

# 비디오 영상에서의 비보정 3차원 좌표 복원을 통한 가상 객체의 비디오 합성

박종승<sup>\*</sup>, 성미영<sup>\*\*</sup>

## 요 약

본 논문에서는 비디오에서 비보정 3차원 좌표의 복원과 카메라의 움직임 추정을 통하여 가상 객체를 비디오에 자연스럽게 합성하는 방법을 제안한다. 비디오의 장면이 부합되도록 가상 객체를 삽입하기 위해서는 장면의 상대적인 구조를 얻어야 하고 비디오 프레임의 흐름에 따른 카메라 움직임의 변화도 추정해야 한다. 먼저 특징점을 추적하고 비보정 절차를 수행하여 카메라 파라미터와 3차원 구조를 복원한다. 각 프레임에서 카메라 파라미터들을 고정시켜 촬영하고 이들 카메라 파라미터는 일정 프레임 동안 불변으로 가정하였다. 제안된 방법으로 세 프레임 이상에서 작은 수의 특징점 만으로도 올바른 3차원 구조를 얻을 수 있었다. 가상객체의 삽입 위치는 초기 프레임에서 특정 면의 모서리점의 대응점을 지정하여 결정한다. 가상 객체의 투사 영역을 계산하고 이 영역에 이음새가 없도록 텍스처를 혼합하여 가상객체와 비디오의 부자연스러운 합성 문제를 해결하였다. 제안 방법은 비보정 절차를 선형으로만 구현하여 기존의 방법에 비해서 안정성과 수행속도의 면에서 우수하다. 실제 비디오 스트림에 대한 다양한 실험을 수행한 결과 여러 증강현실 응용 시스템에 유용하게 사용될 수 있음을 입증하였다.

## Video Augmentation of Virtual Object by Uncalibrated 3D Reconstruction from Video Frames

Jong-Seung Park<sup>\*</sup>, Mee Young Sung<sup>\*\*</sup>

## ABSTRACT

This paper proposes a method to insert virtual objects into a real video stream based on feature tracking and camera pose estimation from a set of single-camera video frames. To insert or modify 3D shapes to target video frames, the transformation from the 3D objects to the projection of the objects onto the video frames should be revealed. It is shown that, without a camera calibration process, the 3D reconstruction is possible using multiple images from a single camera under the fixed internal camera parameters. The proposed approach is based on the simplification of the camera matrix of intrinsic parameters and the use of projective geometry. The method is particularly useful for augmented reality applications to insert or modify models to a real video stream. The proposed method is based on a linear parameter estimation approach for the auto-calibration step and it enhances the stability and reduces the execution time. Several experimental results are presented on real-world video streams, demonstrating the usefulness of our method for the augmented reality applications.

**Key words:** Video Augmentation(비디오 합성), Auto-calibration(비보정 캘리브레이션), Feature Tracking(특징점 추적), Seamless Blending(이음새 없는 합성)

※ 교신저자(Corresponding Author) : 박종승, 주소 : 인천광역시 남구 도화동 177번지(402-749), 전화 : 032)770-8428, FAX : 032)766-6894, E-mail : jong@incheon.ac.kr  
접수일 : 2005년 7월 15일, 완료일 : 2005년 12월 5일

<sup>\*</sup> 종신회원, 인천대학교 컴퓨터공학과 조교수

<sup>\*\*</sup> 종신회원, 인천대학교 컴퓨터공학과 교수  
(E-mail : mysung@incheon.ac.kr)

※ 본 연구는 2005년도 인천대학교 교내연구비 지원에 의하여 수행되었음.

## 1. 서론

최근 실 세계(real world) 비디오 프레임에 가상의 물체를 위치함으로써 실 공간과 가상공간을 혼합하는 증강현실 관련 연구가 활발히 진행되고 있다[1]. 가상 객체는 3차원의 형태로 주어지며 합성을 위해서는 실 세계 비디오의 3차원 정보를 획득해야 한다. 3차원 정보 획득을 위해서는 구조 복원 절차가 필요하며, 이를 위한 여러 방법들이 있다. 간단한 방법으로 3차원 스캐너를 사용할 수 있다. 이는 데이터 획득 시에 실내 고정장치라는 공간적인 제약이 따를 뿐만 아니라 고가의 장비이기에 대중화가 될 수 없다는 것이 가장 큰 단점이다. 3차원 정보를 고려하지 않고 비디오 내의 영역 간의 관계만으로 비디오 혼합을 수행하는 방법도 있다. 예를 들어 움직임 물체를 추적하고 그 물체의 영역을 다른 형태로 변형하거나 배경을 변형하는 응용이 있다[2,3]. 이러한 접근은 촬영된 물체의 실제 3D 형상 정보가 없으므로 가상 카메라의 위치나 방향을 변경할 수 없는 단점이 있다.

특수한 3D 카메라를 사용하지 않고 3차원 구조를 얻는 방법으로 범용 디지털 카메라 같은 저렴한 장비로 촬영한 영상을 이용하여 모델링하는 방법이 있다[4]. 이는 연속적인 영상으로부터 특징이 되는 부분이 존재하면 3차원 복원이 가능하기 때문에 비용이 매우 저렴하다. 또한 보정 패턴이 불필요하기 때문에 실외의 대상에 대해서도 3차원 모델링이 가능하다[5]. 이러한 영상을 이용하는 3차원 모델링 방법에는 보정(calibration) 방법과 비보정(auto-calibration) 방법이 있다[6]. 보정 방법은 획득한 영상으로부터 3차원 좌표까지 복원할 수 있는 장점이 있으나 수동적인 절차로 카메라 파라미터 정보를 미리 계산해야 하는 단점이 있다.

본 논문에서는 카메라의 파라미터 정보 없이도 3차원 모델링이 가능한 비보정 방법을 사용하여 장면을 복원하고 가상물체를 혼합하는 방법을 제시한다. 3차원 복원 과정은 특징점 추적 단계와 비보정 단계로 나눌 수 있다. 특징점 추적 단계에서는 모서리점과 같은 특징점들을 추출하고 이를 연속 프레임에 걸쳐서 추적한다. 추적된 데이터의 정확성을 높이기 위해서 이상치(outlier)를 제거한다. 비보정 단계에서는 특징점 추적 데이터를 사용하여 이에 해당하는 3차원 좌표와 카메라 파라미터를 계산한다. 기존의 구조 계산에 대한 방법들은 2차원과 3차원 사이의

제약식으로부터 유도된 오차식을 반복적 최소화하는 기법에 기반하고 있다[1,6,13]. 이들 방법은 제약식의 가정을 준수하는 입력 데이터에 대해서 정확한 결과를 생성하는 장점이 있으나 많은 미지수 때문에 선형식의 유도가 어렵고 또한 비선형 최적화로 인해 수렴 시간이 오래 걸리는 문제점이 있다. 본 논문에서는 선형식으로만 구조 복원이 가능한 간편한 절차를 제안한다. 구조 복원 시에 고려하는 미지수도 가상 객체의 비디오 합성에 필요한 미지수들만을 대상으로 하여 연속하는 세 프레임으로부터 선형복원이 가능하도록 한다.

비디오의 장면 구조의 복원은 가상 객체를 삽입하기 위한 정보의 획득에 해당한다. 장면 구조가 획득되면 주어진 위치에 가상 객체를 삽입할 수 있다. 삽입 위치의 지정은 가상 객체의 한 면의 모서리점들에 대해서, 각 모서리점에 대응될 영상점들을 추적된 특징점들 중에서 선택 지정하여 삽입 위치를 결정한다. 지정된 위치에 대해 가상 객체를 삽입하는 합성 과정이 실시된다. 비디오 합성이란 실사 비디오에 컴퓨터가 생성한 그래픽 모델을 자연스럽게 어울리도록 넣어 비디오 프레임을 변형하는 것이다[7,8]. 이러한 비디오 합성은 영화, 게임, 방송, 의료 등에 매우 다양하게 사용되고 있으며 매우 중요한 문제로 인식되고 있다. 비디오 합성을 위해서 해결해야 할 난제는 삽입할 정확한 위치를 찾아내는 것과 그 위치에 원래의 비디오 프레임의 주변과 어울리도록 그래픽 객체를 합성하는 문제이다.

가상 객체와 비디오의 합성의 전체적인 절차는 다음과 같다.

단계 1: 비디오 프레임에서 특징점을 추출하고 이를 추적한다.

단계 2: 비보정 절차를 통하여 각 세 이웃 영상에 대하여 장면 구조를 계산한다. 각 프레임에 대한 카메라 파라미터들도 동시에 계산한다.

단계 3: 첫 프레임의 특징점들을 선택하여 삽입 위치를 지정한다.

단계 4: 각 프레임에 대하여 가상 객체의 모든 기하 데이터를 이미 얻어진 카메라 투사 행렬을 사용하여 영상 좌표로 투사한다.

단계 5: 가상 객체의 투사 영역에 대하여 가상 객체의 해당 면의 텍스처를 합성한다.

단계 1과 단계 2는 비디오 장면의 복원에 관한 절

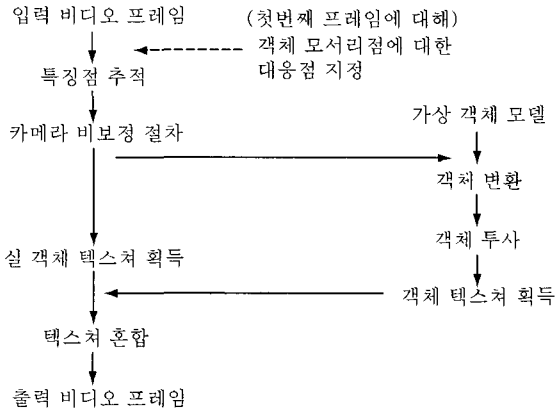


그림 1. 가상 객체의 비디오 합성 시스템 흐름도

차이고 단계 3에서 단계 5까지는 비디오 합성에 관한 절차이다. 각 절차의 흐름도가 그림 1에 있다.

본 논문의 구성은 다음과 같다. 2장에서는 비디오 장면 복원 방법을 설명하고 3장에서는 비디오 합성 방법을 설명한다. 4장에서는 실험 결과를 보이고 5장에서 결론을 맺는다.

## 2. 특징점 추적에 기반한 비보정 복원

복원은 다중 영상에서 특징점의 변이(disparity)에 기반하여 이루어진다. 우선 첫 프레임에서 다수의 특징점들을 검출하고 이를 이우하는 연속 프레임에서 대해 추적한다. 특징점들이 다수의 연속 프레임들에 대해 추적되면 이를 사용하여 비보정 절차를 수행한다. 세 프레임에서의 추적 데이터만으로도 비보정 절차의 실행이 가능하고 그 결과로 3차원 구조와 각 프레임에서의 상대적인 카메라 위치와 자세를 얻는다. 얻어진 카메라 파라미터들은 각 프레임에서의 카메라 투사 행렬을 구성하고 이 행렬들은 가상 객체를 프레임에 삽입하는데 사용된다.

### 2.1 특징점 추출 및 추적

프레임 열에서 특징점을 추적하는 방법으로 Lucas-Kanade 방법[9,10]을 사용하였다. 피라미드식 Lucas-Kanade 매칭 방법은 입력 프레임의 크기를 매우 작은 이미지로 될 때까지 반으로 줄여간다. 가장 작은 이미지에서 매칭을 시도하고 그 결과를 구체화하는 방식으로 진행하므로 매칭을 위한 계산 속도가 매우 빨라지고 탐색 영역을 넓힐 수 있어서 큰 카메라 이

동에도 대처할 수 있다. Lucas-Kanade 방법에서는 한 프레임에서 특징점의 이미지 패치(image patch)를 다음 프레임에서 찾는 방법으로 매칭을 실시한다. 따라서 물체가 고정되어 있고 카메라가 좌우나 상하로만 움직일 때 가장 좋은 결과를 얻을 수 있다. 즉 카메라의 줌 기능을 사용하거나 카메라가 물체에 대해 앞뒤로 움직인다면 매칭의 오차가 증가할 수 있다. 이의 해결을 위해서는 워핑 함수(warping function)의 차원을 높여 affine 변환까지 표현할 수 있도록 하면 오차를 줄일 수 있다. 그러나 이러한 해결 방법은 개선되는 오차에 비해 계산량을 매우 크게 증가시켜 실용적인 시스템에 적용이 어렵게 된다.

본 논문에서의 특징점 추적 방법은 화소의 움직임을 선형으로 가정하여 고속 추적이 가능하도록 하였다. 입력 비디오 프레임에서 나타나는 움직임을 선형 변환으로 추정하여도 연속 프레임에서의 특징점의 변화가 일정 거리 미만이므로 오차가 크지 않고 효율적이다.

### 2.2 카메라 비보정

카메라에 촬영된 공간의 3D 정보를 얻기 위해서는 카메라 파라미터들을 알아야 한다. 카메라 파라미터들을 계산하는 절차를 카메라 보정(calibration)이라고 한다. 카메라 보정은 일반적으로 특정 패턴을 사용하여 번거로운 수동적인 작업으로 이루어진다. 이러한 고전적인 보정 절차를 피할 수 있는 방법으로 비보정(auto-calibration) 접근 방법이 연구되고 있다[1,9]. 비보정 방법은 동일 카메라로 촬영한 연속하는 다중 프레임들로부터 카메라 파라미터를 계산하는 접근 방법이다. 수동적인 보정 작업을 배제할 수 있는 큰 장점이 있으나 많은 미지수를 계산해야 하고 입력데이터에 민감하다는 단점이 있다[1]. 또한 미지수가 많으므로 복원을 위해 필요로 하는 프레임의 수도 증가하게 된다. 미지수를 줄이기 위한 방법으로 카메라의 이동만을 가정하고 두 프레임 이상에서 카메라 이동을 계산하는 방법이 제안되었다[11].

제안하는 방법은 이우하는 세 프레임에서 카메라 파라미터가 불변이라는 가정을 도입하여 각 복원 절차를 선형적으로 수행하도록 하였다. 먼저 비보정 절차를 위한 일반적인 기호 및 수식을 설명한 후 제안하는 세 프레임에서의 카메라 파라미터 고정 방식을 설명한다. 한 3D 점의 좌표  $X_i$ 가  $i$ 번째 이미지 상의

좌표인  $x_j^i$  로 투사된 것은 식 (1)과 같이 표현한다:

$$x_j^i = P^i X_j \quad (1)$$

여기서  $x_j^i$  는  $j$ 번째 점  $X_j$ 가  $i$ 번째 카메라에 맺힌 영상점의 위치이고  $P^i$  는  $i$ 번째 카메라의 투사행렬이다. 다중 영상으로부터의 특징점 일치 관계를 사용하면 투영 복원(projective reconstruction)을 얻을 수 있다[12]. 투영 복원을  $\{P^i, \tilde{x}_j\}$ 라 하면 식 (2)와 같이 표현할 수 있다.

$$x_j^i = \tilde{P}^i \tilde{X}_j = (\tilde{P}^i H) (H^{-1} \tilde{X}_j) = P^i X_j \quad (2)$$

여기서  $P^i = K^i [R^i | t^i]$ 라고 하자. 식 (2)에서  $\{P^i, X_j\}$ 와  $\{\tilde{P}^i, H^{-1} \tilde{x}_j\}$ 는 H에 의한 관계이므로 H 행렬을 알게 되면  $P^i$ 와  $X_j$ 는 구할 수 있다.  $R^1 = I, t^1 = 0, P^1 = K^1 [I | 0], \tilde{P}^1 = [I | 0]$ 로 가정하면, H 행렬은

$$H = \begin{bmatrix} K^1 & 0_3 \\ v^T & 1 \end{bmatrix} \quad (3)$$

와 같은 형태로 표현된다. 여기서  $K^1$ 은 첫 번째 카메라의 보정 행렬(calibration matrix)이고  $v$ 는 무한 원평면(the plane of infinity)의 좌표 정보를 포함하고 있다. H 행렬은 투영 복원  $\{\tilde{P}^i, \tilde{x}_j\}$ 을 메트릭 복원(metric reconstruction)  $\{P^i, X_j\}$ 으로 변환시키는 매우 중요한 정보를 포함하고 있다. H 행렬에 포함되어 있는 미지수인  $K^1$ 과  $v$ 를 추정하는 것이 바로 비보정 문제가 된다.

비보정 문제의 해결을 위해서 최근 여러 연구가 진행되었으나, 일반적으로 미지수의 수가 데이터의 수보다 커서 세 프레임 미만으로부터는 해를 찾기가 어렵다. 해를 찾기 위해서 많은 프레임을 동시에 사용하는 시도가 연구되어왔으나 강건성(robustness)이 떨어지는 경우가 많다[12]. 다양한 반복적 최적화 방법[13]을 사용할 수 있으나 계산 시간이 선형적 방법에 비해 불규칙적이고 매우 길어서 대부분의 대화형 응용 시스템에서 바람직하지 않다.

### 2.3 이웃하는 세 프레임으로부터의 선형 복원

보정 행렬  $K^i$ 에 대한 가정을 통해서 이웃하는 세 프레임으로부터의 선형 복원이 실시될 수 있도록 한다.  $\omega^* = K^i K^{iT}$ 를 DIAC(the dual image of the ab-

solute conic)이라고 한다. DIAC  $\omega^*$ 는 카메라 파라미터들을 인코딩하는 행렬로, 구하고자 하는 카메라 내부 파라미터들을 모두 포함하고 있다. 보정 행렬을 고정시키면 각 영상  $i$ 에 대한 DIAC이 불변이므로  $\omega^* = \omega^*$ 로 표시할 수 있고 미지수를 크게 줄일 수 있다. 첫번째 영상에서는  $\tilde{P}^1 = [I | 0]$ 로 가정하여 세계 좌표계를 첫번째 카메라를 기준으로 설정한다. 미지수는  $\omega^*$ 에 포함된 보정 파라미터들인  $\alpha_x, \alpha_y, s, x_0, y_0$ 의 5개와  $v$ 의 3개의 총 8개이다. 얻을 수 있는 제약식은  $\omega^*$ 가 대칭(symmetric)이고 동차(homogeneous)이므로 5개의 제약식이 주어진다. 따라서 이론적으로 세 개 이상의 프레임으로부터 해를 구할 수 있다.

선형시스템의 유도를 위해서 ADQ(absolute dual quadric) 식을 도입한다[12]. ADQ  $Q^*$ 은 랭크(rank)가 3인  $4 \times 4$  동차 행렬로 식 (4)과 같다.

$$Q^* = H \begin{bmatrix} I_{3 \times 3} & 0_3 \\ 0_3^T & 0 \end{bmatrix} H^T = \begin{bmatrix} \omega^* & v \\ v^T & \|v\|^2 \end{bmatrix} \quad (4)$$

ADQ를 이미지로 투사하면 DIAC이 되므로 식 (5)와 같이 표현할 수 있다.

$$\omega^* = \tilde{P}^i Q^* \tilde{P}^{iT} \quad (5)$$

주점(principal point)이 알려진 경우에는 원점과 주점이 일치하도록 영상좌표들을 미리 이동한 후 이를 사용할 수 있다. 이 경우  $x_0 = y_0 = 0$ 로 둘 수 있다. 또한 비대칭(skew)이 0이라고 가정하자. DIAC은 식 (6)와 같이 간단히 표현될 수 있다.

$$\omega^* = \begin{bmatrix} \alpha_x^2 & 0 & 0 \\ 0 & \alpha_y^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

종횡비(aspect ratio)  $r = \alpha_x / \alpha_y$ 를 알고 있다고 하면  $\omega_{12}^* = 0, \omega_{13}^* = 0, \omega_{23}^* = 0, r^2 \omega_{11}^* = \omega_{22}^*$ 의 4개의 제약식을 얻을 수 있다. 종횡비를  $r=1$ 로 가정하고  $\alpha_x = \alpha_y = f$ 로 두면 더 단순화된 선형시스템을 얻을 수 있다.  $v = [v_1 \ v_2 \ v_3]^T$ 라 하고, 미지수들을 식 (7)과 같이 하나의 벡터  $x$ 로 구성한다.

$$x = [f^2 \ v_1 \ v_2 \ v_3 \ v_1^2 + v_2^2 + v_3^2 \ 1]^T \quad (7)$$

앞의 네 제약식을  $Ax=0$  형태의 선형시스템으

$$A = \begin{bmatrix} p_{11}p_{21}+p_{12}p_{22} & p_{14}p_{21}+p_{11}p_{24} & p_{14}p_{22}+p_{12}p_{24} & p_{14}p_{23}+p_{13}p_{24} & p_{14}p_{24} & p_{13}p_{23} \\ p_{11}p_{31}+p_{12}p_{32} & p_{14}p_{31}+p_{11}p_{34} & p_{14}p_{32}+p_{12}p_{34} & p_{14}p_{33}+p_{13}p_{34} & p_{14}p_{34} & p_{13}p_{33} \\ p_{21}p_{31}+p_{22}p_{32} & p_{24}p_{31}+p_{21}p_{34} & p_{24}p_{32}+p_{22}p_{34} & p_{24}p_{33}+p_{23}p_{34} & p_{24}p_{34} & p_{23}p_{33} \\ p_{11}^2+p_{12}^2-p_{21}^2-p_{22}^2 & 2p_{11}p_{14}-2p_{21}p_{24} & 2p_{21}p_{14}-2p_{22}p_{24} & 2p_{13}p_{14}-2p_{23}p_{24} & p_{14}^2-p_{24}^2 & p_{13}^2-p_{23}^2 \end{bmatrix} \quad (8)$$

로 표현할 수 있다. 여기서 A는 식 (8)과 같다. 여기서 각  $p_{ij}$ 는 주어진 프레임에 해당하는 카메라의 원근 투사행렬의 각 원소들이다.

한 프레임이 4개의 제약식을 제공하고 A는 4x6 행렬이 된다. 한 프레임을 추가하면 A가 8x6 행렬이 된다. 따라서 첫 프레임은 기준 프레임으로 사용되므로, 세 프레임 이상으로부터 해를 구할 수 있다. 이제  $Ax=0$  선형 시스템에 대하여 SVD로 x를 구하면 된다.

x가 구해지면 바로 K와 v를 구하고, 이로부터 H를 구성할 수 있다. 따라서 H를 사용하여 투영 복원을 메트릭 복원으로 바꾸게 된다. 선형 시스템의 해를 구하기 위해서는 세 장 이상의 프레임에서의 추적 데이터가 필요하다. 복원 과정에 대한 시스템의 흐름도가 그림 2에 있다. 이웃하는 세 장의 프레임에서 공통적으로 추적된 특징점들을 사용하여 선형 시스템의 해를 구하고 메트릭 복원을 실시한다. 이들 데이터는 각 프레임마다 저장하여 비디오 합성에서 이용되도록 한다.

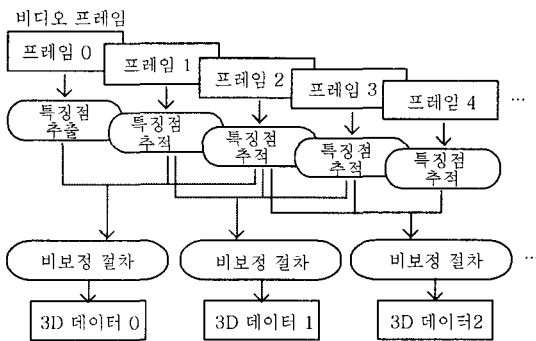


그림 2. 비디오 프레임에 포함된 특징점들의 복원 과정 절차도

### 3. 가상 객체의 비디오 삽입

실 영상을 포함하는 비디오에 가상 객체를 혼합하는 것은 증강현실에서 매우 중요한 과정이다. 가상 객체의 기하 데이터가 비디오의 실 영상들과 기하적으로 일관되게 혼합되도록 하려면 비디오를 촬영할

때의 카메라의 움직임을 정확히 알고 있어야 한다. 카메라의 움직임 정보를 얻는 방법으로 3D 공간에서의 위치와 방향을 측정하는 특별한 센서 (예: Polhemus sensor)를 사용하는 방법이 있다. 그러나 실 세계에서 센서를 사용하여 카메라의 위치와 움직임을 측정하는 데에는 많은 제약이 따른다. 다른 방법으로 비디오의 영상 프레임들을 분석하여 일치 관계를 얻고 이로부터 카메라 움직임을 추정하는 방법이 있다. 비디오에 포함되어 있는 한 물체의 3D 구조가 알려져 있을 때 이 물체를 등록(registration)하여 카메라 움직임을 추정한다. 이러한 접근 방법의 예로서 교각의 3D 기하 구조가 주어진 경우 비디오 프레임에서 교각의 위치를 찾고 교각의 특정 부분에 조명효과를 추가하여 새로운 비디오를 생성하는 연구가 제안되었다[14]. 이러한 방법은 알려진 3D 구조를 비디오에서 찾아 카메라 움직임을 추정하는 방법으로 적용에 있어서 제약이 거의 없으므로 응용의 범위가 넓다.

본 논문에서의 비디오 합성 절차는 그림 3과 같다. 먼저 첫 번째 프레임에서 가상 객체가 삽입될 위치에 대한 특징점을 지정하여 대응점으로 정한다. 세계 좌표계로 표시된 가상 객체의 기하 데이터는 복원된 각 프레임의 투사 행렬을 사용하여 투사된다. 투사 영역을 계산한 다음 영역 내의 각 화소의 칼라값들을 이음새가 없도록 블렌딩하여 합성 프레임들을 생성한다. 가상 객체의 비디오 합성은 전체적으로 다음과

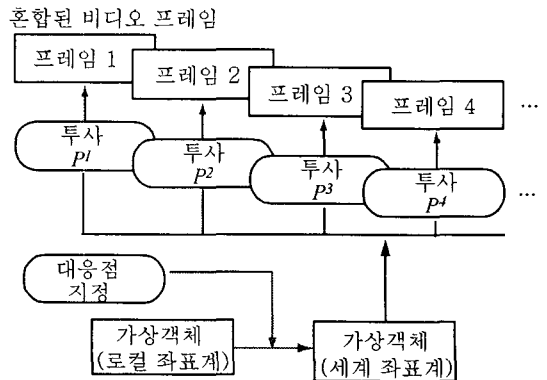


그림 3. 비디오 합성 과정 흐름도

같은 세부 단계로 나누어진다.

- (1) 복원 구조의 융합(integration) 및 카메라 투사 행렬 계산
- (2) 참조점의 설정
- (3) 가상 객체의 좌표 변환
- (4) 각 프레임에 대해 가상 객체를 투사
- (5) 가상 객체의 투사면에 대한 텍스처 혼합

세부 단계 중 (1)-(4)는 가상 객체를 프레임에 투사하기 위한 절차이고, (5)는 가상 객체의 텍스처와 프레임의 텍스처를 혼합하기 위한 절차이다.

### 3.1 가상 객체의 투사

첫 번째 단계인 복원 구조의 융합은 독립적으로 복원된 구조들의 좌표계를 일치하도록 한다. 3D 복원 과정에서 각 복원 결과는 세 장의 프레임으로부터 복원되었으므로 독립적인 좌표계를 가지게 된다. 또한 새로 나타난 특징점들에 대한 3D 구조를 추가해야 한다. 이는 공통된 점들에 대한 3D 좌표 일치 관계로부터 3D 변환 행렬을 추정하여 가장 처음의 복원 구조에 대한 좌표계로 바꾸어 주면 된다. 이때 복원 구조뿐만 아니라 카메라의 이동 및 회전행렬도 동일한 변환 행렬로 바꾸어 주어야 한다. 이 단계가 수행되면 첫 번째 카메라의 좌표계로 표시된 하나의 전체적인 3D 구조를 얻고 각 프레임에 대한 투사행렬도 얻게 된다.

비디오 프레임에서의 가상 객체의 삽입 위치를 지정해야 한다. 삽입 위치는 가상 객체의 한 면에 대해

서 놓여질 위치를 영상 프레임에서 지정하여 결정된다. 참조가 되는 면에 해당하는 사각형의 네 모서리점의 좌표값으로부터 가상객체의 로컬 좌표를 세계 좌표로 변환하는 변환행렬을 계산한다. 구해진 변환행렬로 가상 객체의 모든 좌표들을 세계좌표로 변환하고 이를 프레임의 영상좌표로 투사하여 영상에서의 위치를 계산한다. 따라서 가상 객체의 삽입 위치 지정은 한 사각형의 네 모서리점에 대응되는 참조점만 지정하면 된다. 네 참조점은 첫 번째 프레임에서만 지정해주고 그 이후의 프레임에서는 첫 번째 프레임에서의 참조점을 추적한 위치들이 자동적으로 사용된다. 이 작업은 주어진 특징점들에 대해 단 네 번의 마우스 클릭만으로 이루어지므로 매우 빠르고 간단하게 삽입 위치 지정이 가능하다.

참조점에 해당하는 가상 객체의 면은 일반적으로 바닥면과 같은 가상 객체의 기준이 될 수 있는 면을 선정하면 된다. 설정된 참조점에 의해서 객체의 지역 좌표계로 표시된 가상 객체의 기하 정보는 세계 좌표계로 변환한다.

### 3.2 텍스처 혼합

세계 좌표계로 변환된 가상 객체는 각 프레임에 대하여 해당 카메라 투사 행렬을 사용하여 프레임으로 투사하여 프레임에서의 좌표들을 얻는다. 투사된 프레임 영역에서의 텍스처를 블렌딩하여 합성된 프레임 영상들을 얻는다. 이 과정이 그림 4에 표시되어 있다.

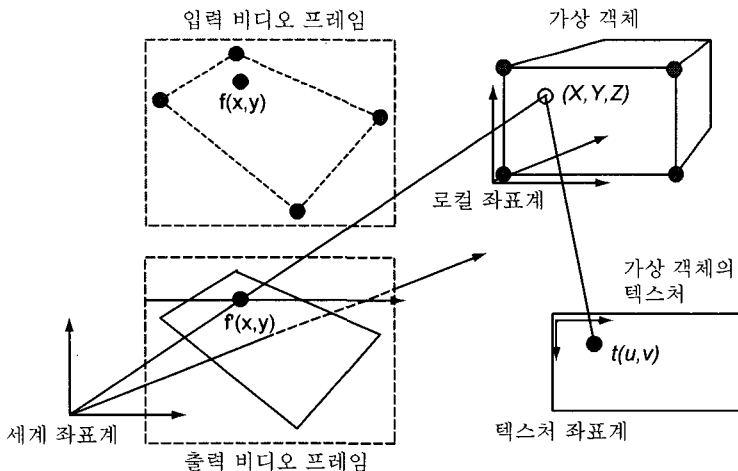


그림 4. 텍스처 혼합을 위한 가상 객체의 텍스처와 입력 비디오 프레임의 텍스처의 좌표 관계

가상 객체의 텍스처를 삽입할 경우 텍스처의 경계 면에서 부자연스러운 혼합이 흔히 발생한다. 이는 이질적인 색 분포의 급격한 변화에 기인한다. 자연스러운 텍스처 블렌딩을 위해서 가중치를 부여한 색 혼합 기법으로 경계 영역에서의 부자연스러운 이음새를 제거하도록 한다. 먼저 블렌딩 상수  $d$ 를 정의한다. 상수  $d$ 는 텍스처 매핑 영역에서의 에지의 영향을 받는 거리이다. 투사된 가상 객체의 에지로부터 거리  $d$  내의 화소들에 대해서는 이음새 제거 처리를 수행하도록 한다. 여기서  $d$ 의 값은 프레임의 크기에 따라 달라질 수 있는 경험적인 값이다. 값이 클수록 넓은 간격에 대해 에지 평활화를 실시하게 된다. 한 위치에서의 색 결정은 외곽선으로부터의 거리가  $d$  이하 일 경우에 해당 화소의 블렌딩 색은 식 (9)와 같이 결정한다.

$$C_b = C_1 * (1 - t/d) + C_2 * (t/d) \tag{9}$$

여기서  $C_1$ 과  $C_2$ 는 프레임의 텍스처 칼라값과 가상 객체의 텍스처 칼라값이고  $C_b$ 는 결정된 칼라값이다.  $t$ 는 해당 화소에서 경계선까지의 거리이다. 위의 식은 경계선으로부터의 거리에 따른 선형적인 가중치 조절을 의미한다.

#### 4. 실험 결과

실내 환경에서 촬영한 다양한 비디오 프레임들에 대해서 실험을 수행하였다. 비모정 기법을 통하여 3D 구조와 카메라 파라미터를 복원하는 부분과 이를 사용하여 가상 객체가 합성된 비디오 프레임을 생성하는 부분을 나누어서 실험을 진행하였다.

##### 4.1 카메라 파라미터와 메트릭 구조 복원

그림 5의 상자 비디오 프레임들을 사용해서 3D 메트릭 구조 및 카메라 파라미터를 복원한 결과가 그림 6에 있다. 상자의 두 면을 복원한 것으로서 각 면이 거의 직각이 되는 결과를 얻을 수 있었다.

그림 7은 미니어처 마을의 프레임들의 세 프레임을 보이고 있다. 프레임 상에 추적된 38개의 특징점들이 표시되어 있다. 추적된 특징점들로부터 복원된 3차원 구조가 그림 8에 있다. 한 평면상의 점들은 비교적 일직선을 이루는 것을 볼 수 있었다.

그림 9는 특징점들이 다양한 여러 면에 분포하는 비디오 프레임들의 예이다. 29개의 특징점들이 각 프레임 상에 표시되어 있다. 복원된 구조의 모습이 그

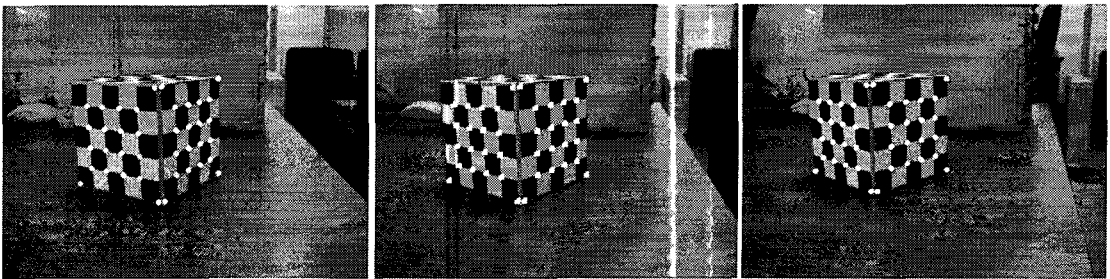


그림 5. 상자 영상열의 세 프레임과 추적된 특징점

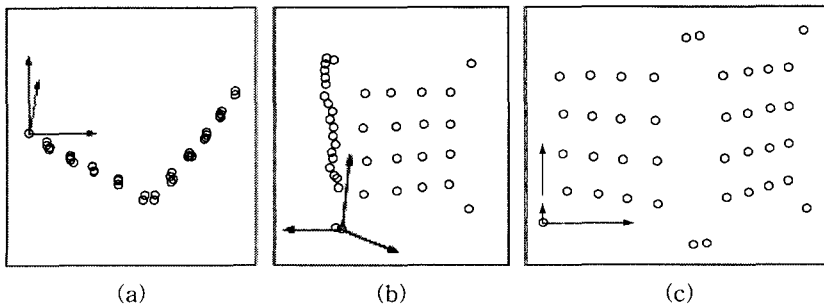


그림 6. 상자 영상열의 세 프레임으로부터 복원된 3차원 데이터: (a) 윗면 모습, (b) 측면 모습(왼쪽 아래), (c) 정면 모습(오른쪽 아래)

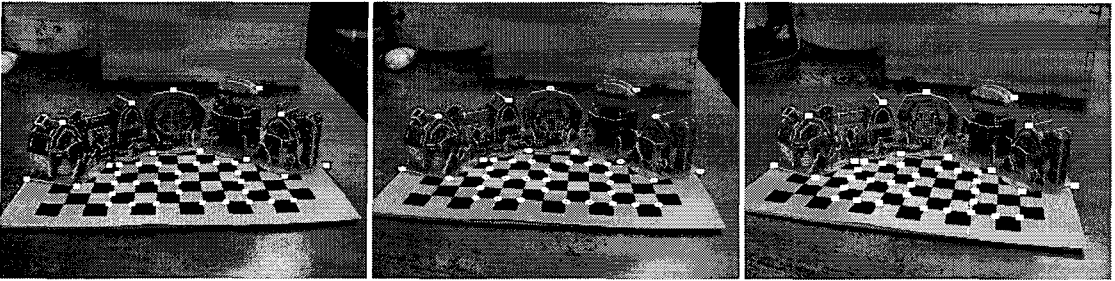


그림 7. 미니어처 마을 영상 열에서의 세 프레임과 추적된 특징점

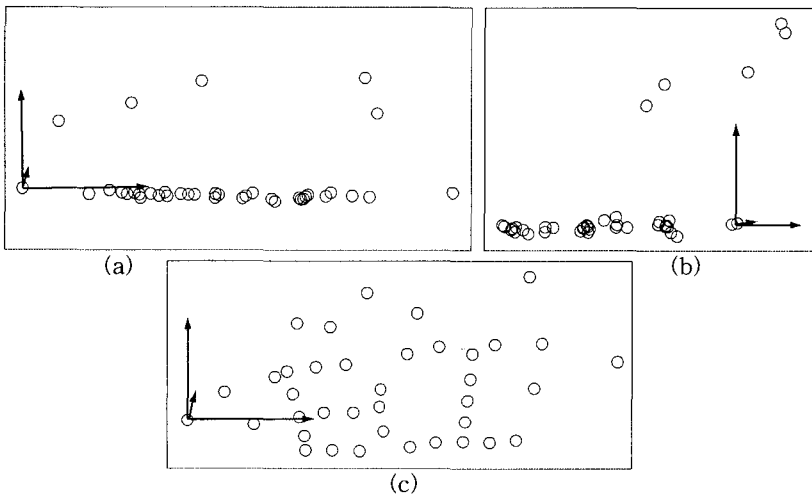


그림 8. 미니어처 마을 영상열의 세 프레임으로부터 복원된 3차원 구조: (a) 정면 모습(위 왼쪽), (b) 측면 모습(위 오른쪽), (c) 윗면 모습(아래)

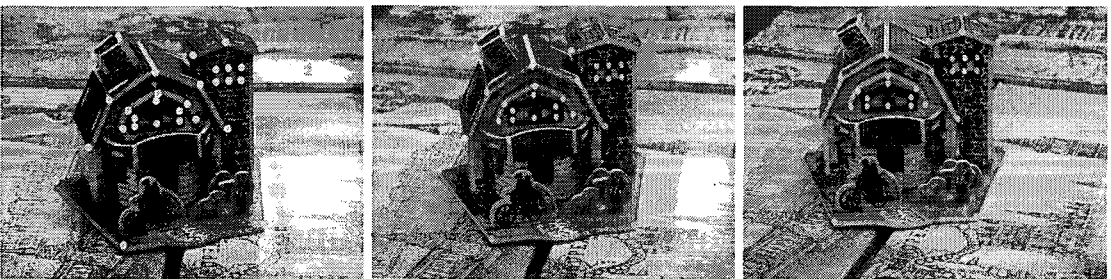


그림 9. 하우스 영상열에서의 세 프레임과 추적된 특징점들

림 10에 있다. 그림 10의 위의 그림들에서는 복원된 3D 좌표점들의 위치를 표시하고 아래의 그림들에서는 구조의 파악이 용이하도록 선으로 연결하여 보여주도록 표시한다.

#### 4.2 가상 객체의 합성

가상 객체는 단순한 형태의 기하 구조를 가정하였

다. 직사각형 모양 또는 직육면체 모양의 가상 객체를 임의로 생성하고 적절한 텍스처 이미지를 가상 객체의 각 면에 지정해 두었다.

각 프레임에서 투사 행렬을 사용하여 가상 객체를 투사한 위치가 원하는 지점과 정확히 일치해야 고품질의 비디오 합성이 가능하다. 추적된 점들을 사용하여 구해진 카메라 투사 행렬로 프레임에 투사한 점과



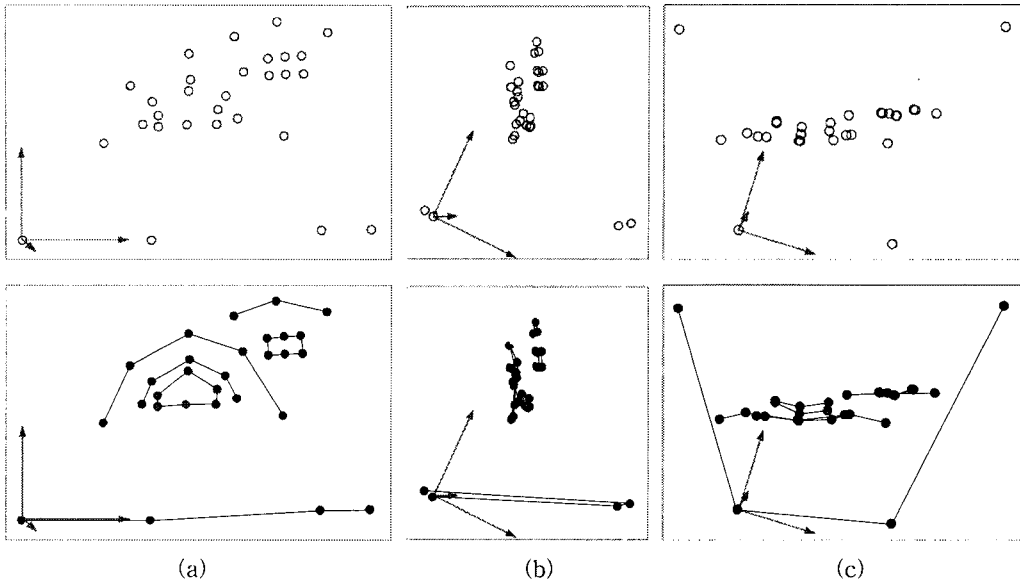


그림 10. 하우스 영상열의 세 프레임으로부터의 복원된 3차원 구조: (a) 정면 모습 (왼쪽), (b) 측면 모습 (중간), (c) 뒷면 모습 (오른쪽)

사용자가 지정한 모서리 점의 거리를 측정해 보았다. 영상이 디지털화된 화소 단위이므로 투사점과 특징 점이 정확히 일치되기는 힘들다. 그림 11은 측정 점과 투사 점의 차이를 나타내고 있다. 원으로 표시된 마크가 특징점의 위치이고 사각형으로 표시된 마크가 투사점의 위치이다. 그림 12는 각 프레임에서의 각 모서리점에 대한 오차를 보여주고 있다. 연속하는 6개의 프레임에서의 4개의 특징점의 추적 오차를 보이고 있다. 각 프레임에서 특징점과 투사점과의 오차의 평균이 화소 단위로 각각 1.7176, 0.9977, 1.0590, 0.8090, 1.6120 이었다. 대부분의 경우 오차가 3 화소 이하를 유지하였다. 프레임 열이 보다 긴 경우에 대해서 특징점과 투사점과의 평균 오차가 그림 13에 있다. 전반적으로 에러가 2 화소 이내이나 특정 부분에서 오차가 다소 증가하는 경우가 발생하였다. 오차

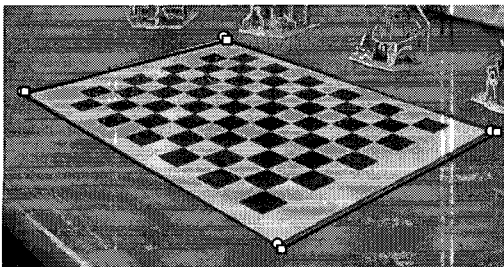


그림 11. 측정된 모서리 점과 투사된 모서리 점의 오차

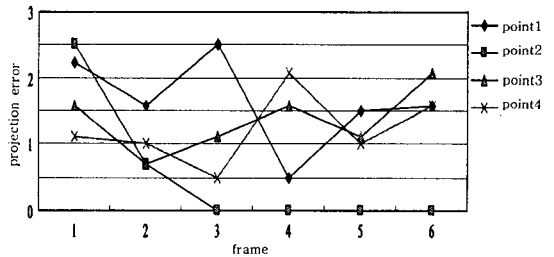


그림 12. 각 프레임에서의 모서리 점들의 오차

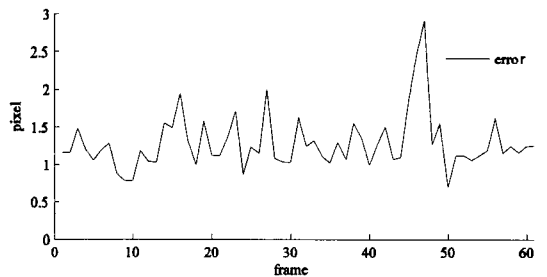
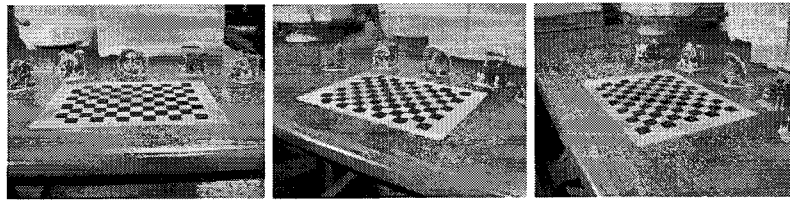
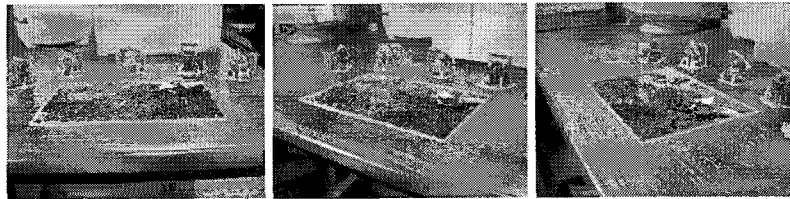


그림 13. 비디오 프레임들에서의 전체 특징점과 해당 투사점과의 평균 오차

증가의 원인은 특징점 추적의 불안정성에 있다. 그림 14, 그림 15, 그림 16은 서로 다른 프레임들에 대한 실험 결과를 보여주고 있다. 그림 14는 경우 바닥의 체크무늬 종이에 한 텍스처를 합성한 모습이고, 그림 15와 그림 16은 캘린더의 배경 그림을 다른 텍



(a)



(b)

그림 14. (a) 입력 프레임과 (b) 합성 프레임



(a)

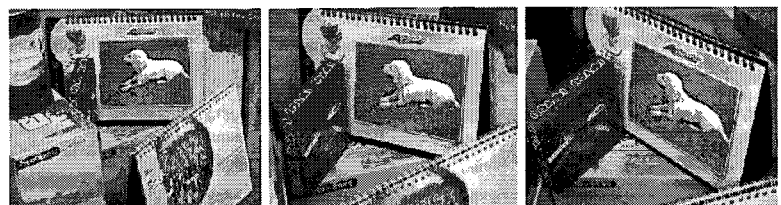


(b)

그림 15. (a) 입력 프레임과 (b) 합성 프레임



(a)



(b)

그림 16. (a) 입력 프레임과 (b) 합성 프레임

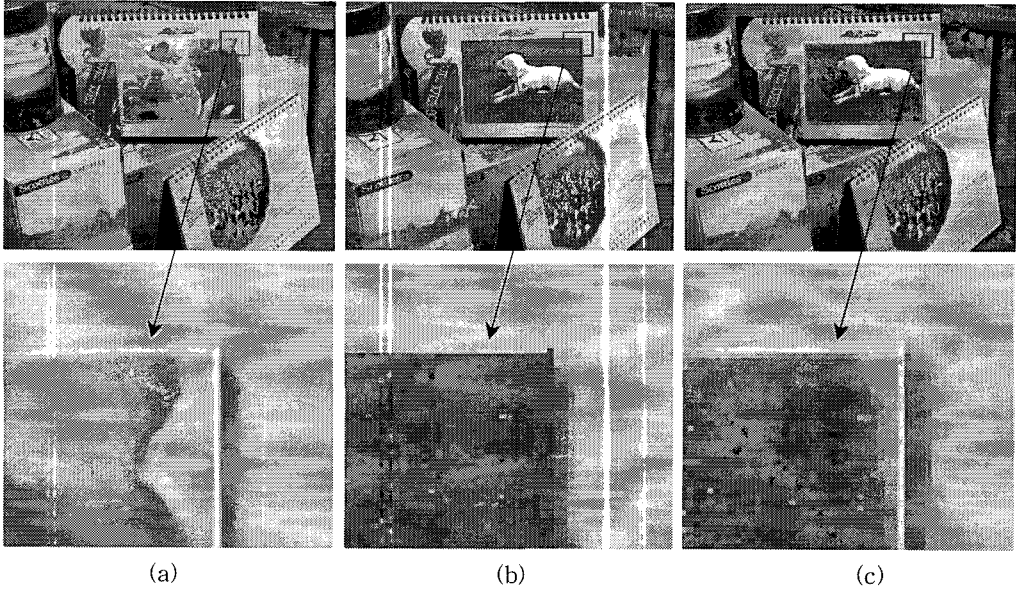


그림 17. 에지에 대한 이음새 제거: (a) 원래의 프레임, (b) 이음새를 제거하기 전의 합성 프레임, (c) 이음새를 제거한 후의 합성 프레임

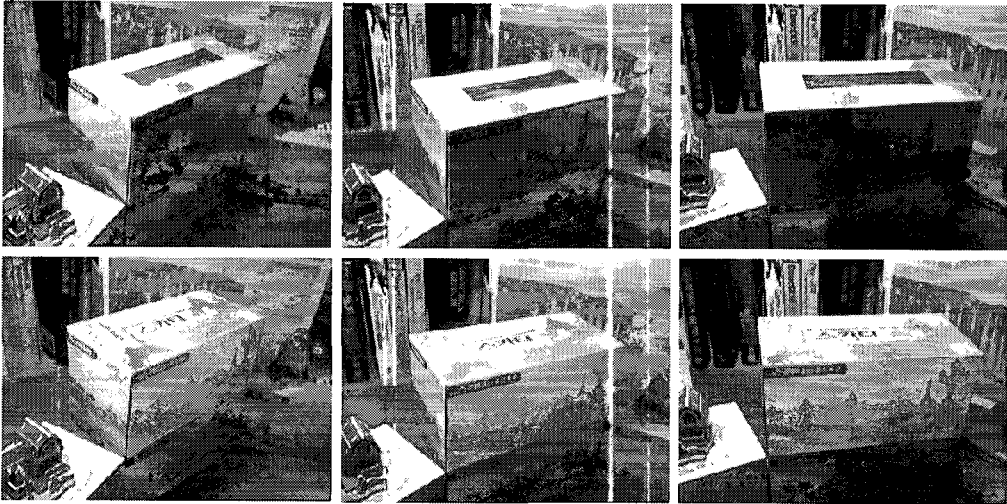


그림 18. 조명 효과를 위해 가상 객체 텍스처의 밝기를 조절한 합성

스처로 합성한 모습이다. 수백 프레임 이상의 비디오 클립에 대해 합성 프레임들을 생성한 결과 매우 자연스러운 합성된 비디오 클립이 생성되었다. 다만 재생 중에 특징점 추적의 오차로 인해서 약간의 떨림 현상이 나타났다. 이러한 현상의 제거를 위해서는 전역적인 특징점 추적 오차를 화소 단위 미만으로 줄일 수 있는 기법의 도입이 필요하다.

그림 17은 텍스처의 블렌딩에 있어서 에지 부분의 이음새 제거 처리를 보여주고 있다. 왼쪽은 입력 프레임이고, 중간은 이음새를 제거하지 않은 결과이며, 오른쪽은 이음새를 부드럽게 제거한 결과이다. 실험에 의해서 블렌딩 상수  $d$ 를 프레임의 너비의  $1/40$ 에 해당하는 값으로 지정할 경우 그 간격이 적절함을 알 수 있었다. 아래의 그림은 위의 그림의 모서리 영

역의 확대 그림이다.

합성 시에 가상 객체의 텍스처에 대해서 전체적인 밝기를 적절히 변화시켜 준 후에 텍스처 블렌딩을 할 수 있다. 그림 18에서와 같이 티슈 상자의 텍스처 뿐만 아니라 밝기의 조절도 가능하다. 티슈 상자의 윗면과 앞면의 텍스처를 다른 모양의 밝은 텍스처로 대체한 합성 프레임이다. 이 경우는 밝기를 임의로 조절한 예로써 실제 상황과 일치하지 않을 수 있다. 조명의 위치를 인식하여 밝기를 자동으로 조절하는 것은 추후의 연구 과제가 될 수 있다.

## 5. 결 론

본 논문에서는 카메라의 내부 파라미터에 대한 보정이 없는 환경에서 영상 프레임만 사용해서 3차원 좌표를 복원하고 가상 객체를 비디오에 합성하는 방법을 제안하였다. 복원 방법으로 비보정 접근 방법을 사용하였다. 기존의 비선형 반복적 비보정 방법과는 달리 제안한 방법은 연속하는 세 프레임에 대해서 선형적으로 비보정 접근이 가능하도록 하였다. 비보정 절차 수행에서 미지의 카메라 파라미터를 비디오 합성에 필요한 파라미터만으로 축소하여 세 프레임에 대해서 선형적인 구조 계산이 이루어지도록 하여 빠르고 안정적인 합성이 가능하였다. 다양한 비디오 데이터를 가지고 복원 및 합성 실험을 수행한 결과 작은 수의 특징점 만으로도 비교적 정확도가 높은 결과를 얻을 수 있었다. 동일 장면의 세 프레임에서의 추적 데이터를 사용하여 선형으로 복원이 가능한 방법을 제시하였다. 가상 객체의 삽입 위치의 지정을 위해서 첫 프레임에서의 특징점들 중 4개를 선택하도록 하였다. 가상 객체의 기하 데이터는 추정된 각 프레임에 대한 카메라 투사 행렬을 통하여 각 프레임으로 투사된다. 프레임에서의 해당 영역이 결정되면 영역 내의 각 화소의 칼라값을 텍스처 블렌딩으로 채운다.

복원의 정확성이 특징점의 추적 결과에 상당히 의존적인 측면은 여전히 시스템의 강건성에 장애 요소이다. 이 문제의 해결을 위해서는 추적의 정확도를 높이는 방법의 개선이 바람직하다. 또한 추적 결과를 사용자가 확인하고 드물게 발생하는 오류 데이터를 제거하도록 하는 방법도 고려되어야 한다.

향후 연구 과제로 특징점 추적의 정확도를 높임으

로써 더 정확한 복원이 가능한 알고리즘의 설계, 실시간 복원을 위한 빠른 선형 알고리즘 개발, 조명의 위치와 방향을 자동으로 검출 및 조명 효과를 고려한 가상 객체 합성 등이 있다. 또한 텍스처로 인한 에지와 occlusion으로 인한 에지를 구분하고 각각에 대한 블렌딩 기법의 연구가 필요하다.

## 참 고 문 헌

- [1] A. Fusiello, "Uncalibrated Euclidean Reconstruction: A Review," *Image and Vision Computing*, Vol. 18, pp. 555-563, 2000.
- [2] J. M. Rehg, S. B. Kang, T.-J. Cham, "Video Editing Using Figure Tracking and Image-Based Rendering," *Proc. ICIP*, Vol. 2, pp. 17-20, 2000.
- [3] J. Blanc, S. Livatino and R. Mohr, "Fast and Realistic Image Synthesis for Telemenuipulation Purposes," *European Workshop on Hazardous Robotics*, Spain, pp. 77-83, 1996.
- [4] 김효성, 남기근, 주재홍, 이철현, 설성욱, "스테레오 보정 카메라를 이용한 영상 기반 3차원 모델링에 관한 연구," *신호처리 시스템학회 논문집*, Vol. 4, No. 3, pp. 27-33, 2003.
- [5] 김성용, 한준희, "제약조건을 적용한 셀프 캘리브레이션 방법," *정보과학회논문지: 소프트웨어 및 응용*, Vol. 28, No. 4, pp. 358-368, 2001.
- [6] M. Han and T. Kanade, "Creating 3D Models with Uncalibrated Cameras," *Proc. IEEE Computer Society Workshop on the Application of Computer Vision*, Vol. 9(2), pp. 137-154, 2000.
- [7] Paul Debevec, "Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography," *Proceedings of SIGGRAPH 98*, Vol. 32(4) pp. 189 - 198, 1998.
- [8] C. Loscos, M.-C. Frasson, and G. Drettakis, "Interactive Virtual Relighting and Remodeling of Real Scenes," *Rendering techniques '99 (Proceedings of the 10th Eurographics*

*Workshop on Rendering*), Vol. 10, pp. 235-246, Jun 1999.

[9] B D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, Vancouver, pp. 674-679, 1981.

[10] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *International Journal of Computer Vision*, Vol. 56, No. 3, pp. 221-255, 2004.

[11] G. Sharma, A. Kumar, S. Kamal, S. Chaudhury, and J.B. Srivastava, "Novel View Synthesis Using a Translating Camera," *Pattern Recognition Letters*, Vol. 26, pp. 483-492, 2005.

[12] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, 2<sup>nd</sup> ed, Cambridge Press, 2003.

[13] K.H. Wong and M.Y. Chang, "3D Model Reconstruction by Constrained Bundle Adjustment," *Proc. ICPR*, Vol. 3, pp. 902-205, 2004.

[14] M. Berger, C. Chevrier, G. Simon, "Compositing Computer and Video Image Sequences: Robust algorithms for the reconstruction of the camera parameters," *Computer Graphics Forum*, Vol. 15, No. 3, pp. 23-32, 1996.



**박 종 승**

1992년 경북대학교 전자계산학과 (이학사)  
 1994년 포항공과대학교 컴퓨터공학과 (공학석사)  
 1999년 포항공과대학교 컴퓨터공학과 (공학박사)  
 1999년~2001년 한국전자통신연구원 가상현실연구부 선임연구원  
 2001년~2003년 디지털아리아 시각정보처리연구소 책임연구원  
 2004년~현재 인천대학교 컴퓨터공학과 조교수  
 관심분야: 영상처리, 3D 비전, 증강현실, 게임공학



**성 미 영**

1982년 서울대학교 학사  
 1987년 프랑스 INSA de Lyon 컴퓨터공학 석사  
 1990년 프랑스 INSA de Lyon 컴퓨터공학 박사  
 1990년~1993년 한국전자통신연구원 인공지능연구실 선임연구원  
 1993년~현재 인천대학교 컴퓨터공학과 교수  
 관심분야: 멀티미디어 협동 컴퓨팅, 멀티미디어 저작, 인터넷 멀티미디어, 가상 협동 환경