실사/CG 합성을 위한 3 차원 안정화

최성인 O, 박순용

경북대학교 전자전기컴퓨터학부, 경북대학교 IT 대학 컴퓨터학부

ellim5th@naver.com, sypark@knu.ac.kr

요 약

증강현실은 3 차원 가상물체를 비디오 영상과 합성하여 가상물체가 실제 존재하는 것과 같은 효과를 제공하는 기술이다. 증강현실 분양에서 현실감은 매우 중요한 문제이며 이는 얼마나 정확하고 적절하게 실사와 CG 를 합성할 수 있는가에 대한 문제로 볼 수 있다. 가상의 3 차원 물체를 영상에 안정적으로 합성하기 위해서는 실사 영상을 촬영하는 카메라의 위치와 방향을 정확히 추적하는 것이 중요하다. 카메라의 모션을 추적하기 위하여 일반적인 경우 GPS 또는 IMU 와 같은 항법센서를 사용한다. 하지만 항법센서에서 출력되는 결과는 오류 값을 포함할 수 있다. 본 논문에서는 카메라의 모션을 추정하는 항법센서의 오차를 줄이고 실사와 CG 를 안정적으로 합성하기 위한 컴퓨터비전 기술을 제안한다.

1. 서론

증강현실(augmented reality)은 3 차원 가상물체(virtual object)를 비디오 영상과 합성하여 가상물체가 실제 존재하는 것과 같은 효과를 제공하는 기술이다. 실사와 CG(computer graphics)를 합성하는 이 기술은 스포츠 중계나 음악 전문 채널 같은 방송이나 게임, 교육, 위치인식 서비스 분야 등에서 이미 널리 사용되고 있으며 최근에는 군사 분야에서 가상 훈련을 목적으로 증강현실 기술을 도입하려는 연구가 활발하게 진행되고 있다.

종전에 군사용 훈련 시뮬레이터를 구성하는 방식은 주로 가상현실(virtual reality) 기반이었다. 하지만 이 방식은 CG 로 제작된 가공의 상황과 환경에서 사람의 시각을 통해 느끼게 하고 상호작용을 유도하므로 시스템의 완성도에 따라 시뮬레이터를 접하는 훈련생이 느끼는 현실감의 편차가 매우 심한 단점이 있다. 또한 가상 공간에서만 훈련이 실시된다는 점 때문에 훈련 성과 면에서도 역시 현실성이 떨어진다는 문제점이 있다. 하지만 실 세계(real world)를 바탕으로 하는 증강현실 기술을 적용하면 이전보다 더 나은 현실감을 제공할 수 있기 때문에 실전과 같은 환경에서 훈련을 실시할 수 있는 장점을 가지게 된다.

증강현실 분양에서도 역시 현실감은 매우 중요한 문제이며 이는 얼마나 정확하고 적절하게 실사와 CG 를 합성할 수 있는가에 대한 문제로 볼 수 있다. 가상의 3 차원 물체를 영상에 안정적으로 합성하기 위해서는 실사 영상을 촬영하는 카메라의 위치와 방향을 정확히 추적하는 것이 중요하다. 추적된 카메라의 모션 정보를 이용하면 카메라 좌표계를 기준으로 가상의 물체를 원하는 위치에 합성할 수 있다. 반면 카메라 자세추정이 불안정하면 합성된 가상물체의 자세 또한 매우 불안정하게 된다.

본 논문에서는 전차 포 사격 훈련 시뮬레이터에서 실사와 CG 를 안정적으로 합성하기 위한 카메라 모션 추정 방법에 대해 기술하고자 한다. 제안하는 시스템은 그림 1 과 같은 구조를 가지고 있다.

카메라의 모션을 추정하기 위한 센서로서 제안한 시스템에서는 NAVCOM 사의 SF-2050 GPS 와 InterSense 사의 InertiaCube3 IMU 를 기본적으로 탑재하고 있다. GPS 와 IMU 는 각각 오차범위가 약 3m 이내에 0.25~1 도 급으로 매우 정확한 편에 속한다. 하지만 실제 주행상황에서는 원인을 알 수 없는 다양한 상황에 의해 오차범위를 크게 벗어나는 경우가 다반사며 이로 인해 영상 속 합성된 CG 는 떨림 현상을 보이게 된다. 이를 보정하기 위하여 컴퓨터 비전 알고리즘을 사용하여 항법 센서의 오차를 최소화하기 위한 방법을 제안한다.

본 논문은 다음과 같이 구성되어있다. 이어지는 2 장에서는 제안된 시스템을 설명하고, 3 장에서는 제안된 방법에 대한 실험결과를 제시한다. 마지막 4 장에서는 결론으로 끝을 맺는다.

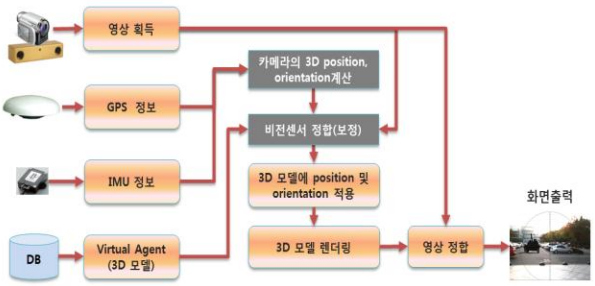


그림 1. 전차 포 사격 훈련 시스템 구성도

2. 비전 센서를 이용한 항법센서 안정화

카메라의 움직임을 추정하기 위해서는 카메라로부터 획득한 영상에서 특징점을 추출하고 이들 사이를 매칭하기 위한 기법이 요구된다. 또한 비전 센를 이용하여 항법센서의 오차를 최소화하는데 주목적이 있기 때문에 이전 상태에서 현재 상태 사이의 상대적 움직임에 대한 오류만을 보정하도록 한다.

2.1 특징점 추출 및 매칭

일반적으로 30fps 속도로 동작하는 CCD 센서를 사용하는 경우 연속하는 이미지에서 장면의 변화는 크지 않다. 또한 실시간 비전 시스템에서는 빠른 처리 속도가 결국 작은 모션에 대한 처리로 이어지기 때문에 정확도 향상에도 도움이 될 수 있다. 그래서 본 논문에서는 빠른 특징점 추출을 위하여 FAST 코너점[1]을 사용하였으며 코너점 주변에서 템플릿을 추출한 뒤 이를 기술자(descriptor)로 사용하여 매칭하는 방법을 사용하였다.

전통적인 템플릿 기반 특징점 매칭 방법은 두 이미지의 각 코너점에서 각각 독립적으로 템플릿을 추출한 뒤 이 템플릿들을 서로 비교하는 방법을 사용한다[2][3]. 하지만 본 논문의 시스템과 같이 주행하는 차량에서 영상을 획득하는 경우 차량의 흔들림이나 실시간으로 변화하는 카메라 화이트밸런스의 차이로 인해 추적에 용이한 일관성 있는 특징점을 추출하는 것이 어려워지게 된다.

이러한 문제를 해결하기 위하여 본 논문에서는 특징점을 매칭하는데 있어서 매칭 후보군이 쌍방유일성(biuniqueness)을 만족해야 한다는 것을 제약으로 하여 매칭을 실시하였으며 자세한 내용은 다음과 같다.

- 현재 프레임 영상 I_t 에서 FAST 특징점 C_t 를 추출한 뒤 C_t 를 중심으로 하는 m x m 크기의 템플릿 D_t 를 추출
- 이전 프레임 I_{t-1} 에서 코너점 C_t 를 중심으로 하는 j x k (j >= k) 크기의 관심영역 (region of interest) R_{t-1} 를 설정
- NCC(normalized cross correlation)를 이용하여 R_{t-1} 에서 D_t 와 가장 유사한 매칭 후보점 M_t 획득
- 이전 프레임 I_{t-1} 에서 매칭 후보점 M_t 를 중심으로 하는 m x m 크기의 템플릿 D_{t-1} 를 추출
- 현재 프레임 I_t 에서 j x k (j >= k) 크기의 관심영역 R_t 를 설정
- NCC 를 이용하여 R_t 에서 D_{t-1} 과 가장 유사한 매칭 후보점 M_{t-1} 획득
- 만약 매칭 후보점 M_{t-1} 과 M_t 가 동일한 좌표를 가지면 매칭 성립

관심영역 설정 시 j 를 k 보다 크게 설정한 이유는 차량의 모션이 주로 heading(또는 y 축)에 대해 크게 발생한다는 것에서 기인한 것이다. 본 논문에서는 640 x 480 크기의 실험 영상에 대해서 m 의 크기를 13, j 와 k 의 크기를 각각 201 과 31 로 설정하여 사용하였다. 또한 고속 특징점 매칭을 위하여 인텔의 IPP 라이브러리 [4]를 사용하였다.

2.2 매칭 오류 제거

2 차원 특징점 매칭시 쌍방유일성 제약을 적용하더라도 템플릿 비교 방법의 한계로 인해 반복되는 텍스쳐 영역 또는 텍스쳐 정보가 전무한 영역에서는 잘못된 매칭이 이뤄질 수 있다. 이러한 매칭 오류를 제거하기 위하여 본 논문에서는 매칭점 군에 대해서 에피폴라 제약(epipolar constraint)을 사용하여 RANSAC[5] 알고리즘을 적용한 뒤 이를 만족하는 매칭점들에 대해서만 카메라 모션을 추정하는데 사용하였다. 에피폴라 제약을 적용하기 위한 fundamental 행렬은 8-점 알고리즘[6]을 사용하여 계산하였으며 화소 재투영(pixel reprojection) 오차범위가 1 화소(pixel) 밑으로 떨어지면 RANSAC 알고리즘을 중지시켰다.

8-점 알고리즘은 SVD 를 이용하여 선형으로 해를 구하기 때문에 최적 해(optimal solution)를 구하기 위하여 비선형 방식으로 fundamental 행렬을 보정해야 한다. 본 눈문에서는 비선형 최적화 방법으로 잘 알려진 Levenberg-Marquardt[7] 알고리즘을 이용하여 아래 수식 (1)의 에너지 값 e 가 최소화 하도록 fundamental 행렬을 업그레이드 하였다.

$$e = \sum (x_t - P_t X_t)^2 + (x_{t-1} - P_{t-1} X_t)^2 \qquad (1)$$

수식 (1)에서 x 와 X 는 각각 2 차원 영상 점과 3 차원 월드 점을 말하며 P 는 투영행렬을 뜻한다. 투영 행렬을 만들기 위한 내부 파라미터는 카메라 보정을 통해 사전에 구하였으며 외부 파라미터는 연속하는 두 이미지 프레임에 대해서만 모션을 추정하는 규칙에 따라 t-1 에서는 $[I|0]$ 행렬, t 에서는 essential 행렬에서 추출한 $[R|t]$를 사용하였다. Essential 행렬은 앞서 구한 fundamental 행렬로부터 계산하였으며 이에 대한 내용은 이어지는 2.3 절에서 자세하게 설명하도록 한다.

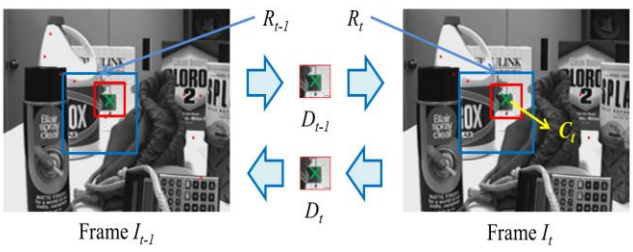


그림 2. 쌍방유일성 제약을 이용한 템플릿 매칭

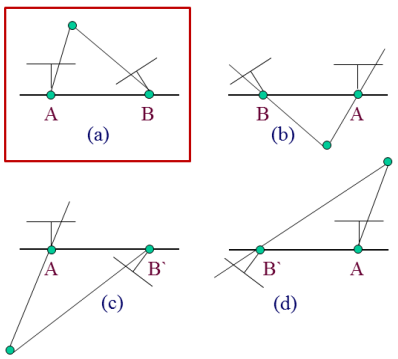


그림 3. E 행렬로 복원한 4 가지 카메라 모션 관계

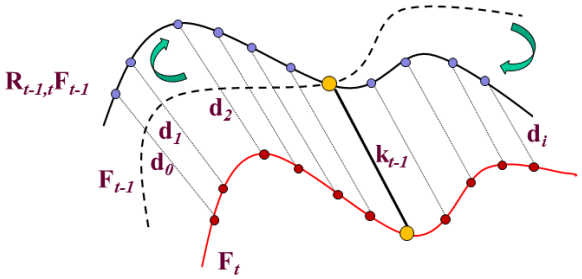


그림 4. 3 차원 거리영상을 이용한 스케일 계산

2.3 모션추정

카메라의 모션을 추정하기 위하여 본 논문에서는 [8]의 방법을 사용하였다. Essential 행렬은 두 카메라의 상대적 위치관계를 기술하는 3x3 행렬이며 아래 수식 (2)와 같이 회전을 나타내는 R 과 이동성분 $[t]_\times$로 표현된다. 여기서 $[t]_\times$는 3 차원 이동 벡터 t 에 cross product 를 사용하여 행렬 형태로 나타낸 것이다.

$$E = [t]_\times R \tag{2}$$

Fundamental 행렬과 essential 행렬과의 관계는 수식 (3)과 같다. K 는 카메라 내부 파라미터를 나타내며 사전에 알려진 값으로 이를 이용하여 우리는 F 로부터 E 를 구할 수 있다.

$$F = K^{-T} E K^{-1} \tag{3}$$

구해진 E 로부터 회전 R 과 이동 t 를 분리(decomposition) 하기 위하여 E 를 SVD 하면 아래와 같이 $U, V, \sum$로 분리된다.

$$E = U \sum V^T \tag{4}$$

분리된 U 와 V, 그리고 수식(5)의 W 를 이용하면 총 4 개의 해를 구할 수 있다.

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$[R \mid t] = \begin{bmatrix} UWV^T \mid -U(3) \end{bmatrix} \;\|\; \begin{bmatrix} UWV^T \mid U(3) \end{bmatrix} \;\|\; \begin{bmatrix} UW^TV^T \mid -U(3) \end{bmatrix} \;\|\; \begin{bmatrix} UW^TV^T \mid U(3) \end{bmatrix} \tag{5}$$

구해진 4 개의 해는 그림 3 과 같은 기하구조를 가진다. 이중 그림 3-(a)와 같이 두 카메라의 영상면(image plane) 앞에 3 차원 포인트가 존재하도록 하는 $[R|t]$가 최적 해가 된다. 최적 해를 선택하기 위해서 4 개의 $[R|t]$ 각각에 대해서 삼각측량(triangulation)을 실시하고 복원된 3 차원 점의 z 성분이 양의 값을 가지는 경우의 수를 카운트하였다. 그리고 가장 많은 수가 측정된 $[R|t]$를 최적 해로 사용하였다.

2.4 스테레오 정보를 이용한 스케일 계산

Essential 행렬 이용하여 카메라의 모션을 추정하는 방법은 원근 투영에 기반한 카메라 기하 모델을 사용하기 때문에 두 영상 시점 사이의 실제 거리는 알 수는 없으며 추정된 이동벡터는 항상 길이가 1 인 단위 방향 벡터가 된다. 하지만 GPS 의 위치 오차를 보정하기 위해 추정된 방향 벡터는 실제 월드를 반영할 수 있어야 하며 이를 위해서는 추정된 방향 벡터에 실제 스케일 값을 적용시켜야 한다. 실제 스케일 값을 반영하기 위해 본 논문에서는 스테레오 거리 정보를 사용한다.

앞서 소개한 모션 추정 과정을 통해 우리는 좌측 카메라 기준으로 현재 프레임 t 에서의 특징점을 알고 있다. 이 특징점을 이용하여 현재 프레임 t 에서 좌우 영상에 대한 스테레오 매칭을 실시 함으로써 시차(disparity)를 계산한고, 3 차원 깊이 정보를 추출하여 거리영상을 생성한다. 빠른 속도로 깊이 정보를 추출하기 위해서 본 논문에서는 CUDA 를 사용하여 스테레오 매칭 과정을 가속화 하였다.

현재 프레임과 이전 프레임의 거리 영상을 각각 F_t 와 F_{t-1} 이라고 하였을 때 스케일 k_{t-1} 을 구하는 과정은 다음과 같다. 먼저 essential 행렬을 분해하여 구한 R_{t-1} 을 F_{t-1} 에 적용한다. 그리고 모션추정 과정에서 획득한 2 차원 매칭 관계와 거리영상을 이용하여 각 특징점에 대한 3 차원 매칭 관계를 구하고 이 3 차원 매칭점들에 대한 유클리디언 거리 D 를 구한다. 모든 3 차원 매칭점에 대한 거리가 구해지면 스케일 k_{t-1} 은 아래 수식 (6)을 이용하여 결정한다.

$$k_{t-1} = median(d_0, d_1, d_2, ...d_i) \tag{6}$$

실제 주행상황에서는 반복되는 패턴, 가려짐(occlusion) 등에 의해 잘못된 스테레오 매칭 결과가 발생할 수 있으며 이는 곧 오류를 포함한 3 차원 깊

이 정보가 생성될 수 있음을 의미한다. 또한 스테레오 비전 센서를 이용하여 계산한 3 차원 깊이 정보는 구조상의 한계로 인해 원거리에 속한 점들이 근거리에 속한 점들에 비해 정확도가 현저하게 떨어지는 문제점을 가지고 있다. 이러한 제약사항들은 불안정한 스케일 값의 원인이 될 수 있다.

본 논문에서는 스테레오 깊이 정보의 불확실성으로 인한 스케일 값 오류를 최소화 하기 위하여 수식 (7)과 같이 타임슬롯을 이용한 필터링 방법을 사용한다. S_n 은 슬롯의 값을 나타내며 n 은 슬롯의 수를 말하며 실험에서는 이 값을 10 으로 설정하였다. 그림 5 는 타임슬롯을 이용한 필터링 방법을 보여준다.

$$k'_t = \frac{1}{n}\left(k_t + \sum_{2}^{n} s_n\right) \tag{7}$$

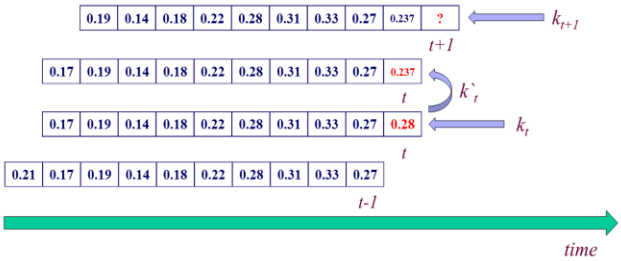


그림 5. 타임슬롯 슬라이딩을 이용한 k 값 필터링

2.5 GPS 및 IMU 센서 오차 보정

GPS 및 IMU 센서의 오차를 보정하기 위하여 본 논문에서는 비전 센서를 사용하여 영상 프레임간의 모션을 추정하고 그 결과를 센서 정보에 반영하도록 한다. GPS 및 IMU 센서에 의해 획득한 자세정보를 P_G 라고 하였을 때 항법 센서를 이용한 프레임간의 모션은 수식 (8)과 같이 나타낼 수 있다.

$$[R_{Gi} \mid t_{Gi}] = P_{G(i+1)} P_{Gi}^{-1} \tag{8}$$

비전 센서를 이용하여 획득한 모션을 $[R_{Vi} \mid t_{Vi}]$ 라고 하였을 때, 보정된 3 차원 위치 좌표 P_F 는 다음과 같다.

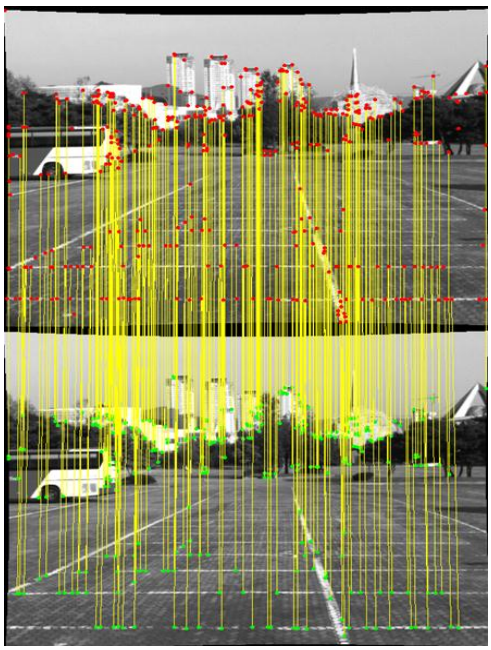
$$P_{Fi} = w_{Vi}[R_{Vi} \mid t_{Vi}]P_{Vi} + w_{Gi}[R_{Gi} \mid t_{Gi}]P_{Gi} \tag{9}$$

w_{Vi} 와 w_{Gi} 는 실험에 의해 결정된 가중치 값이다. 본 논문에서는 w_{Vi} 와 w_{Gi} 를 각각 0.3 과 0.7 로 설정하여 사용하였다.

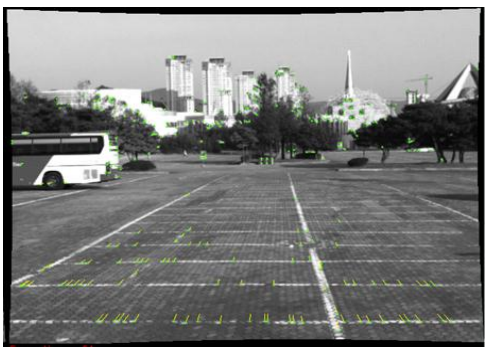
3. 실험 결과

실험을 위해 CCD 센서는 PointGrey 사의 Flea2 를 사용하였으며 베이스라인을 50cm 로 하여 스테레오 시스템을 구성하였다. 구성된 스테레오 센서는 그림 7 과 같이 RV 차량의 지붕 위에 설치하였으며 전 방향을 바라보도록 하였다. 영상은 640x480 해상도에 30fps 의 속도로 획득하였다. 메인 처리 시스템은 인텔 Q9550(2.83 GHz) CPU 에 4GB 메인 메모리를 탑재하고 있다.

그림 7. 스테레오 비전 시스템



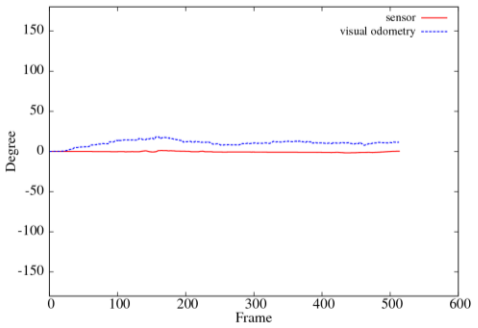
(a)



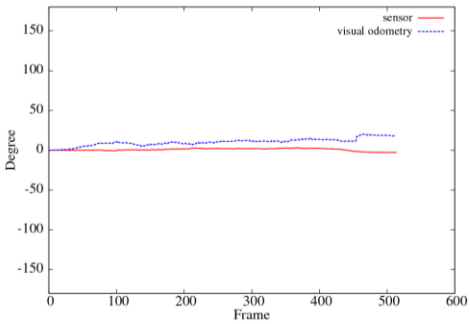
(b)

그림 8. 쌍방 유일성 제약을 적용한 특징점 추적 (a) 이전 프레임과 현재 프레임의 매칭 관계 (b) 모션 벡터 계산

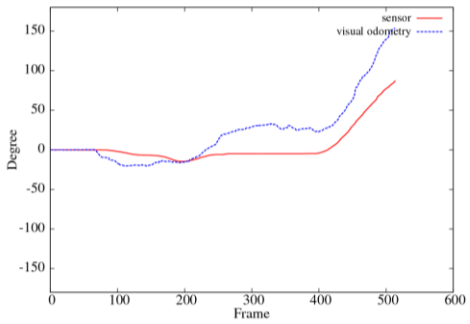
그림 8 은 쌍방 유일성 제약을 적용한 특징점 추적 결과를 보이고 있다. 매 프레임에 대해 약 400~600 개의 특징점을 추출하고 매칭하였으며 RANSAC 을 이용한 매칭오류제거 과정을 포함하여



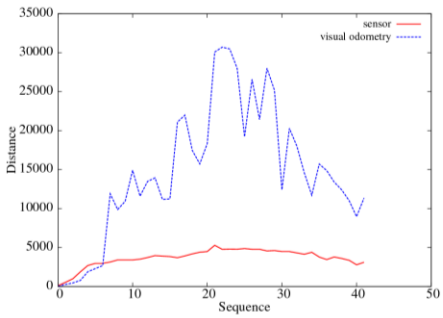
(a)



(b)



(c)



(d)

그림 9. 비전 센서와 항법 센서의 모션 추정 결과 (a) pitch 각 비교 (b) head 각 비교 (c) roll 각 비교 (d) 이동 벡터의 크기 비교

약 300ms 의 성능을 보이고 있다.

그림 9 은 비전 기술을 사용하여 추정한 모션과 항법 센서의 모션 추적 결과를 비교한 것이다. 특히 그림 9-(d)의 경우 GPS 와 비전센서 간의 이동벡터 차이가 매우 큰 것으로 나타난다. 이는 스테레오 복원 과정에서 잘못된 매칭이 이루어져 복원한 깊이 정보의 오류가 매우 크기 때문인 것으로 판단된다.

4. 결론

본 논문에서는 카메라의 모션을 추정하는 항법 센서의 오차를 줄이고 실사와 CG 를 안정적으로 합성하기 위한 컴퓨터비전 기술을 제안하였다. 우리는 연속하는 영상 사이의 상대적인 모션을 추정하고 그 결과를 항법센서 정보의 오차를 보정하는데 사용하였다. 상대적인 모션을 추정하기 위하여 쌍방 유일성을 만족하는 특징점 추출 및 추적, 에피폴라 기하 제약을 이용한 매칭 오류 제거, essential 행렬을 이용한 모션 추정 방법이 소개되었다. 그리고 essential 행렬로 구한 이동벡터의 스케일을 결정하기 위해서 스테레오 매칭 정보를 이용하였으며 거리영상으로 복원한 뒤 이동 벡터의 실 스케일을 획득하였다.

제안하는 시스템은 현재 초당 3.3 프레임 정도의 처리 속도를 보이고 있으며 실시간 시스템에 사용하기에는 아직까지 무리가 있는 실정이다. 향후 연구과정에서 보다 빠른 속도 처리를 위한 알고리즘 개선이 필요하다.

감사의 글

본 연구는 국방과학연구소의 "소형무인로봇의 자율복귀를 위한 스테레오 비전 기반 위치인식 기술 개발"과제의 지원으로 수행되었음.

참고문헌

[1] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” In Proc. 9th European Conference on Computer Vision (ECCV), 2006

[2] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, “Real time localization and 3d reconstruction,” In Proc. 19th Computer Vision and Pattern Recognition (CVPR) , 2006

[3] A. Howard, “Real-time stereo visual odometry for autonomous ground vehicles,” International Conference on Intelligent Robots and Systems (IROS), pp.3946-3952, 2008

[4] “Integrated Performance Primitives (IPP),” Intel, http://software.intel.com/en-us/articles/intel-ipp

[5] M. Fischler and R. Bolles, “Random Sample Consensus: a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography,” Communications of the ACM, vol. 24, no.6, pp. 381-395, 1981

[6] R. Hartley, "In Defense of the Eight-Point Algorithm". IEEE Transaction on Pattern Recognition and Machine Intelligence, vol.19, no. 6, pp. 580–593, 1997

[7] J. Nocedal and S. Wright, “Numerical Optimization,” Springer. ISBN 0-387-30303-0., 2006

[8] R. Hartley and A. Zisserman, “Multiple View Geometry in computer vision,” Cambridge University Press. ISBN 978-0-521-54051-3., 2003