

합성 콘텐츠 저작을 위한 반자동 정합 기술에 관한 연구

정세윤, 김규헌

한국전자통신연구원 무선방송연구소 방송미디어연구부

e-mail : jsy@etri.re.kr

A Study on Semi-Automatic Registration for Synthesizing Natural Video and Virtual Objects

Seyoon Jeong, Kyuheon Kim

Dept. of Broadcasting Media Technology

Radio & Broadcasting Laboratory

Electronics and Telecommunications Research Institute

요 약

실사 영상에 가상 객체를 합성하기 위해서는 실사 영상 촬영 당시의 카메라 정보가 필요하다. 본 논문에서는 이러한 카메라 정보를 구하기 위하여 가상현실 분야에서 사용하고 있는 캘리브레이션 프리 정합 (Calibration-Free Registration) 기술을 기반으로 한 반자동 정합 기술을 제안하였다. 가상 현실은 실시간 응용인데 반하여 본 논문에서 제안하는 반자동 정합 기술은 합성 콘텐츠 저작을 위한 오프라인 응용에 적합한 방법으로 캘리브레이션 프리 정합기술의 합성 결과는 사용자의 입력 정보와 밀접한 관계가 있다. 캘리브레이션 프리 정합기술은 두가지 사용자 입력을 필요로 한다. 첫 번째 입력은 아파인공간 (Affine space)의 기저 (Basis vector)를 위한 특징점 정보이고, 두번째 입력 정보는 가상객체의 영상 투영점 입력이다. 본 논문에서는 이 두가지 사용자 입력중 기저를 위한 특징점 정보입력을 사용자가 쉽게, 정확한 정보를 입력할 수 있게하기 위하여, 사용자가 특징점을 개략적으로 입력하게 하고, 주변 영역에서 코너점 검출을 수행하여 사용자 입력을 수정하여 받아들이는 방법을 제안하였다. 실험결과 제안한 방법을 사용하여 구한 카메라 정보로 만족할 만한 합성 영상을 얻을 수 있었다.

1. 서론

실사(Natural) 동영상과 CG(Computer Graphics) 가상(Virtual) 객체를 합성한 콘텐츠 제작 및 이에 대한 수요가 날로 증가 하고 있다. 이러한 실사 동영상과 CG 객체 합성 콘텐츠 제작에 있어서 중요한 일 중 하나는 CG 객체를 실사 배경내에 원하는 위치에 자연스럽게 합성시키기는 것인데, 이를 위해서는 실사 배경 영상의 카메라 정보, 즉 실제공간과 영상 평면간의 투영(Projection) 관계를 구해야 한다. 이 투영 정보를 이용하여 가상 객체 공간과 실제 공간의 좌표계를 일치시키고, 원하는 위치에 가상 객체를 등록하고 영상 평면에 투영하여 자연스러운 합성 결과를 얻게 된다. 이

러한 배경 영상의 카메라 투영 정보를 구하여, 가상 공간 좌표계와 카메라 좌표계를 일치 시킨 후, 가상 객체를 카메라 좌표계에 등록하고, 영상 평면의 원하는 위치에 합성되게 하는것을 정합(Registration)이며, 이 정합 기술은 증강현실 분야에서 연구되는 기술이다 [1]. 본 논문에서는 증강현실의 정합 기술 중 캘리브레이션 프리 정합 (Calibration-Free Registration) 기술을 기반으로 한 합성 콘텐츠 저작을 위한 반자동 정합 기술을 제안한다 [2].

본 논문의 구성은 다음과 같다. 2 절에서는 증강현실의 캘리브레이션 프리 정합기술에 대한 설명과 해당 기술을 합성 콘텐츠 저작에 바로 적용할 경우의

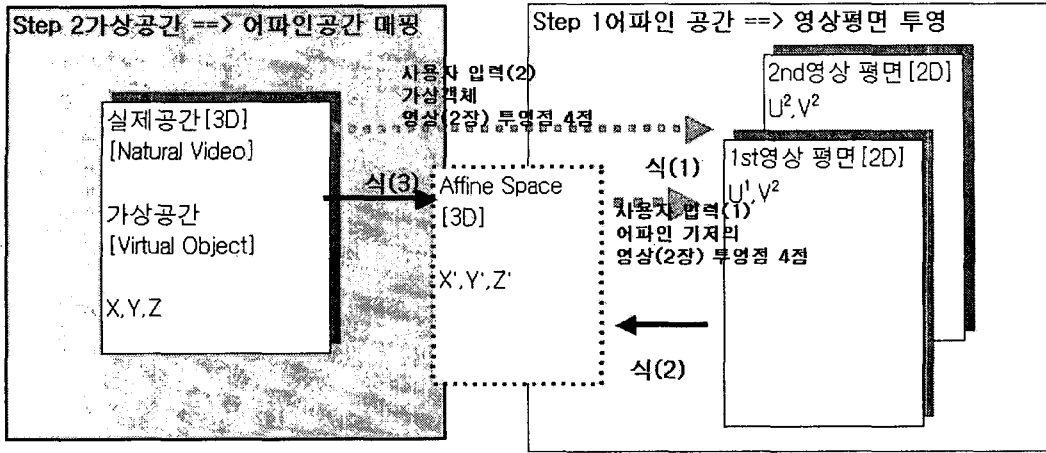


그림 1. 캘리브레이션 프리 정합 기술 개요도

문제점에 대해서 논하고, 3 절과 4 절에서는 컨텐츠 합성을 위해 제안된 반자동 정합 기술에 대한 설명 및 실제 구현과 합성 결과에 대해서 논하고 끝으로 5 절에서 결론 및 향후 연구 과제에 대하여 설명한다.

2. 캘리브레이션 프리 정합 기술

정합(Registration)은 가상 객체가 실제로 존재하는 것처럼 가상 객체를 실제 공간에 등록시키는 과정이다. 이를 위해서는 실사 배경 영상의 카메라 투영 정보를 구해, 실제 공간의 좌표계 정보를 구하고, 이를 가상 객체의 CG 좌표계와 일치시켜야 한다. 이는 실사 카메라와 CG 카메라를 일치시키는 것이다. 합성 결과는 실사 배경 영상위에 CG 영상을 오버레이(Overlay) 처리하여 얻게 된다. 이때, CG 카메라의 투영 정보가 실사 카메라의 투영 정보와 일치하면 이 오버레이 합성 장면은 실제 공간에 실제로 가상 객체가 존재한 장면을 찍은 영상과 일치하게 된다.

캘리브레이션 프리 정합기술은 촬영전에 카메라 캘리브레이션 작업없이, 실사 영상에 가상 객체를 합성하는 증강 현실 시스템을 구현하기 위해 Kiriakos[2]가 제안한 방법으로, 이에 대한 개요도가 그림 1에 나타나 있다. 캘리브레이션 프리 정합기술은 실제 공간과 영상평면과의 투영 관계를 정의하는데 있어서 그림 1에 나타난 바와 같이 중간 매개체로 어파인 공간을 도입하여 투영관계를 2 단계로 정의하고 있다.

이 투영관계를 구하는 순서는 다음과 같다. 먼저 어파인 공간과 영상 평면과의 투영관계를 구하기 위해서 (그림 1의 Step1에 해당), 해당 공간의 4 점 이상의 정보가 필요하다. 이는 3 차원 공간은 동일 평면상이 아닌 임의의 4 점의 정보를 통해 기저 (Basis vector)를 구성할 수 있기때문에 4 점 이상의 정보가 필요하다[2]. 어파인공간과 영상평면간의 투영식은 식(1)과 같다[2].

$$\begin{bmatrix} u^m_p \\ v^m_p \\ 1 \end{bmatrix} = \begin{bmatrix} u^m_{b1} - u^m_{p0} & u^m_{b2} - u^m_{p0} & u^m_{b3} - u^m_{p0} & u^m_{p0} \\ v^m_{b1} - v^m_{p0} & v^m_{b2} - v^m_{p0} & v^m_{b3} - v^m_{p0} & v^m_{p0} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} \quad (1)$$

x', y', z 는 어파인 공간상의 임의의 점 P의 좌표이고 u^m_p, v^m_p 는 점 P의 m 번째 영상평면 상의 투영점이다. $p0, b1, b2, b3$ 는 미리 알고 있는 4 점, 즉 어파인 공간의 기저를 구성하는데 요구되는 투영점을 나타낸다. 상기에 나타난 식(1)을 통하여 영상평면 상에 어파인 공간의 기저를 구성하는 4 개의 투영점 좌표를 구하면 어파인 공간의 임의의 점에 대한 영상평면상의 투영점을 구할 수 있다. 이는 역으로 영상평면상의 점으로부터 어파인 공간의 기저를 구할 수 있으며, 이를 어파인 재구성(Reconstruction)이라고 하며, 일례로 영상 2 장에 대한 어파인 재구성을 아래 식 (2)와 같이 표현할 수 있다[2].

$$\begin{bmatrix} u^1_p \\ v^1_p \\ u^2_p \\ v^2_p \end{bmatrix} = \begin{bmatrix} u^1_{b1} - u^1_{p0} & u^1_{b2} - u^1_{p0} & u^1_{b3} - u^1_{p0} & u^1_{p0} \\ v^1_{b1} - v^1_{p0} & v^1_{b2} - v^1_{p0} & v^1_{b3} - v^1_{p0} & v^1_{p0} \\ u^2_{b1} - u^2_{p0} & u^2_{b2} - u^2_{p0} & u^2_{b3} - u^2_{p0} & u^2_{p0} \\ v^2_{b1} - v^2_{p0} & v^2_{b2} - v^2_{p0} & v^2_{b3} - v^2_{p0} & v^2_{p0} \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} \quad (2)$$

따라서, 실제 공간상의 임의의 점에 대한 두 영상 평면에서의 투영점 좌표를 알면 식(2)를 사용하여 어파인 공간상의 좌표를 계산할 수 있으며, 또한 식(1)과 (2)에 의해 어파인 공간과 두 영상평면상의 관계가 완전하게 정의된다[2].

다음으로 그림 1의 Step2에 나타난 바와 같이 가상공간과 어파인 공간의 투영관계를 구하기 위해서는 가상 객체의 동일 평면상이 아닌 4 점의 영상 평면상의 투영점을 필요로 한다. 역시 영상 2 장에 대한 투영점 정보가 필요하다. 이 점들의 정보를 식(2)를 이용하여 4 점의 어파인 공간상의 대응 좌표를 구하게 된다. 가상공간의 4 점의 좌표와 어파인 공간의 대응

좌표 4 점의 정보를 사용하여 식(3) 과 같이 가상공간 과 어파인 공간의 관계를 정의할 수 있다[2].

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = A_{3 \times 3} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (3)$$

식(3)을 풀어 행렬 A 를 구하면, 가상 공간과 영상 평면간의 모든 투영 관계가 정의되는 것이다.

즉, 식 (3)을 사용하여 가상 공간의 임의의 점의 어파인 공간상의 좌표를 구한 후, 식(1)을 사용하여 이점의 영상 투영점을 구할 수 있는 것이다.

이상과 같이 캘리브레이션 프리 정합기술은 가상 객체를 합성하고자 하는 배경 영상 평면과 가상 공간의 투영 관계를 어파인 공간이라는 매개 공간을 도입하여 정의하고, 2 가지의 사용자 입력, 기저를 구하기 위한 투영점과 가상 객체의 투영점을 입력 받아서 투영관계를 정의하고 있다.

3. 콘텐츠 합성을 위한 반자동 정합 기술

캘리브레이션 프리 정합 기술은 증강현실에 사용하기 위해 개발된 방법으로 실시간 처리가 보장 되어야 한다. 실시간 처리를 위해 사용자 입력 대신에 캘리브레이션 객체가 존재하는 장면을 촬영하고, 캘리브레이션 객체의 점들을 사용자 입력으로 대체하는 방법을 사용하고 있다. 이 방법은 합성 콘텐츠 제작 방법에서는 사용할 수 없는 방법이다.

합성 콘텐츠 제작에서는 미리 녹화된 동영상을 사용하므로, 영상내에 캘리브레이션 객체가 존재하지 않으므로, 사용자가 직접 특징점 정보를 입력해주는 방법을 사용해야 한다. 영상내의 특징점은 트래킹을 쉽게 하기위해 주로 코너점을 사용하게 된다. 사용자가 정확하게 코너점을 지정해 주어야 정확한 투영관계를 구할 수 있다. 사용자가 마우스를 이용하여 정확하게 코너점을 지정하기 위해서는 매우 많은 노력을 필요로 한다. 이에 본 논문에서는 사용자가 특징점 지정을 쉽게 할 수 있고, 정확한 정보를 입력받을 수 있는 방법을 제안하고자 한다.

캘리브레이션 프리 방법에서는 두 가지 사용자 입력을 요구한다. 첫번째 입력은 어파인 공간과 영상 평면에 대한 관계를 구하기 위한 기저 특징점을 위한 동일 평면상이 아닌 4 점의 영상에서의 위치이다. 완전한 관계를 정의하기 위해 다른 카메라 위치에서 찍은 영상 2 장에 대한 입력을 해주어야 한다. 이때, 합성 콘텐츠 제작 과정을 위해 본 논문에서 제안된 반자동 정합기술에서는 이 특징점들을 영상내의 코너점 중에서 지정하게 하는 방법을 사용하였다. 코너점을 특징점으로 사용하면 사용자가 특징점 지정을 쉽게 지정하게하는 것이 가능하다. 사용자가 대략적으로 특징점 근처를 마우스로 지정해주면, 이 지정한 점 주변

영역에 대해 코너점 검출을 수행하여 가장 가까운 코너점을 사용자 입력으로 받아들이게 하였다. 이때 코너점 검출 방법으로는 SUSAN[5] 코너 검출 방법을 사용하였다.

두번째 사용자 입력은 가상객체 공간과 어파인 공간의 매핑 관계를 구하기 위한 가상 객체상의 4 점의 투영점을, 역시 동일 평면상이 아닌, 동일 영상 2 장에 대해서 지정해 주어야 한다. 가상 객체의 영상 상에 투영 위치를 지정하는 것은 실제로 합성을 하기전에는 상당히 어려운 일이다.

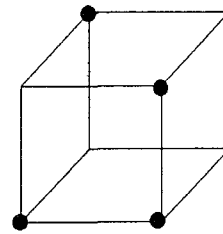


그림 2. 가상정육면체에서의 특징점의 위치

본 논문에서는 합성하기전에 입력할 수 있게하기 위해 다음과 같은 방법을 사용하였다. 정육면체를 가상 물체로 가정하고 그림 2 에 표시된 점들의 투영점을 사용자가 입력하게 하는 방법을 사용하였다. 이 두번째 입력에 대한 보다 쉬운 입력방법은 향후 연구 과제로 한다.

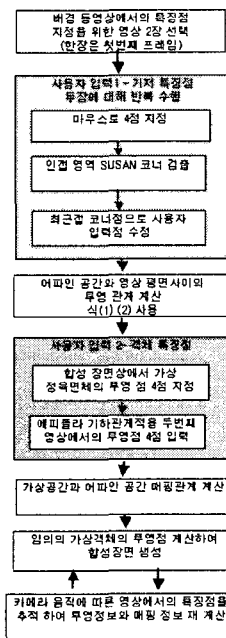


그림 3 합성콘텐츠 제작을 위한 반자동 정합 정보 생성 과정

이 가상객체의 투영점 정보 입력 과정은 먼저 영상 한 장에 대해서 사용자가 4 점을 입력하게 한 후, 다른 영상에 대해서는 어파인 공간과 두 영상 평면의 투영 관계로 부터 설정된 에피폴라(Epipolar) 기하관계를 적용하여 에피폴라 선(line)상에서 입력 하게하는 Kiriakos[2]의 방법을 따랐다.

상기에서 설명한 바와 같이 본 논문에서는 콘텐츠 합성 저작 작업에 적합한 사용자 입력 방법을 제안하

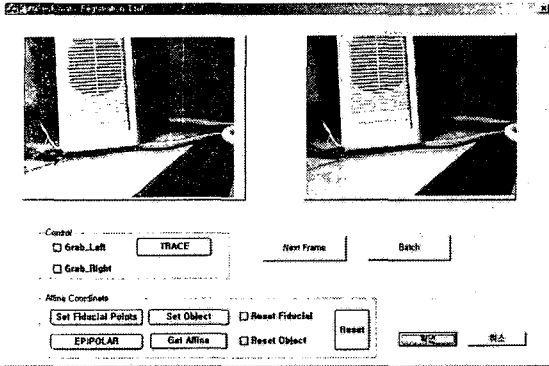


그림 4. 반자동 정합 정보 추출 도구 GUI

였다. 제안된 사용자 입력 방법을 적용한 콘텐츠 합성을 위한 반자동 정합 정보 추출과정은 그림 3에 나타난 바와 같다. 반자동 정합도구는 배경 동영상의 매 프레임의 카메라 투영정보 및 정합을 위한 정보를 파일로 출력하며 이 정보는 합성 콘텐츠 저작 도구에서 가상객체를 합성할 때 사용하게 된다.

4. 구현 및 합성결과

본 논문에서 제안한 반자동 정합 정보 추출 도구는 그림 4에 나타난 바와 같으며, MS Windows 2000 환경에서 VC++6.0을 개발도구를 이용하여 구현되었으며, 해당 도구를 이용한 합성 실시 동영상으로는 720x480 DV 포맷을 사용하였다. 해당 반자동 정합 정보 추출 도구에서 동영상을 오픈하면 첫번째 프레임을 좌측 뷰에 로드하고 오른쪽 뷰에 다른 위치에서 찍은 영상으로 기본적으로 10 번째 프레임을 로드하게 하였고, 필요할 경우 다른 프레임을 선택할 수 있게 하였다. 3절에서 설명한 바처럼, 첫번째 사용자 입력인 기저 특징점 입력을 위해서 Set Fiducial Points 버튼을 클릭한 후 좌측 뷰와 오른쪽 뷰 상에서 입력 받으며 코너 검출로 입력값을 보정하여 입력받게 하였다. 이때 코너점 검출은 마우스 입력점 주변의 블러영역내에서만 수행하게 하였다. 실험결과 블록의 크기는 32X32를 할 때 가장 효과적이었다.

본 논문에서 제안한 방법으로 구한 카메라 정보를 사용하여 합성한 동영상에서 5 번째, 15 번째, 25 번째, 35 번째 프레임을 그림 5에서 보여주고 있다. 배경영상의 그림 액자 밑에 텍스트를 적용한 타원 객체가 합성된 장면이다. 이 합성객체는 동영상의 각 프레임의

반자동 정합 도구가 미리 구한 카메라 투영 정보에 의해 매 프레임마다 계속 액자 밑에 위치에 합성되고 있다.

5. 결론

본 논문에서는 배경 동영상에 가상 객체를 합성한 콘텐츠 저작에 적합한 반자동 정합 정보 추출 기술에 대하여 제안하였고, 제안한 방법으로 정합에 필요한 카메라 정보를 추출하여, 이를 사용하여 합성 콘텐츠를 시험 저작하여, 제안한 정합 정보 추출 방법이 합성 저작도구에서 사용 가능함을 보였다.

향후 과제로는 현재 사용자 입력이 편리하지 않은 가상 객체 특징점 입력 부분의 개선과 배경 동영상의 매 프레임에 대해서 사용자 입력 특징점들의 위치를 자동으로 추적(tracking)하는 방법 개발이 필요하다.

참고문헌

- [1] Youngmin Lee, "Registration of 3D Models and Video Images in the Affine Space for Augmented Reality", KAIST VR-Lab TM 2000-4, Mar. 2000.
- [2] Kiriakos N. Kutulakos, James R. Vallino, "Calibration-Free Augmented Reality", IEEE Trans. Visualization and Computer Graphics, Vol. 4, No. 1 Jan.-Mar. 1998.
- [3] Yongduek Seo, Ki Sang Hong, "Calibration-Free Augmented Reality in Perspective", IEEE Trans. Visualization and Computer Graphics, Vol. 6, No. 4, Oct.-Dec. 2000.
- [4] M.Uenohara and T.Kanade, "Vision-Based Object Registration for Real-Time Image Overlay", *Proceedings of CVRMed'95*, pp. 75-85, July 1993
- [5] S.M. Smith, J.M. Brady, "SUSAN -- A New Approach to Low Level Image Processing" <http://www.fmrib.ox.ac.uk/~steve/susan/susan/susan.html>

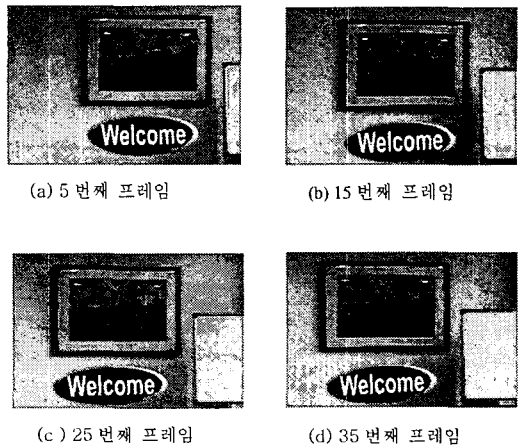


그림 5. 합성결과 예