 좁게 하여 대응 픽셀을 찾을 확률을 높여 주는 것이다. 하지만 단순히 키프레임 간격을 좁게 하면 영상 데이터 크기가 지나치게 커져 하이브리드 카메라의 장점이 사라진다. 따라서 바람직한 것은 영상의 움직임에 따라 실시간으로 키프레임 간격이 가변적으로 바뀌게 하여 데이터 크기가 불필요하게 커지는 것을 피하는 것이다. 하지만 실시간으로 키프레임 간격을 최적으로 결정하는 것은 쉽지 않은 문제로 향후 추가 연구가 필요하다.

본 논문에서 사용된 빈곳 채우기 알고리즘은 매우 간단하고 빠른 반면 unknown 영역이 커지는 경우 블러링이 심해지는 단점이 있다. 이런 영역에 대해서는 좀 더 세련된 이미지 인페

인팅 알고리즘을 적용하여 합성된 영상의 품질을 향상할 수 있다[13, 14, 15]. 여기서 주의할 점은 스틸 이미지에서 잘 동작하는 알고리즘을 비디오에 바로 적용할 경우 시간적 앨리어싱(temporal aliasing)이 발생할 수 있다는 것이다. Kokaram 등 [13]이 제안한 방법은 이러한 문제를 고려하여 비디오의 손상된 부분을 채워줄 수 있을 것으로 기대한다.



그림 8: 합성된 고해상도 이미지들.

감사의 글

본 연구는 아주대학교 교내 정착연구비 지원을 받아 수행되었습니다.

참고 문헌

- [1] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 1981, pp. 674-679.
- [2] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 2, pp. 185-204, 1981.
- [3] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communication*, vol. 31, no. 4, pp. 532-540, 1983.

- [4] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proceedings of the Second European Conference on Computer Vision*, 1992, pp. 237–252.
- [5] R. Szeliski and J. Coughlan, "Spline-based image registration," *International Journal of Computer Vision*, vol. 22, no. 3, pp. 199–218, March/April 1997.
- [6] M. Unser, "Splines: a perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, Nov. 1999.
- [7] T. Kondo, T. Fujiwara, Y. Okumura, and Y. Node, "Picture conversion apparatus, picture conversion method, learning apparatus and learning method," Nov. 2001, uS-patent 6323905.
- [8] J. A. Leitaó, M. Zhao, and G. de Hann, "Content-adaptive video up-scaling for high definition displays," in *Proceedings of Image and Video Communications and Processing 2003*, 2003, pp. 612–622.
- [9] T. S. Huang and R. Y. Tsai, *Multiple frame image restoration and registration*. JAI Press, 1984, vol. 1, pp. 317–319.
- [10] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *Journal of Optical Society A*, vol. 6, pp. 1715–1726, 1989.
- [11] R. Schultz and R. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, no. 6, pp. 996–1011, June 1996.
- [12] H. S. Sawhney, Y. Guo, K. Hanna, R. Kumar, S. Adkins, and S. Zhou, "Hybrid stereo camera: an ibr approach for synthesis of very high resolution stereoscopic image sequences," in *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM, 2001, pp. 451–460.
- [13] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald, and P. J. W. Rayner, "Interpolation of missing data in image sequences," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 1509–1519, 1995.
- [14] A. N. Hirani and T. Totsuka, "Combining frequency and spatial domain information for fast interactive image noise removal," in *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM, 1996, pp. 269–276.
- [15] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 417–424.
- [16] H. Johan and T. Nishita, "A progressive refinement approach for image magnification," in *Proceedings of Pacific Graphics 2004*, May 2004, pp. 91–96.