시공간 정합을 이용한 비디오 시퀀스에서의 가려진 객체의 복원

허 미 경[†] · 문 재 경[††] · 박 순 용[†††]

요 약

비디오 복원(video completion) 기술은 비디오 영상에서 색상 정보가 없는 픽셀에 적절한 색을 채워 영상을 복원하는 기술이다. 본 논문에서는 움직이는 물체가 서로 교차하는 비디오 영상에서 원하지 않는 물체를 제거하고 이때 발생한 영상 홀(image hole)을 채우는 비디오 복원 기술을 제안한다. 움직이는 카메라에서 획득한 비디오 영상에서 이동하는 두 물체 중 카메라와 가까운 물체를 제거함으로써 가려진 이동물체와 배경에 홀이 발생하게 되고, 이 홀을 다른 프레임들의 정보를 이용하여 채움으로써 새로운 비디오를 생성한다. 입력 영상의 모든 프레임에 대해 각 물체의 중심을 추정하여 물체의 중심을 기준으로 시-공간 볼륨(spatio-temporal volume)을 생성하고, 복셀 매칭(voxel matching)을 통한 시간적 탐색(temporal search)을 수행한 후 두 물체를 분리한다. 가리는 물체 영역으로 판단된 부분을 삭제하고 공간적 탐색(spatial search) 방법을 이용하여 홀을 채워 가려짐이 있는 이동 물체 및 배경을 복원한다. 복원된 영상에서 블렌딩을 통해 솔기(seam)를 제거한다. 비디오카메라로 획득한 두 실영상을 이용하여 실험을 수행한 결과 가려진 물체를 복원한 새로운 비디오 영상을 생성할 수 있었다.

키워드 : 비디오 복원, 복셀 매칭, 시-공간 볼륨, 디지털 비디오 특수 효과

Completion of Occluded Objects in a Video Sequence using Spatio-Temporal Matching

Heo, Mi Kyoung[†] · Moon, Jaekyoung[††] · Park, Soon-Yong[†††]

ABSTRACT

Video Completion refers to a computer vision technique which restores damaged images by filling missing pixels with suitable color in a video sequence. We propose a new video completion technique to fill in image holes which are caused by removing an unnecessary object in a video sequence, where two objects cross each other in the presence of camera motion. We remove the closer object from a camera which results in image holes. Then these holes are filled by color information of some others frames. First of all, spatio-temporal volumes of occluding and occluded objects are created according to the centroid of the objects. Secondly, a temporal search technique by voxel matching separates and removes the occluding object. Finally, these holes are filled by using spatial search technique. Seams on the boundary of completed pixels are removed by a simple blending technique. Experimental results using real video sequences show that the proposed technique produces new completed videos.

Key Words : Video Completion, Voxel Matching, Spatio-Temporal Volume, Digital Video Special Effects

1. 서 론

비디오 복원은 연속적인 영상 프레임에서 색상 정보가 없는 픽셀에 대해 적절한 색을 채워 복원하는 기술을 말한다. 비디오 영상은 손상이나 자연적인 퇴락으로 인해 홀(hole)이 생길 수 있다. 또는 영상에서 원하지 않는 물체를 제거함으로써 인위적인 손상을 줄 수도 있다. 이러한 홀은 다른 여러 프레임들로부터 보간법을 이용해서 채우는 방법이나 다른 프레임으로 대체하는 방법들이 있는데 이를 비디오 복원(video completion)이라 한다[1-4]. 비디오 복원 기술은 오래된 영상이나 손상된 영상을 복원하는 분야나 영화 특수 효과 분야에서 많이 응용되고 있다[5-7].

이미지 복원(Image Completion)은 정지된 하나의 영상에 대한 복원을 하기 때문에 공간적인 일관성(spatial consistency)을 유지하는 것만 중요한 반면, 비디오 복원은 공

※ 본 연구는 경북대학교 정보기술연구인력양성 사업단(BK21)의 지원으로 수행되었습니다.
† 준 회 원 : 경북대학교 컴퓨터공학과 석사과정
†† 정 회 원 : 경북대학교 전자전기컴퓨터학부 BK21 계약교수
††† 정 회 원 : 경북대학교 컴퓨터공학과 교수(교신저자)
논문접수 : 2007년 3월 26일, 심사완료 : 2007년 6월 1일

간적 일관성뿐만 아니라 시간적 일관성(temporal consistency)도 반드시 유지해야한다[8]. 인간의 눈은 아주 예민하기 때문에 비디오 복원에서는 공간적인 일관성보다 시간적 일관성을 유지하는 것이 더 중요하다. Jiaya Jia[1]는 비디오 리페어링(video repairing) 기술을 이용하여 홀이 있는 비디오를 복원하였다. 이들은 적어도 한 주기의 모션 정보를 가진 샘플 모벨(sample movel)을 정렬하여 손상된 모벨의 홀을 복원하는 방법을 제안하였다. Yun-Tao Jia[2]는 텍스쳐 합성(texture synthesis)을 이용한 비디오 복원 방법을 제안하였다. 홀의 에지부분에서 가장 유망한 목표 픽셀(target pixel)을 선택하고 이 목표 픽셀을 포함한 이웃한 픽셀들을 하나의 목표 플레그먼트로 두고 목표 플레그먼트에서 이미 알고 있는 정보인 이웃한 픽셀과 가장 유사한 소스 플레그먼트를 찾아 두 플레그먼트를 합성하여 홀을 채우는 방법을 사용하였다. 이 방법은 한 픽셀씩 복원하는 방법이 아니라 플레그먼트 대 플레그먼트로 복원하기 때문에 빠르다는 장점이 있다. Y. Wexler[5]는 전역 시-공간(spatio-temporal) 일관성을 갖기 위한 전역 최적화 문제로서 비디오 복원을 다루었다. Y. Zhang[6]은 이동하는 물체와 배경, 그리고 이동물체를 가리는 또 다른 물체 등, 카메라로부터 물체까지의 거리에 따라 여러 계층(layer)으로 영상을 분할하고 이 분할된 영상정보를 이용하여 물체를 삭제였다. 이 기술의 장점은 물체까지의 거리를 추정하여 여러 계층으로 물체를 분할하기 때문에 이동하지 않고 정지되어 있는 물체도 삭제가 가능하다.

본 논문에서는 움직이는 물체가 교차하는 연속적인 비디오 영상에서 하나의 움직이는 물체를 제거하고 이때 발생한 영상 홀을 채우기 위한 비디오 복원 기술을 소개한다. 예를 들어, 두 움직이는 물체가 교차하는 영상에서는 카메라로부터 더 가까이 있는 물체가 더 멀리 있는 물체를 가리게 되고, 가리는 물체를 제거하는 경우 가려진 물체 및 배경에 홀이 발생한다. 따라서 이동 물체의 반복적인 동작 주기를 이용하여 가려진 물체 및 배경의 홀을 채워 비디오를 복원한다. 본 논문에서 제안하는 알고리듬은 시-공간 볼륨(spatio-temporal volume)의 구성과 비디오 영상 복원 두 부분으로 구성되어 있다. 시-공간 볼륨은 입력 영상으로부터 움직이는 물체를 분리하고 움직이는 물체의 중심을 기준으로 분리한 영상을 모든 프레임에 대해 모은 직사각형 형태의 3차원 볼륨이다. 시-공간 볼륨을 구성하기 위해서는 각각의 프레임에서 움직이는 물체에 대한 중심을 구하는 것이 무엇보다 중요하며, 또한 시-공간 볼륨은 시간적 탐색 혹은 공간적 탐색을 위한 데이터베이스가 된다. 두 움직이는 물체가 교차하는 목표 프레임(target frame)에서의 원하지 않는 물체를 삭제하기 위해 두 물체에 해당하는 영역을 정확히 구분해야한다. 이를 위해 복셀 매칭(voxel matching) 방법을 이용한 시간적 탐색을 수행하여 목표 프레임과 가장 유사한 프레임을 결정하고, 이 두 프레임을 비교하여 목표 프레임에서 가리는 물체 영역만을 삭제하여 홀을 만든다. 또한 시간적 탐색 과정의 결과로 정합 후보 프레임들이 얻어지는데, 이 정합 후보 프레임들에 대해서 공간적 탐색 과정을 수행하여 최종 정합 프레임을 결정 하였다. 우리는 공간적 탐색 시 목표 프레임의 주변 프레임을 정합 템플릿(matching template)으로 사용하지 않는다. 목표 프레임과 근접한 프레임에서는 두 물체가 교차하고 있기 때문에 가리는 물체가 포함된 영상을 목표 프레임의 최종 정합 프레임 탐색에 사용할 수 없다. 그러므로 시간적 탐색과 공간적 탐색을 분리하여 수행되고, 시간적 탐색 후 가려진 물체만 분리된 영상과 정합 후보 프레임들 사이의 공간적 탐색을 수행함으로써 더욱 정확한 정합 프레임을 결정 할 수 있다.

결정된 최종 정합 프레임에서 복원 영상의 홀에 해당하는 영역의 색상 정보를 이용하여 비디오를 복원한다. 또한 복원 영상의 기존에 남아있는 물체 영역과 복원된 영역 사이에 솔기를 제거하기 위해 가중치 평균(weighted average) 방법으로 블렌딩을 수행한다. 마지막으로 복원된 물체와 배경을 합성하여 가려진 물체가 복원된 최종 영상을 얻는다. (그림 1)은 본 논문에서 제안한 알고리듬의 순서도이다.

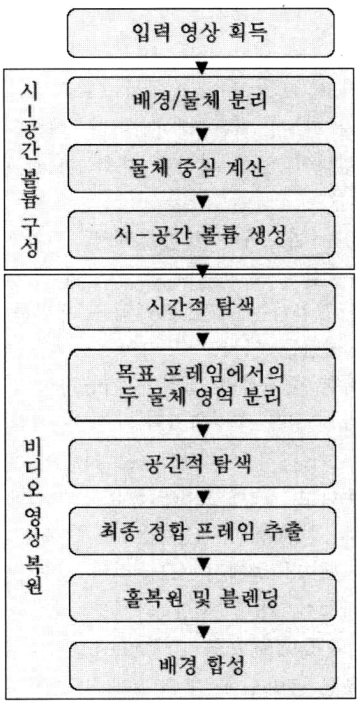
본 논문의 구성은 다음과 같다. 2절에서는 시-공간 볼륨을 구성하는 과정을 설명하고 3절에서는 복셀 매칭을 이용한 시간적 탐색과 공간적 탐색 방법을 설명한다. 4절에서는 두 개의 비디오를 사용한 실험 결과에 대해서 기술하며, 5절에서는 본 논문의 결론을 맺는다.

2. 시-공간 볼륨의 생성

본 논문에서 사용한 비디오 영상은 움직이는 두 물체(사람) 중 가려지는 물체를 따라 회전하는 카메라를 이용하여 획득하였다. 카메라를 렌즈의 중심을 기준으로 회전시켜 영상 프레임간의 2차원 변환을 호모그래피(homography) 변환으로 정의하였고 따라서 프레임 배경을 쉽게 합성할 수 있다. 이러한 특성을 이용하여 배경에서 물체를 추출하고 시-공간 볼륨을 생성하였다.

2.1 움직이는 물체의 추출

동영상에서 움직이는 물체를 추출하기위하여 입력 영상으로부터 배경 영상 추적(background tracking)과 이동 물체에 대한 누적 차영상(accumulated image difference)을 이용하여 움직이는 물체를 분리하였다[9,10]. 즉 영상 프레임간의 배경 이동 정보를 정확히 추적할 수 있다면 우리는 배경에 대하여 움직이는 물체를 추출할 수 있다. 이 과정에서 영상 모자익(image mosaics)기술을 이용하는데, 서로 다른 시점에서 촬영된 영상을 하나의 정지 영상으로 만들기 위해서는 프레임간의 특징점을 추적하고 이들을 이용하여 프레임간의 2차원 투영변환 매트릭스를 구해야한다[11]. 모든 프레임에 대하여 하나의 기준 프레임으로 투영되는 변환 매트릭스를 구하면 하나의 파노라마 영상을 생성할 수 있다. 파노라마 영상이 생성된다는 것은 정지되어 있는 하나의 배경을 생성한다는 것으로 간주할 수 있으므로 배경영상과 현재의 프레임의 차영상(difference image)을 구하면 이동하는 물체를



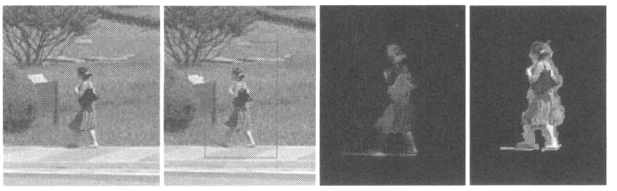
(그림 1)비디오 복원 과정의 순서도

쉽게 찾을 수 있다.

물체의 추적을 위하여 KLT(Kanade-Lucas-Tamosi) 추적기를 사용하였다[12]. KLT 추적기는 물체의 이동을 프레임 사이에서 영상 밝기값의 변화로 나타나며 이것은 두 프레임 사이의 밝기값의 기울기(gradient) 차이로 나타남을 이용한 것이다. 이를 위하여 영상에서 특징점을 찾아내고 다음에 오는 영상 프레임에서 그 특징점과 일치하는 점을 탐색한다. 영상의 특징점은 밝기값의 변화가 가장 많은 픽셀을 선택하고 영상의 특징점이 정해지면 KLT 알고리듬은 두 프레임 사이에서 특징점의 이동을 추적한다. 원래의 KLT 추적기는 범용적인 목적으로 구현되었기 때문에 영상에서 이동하는 물체가 있는지의 여부를 고려하지 않는다. 영상의 특정 영역에 대하여 KLT 알고리듬을 적용하면 그 영역 내에 존재하는 이동물체, 정지물체들의 모든 특징점들이 추적 대상이 된다. 그러므로 KLT 알고리듬만으로는 이동물체를 분리해 낼 수 없다. 이러한 KLT 알고리듬을 보완하기 위해 이동하는 물체의 추적창을 두어 배경픽셀과 전경 픽셀을 구분하였다. 즉, (그림 2(b))에서 추적창은 붉은색으로 표시되어있다. 배경 영역에서 특징점들은 붉은색 점들로 물체 영역의 특징점들은 녹색점들로 표시하였다.

시간 $t-1$에서의 프레임 영상 I_{t-1}의 특징점 (x,y)가 시간 t에서의 프레임 영상 I_t의 (u,v)와 대응점이라 한다면 우리는 2차원 투영변환 매트릭스 M을

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = M \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ m_{20} & m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{1}$$



(a) (b) (c) (d)

(그림 2) 이동물체의 추출. 좌로부터 입력영상의 일부, 물체 추적창 및 특징점, 메디안필터링영상, 추출된 물체영역

으로 정의할 수 있다. 여기서 m_{22}를 1로 가정하고 위의 식을 정리하면

$$u = \frac{m_{00}x + m_{01}y + m_{02}}{m_{20}x + m_{21}y + 1}$$
$$v = \frac{m_{10}x + m_{11}y + m_{12}}{m_{20}x + m_{21}y + 1} \tag{2}$$

$$\begin{pmatrix} x & y & 1 & 0 & 0 & 0 & -xu & -yu \\ 0 & 0 & 0 & x & y & 1 & -xv & -yv \end{pmatrix} \begin{pmatrix} m_{00} \\ m_{01} \\ m_{02} \\ m_{10} \\ m_{11} \\ m_{12} \\ m_{20} \\ m_{21} \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} \tag{3}$$

로 나타낼 수 있다[13]. KLT 추적기를 사용하여 얻은 두 프레임 사이의 대응되는 특징점들을 이용하여 프레임간의 2D 호모그래피를 얻을 수 있다. 이 2D 호모그래피를 이용하여 모든 프레임을 공통 좌표계로 등록(registration)하여 하나의 배경 파노라마를 만든다. 움직이는 물체의 픽셀들은 파노라마 배경에 대하여 메디안 필터링(median filtering)으로 추출하면 우리는 (그림 2(c))와 같은 영상을 얻을 수 있고 이 메디안 필터링된 영상을 문턱값(threshold)을 이용하여 움직이는 물체에 대한 이진화 영상을 얻을 수 있다.

2.2 시-공간 볼륨

각 프레임에서 이동 물체를 추출하면 움직이는 물체 영역을 중심으로 140×220 크기의 물체 영상을 생성한다. 이때 물체의 중심이 영상의 중심이 되도록 하고, 물체 영역이 아닌 영역의 색상값은 0으로 채운다. I_t를 시간 t에서 영상 프레임이라 하고, 픽셀 $v(x,y,t)$ 중 움직이는 물체에 대한 픽셀 면적을 M_t라고 하면, t에 대한 움직이는 물체의 중심 C_t는 식(4)로 구할 수 있다

$$C_t = \sum_{v \in I_t} v / M_t \tag{4}$$

모든 프레임에 대해 물체의 중심을 기준으로 움직이는 물체 영역을 추출한 후, 첫 번째 프레임과 다른 모든 프레임에 대한 SAD(Sum of Absolute Differences)를 계산한다. I_0

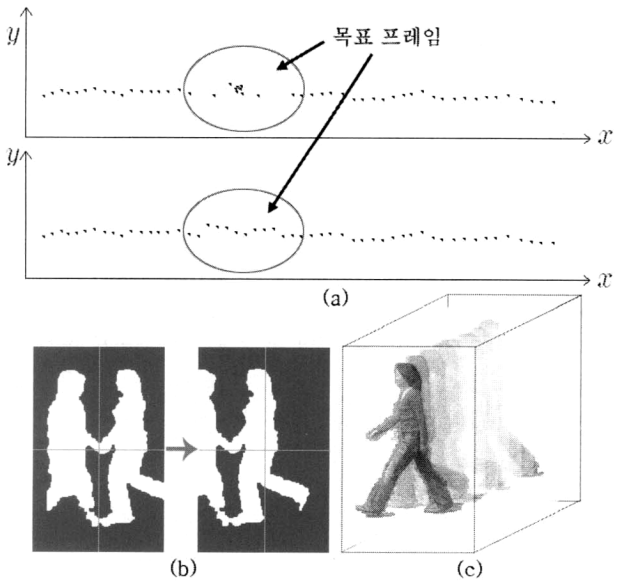
와 I_t의 SAD 오차 e_t는

$$e_t = \sum_{v \in I_t, v' \in I_t} |v(x,y) - v'(x-i, y-j)| \quad (5)$$

으로 계산할 수 있다. 이때 5×5 크기의 탐색창에 대해 식 (5)를 모두 계산하여 그 중 SAD 오차가 최소가 되는 (i,j)를 얻었다. 물체의 중심 추정 시 배경 분리오차, 그리고 사람과 같이 형태가 변하는 물체의 영향으로 인해 오차가 생길 수 있으므로 물체 중심의 이동 변화값 (i,j)를 이용하여 물체의 동작 주기를 파악하고 보다 정확한 물체의 중심을 구하였다. 분리된 중심을 기준으로 모든 프레임의 물체 영역을 이용하여 시-공간 볼륨을 생성하였다.

2.3 파노라마 영상을 이용한 물체의 중심 추정

앞의 2.2절에서 설명한 방법을 이용하여 움직이는 물체의 중심을 추정할 경우 (그림 3(b))처럼 두 물체가 서로 교차하는 프레임에서는 움직이는 두 물체 모두 추적되기 때문에 한 물체에 대한 중심을 정확하게 계산 할 수 없다. 이처럼 두 물체가 모두 나타나는 프레임이 우리가 복원해야할 목표 프레임(target frame)들이고, 이러한 목표 프레임에서 한 물체에 대한 중심을 추정하는 것은 중요한 문제가 된다. 본 논문에서는 한 물체에 대한 중심을 정확하게 구하기 위해 프레임 간의 2D 호모그래피를 이용하여 파노라마 영상을 생성하여 공통의 좌표계에서 물체의 모션을 추정하였다. 파노라마 영상은 하나의 기준 프레임의 좌표계로 다른 모든 프레임들을 등록시킴으로써 얻을 수 있다. 시간 $t-1$에서의 프레임 영상을 I_{t-1}라 하자. I_{t-1}를 프레임 I_t의 좌표계로



(그림 3) (a)물체 중심 추정 전(상)/후(하)의 파노라마에서 물체 중심의 궤적 (b)목표 프레임에서의 물체 중심 추정 (c) 생성된 시-공간 볼륨

등록시키기 위해서는 M_{t-1}을 I_{t-1}에 곱해주면 된다. 이때, M_{t-1}은 I_t 좌표계에 대한 2차원 호모그래피로 $M_{t-1} = M_t H_{t,t-1}$로 구할 수 있고, $H_{t,t-1}$은 I_{t-1}에서 I_t로의 호모그래피로 움직이는 물체 분리 시 얻은 데이터이다. 이 과정을 모든 프레임에서 수행하여 입력 영상에 대한 파노라마 영상 I_p를 얻었다.

(그림 3(a))는 파노라마 영상에서 각 프레임의 움직이는 물체 영역의 중심 좌표를 관찰한 그림이다. 그림에서와 같이 두 사람이 교차하는 목표 프레임을 제외하고 사람의 중심이 일정한 패턴을 가지고 주기적으로 이동하는 것을 알 수 있다. 이러한 주기적인 성질을 이용하여 파노라마 영상에서의 물체의 중심을 추정하였다. I_p에서 시간 t에 대한 물체의 중심을 $c_t(x,y)$라고 하면, x는 사람이 일정한 속도를 유지하며 걷는다고 가정하고 각 목표 프레임에서의 x를 일정한 간격으로 설정하였다 y는 앞, 뒤 주기의 y의 평균값을 취함으로써 (그림 3(b))에서 보는 것과 같이 I_p에서의 목표 프레임의 물체들에 대한 중심 좌표를 결정하였다. 결정된 물체의 중심 좌표 $c_t(x,y)$는 $M_t^{-1} c_t(x,y)$를 취해 원영상으로 역매핑시켜 목표 프레임에 대한 물체의 중심을 추정하였다. 이와 같은 과정을 카메라에서 더 가까이 있는 가리는 물체와 이 물체에 의해 가려진 물체에 대해 각각 수행한다. (그림 3(c))는 복원하려는 물체인 가려진 물체의 시-공간 볼륨을 보여준다.

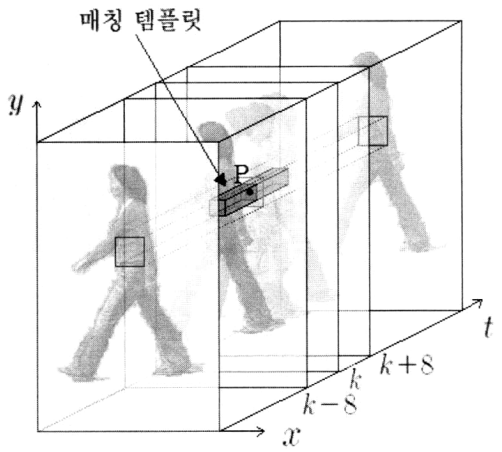
3. 비디오 복원(Video Completion)

전체 영상에서 복원되어야할 목표 프레임들은 두 물체(이동하는 사람)가 교차되는 시점의 연속된 여러 프레임들이다. 다시 말해, 가려진 물체에 대한 시-공간 볼륨에서 카메라에 더 가까이 있는 물체인 가리는 물체가 출현하는 프레임들을 말한다. 이 절에서는 목표 프레임에서 가리는 물체를 삭제하고 가려진 물체에 대한 홀을 복원하는 과정을 설명한다. 복셀 매칭을 이용한 시간적 탐색 방법으로 두 물체를 분리하고 목표 프레임과 정합 후보 프레임들과의 공간적인 탐색 방법을 통해 홀을 대체함으로써 가려진 물체를 복원하였다.

3.1 시간적 탐색(Temporal Search)

시-공간 볼륨에서 목표 프레임에서의 물체 동작과 가장 유사한 프레임을 찾는 과정으로 복셀 매칭 방법을 이용한다. 목표 프레임 중 임의의 k번째 목표 프레임 f_k의 한 복셀 $p(x,y,k)$와 가장 유사한 복셀을 시-공간 볼륨에서 찾는다. $p(x,y,k)$를 중심으로 하는 3×3×17 크기의 복셀들을 정합 템플릿으로 두고 전체 t축을 대해 SAD 오차가 가장 적은 복셀을 찾는다. 이때 f_{k-1}, f_k, f_{k+1}의 복셀 정보는 사용하지 않았다. 앞서 설명했듯이 목표 프레임에서는 두 물체가 교차하고 있기 때문에 한 물체의 동작에 대해 가장 유사한 프레임을 찾는데 방해가 되는 정보를 포함 하고 있다. 즉, k 프레임에서 가려진 물체와 가장 유사한 프레임을 탐

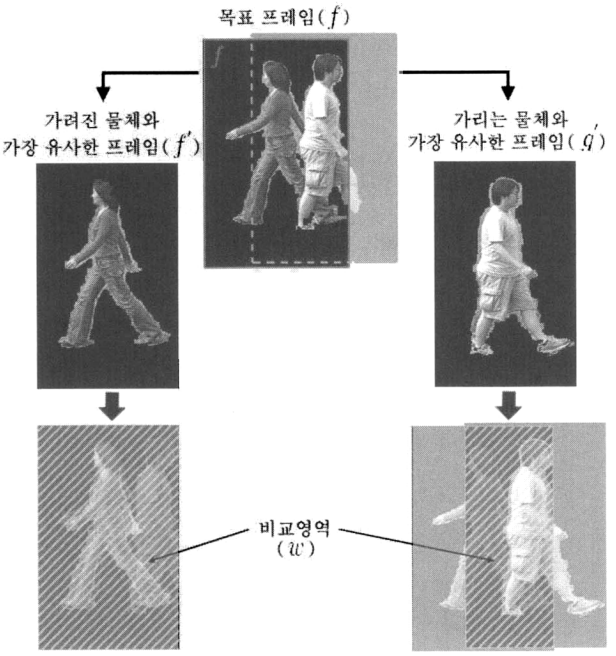
색할 때 k 프레임과 인접한 프레임에서는 가리는 물체에 해당하는 영역이 많이 포함되어있기 때문에 정합 템플릿에서 제외함으로써 보다 정확한 결과를 얻을 수 있다. 또한 탐색 창을 3×3크기로 두어 시간적인 정보뿐만 아니라 공간적인 정보도 이용하도록 하였다. k번째 프레임의 전체 픽셀에 대해 시-공간 상에서 가장 유사한 복셀을 모두 찾는다. 이때 가장 유사한 복셀들의 해당 프레임을 카운트하여 오름차순으로 정렬하고, 참조한 횟수가 가장 많은 프레임을 k번째 프레임과 가장 유사한 프레임으로 결정하였다. 또한 참조 횟수에 따라 정렬된 프레임 정보는 공간적 탐색 시 3.3절에서 후보 프레임 결정에 사용된다. 위 과정을 모든 목표 프레임에 대해 수행하여 가장 유사한 프레임을 결정하였고, 이 과정 역시 두 물체에 대해 수행 한다. (그림 4)는 가려진 물체의 시-공간 볼륨에서의 복셀 매칭을 이용한 시간적 탐색 방법을 보여준다.



(그림 4) 시-공간 볼륨에서의 복셀 매칭

3.2 가려진 이동물체의 분리

이 절에서는 목표 프레임에서 두 물체의 대한 영역을 결정하고 카메라에 더 가까이 있는 물체에 해당 하는 영역을 삭제하는 과정을 설명한다. 우리는 앞서 3.1절에서 시간적 탐색을 통해 시간 t에 대한 목표 프레임 f에서 가리는 물체와 가려지는 물체 각각의 동작과 가장 유사한 프레임을 결정하였다. 여기서 가리는 물체에 대한 가장 유사한 프레임은 가리는 물체의 시-공간 볼륨에서 시간 t에 대한 프레임 g와 가장 유사한 프레임으로 결정하였다. 이를 (그림 5)와 같이 g'라고 하고 가려진 물체에 대한 가장 유사한 프레임을 f'라고 하자. 여기서 g는 f와 같이 시간 t에 대한 영상이지만 가리는 물체를 중심으로 생성한 영상이다. f와 g에 대한 유사 프레임인 f'와 g'사이에 각각의 SSD(Sum of Squared Differences)를 계산 하여 두 물체의 영역을 판단하였다. 목표 프레임 f와 f', g와 g'사이의 오차 e_f, e_g는 각각 다음 식(6)과 (7)을 이용하여 계산 하였다.



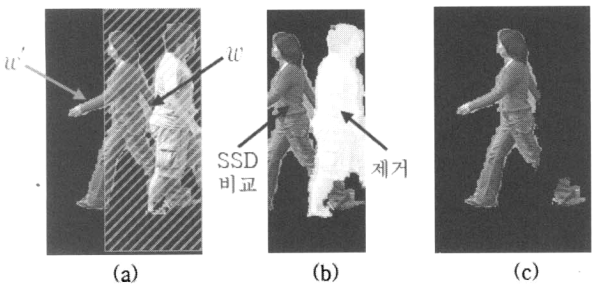
(그림 5) 목표 프레임에 대한 두 물체의 동작과 가장 유사한 프레임 탐색 및 두 프레임 사이의 비교 영역 결정

$$e_f = \sum (f(x,y) - f'(x,y))^2 \qquad (6)$$

$$e_g = \sum ((f(x,y) + d(i,j)) - g'(x,y))^2 \qquad (7)$$

여기서 주의할 점은 두 물체에 대한 각각의 유사 프레임인 f', g'의 물체 중심을 고려해야 한다는 점이다. 가려지는 물체에 대한 유사 프레임 f'의 경우는 목표 프레임 f가 가려진 물체의 시-공간 볼륨에 해당함으로 특별한 물체의 중심의 고려 없이 140×220 영상 전체에 대해 SSD를 계산하면 된다. 하지만 가리는 물체에 대한 유사 프레임 g'는 목표 프레임과는 달리 가리는 물체의 시-공간 볼륨에 해당하는 프레임이므로 f와 g사이의 물체 중심의 차이만큼 g'의 픽셀을 이동시켜 SSD를 계산해야한다. 각각의 시-공간 볼륨에서 시간 t에 대한 f의 물체의 중심 좌표값을 $c_f(x,y)$, g의 물체 중심의 좌표값을 $c_g(x,y)$라 하면, 두 좌표값의 차이로 $d(i,j)$를 구할 수 있다. 즉, $g(x,y) = f(x,y) + d(i,j)$이므로 가려진 물체에 대한 SSD 오차도 식(7)과 같이 계산 할 수 있다. 이때 $c_f(x,y)$와 $c_g(x,y)$의 좌표값은 2절에서 얻어진 값으로, 입력 영상에 대한 움직이는 물체의 중심 좌표를 타나낸다. 또한 이로 인해 두 영상의 비교해야 할 영역 역시 얻을 수 있다. 앞으로의 f와 g'의 비교 시에는 w에 대해서만 수행하면 된다. 역시 SSD 계산도 영역 w에서만 수행하였다. 더 정확한 $d(i,j)$를 구하기 위해 f와 g'가 겹치는 영역에 대해서만 SAD를 다시 한 번 수행한다.

위 과정을 수행 후 목표 프레임은 (그림 6(a))에서 보는 것과 같이 크게 영역 w와 영역 w가 아닌 부분(w')으로 나뉜다. 영역 w는 가리는 물체가 존재할 가능성이 있는 영역이 되고, 영역 w를 제외한 영역은 가려진 물체에 대한 영역으로 결정할 수 있다. 이미 언급했듯이 영역 w는 목표 프레



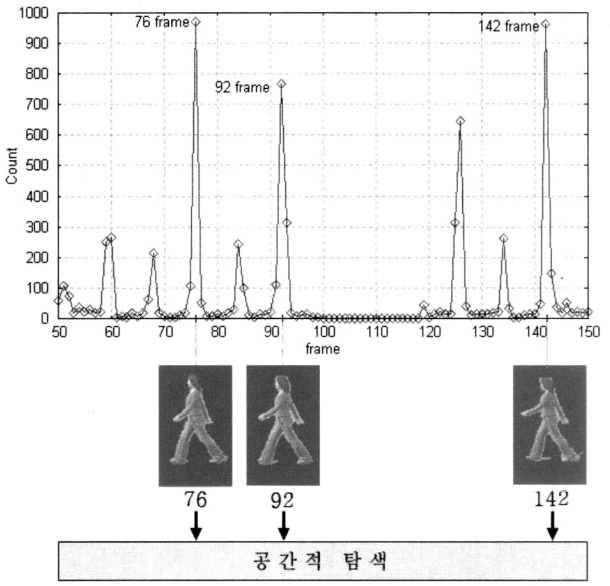
(그림 6) (a)목표 프레임에서의 영역 w (b)영역 w에서의 두 물체 영역 판별 (c) 가리는 물체를 제거한 영상 h

임 f과 가리는 물체에 대한 유사 프레임 g'를 비교하는 영역으로 가리는 물체가 가장 일치하도록 영상을 이동시켜 두 영상을 포개서 비교했을 때 서로 겹치는 영역이 된다. 그러므로 g'에서 물체 영역에 해당하는 f의 영역을 제거 한다. 즉, f에서 가리는 물체 영역이라고 판단된 해당 픽셀의 색상값을 0으로 채운다. 그 결과 영역 w의 나머지 영역에서는 위에서 계산한 SSD 오차를 비교한다(그림 6(b)). SSD 비교 영역 중 한 픽셀이 가려진 물체에 대한 영역이면 e_f가 e_g보다 상대적으로 적은 값을 가질 것이고 가리는 물체에 대한 영역이면 e_f가 e_g보다 상대적으로 큰 값을 가질 것이다. 그러므로 $e_f<e_g$인 영역은 가려진 물체에 대한 영역으로 해당 픽셀의 색상값을 그대로 유지하고, $e_f>e_g$일 경우는 픽셀의 색상값을 0으로 채워 제거하였다. 이 결과 (그림 6(c))에서 보는 것과 같이 목표 프레임에서 가려진 물체 영역이라고 정확히 판단되는 영역만 남는다.

$$\text{픽셀분류}\begin{cases} w' : \text{가려진 픽셀} \\ w : \begin{cases} e_f < e_g : \text{가려진픽셀} \\ e_f > e_g : \text{가리는픽셀} \end{cases} \end{cases}$$

3.3 공간적 탐색 및 홀(hole) 복원

목표 프레임에서 카메라와 더 가까이 있는 물체를 삭제함으로서 영상 프레임에 홀이 발생하였고, 홀을 채우기 위해서 가려진 물체의 주기적인 동작 패턴을 이용하였다. 앞의 3.1절에서 목표 프레임의 물체의 동작과 가장 유사한 프레임을 찾는 과정에서 목표 프레임의 전체 픽셀에 대해 참조하는 프레임을 카운트하여 정렬하였다. 정렬된 데이터에서 가장 카운트가 많은 3개의 프레임을 정합 후보 프레임으로 결정하고, 이들 세 후보 프레임에 대해 공간 탐색을 다시 수행한다. (그림 7)은 109번째 목표 프레임에 대한 예로, 복셀 매칭을 이용하여 찾은 복셀의 프레임을 카운트한 결과를 보여준다. 시간적 탐색 방법으로 109번 프레임과 가장 유사한 프레임으로 참조 횟수가 가장 많은 76, 92, 142번 프레임을 유사 후보 프레임이 결정하였다. 목표 프레임에서 가리는 물체를 제거한 영상을 I_h라 하고, I_h와 세 개의 프레임에 대해 다시 SAD 오차를 계산하여 SAD 오차가 가장 작은 프레임을 최종 정합 프레임으로 결정한다. 목표 프레임은 최종 정합 프레임으로부터 동일한 픽셀의 위치에 있는 색상



(그림 7) 시간적 탐색 후 공간적 탐색 수행

값을 가져와 홀을 채워 가려진 물체를 복원한다.

3.4 복원 영역의 블렌딩

본 논문에서 제안한 비디오 복원 기술은 목표 프레임의 홀을 채우기 위해 다른 프레임으로부터 적절한 영역의 색상값으로 대체하는 방식을 사용하였다. 이때우리는 기존의 남아있는 물체 영역과 새롭게 채워진 영역 사이의 솔기(seam)가 존재하여 복원된 영상에서 부자연스러움이 발견될 수 있다. 목표 프레임에서 두 물체의 영역을 구분하여 가리는 물체를 삭제할 때 두 물체 영역의 경계 부분에 색상값의 모호함 때문에 가리는 물체가 완벽하게 제거되지 않을 수도 있다. 또한 기존의 남아 있는 영역과 새롭게 채워진 영역은 서로 다른 프레임이다. 즉, 두 영역 사이의 시간적 차이 때문에 이들 사이에는 밝기차가 발생할 수 있다. 이러한 문제점을 극복하기 위해 우리는 솔기 주위의 영역을 가중치 평균(weighted average) 방법을 사용하여 블렌딩하였다.

목표 프레임에서 가리는 물체를 제거하고 가려진 물체만 분리한 영상인 I_h에서 모든 픽셀에 대해 홀 영역에 해당하는 픽셀 중 가장 가까운 픽셀을 찾는다. 두 픽셀 사이의 거리가 10픽셀 이하인 영역에서 아래 식을 수행하여 I_h와 최종 정합 프레임을 블렌딩하였다.

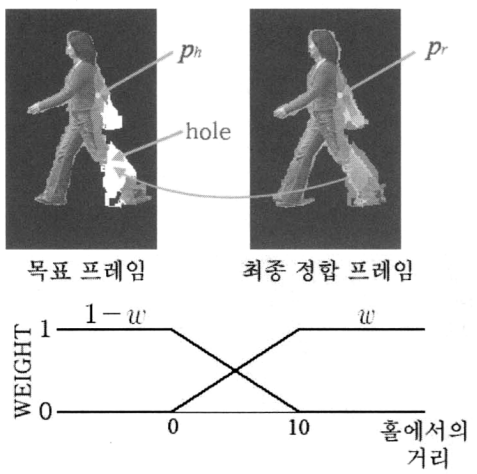
$$P_{comp} = w_b P_h + (1 - w_b) P_r \tag{8}$$

위 식에서 P_{comp}는 블렌딩된 한 픽셀의 색상값을 나타내고, P_h는 영상 I_h, P_r는 최종 정합 프레임에서의 블렌딩할 픽셀의 색상값이다. w_b는 가중치(weighted function)로 블렌딩할 픽셀과 홀 중 가장 가까운 픽셀 사이의 거리에 따라 정규화 시켜 0~1 사이의 값을 가진다.

4. 실험결과 및 분석

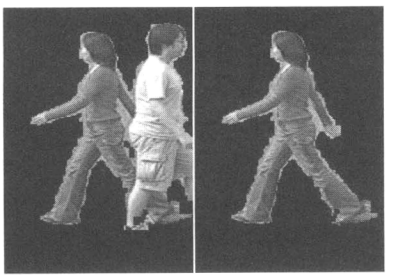
4.1 실험 결과

첫 번째 실험 영상은 SONY 캠코더를 사용하여, 가려지는 물체를 추적하면서 카메라를 회전하여 영상을 획득하였다. 초당 15 프레임으로 촬영했으며, 비디오의 원영상에서 50 프레임에서 150 프레임까지 총 100 프레임의 연속적인 영상을 입력으로 사용하였다. 실험에 사용한 영상에서 102 프레임에서 111 프레임까지 10개의 목표프레임을 복원하였다. 다음 (그림 9)은 목표 프레임 중 결과의 일부를 보여준다. 첫 번째 행은 입력 영상으로부터 움직이는 물체만 분리해 낸 영상으로 두 물체가 교차하는 목표 프레임들이다. 이 두 물체 중 카메라와 더 가까운 물체인 가리는 물체를 삭제하고 홀을 포함한 가려진 물체만 분리된 상을 두 번째 행에서 보여주고 있다. 세 번째 행은 제안된 기술을 이용하여



(그림 8) 목표프레임 복원의 블렌딩

(그림 9) 실험1:움직이는 두 물체가 교차하는 영상, 위로부터 목표 프레임 영상 가려진 물체 영상, 복원된 영상

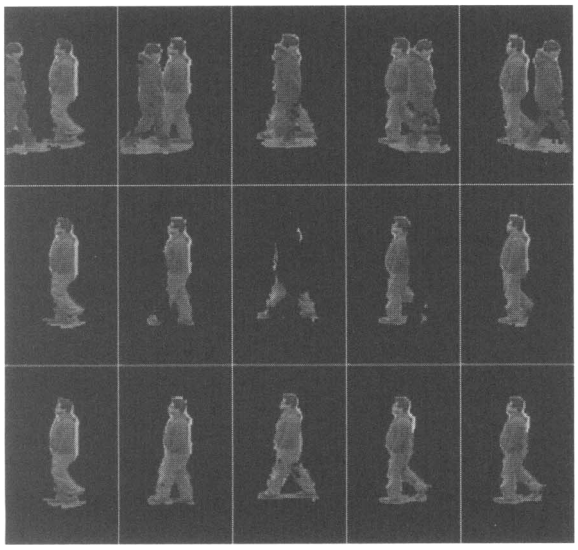


(그림 10) 실험1의 109프레임에 대한 결과. 좌로부터 원영상과 복원된 영상

(그림 11) 실험1: 배경과 합성된 새로운 비디오시퀀스 좌로부터 입력 원영상, 복원된 결과영상

복원된 영상이다. 가려진 부분의 홀을 자연스럽게 채운 결과를 볼 수 있다. (그림 9)에서 물체의 윤곽에 배경이 나타나는 것은 물체 분리 시 발생하는 오차로 추후 개선할 예정이다. (그림 10)은 실험 1의 한 예로 109번째 프레임을 확대한 영상을 보여주고 있다.

다음 (그림 11)은 복원한 물체와 배경을 합성하여 새로운 비디오를 생성한 결과를 보여준다. 목표 프레임이 아닌 프레임에서의 가리는 물체는 이동물체 삭제 기술로 제거하였



(그림 12) 실험2: 두 물체가 교차하는 영상. 위로부터 목표 프레임 영상, 가려진 물체 영상, 복원된 영상

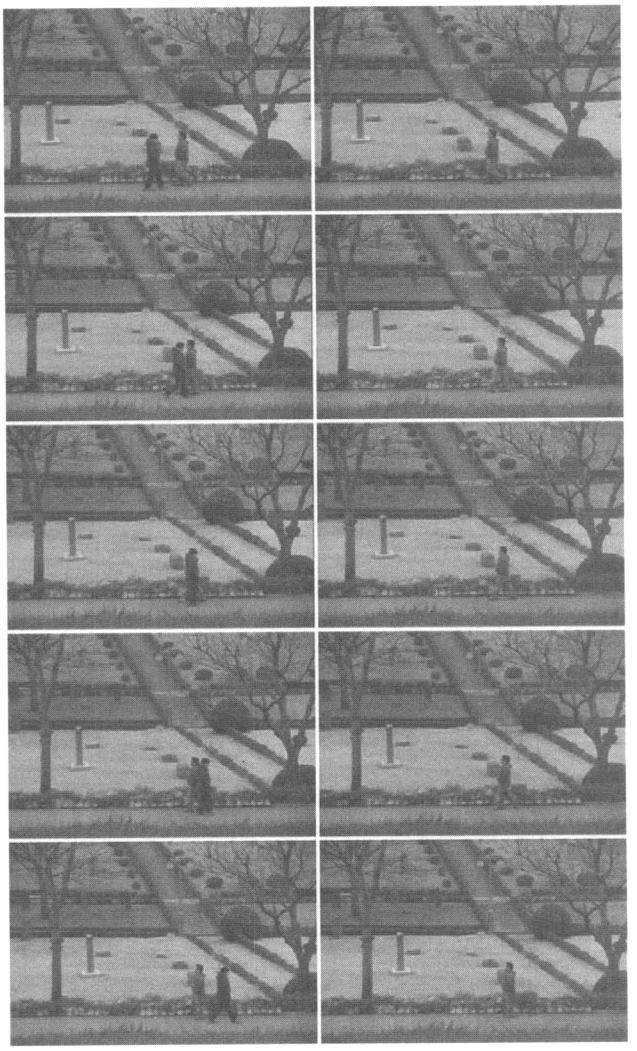
〈표 2〉 알고리듬 수행시간

	실험1	실험2
전체 프레임 수	100	200
목표 프레임 수	10	28
물체창 크기(pixel)	140×220	140×220
총 수행시간(min)	39.72	79.34

다[10]. 새로운 비디오에서는 마치 가리는 물체가 처음부터 없었던 것과 같은 효과를 볼 수 있다.

두 번째 실험은 Canon GL-2 비디오 캠코더를 이용하여 촬영한 영상으로 역시 움직이는 두 물체가 교차하는 영상이다. 이 카메라에서는 초당 30 프레임의 영상을 획득할 수 있었다. 프레임 사이즈가 720×480으로 촬영한 영상에서 300 프레임에서 500 프레임까지 총 200 프레임을 입력 영상으로 사용하였다. 목표 프레임은 404프레임에서 431프레임까지의 28 프레임을 복원하였다. 다음 (그림 12)에서 복원된 결과를 보여준다. 첫 번째 행은 입력 영상에서 움직이는 물체 영역만 분리한 원영상, 두 번째 행은 홀을 포함한 가려진 물체만 분리한 영상, 세 번째 행은 복원된 영상을 보여준다. 세 번째 열의 몇 개의 목표 프레임에 남아 있는 홀은 물체 분리 시 이미 발생된 것이다. (그림 13)은 복원한 물체와 배경을 합성하여 새로운 비디오를 보여준다.

<표 1>은 위 두 실험에 대한 수행 시간을 보여준다. 전체 프레임 수는 입력영상으로 사용한 프레임의 수를 나타내고 목표 프레임 수는 전체 프레임 수에서 복원한 프레임의 수를 나타낸다. 또 입력 영상으로부터 물체만 분리해 낸 물체 창의 크기를 픽셀로, 전체 목표 프레임에 대한 알고리듬 수행 시간을 분으로 표시하였다. 수행 시간은 전체 프레임 수, 목표 프레임 수 그리고 물체창의 크기에 따라 비례하게 증가하는 것을 확인할 수 있다. 비디오 복원 시간이 비교적

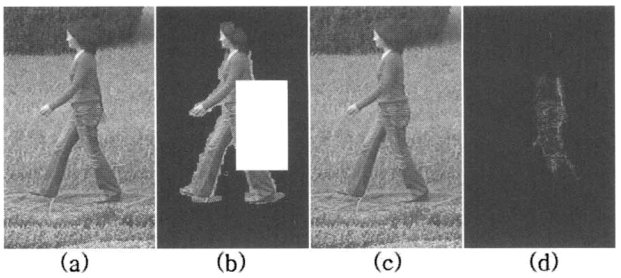


(그림 13) 실험2:배경과 합성된 새로운 비디오시퀀스. 좌로부터 입력 원영상, 복원된 결과영상

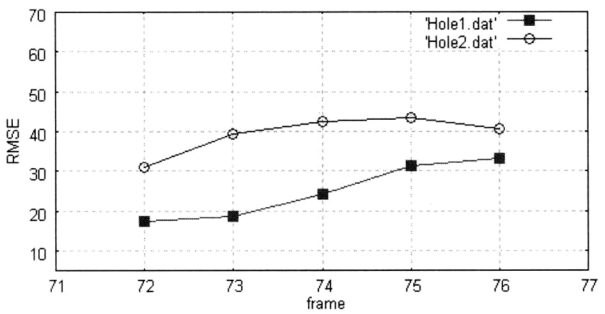
많이 걸리는 것은 목표 프레임에 대한 정합 프레임을 탐색하기 위한 시간 및 공간 탐색에 시간이 소요되기 때문이다. 실험 1은 한 프레임 당 평균 수행 시간은 약 4분, 실험 2는 약 3분 정도 소요되다.

4.2 오차 분석

제안한 알고리듬을 평가하기 위해서 오차 분석을 수행하였다. 우리가 수행한 실험에서는 복원된 결과 영상과 비교할 영상이 없기 때문에 결과의 오차를 계산하는 것은 불가능하다. 따라서 우리는 하나의 물체가 움직이는 비디오 영상에서 임의의 홀을 만들고 우리가 제안한 알고리듬을 수행하여 홀을 복원한 영상과 원영상 사이의 RMSE(root mean-square error)를 계산하였다. 입력 영상은 실험1에서 획득된 영상으로 46 프레임에서 100 프레임까지 총 55 프레임을 사용하였고, 72~76 프레임에 임의의 홀을 만들어 우리가 제안한 알고리듬을 수행하였다. (그림 14)는 오차 분석을 위한 실험의 결과를 보여주는 그림으로 (a)는 원영상, (b)는



(그림 14) 오차 분석:(a)원영상,(b)영상 홀,(c)복원된 영상, (d) 영상(a)와 영상(c)의 밝기 차영상



(그림 15) RMSE 계산 결과

임의의 홀을 정의한 영상을 나타낸다. (c)는 복원된 결과 영상을 보여 주고, (d)는 (a)와 (c)사이의 밝기값의 차이를 나타내는 영상이다. 우리는 홀 영역에 대해 RMSE를 계산하였고 (그림 15)는 그 결과를 보여주는 그림으로, 동일한 입력 영상에 대해 두 종류의 홀을 정의하여 실험을 수행하였다. RMSE는 15에서 45사이의 분포를 가지고 한 프레임 당 평균 약 32.1의 값을 가진다.

5. 결 론

본 논문에서는 움직이는 두 물체가 교차하는 비디오 영상으로부터 하나의 물체를 제거함으로써 생기는 홀을 채우기 위한 비디오 복원 기술을 제안하였다. 비디오 영상에서 움직이는 두 물체의 영역을 추출하고 각 물체의 시-공간 볼륨을 생성하였다. 가리는 물체와 가려진 물체의 유사 프레임을 결정하기위하여 시-공간 볼륨에서 복셀 정합 방법을 사용하여 두 물체에 대해 가장 유사한 프레임들을 결정 하였다. 이 정보를 바탕으로 목표 프레임에서 두 물체를 분리하고, 가리는 물체를 삭제함으로써 가려진 물체에 대한 영역만 남게 되고 나머지 영역에서는 홀이 발생하였다. 가려진 물체의 동작 주기를 바탕으로 홀에 적절한 색상값을 채움으로서 손상된 목표 프레임들을 복원하였다. 또한 복원된 영상에서 기존의 남아있던 물체 영역과 다른 프레임으로부터 새롭게 채워진 영역사이의 블렌딩을 통해 솔기를 제거하여 시간적, 공간적 일관성을 높였다. 그 결과 비교적 자연스럽게 새로운 복원된 비디오 영상을 얻을 수 있었다. 제안된 기술은 움직이는 물체가 두 개 이상인 경우에도 물체 영역에 대한 추적 및 추출만 가능하다면 항상 적용이 가능하다.

지금까지의 연구는 인위적인 홀을 만들고 다른 프레임으로부터 시-공간상에 가장 적절한 영역을 그대로 가져와서 홀을 채우고 블렌딩하여 영상을 복원하였다. 앞으로의 연구는 홀을 복원할 때 움직이는 물체에 대한 정확한 동작을 추정하여 더 자연스러운 결과를 얻고자 한다. 이를 위하여 홀 영역과 홀의 주위 영역에 대한 목표 프레임과 정합 프레임의 모션 정보를 추출하는 기술이 적용되어야 할 것이다.

참 고 문 헌

[1] Jiaya Jia, Tai-Pang Wu, Yu-Wing Tai, Chi-Keung Tang, "Video Repairing: Inference of Foreground and Background under Severe Occlusion," IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 364-371, 2004.

[2] Yun-Tao Jia, Shi-Min Hu, Ralph R. Martin,"Video completion using tracking and fragment merging," The Visual Computer, Vol. 21, No. 8-10, pp. 601-610, 2005.

[3] K. A. Patwardhan, G. Sapiro, and M. Bertalmio, "Video inpainting of occluding and occluded objects," Proceedings of IEEE. International Conference on Image Processing Genoa, Italy, 2005.

[4] Takaaki Shiratori, Yasuyuki Matsushita, Sing Bing Kang, and Xiaoou Tang,"Video completion by motion field transfer," IEEE Computer Vision and Pattern Recognition, Vol 1, pp. 411-418, 2006

[5] Y. Wexler, E. Shechtman, M. Irani,"Space-Time Video Completion," IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 120-127, 2004.

[6] Y. Zhang, J. Xiao, M. Shah, "Motion Layer Based Object Removal in Videos," IEEE Workshop on Application on Computer Vision, Jan 5-6, Breckenridge,Colorado, 2005.

[7] A. Schodl, R. Szeliski, D. Salesin, and I. Essa, "Video Textures," Proc. SIGGRAPH, pp. 489-498, 2000.

[8] M. Bertalmio, G. Sapiro, C. Ballester, and V. Caselles, "Image Inpainting," Proc. SIGGRAPH, pp. 417-424, 2000.

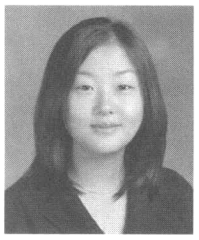
[9] Soon-Yong Park,"Moving Object Removal Based on Global Feature Registration," ACIVS 2006, LNCS 4179, pp.275-286, 2006.

[10] Soon-Yong Park, Chang-Joon Park, and Inho Lee,"Moving Object Removal and Background Completion in a Video Sequence," IVCNZ2005, 2005.

[11] R. Szeliski, "Video Mosaics for Virtual Environments," IEEE Computer Graphics and Applications, pp. 22-30, 1996.

[12] J. Shi and C. Tomasi."Good features to track," In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1994.

[13] Richard Hartley and Andrew Zisserman, 'Multiple View Geometry in Computer Vision', 2nd Ed., Cambridge University Press, 2004



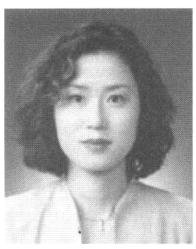
허 미 경

e-mail : mkheo@vision.knu.ac.kr

2005년 대구대학교 멀티미디어 공학과(학사)

2006년~현재 경북대학교 컴퓨터공학과 석사과정

관심분야 : Video completion, Video repairing, Digital video spacial effect, Image analysis, Computer graphics



문 재 경

e-mail : jkmoon@ee.knu.ac.kr

1991년 경북대학교 전자공학과(학사)

1993년 경북대학교 전자공학과(공학석사)

1999년 경북대학교 전자공학과(공학박사)

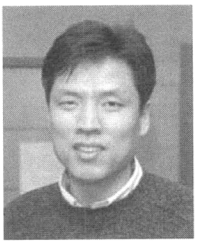
1993년~2001년 한국전자통신연구원 선임연구원

2000년~2004년 뉴욕주립대 스토니브룩 방문연구원

2005년~2006년 경북대학교 센서기술연구소 연구원

2006년~현재 경북대학교 전기전자컴퓨터학부 BK21 계약교수

관심분야 : Image Processing, Depth-from-Defocus



박 순 용

e-mail : sypark@knu.ac.kr

1991년 경북대학교 전자공학과(학사)

1993년 경북대학교 전자공학과(공학석사)

2003년 뉴욕주립대 스토니브룩(공학박사)

1993년~1999년 한국원자력연구소 선임연구원

2003년~2004년 뉴욕주립대 스토니브룩 박사후연구원

2004년~2005년 한국전자통신연구원 디지털컨텐츠연구단 선임연구원

2005년~현재 경북대학교 컴퓨터공학과 조교수

관심분야 : 3D Computer Vision, Image Processimg, Digital Video Effect, Stereo Vision