

다시점 혼합현실 시스템을 위한 강인한 3차원 정합영역 추적기법

김장현⁰ 손광훈

연세대학교 전기전자공학과 디지털영상미디어연구소
(jangheon⁰, khsohn)⁰@yonsei.ac.kr

Robust 3D region tracking for multiview mixed-reality system

Jang-heon Kim⁰ Kwang-hoon Sohn

Digital Image Media Lab, Dept. of Electrical and Electronic Eng., Yonsei University

요 약

다시점 영상 부호화와 함께, 가상물체를 지능형으로 합성하기 위해서는 정합할 위치를 강인하게 추적(tracking)하여야 한다. 기존의 특징추적기법은 시점이 확장된 시스템에서, 광원의 영향과 가려진 영역의 발생으로 인하여, 추적할 수 없게 되거나 시각적 혼동이 발생하는 등의 많은 문제점을 유발한다. 뿐만 아니라 다시점에 의한 정보의 급격한 증가는 시점간 움직임 추적을 추적하여 정합하는 연산량의 추가적 발생을 초래한다. 이에 본 논문은 광원이나 가려진 영역(occlusion)의 영향이 적은 형상과 정보량이 많은 텍스처를 동시에 고려하여 영역의 변형에 따라 발생하는 최적 움직임을 예측하는 파라미터 모델을 결합함으로써 시각영역을 추적한다. 이 때 미리 연산된 카메라 파라미터와 투영관계를 사용해 시점간에 가상 객체를 지속적으로 3차원 정합한다. 결과적으로, 제안된 알고리즘은 기존의 특징추적 알고리즘에 비해 광원의 영향과 가려진 영역에 강인하며 정합의 오류가 적고 추적할 때 시점간의 대응을 위한 연산을 줄이기 때문에 실감 혼합현실 시스템의 구현에 매우 효율적임을 실험을 통해 증명하였다.

1. 서 론

오늘날의 멀티미디어 및 방송시스템은 MPEG-4, MPEG-7 등의 표준화와 함께 각종 전달매체가 통합되고, 사용자가 고품질의 다양한 콘텐츠를 요구하면서, 실 영상과 3차원 가상 객체를 다시점으로 자동적으로 혼합하여 현장감 있게 전달할 수 있는 지능형 실감 합성 기술을 필요로 한다.

기존의 증강현실(augmented reality) 기법은 가상 객체를 현실의 추가정보로 사용하게 하는 대표적인 예이나, 특징기반의 추적 기법을 주로 사용하고 있어 다시점 영상기술을 적용하였을 경우 다음과 같은 심각한 문제점이 발생한다. 시점이 늘어날수록 영상정보량의 증가하기 때문에, 추적을 위해 사용되는 특징추출 및 대응의 연산량은 급격히 증가한다. 또한, 넓어진 FOV(field-of-view)는 대응을 어렵게 하여 특징의 흔들림을 초래하고 광원과 가려진 영역의 영향을 증대시켜 시점간의 추적을 어렵게 한다. 따라서 본 논문은 카메라간의 관계와 투영을 이용하여 효율적으로 연산량을 감소시키고, 많은 노이즈에도 인위적 조작 없이 시점간의 흔들림 없는 강인한 정합을 유지하는 영역기반 다시점 영상 합성 추적 기법을 제안하고자 한다.

2. 기존 방식

2.1 기존 영상 합성 추적 기법의 문제점

기존의 많은 영상기반 합성 기법의 연구에서 3차원 물체의 시점과 움직임에 따른 자세 추정과 시점분석에 대한 광학할 만한 수학적 해법의 발전이 있었다.

또한 증강현실에서는 실현실과 가상현실의 시점간 움직임간의 상관성 있는 정합을 위해 다양한 추적기법이 연구되었고, 더욱 용이한 검출과 추적을 위해 기점 마커(fiducial marker)나 표지물(landmark)을 영상에 삽입하였다.[1]

이들은 LED, 컬러, 반사형(예. 거울)과 같이 광학적, 시각적 특징을 고려하거나 사각, 점, 원마커, 바코드패턴과 같이 형상의 대응을 이용한다. 그리고, 이러한 것들은 Shi and Tomasi가 연구를 통해 검증한 것처럼 검출이 용이하게 디자인 되었기 때문에 시점이나 광원에 민감하지 않은 강인한 특성을 띄며 좋은 추적 성능을 가진다. [2]

하지만, 다양한 콘텐츠에 적용하기 위해서는 인위적인 조작 없는 자연영상에 기반하여 추적하여야 하며, 물체표면의 상호반사, 조도(illumination), 그림자, 영상의 밝기, 가려진 영역 등에 강인하여야 한다.

하지만, 위에 제시한 기존 합성추적방법에 주로 많이 사용되는 임계화와 같은 특징추적기법은 FOV가 넓고 광원의 영향이 큰 다시점에 적용하였을 경우, 노이즈나 각종 환경적 영향으로 식별이 어려워 질 때 추적할 수 없거나, 특징의 흔들림을 초래할 수 있다. 이는 정합의 부자연스러움과 시각적인 혼동, 그리고 복시 발생에 의한 깊이감의 상실이 발생할 수 있기에 적절하지 못하다.

2.2 영역 추적 기반 합성 기법

2.1절에서 언급했던 특징의 흔들림 및 불인식의 문제점을 개선하기 위해 영역 추적(region tracking) 기법이 연구되었다.[3] 이는 변환되는 영역과 참조영역간의 에너지를 최소화하기 위하여, L_2 (sum-of-squares) norm을 사용한다. 이는 에러의 측정수단으로 사용되는 방법이므로 수치적 최적화 기법으로써 적합할 뿐만 아니라, 카메라 투영과 시점의 성질 등 기하학적 특성에 기반한 최적화에도 잘 적용할 수 있다.

본 논문에서는 SSD(sum of squared difference)를 사용하며 이를 이용한 추적기법은 기존의 많은 연구에서 유사변환에 의한 빠른 변화나 복잡한 투영에서도 좋은 성능을 가질뿐만 아니라, 불연속적인 영역의 복잡한 변형에도 효율적이고 강인함을 보였다.[4,5]

3. 강인한 시각영역 3차원 추적기법

3.1 시각영역 추적

다시점 혼합현실에서 자연영상이나 마커에 가상 영상을 정합하기 위해서는 지표(beacon)와 같은 자연영상의 부분이나 인위적인 마커와 같은 정합할 영역이, 시점이나 광원의 영향에도 정확하게 같도록 추적되어야 한다.

이는 검출된 특징간에 유사변환(affine transform)에 의한 대응에 대한 연관성 있는 이동을 갖는 것을 의미한다. 본 논문은 자연영상에서 이러한 연관성 높은 강인한 추적을 위하여 형상과 텍스처 정보를 동시에 고려한다.

경계와 같은 형상 정보는 광원의 영향을 덜 받지만, 움직임에 의해 자주 발생하거나 사라질 수 있는 특징점으로 인해 잘못된 추적을 할 수 있다. 그리고, 텍스처 정보는 보이지 않는 영역의 발생 및 광원과 노이즈에 민감하고 분석해야 할 데이터의 양이 많다.

이에 비하여, 두 가지를 모두 고려하여 추적하는 영역추적기법은 형상과 텍스처 중 한 가지만을 사용한 방법에 대해, 다음과 같이 서로의 단점을 보완해 줄 뿐만 아니라 효율적이다.

무엇보다도, 움직임에 의해 쉽게 발생하거나 없어질 수 있는 형상의 특징점에 의한 잘못된 추적을 텍스처 정보를 사용함으로써 방지할 수 있으며, 대응되는 영역을 영상 영역 내부로 제한함으로써 각종 환경적 변화에 의해 발생하는 에러의 범위를 한정할 수 있다. 뿐만 아니라, 영상 정보가 가진 곡률, 길이, 법선 성분은 자세 추정, 텍스처 정보는 합성될 물체의 분류에 효율적으로 사용될 수 있다.

본 논문이 제안하는 영역기반 방식은 이전 프레임의 영역 R_{i-1} 에 대하여 유사 변환한 현재 프레임의 영역을 예측하여 현재 프레임의 영역 R_i 를 결정하며, 추적은 이전 프레임의 유사변환 A_{i-1} 에 의한 식 (1)의 예측치를 반영함으로써 이루어진다.

$$\hat{A}_i = R_{i-1} A_{i-1} \quad (1)$$

먼저, 전 영상에 대해 경계의 방향과 사각형의 모서리가 검출되면 본 논문이 제안하는 사각 영역기반 트래커가 동작한다. 모서리 검출에 의해 사각 영역에 제한되어 초기화되며 각 영역은 다음과 같이 추적 된다.

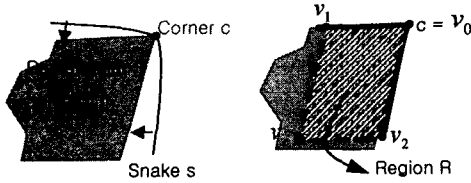


그림 1. 사각 영역 추적

그림 1과 같이 사각형 영역을 결정하기 위하여, 검출된 모서리 c 를 통과하고 두 경계의 방향이 교차하는 사각 영역의 두 변을 감싸는 스네이크(Snake) S 를 취한다. [6] 스네이크 S 는 모서리 c 에 의해 결정되는 점 v_0 와 각 두 변상의 점 v_1, v_2 를 포함한다.

초기 검출된 모서리 c 에 대해 스네이크의 중심이 되는 점 v_0 가 결정되고 식(2)의 기울기의 중합을 최대화 하기 위해 제어점의 이동을 추정함으로써 세 점은 움직이게 된다.

$$E_{Snake}(v_0 = c, v_1, v_2) = \sum_{P \in v_0 v_1} |\nabla I(p)| + \sum_{P \in v_0 v_2} |\nabla I(p)| \quad (2)$$

또한, 스네이크에 의한 컨트롤 방향으로의 v_1, v_2 의 이동과 함께 스네이크가 감싸는 텍스처 정보의 프레임간의 최적 정보를 하여 선 위에서의 v_1, v_2 와 v_0 의 최적 위치를 추정함으로써, 사각영역이 결정된다. 기하학적 이동은 이전 프레임의 유사변환 A_{i-1} 에 의해 변형(warping)되는 영역 R 를 결정하여야 하므로, v_1, v_2 는 스네이크 라인을 따라서 움직이도록 구속조건을 둔 상태에서, 파라미터 모델 γ_q 에 의한 v_1, v_2 그리고 v_0 의 최적 위치를 구한다. [3]

$$E_{texture}(q) = \sum_{x,y \in R, R_{i-1}} (R_{i-1} - A(\gamma_q)R_{i-1})^2 \quad (3)$$

이에 의해 제한된 범위 안에서의 최적의 가능한 정합 영역을 결정하도록 함으로써, 높은 신뢰성과 대응 실패에 대한 강인함을 갖는다. 뿐만 아니라, 부분적인 가려짐에 의해 발생하는 에러가 입체적 보다 를 경우 추적 되는 스네이크에 대해 점 v_0 를 기준으로 이전 프레임을 참조하여 v 를 선형적으로 재설정함으로써 기준 모서리인 한 점 v_0 가 사라지거나, 신뢰성이 없을 정도의 한쪽 모서리가 사라질 때 까지 추적을 유지할 수 있다.

3.2 사각 영역의 3차원 물체 정합

스네이크는 초기화될 때 v_0 에 고정되어 있으므로, v 에 의한 사각 영역이 결정 될 때 영역의 유사변환에 의해 변형 되는 텍스처 영역은 그림2와 같이 실제 3차원 공간의 투영 변환이며 식 (4)로 표시될 수 있다.

$$\begin{aligned} x'(h_{31}x + h_{32}y + h_{33}) &= h_{11}x + h_{12}y + h_{13} \\ y'(h_{31}x + h_{32}y + h_{33}) &= h_{21}x + h_{22}y + h_{23} \end{aligned} \quad (4)$$

주어진 사각영역의 대응에 의해, 식 (5)와 같이 투영 변환 행렬의 요소를 풀 수 있는 8개의 선형 방정식을 얻을 수 있다.

여기서 h 는 투영변환의 3×3 행렬의 h_{ij} 요소를 포함하는 9개 요소이고, $Ah=0$ 형태를 취하기 때문에, h 는 A 의 무효공간이고, SVD(singular value decomposition)의 알려진 해법을 사용하여 계산될 수 있다. [6]

$$\begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1'x_1 & -x_1'y_1 & -x_1' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y_1'x_1 & -y_1'y_1 & -y_1' \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2'x_2 & -x_2'y_2 & -x_2' \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -y_2'x_2 & -y_2'y_2 & -y_2' \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3'x_3 & -x_3'y_3 & -x_3' \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -y_3'x_3 & -y_3'y_3 & -y_3' \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4'x_4 & -x_4'y_4 & -x_4' \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -y_4'x_4 & -y_4'y_4 & -y_4' \end{pmatrix} h = 0 \quad (5)$$

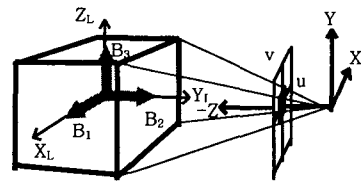


그림 2. 투영변환(perspective transform)

3.3 효율적인 다시점 확장 기법

다시점으로 추적을 수행하기 위해서는 Fundamental matrix나 Trifocal tensor와 같은 대응관계를 미리 연산하고 이의 에피폴라(epipolar) 구속조건하에 추적 관계를 연산하는 것이 일반적이다[7,8], 본 논문이 제안하는 최적 영역 추적과 함께 이루어지기에는 연산량이 너무 많다. 그러므로, 본 논문에서는 카메라간의 파라미터를 추적된 영역의 투영에 적용하여 정합되는 시점 V 을 예측함으로써, 3차원 가상 객체를 효율적으로 다시점 정합하는 기법을 제안한다.

먼저 양쪽 카메라(C, C')의 기지패턴(known-pattern) P 의 상대적인 관계에 의해 식 (6)과 같은 카메라 외부파라미터를 통한 카메라 C 간의 변환 T 를 얻을 수 있다.

$$T_{CC} = T_{CP} * T_{CP}^{-1} = (R_{CP}, t_{CP}) * (R_{PC}, t_{PC}) \quad (6)$$

그리고 동일평면상의 사각 영역 R 이 회전과 이동에 의해 가상영상을 정합할 시점 V 으로 투영되는 변환은 식 (7)과 같이 이루어진다.

$$T_{VR} = T_{VP} * T_{PR} \quad (7)$$

이 때, 가상물체를 정합시키기 위한 사각 영역 R 은 캘리브레이션 과정과 투영방정식에 의한 변환으로부터 카메라와 시점의 구속에 의해, 영역 R 의 패턴 P 에 대한 상대적 변환으로 구할 수 있다.

$$T_{PR} = T_{CP}^{-1} T_{CR} \quad (8)$$

식 (8)로부터, 기준카메라의 사각영역에 대한 다른 카메라의 영역은 캘리브레이션으로부터 얻은 카메라의 상대적 변환을 추적에 의해 얻어지는 영역 R 에 적용함으로써 투영된다. 결과적으로, 이러한 추적과 투영 관계에 의해 3차원 가상 객체는 다시점으로 효율적으로 정합 될 수 있다.

4. 구현

본 논문에서는 제안한 3차원 합성영역 추적 알고리즘을 사용하여 다시점 영상에 컴퓨터 그래픽 물체를 정합하는 실시간합성실 시스템을 구현함으로써 성능과 실용성을 검증하였다. 가상 물체를 영상기반으로 실시간 합성하기 위하여, Pentium 4 1.4GHz 시스템과, 2대의 IEEE1394 CCD 카메라를 사용하였으며, 가상 물체를 빠르게 렌더링하기 위하여, 하드웨어 가속된 OpenGL을 사용하였다. 사용된 IEEE1394 인터페이스의 전송 속도는 최대 63대까지 초당 100MB이상으로 접속이 가능하고, 데이터의 동시전송과 비동기 전송을 모두 지원하므로, 고화질의 다시점 영상을 실시간으로 처리할 수 있었다.

3차원 그래픽스 객체로 그림 3의 텍스처 매핑된 boat 모델과 텍스처가 없는 teapot 모델을 사용하였다.



그림 3. 실험에 사용된 3차원 모델 (순서대로 boat, teapot)

또한, 그림 4와 같은 100x100 해상도의 작은 정사각형 2차원 영상을 텍스처로 동시에 합성 영역에 추적 정합하였다.



그림 4. 실험에 사용된 2차원 텍스처 (순서대로 portrait, logo, mr)

5. 실험 결과

4절의 구현에서 서술한 시스템으로 카메라에 들어오는 물체에 대해 그림 3과 그림 4의 3차원 모델과 2차원 영상을 추적하며 실시간 정합하였다. 그림 5의 실험 영상에서 보는 것과 같이, 인위적 기점 마커나 표지물이 없는 자연영상의 경우 무수히 많은 모서리와 노이즈가 존재한다.

특히, 기존의 특징기반 추적기법의 경우 시점이 확장된 시스템에서 그림 6과 같은 카메라의 불균형, 광원의 영향, 가려진 영역 등에 의한 움직임과 시점에 대한 특징의 왜곡은 추적하지 못하거나, 잘못된 정합을 하는 결과를 가져올 수 있다.



그림 5. 제안한 알고리즘을 사용한 두 시점 동시 정합추적



(a) teapot 모델 정합영역 (b) boat 모델 정합영역
그림 6. 동일한 정합 영역에 미치는 광원의 영향

이에 비하여, 본 논문이 제안한 정합영역 추적 알고리즘을 사용한 실감혼합현실 시스템에서는, 두 카메라에 대해 그림 6과 같은 텍스처 정보의 변형이 큰 정합할 시각 영역에 대해 변형에 따라 발생하는 꺾적 움직임 파라미터를 추정하여, 그림 7과 같은 효율로 광원의 영향과 가려진 영역에 그림 5에서의 같이 합성상태를 강인하고 자연스럽게 유지함을 확인할 수 있었다.

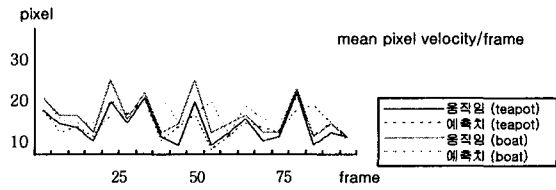


그림 7. 정합 영역 추적 성능

6. 결론

기존의 특징에 기반한 정합방식보다 본 논문이 제안한 영역기반 정합방식은 결과에서 제시한 바와 같이 정합 영역에 영향을 미치는 특징의 흔들림이 없고 광원의 영향이나 가려진 영역에 강인한 영역의 강인한 추적 성능을 보여주었다. 또한 시간의 전처리로써 수행되는 캘리브레이션을 다시점간의 가상영상의 정합에 이용함은 그림 6과 같이 잘못된 정합을 유발할 수 있는 영역에 대해 시각적 에러를 최소화 할 수 있다. 뿐만 아니라, 미리 연산된 카메라 파라미터의 사용은 트래킹시의 연산을 줄임으로써 시스템 비용을 감소시킨다.

하지만, 이렇게 카메라 파라미터를 사용하여 정합을 잘 수행하려면, 정합영역의 신뢰성 있는 움직임 추적이 이루어져야 한다. 본 논문은 영역기반의 추적기법을 사용하여 기존의 특징기반 추적기법에서 문제가 되었던 유사 기하와 투영으로의 정합 추정의 신뢰도를 높임으로써, 카메라간의 상대적인 관계를 투영관계에 적용하여 가상 객체를 정합할 수 있는 기반을 제공하였다. 이로써 시간간의 에러의 발생시에도 실험 결과와 같이 상대적인 정합으로 강인한 합성상태를 유지하는 다시점 혼합현실 시스템을 구현할 수 있었다. 이러한 자연영상 기반의 강인한 합성추적 기법은 다시점의 지능형 실감방송시스템의 구현을 가능하게 한다. 이러한 지능형 실감방송시스템을 위한 다시점 혼합현실 추적기법은 기존의 중강현실의 적용범위 뿐만 아니라, 차세대 임체 방송시스템의 콘텐츠의 제작이나, 가상광고 및 안내 시스템 등의 넓은 응용분야에 효율적으로 적용될 수 있다.

7. 참고문헌

- [1] R. Azuma, "A survey of augmented reality, Presence: Teleoperations and Virtual Environments", p355-385, 1997
- [2] W. Hoff, "Fusion of data from head-mounted and fixed sensors", 1st Int. Workshop on Augmented Reality, 1998
- [3] Gregory D. Hager "Efficient Region Tracking With Parametric Models of Geometry and Illumination", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998
- [4] J. Shi and C. Tomasi, "Good features to track", In IEEE Conf. on Comp. Vis. and Pat. Rec., pp.593-600, 1994
- [5] Charles V. Stewart, "Robust parameter estimation in computer vision", Society for Industrial and Applied Mathematics, SIAM Review Vol.41, No.3, pp.513 - 537 1999
- [6] A. K. Jain, Y. Zhong, and S. Lakshmanan, "Object matching using deformable templates. IEEE Trans. on Pattern Analysis and Machine Intelligence, pages 267-277, 1996.
- [7] J.J. Koenderink and A.J. vanDoorn. "Affine structure from motion". JOSA-A, 8(2), 1991.
- [8] A. Sashua and M. Werman. "Trilinearity of three perspective views and its associated tensor", In ICCV, 1995.