

3차원 기하정보 및 특징점 추적을 이용한 다시점 거리영상의 온라인 정합

Online Multi-view Range Image Registration using Geometric and Photometric Features

백재원, Jaewon Baek, 박순용, Soon-Yong Park

경북대학교 컴퓨터공학과 컴퓨터비전 연구실
depia@vision.knu.ac.kr, sympark@knu.ac.kr

요약 본 논문에서는 실물체의 3차원 모델을 복원하기 위해 거리영상 카메라에서 획득된 3차원 점군에 대한 온라인 정합 기법을 제안한다. 제안하는 방법은 거리영상 카메라를 사용하여 연속된 거리영상과 사진영상을 획득하고 문턱값(threshold)을 이용하여 물체와 배경에 대한 정보를 분류한다. 거리영상에서 특징점을 선택하고 특징점에 해당하는 거리영상의 3차원 점군을 이용하여 투영 기반 정합을 실시한다. 초기정합이 종료되면 사진영상간의 대응점을 추적하여 거리영상을 정제하는 과정을 거치는데 대응점 추적에 사용되는 KLT(Kanade-Lucas-Tomasi) 추적기를 수정하여 초기정합의 결과를 대응점 탐색에 이용함으로써 탐색의 속도와 성공률을 증가시켰다. 특징점과 추적된 대응점에 해당하는 3차원 점군을 이용하여 거리영상의 정제를 수행하고 정합이 완료되면 오프라인에서 3차원 모델을 합성하였다. 제안한 알고리즘을 적용하여 2개의 실물체에 대하여 실험을 수행하고 3차원 모델을 생성하였다.

핵심어: multi-view range image, KLT registration, stereo, online registration

1. 서론

3차원 컴퓨터비전기술은 카메라로 획득한 영상 정보로부터 실환경(real scene)의 3차원 정보를 정확하고 빠르게 추출하려는 노력에서 비롯되었다. 국내외의 3차원 컴퓨터비전 기술은 주로 로봇, 디스플레이, 그리고 디지털 콘텐츠(contents) 등의 응용분야에 적용 되고 있다. 3차원 정보를 빠르고 정확하게 추출함으로써 로봇이나 자동차가 카메라를 이용하여 자율적인 주행이나 위험감지, 자동주차 등에 사용할 수 있다. 또한 역사적인 유물 또는 유적을 3차원으로 디지털화 하여 문화원형을 보존 하고 특수효과 및 가상현실을 구현하는 등의 다양한 분야에 활용할 수 있다. [1,2,3]

일반적으로 실환경의 3차원 정보를 빠르게 추출하기 위해서는 거리영상 카메라(range imaging camera)를 사용한다. 하지만 사람의 눈과 마찬가지로 물체의 가려진 부분을 볼 수 없기 때문에 물체의 여러 방향에서 거리영상을 획득하여야 물체의 완전한 3차원 정보를 획득 할 수 있다. 한편 다시점(multi-view)에서 획득된 거리영상들은 서로 독립적인 좌표계를 가지기 때문에 이들을 하나의 좌표계로 이동시켜야 하는데 이것을 3차원 정합(registration)이라 한다. 3차원 정합에 대한 초기 연구는 3차원 점군(point clouds) 또는 곡면(surface)을 정확하게 정합하고자 하였다. [4]

주) 이 논문은 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임. (KRF-2006-331-D00518)

반면 최근에는 고속으로 거리영상을 정합하려는 시도가 많이 진행이 되고 있고 또한 실시간으로 3차원 정합을 구현하려는 기술이 연구 되고 있다.

Liu와 Heidrich는 스테레오 기반의 다시점 거리영상을 실시간으로 정합하는 연구를 수행하였다. [5] Jaeggli와 Koninckx는 패턴 프로젝터를 이용하여 다시점 거리영상을 획득하고 이를 온라인으로 정합하고 오프라인에서 정합결과를 정류하는 연구를 수행하고 있다. [6] 또한 Popescu 등은 레이저 포인트 및 비디오카메라를 이용한 실시간 3차원 거리영상 센서를 이용하여 실물체 및 실내 환경 복원에 대한 연구를 수행하고 있다. [7]

본 논문에서는 거리영상(range image)과 사진영상(photometric image)만을 이용하여 온라인으로 정합할 수 있는 시스템을 제안하고자 한다. 투영기반 정합 기술을 통해 거리영상을 정합하는 과정을 거친 후, 수정된 KLT 추적기를 사용하여 영상 특징점들의 대응점을 찾아 투영기반 정합 기술의 결과를 정제 하였다. 이때 초기정합의 결과를 이용하여 사진영상의 특징점을 탐색범위를 한정할 수 있어 정합을 빠르게 구현할 수 있다. 또한 정합된 3차원 점군을 하나의 3차원 모델로 합성(integration) 하였다. 제안된 시스템을 사용하여 실험실 환경과 외부 환경에서 실물체(real object)에 대하여 실험을 수행하였다.

본 논문의 구성은 다음과 같다. 2절에서는 온라인 정합의 전체적인 구조를 다루고 투영기반 거리영상 정합 기술은 3절에서 다룬다. 4절에서는 영상기반 정합기술을 보이며 5절에서는 실험 결과를 다룬다. 6절에서는 본 연구의 결론을 제시한다.

2. 온라인 정합

본 논문에서 다시점 거리영상의 정합은 그림 1과 같은 과정으로 이루어진다. 거리영상 카메라를 통해 거리영상 및 사진영상을 획득한다. 획득된 거리영상은 배경과 물체의 구분이 없는 상태인데 일반 환경에서는 스테레오 알고리즘의 특성상 특징이 없는 벽 등에는 잡음이 많아서 정합에 사용하기에 좋지 않다. 거리영상에서 배경에 대한 부분을 제거하기 위해 문턱값을 기준으로 배경이라고 판단되는 부분을 제거한다. 우선 두 개 이상의 거리영상을 획득하면 KLT 추적기를 사용하여 n번째 영상의 특징점을 선택한다. 기존의 KLT 추적기는 특징점을 선택하는데 사진영상만의 정보를 사용하지만 본 논문에서는 특징점에 해당하는 3차원 좌표가 없는 경우에는 선택하지 않았다. 이를 통해 투영기반 거리영상 정합과 영상기반 정합에 사용되는 특징점은 배경을 제외하고 물체에 대해서만 선택된다.

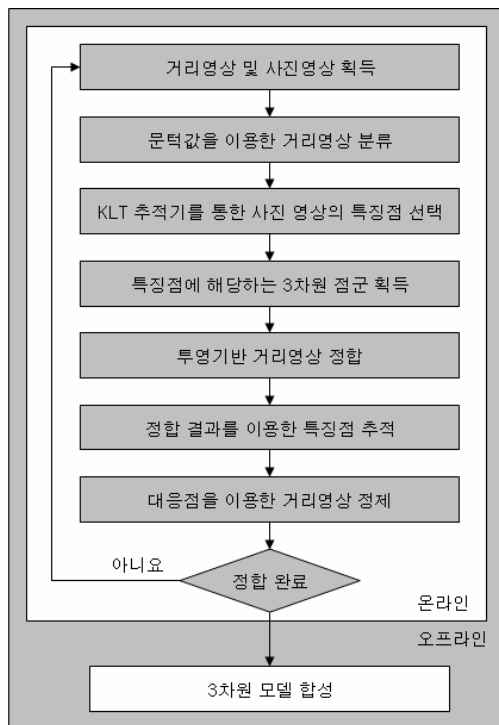


그림 1. 다시점 거리영상의 정합과정

선택된 특징점들에 해당하는 3차원 좌표를 투영기반 거리영상 정합의 입력 좌표로 사용하여 $[n, n-1]$ 한 쌍으로 정합을 수행한다. 투영기반 거리영상 정합에 대해서는 3절에서

자세히 설명한다. 투영기반 거리영상 정합의 결과를 정제하기 위해 수정된 KLT 추적기를 사용한다. n번째 영상의 특징점 p에 해당하는 3차원 좌표 P가 투영기반 거리영상 정합을 통해 p의 대응점 q에 해당하는 3차원 좌표 Q와 가까워졌다고 가정하고 정합된 P점을 n-1번째 영상에 투영하여 KLT 추적기의 시작 좌표로 사용한다. 수정된 KLT 추적기를 사용하여 특징점을 n-1번째 영상에서 추적하여 대응점을 구한다. 특징점과 대응점에 해당하는 3차원 점군을 획득하여 투영기반 정합의 결과를 정제하였다. 거리영상 카메라의 시점을 이동하고 위 과정을 반복하면서 충분한 데이터를 얻으면 정합을 종료하고 3차원 합성 과정을 통해 하나의 모델을 구성한다.

3. 투영기반 거리영상 정합

거리영상의 정합은 크게 점대점(point to point), 점대면(point to plane), 그리고 점대투영점(point to projection)의 세 가지로 나눌 수 있다.[8,9] 일반적으로 알려져 있는 ICP(iterative closest point) 방법은 점대점 또는 점대면 방법을 말하며 정확성이 높지만 속도가 느린 단점이 있다. 점대투영점 방식은 2.5차원의 거리영상과 3차원 곡면사이의 투영관계를 이용한다. ICP와는 달리 속도가 빠르지만 두 곡면의 초기 자세가 큰 오차를 가지고 있을 때는 ICP에 비해 정확성이 떨어지는 단점이 있다.

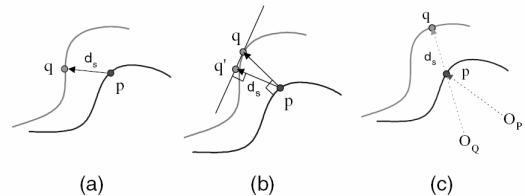


그림 2. 3가지 일반적인 거리영상 정합 기술 (a)점대점 방법 (b)점대면 방법 (c)점대투영점 방법

본 논문에서는 이러한 문제를 해결하기 위해 점대면과 점대투영점 방식의 장점을 결합한 정합 기술인 IPP(iterative-projection-point) 방식을 사용하였다.[10] 이 방식은 점대투영점 기술을 변형하여 점 P와 Q를 각각 영상면과 P의 법선에 연속적으로 투영함으로써 점대면 방식을 고속으로 구현할 수 있다.

3.1 Iterative Project Point(IPP) Algorithm

그림 3에서 IPP 방식을 간단히 설명하고자 한다. S 곡면상의 점 P_0 의 정합점을 곡면 D에서 찾는 것이 정합의 주목적이다. 우선 P_0 를 Q 곡면의 2차원 영상으로 투영하여 좌표 p_D 를 구한다. 좌표 p_D 에 해당하는 D 곡면상의 점 Q_r 을 거리영상을 이용하여 구할 수 있다. 이 점을 다시 P_0 의 법선에

투영하여 P_i 를 구하고 위의 과정을 P_k 가 수렴할 때까지 반복하여 Q 를 구한다. 이러한 방법으로 충분한 수의 대응점의 집합 $\{P\}$ 와 $\{Q\}$ 를 구하였을 때 두 점집합 사이의 정합 오차 ϵ 를 최소화하는 3차원 좌표변환 행렬 T 를 다음 선형 방정식을 이용하여 구한다. 본 논문에서는 수식 (1)을 풀기 위해 SVD(singular value decomposition) 방법을 사용하였다.

$$\epsilon = \sum_i \| P_i - TQ_i \| \quad (1)$$

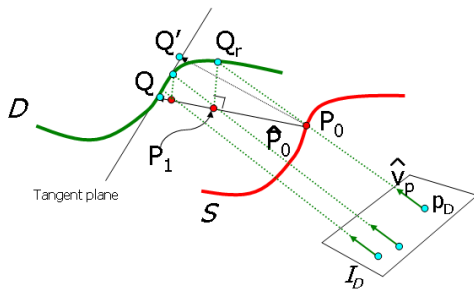


그림 3. IPP 기술을 이용한 정합점 탐색

위 과정을 반복하여 최종적인 변환 행렬을 계산한다. 그림 4는 두 개의 거리영상에 대하여 IPP 알고리즘을 적용한 결과를 보여준다. 거리영상은 본 논문의 실험에 사용된 것이며 이 그림은 IPP 알고리즘 적용에 따른 정합의 RMS 오차를 보여준다. 그림에서와 같이 일반적인 경우 20회 이상의 정합을 수행하여도 정합 결과가 변하지 않았으며, 따라서 본 논문에서는 투영기반 거리영상 정합은 20회만 수행하였다.

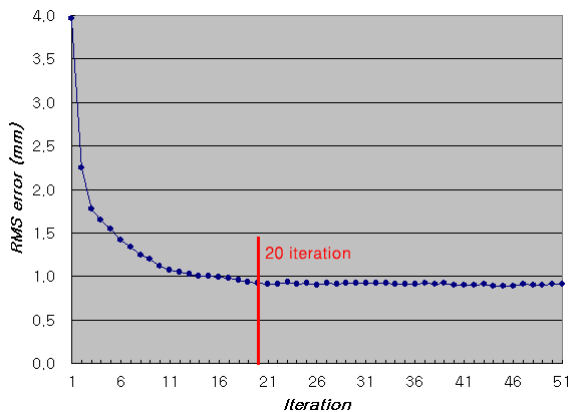


그림 4. IPP 정합 반복에 따른 RMS 오차의 변화

3.2 IPP 입력 좌표 선택

본 논문에서는 IPP의 입력 좌표 집합 P 를 선택하기 위해 여러 방법을 사용한다. 먼저 사진영상에서 KLT 추적기를 사용하여 특징점을 선택한다. 선택한 특징점들에 해당하는 3차원 좌표가 있는지를 확인하여 특징점이 물체위에 존재하

는가를 판단하고 배경에 해당하는 특징점은 사용하지 않는다. 위 과정을 통해 그림 5에서 보는 것과 같이 물체에 해당하는 특징점만을 추출 할 수 있는데 이것은 영상기반 정제 과정에서도 사용된다. 선택된 3차원 점의 법선(normal)과 카메라의 시점(view point)을 내적 하여 IPP에서 정합점을 찾기 힘들다고 판단되는 점들은 사용하지 않는다. 실험에서는 내적이 -0.5 이상인 3차원 점들은 정합에 사용하지 않았다.

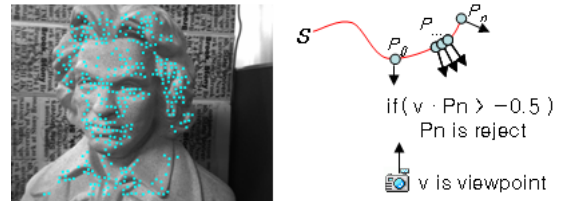


그림 5. 유효한 특징점과 IPP에 적합한 3차원 점 선택

3.3 쌍정합과 전역 정합

본 논문에서 투영기반 정합은 한 쌍(pairwise)의 거리영상 사이에서 이루어진다. 다수의 연속된 영상에 대한 정합을 실시할 때 쌍정합의 문제점은 그림 6(a)를 통해 알 수 있다. 다수의 영상에서 0번과 1번 곡면을 정합하면 1번과 2번 곡면은 실제로 획득된 거리보다 멀어지게 되므로 정확한 정합점을 찾기 힘들다.

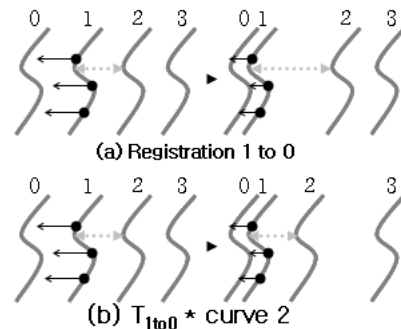


그림 6. (a) 쌍정합의 문제점 (b) 해결방안

이러한 문제점을 해결하기 위해 1번과 2번 곡면을 정합하기 전에 0번과 1번 곡면 사이의 변환 행렬을 2번 곡면에 곱해주면 1번과 2번 곡면은 다시 원래의 변환 관계로 돌아간다. 또한 제안된 시스템에서는 정합 속도에 무관하게 정합 성능을 최대도 해야 하는 경우를 생각하여 쌍정합이 아닌 전역 정합을 구현하였다. 그림 7(a)와 같이 쌍정합은 $[n-1, n]$ 사이에서만 IPP를 이용하여 정합점을 탐색하지만 전역 정합은 그림 7(b)와 같이 $[0, \dots, n]$ 사이에서 정합점을 탐색하므로 상대적으로 높은 정합 성능을 보이지만 정합해야 할 영상의 수가 늘어나면 정합에 걸리는 시간이 증가하는 단점이 있다. 전역 정합은 사용 여부를 선택할 수 있

다.

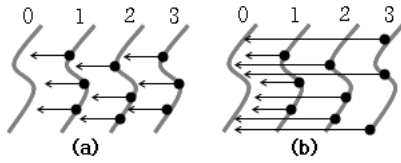


그림 7. (a) 쌍정합 (b) 전역 정합

4. 영상기반 정합기술

기존의 영상기반 정합기술은 거리영상이 존재할 때 사진 영상간의 대응점을 추적하여 해당하는 3차원 점군을 정합에 이용하였다. 하지만 좌표변환이 큰 거리영상 사이의 대응점을 추적 할 경우에는 탐색 범위의 크기에 따라 추적의 성공률과 추적에 걸리는 시간이 상충(trade off)한다. 실험에서는 보정(recitification)된 거리영상을 사용 하였다.

본 논문 에서는 투영기반으로 정합된 거리영상을 이용하여 KLT 추적기의 탐색 시작 좌표를 결정 하는 방법으로 KLT 추적기를 수정하여 사용함으로써 상대적으로 좁은 탐색 범위로 대응점 추적 성공률을 높였고 추적 속도를 개선 하였다.

그림 8은 기존 KLT 추적기가 사진영상 I_S 상의 특징점 p_S 의 대응점을 I_D 에서 추적하는 과정을 나타낸다. 기존 KLT 추적기는 p_S 와 동일한 좌표를 가지는 p_S' 를 탐색 시작 좌표로 사용하기 때문에 넓은 탐색범위를 필요로 한다.

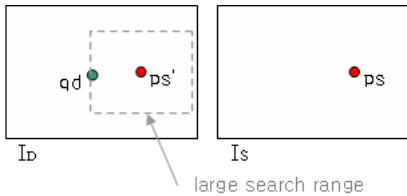


그림 8. 기존 KLT의 대응점 탐색 방법

그림 9는 본 논문에서 제안하는 방법을 나타낸다. 특징점 p_S 에 대응되는 3차원 거리를 $r_S(p_S)$ 라 할 때 수식 (2)와 같이 I_D 에 투영하여 p_S' 를 구한다. 이 때 모든 3차원 점군 데이터는 초기정합 과정을 통해 0번 거리영상에 정합되어 있으므로 I_D 거리영상의 역변환행렬 T_{n-1}^{-1} 를 I_S 거리영상에 곱해주었고 투영변환행렬 M 을 곱하여 I_D 사진영상에 투영하였다.

$$p_S' \cong M(T_{n-1}^{-1}(r_S(p_S))) \quad (2)$$

정합 전에는 투영된 $r_S(p_S)$ 가 기존 KLT의 탐색 시작 좌표와 동일하지만 정합 후에는 추적해야할 대응점 q_d 에 근접

하는 것을 알 수 있다. 이를 통해 상대적으로 작은 탐색 범위를 사용하여 대응점을 추적하고 p_S 와 q_d 의 집합을 구한다. p_S 와 q_d 에 해당하는 3차원 점들의 집합을 수식 (1)에 대입하여 기하정합 결과를 정제하였다.

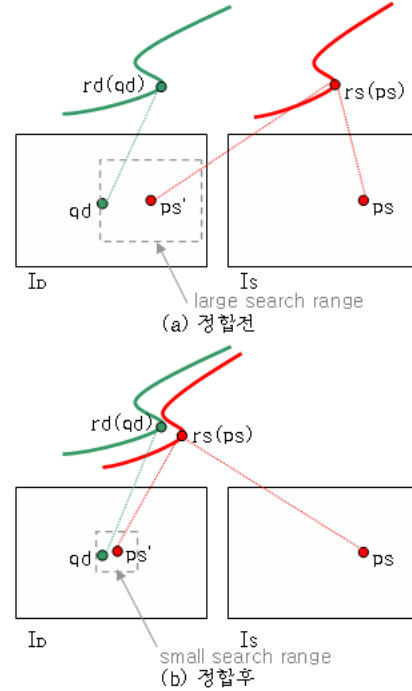


그림 9. 정합전(a)과 정합후(b)의 탐색영역 비교

5. 실험 결과

본 논문에서는 캐나다 PointGrey에서 개발된 BumbleBee 스테레오 카메라를 사용하여 320x240 해상도의 사진영상과 거리영상을 획득하였다. 실험은 실험 환경과 일반 환경에서 서로 다른 물체를 사용하여 진행되었다. 정합에 사용된 컴퓨터는 Pentium 3.4 GHz 이고 제안된 방법은 위 시스템에서 두 장의 거리영상을 정합하는데 색상 정보를 포함할 경우 약 800ms, 포함하지 않을 경우 약 600ms의 정합 속도를 보였다.



그림 10. 거리영상 카메라와 실험 환경

5.1 기존 방법과의 성능 비교

실험에서는 동일한 KLT 탐색 범위를 사용하여 제안한 방법과 기존의 KLT를 사용했을 경우에 추적 성공률 및 정

합 결과에 대한 비교를 수행하였다.

그림 11은 특징점의 이동을 운동 벡터로 표현한 것이다. 회색 선은 추적이 성공한 경우라고 판단된 것이고 흰색 선은 검색이 실패한 경우를 나타낸다. 기존의 KLT에서 좁은 탐색 범위로 인해 추적할 수 없는 대응점을 제안된 방법을 사용했을 경우 비교적 좋은 추적 결과를 볼 수 있었다.

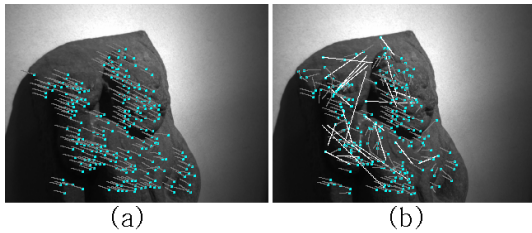


그림 11. 추적 결과 비교 (a) 제안된 방법 (b) 기존 KLT

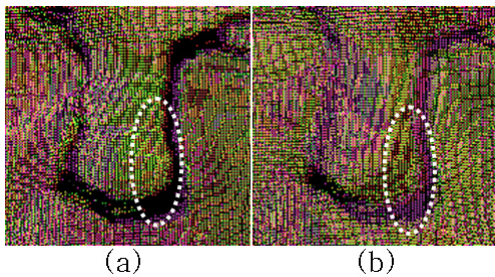


그림 12. 정합결과 비교 (a) 제안된 방법 (b) 기존 KLT

그림 12는 수정된 KLT와 기존 KLT를 사용하여 영상기반 정합을 수행한 결과를 나타낸 것이다. 기존 KLT의 경우 그림 11(b)와 같이 추적이 실패한 경우가 영상기반 정합의 입력으로 사용되므로 그림 12(b)와 같이 좋지 않은 정합 결과를 보였다.

5.2 실험 결과

다음은 실험실 환경에서 베토벤상에 대해 40장의 연속된 사진영상 및 거리영상을 획득하여 정합을 수행한 결과이다. 데이터를 획득 할 때 배경과 카메라는 고정시키고 물체만 회전판위에서 회전시키는 방법을 사용하였다. 획득된 모든 거리영상을 정합하고 메쉬(mesh) 구조로 3차원 모델이 생성된 것을 확인할 수 있다. 메쉬 모델의 생성은 Marching Cubes 알고리즘을 사용하였다.

그림 14는 일반 환경의 박물관에 소장중인 금강역사 부조에 대해 60장의 연속된 사진영상 및 거리영상을 획득하여 정합을 수행한 결과이다. 데이터를 획득 할 때 물체는 고정되어 있고 카메라를 손에 휴대하는 방법으로 촬영하였다. 실험 환경에 비해 카메라의 움직임이 불안정하고 3차원 점군

에 약간의 잡음도 포함되어 있다. 또한 박물관의 조명 특성상 물체에 그늘진 곳이 생겨 다시점에서 데이터를 받았지만 3차원 점군의 데이터를 획득할 수 없는 곳이 있었다.



그림 13. (a)보정된 사진영상 (b)거리영상 (c) 특징점 선택 (d)추적된 특징점의 운동 벡터 (e)정합된 점군 (f)합성된 모델

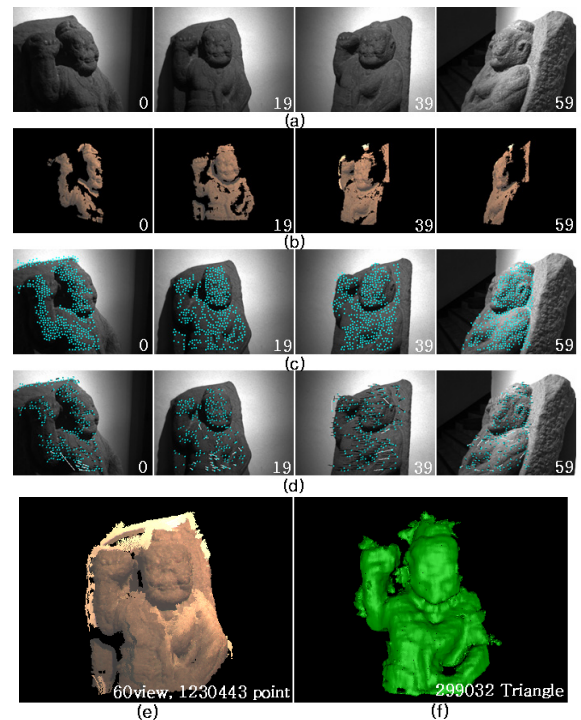


그림 14. (a)보정된 사진영상 (b)거리영상 (c) 특징점 선택 (d)추적된 특징점의 운동 벡터 (e)정합된 점군 (f)합성된 모델

표 1에서는 2가지 실험에 대한 정합 속도를 제시한다. 정합 속도는 쌍정합의 속도를 전체 정합 횟수로 나누어 평균을 계산하였고 합성속도는 전체 영상을 사용해서 합성했을 경우, 일부 영상을 사용했을 경우로 나누어서 실험하였다.

표 1. 정합 및 합성 속도 비교 (sec)

실험물체	거리영상정합 (평균)	모델합성 (voxel=5mm)		
		20frame	40frame	60frame
베토벤상	0.71/frame	350	720	N.A.
금강역사 부조	0.85/frame	220	N.A.	660

베토벤상은 총 40장의 거리영상을 정합하였고 메쉬 모델의 합성은 전체 거리영상을 사용한 경우와 20장의 영상의 사용한 경우로 실험하였다. 그림 13(f)의 메쉬 모델을 합성할 때의 복셀 크기는 3mm로 지정하였다. 금강역사 부조는 평균 정합 속도가 약간 느린 것을 볼 수 있는데 일반 환경에서 실험했기 때문에 조명 등의 문제로 거리영상에 잡음이 일부 포함되어 있었기 때문이라고 추정된다. 그림14(f)의 금강역사 부조 모델의 합성에는 복셀 크기를 2mm로 지정하였다.

6. 결론 및 향후과제

본 논문에서는 다시점 거리영상을 온라인으로 정합하기 위한 시스템을 제안하였다. 제안된 방법을 사용함으로써 온라인으로 시점을 이동하며 획득된 거리영상이 성공적으로 정합됨을 알 수 있다. 하지만 사용자가 거리영상 카메라에 카메라의 셔터 속도이상의 움직임을 보이는 경우와 연속된 거리영상간의 변환이 많은 경우 정합이 잘 되지 않음을 알 수 있었다. 향후에는 이러한 문제점들을 해결하고 사용자가 정합의 진행 정도를 보고 시점을 이동할 수 있도록 대화형 인터페이스에 대한 연구를 진행할 예정이다.

참고문헌

[1] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project : 3D scanning of large statues. In SIGGRAPH, pages 131-144, 2000.

[2] Daniel Huber and Martial Hebert. 3-D Modeling Using a Statistical Sensor Model and Stochastic Search. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 858-865, June 2003

[3] P. Allen, S. Feiner, A. Troccoli, H. Benko, E. Ishak, B. Smith, "Seeing into the Past: Creating a 3D Modeling Pipeline for Archaeological Visualization," 3dpvt, pp. 751-758, Second International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT'04), 2004.

[4] Rusinkiewicz, S., Levoy, M., 2001. Efficient variants of the ICP algorithm, In: Proceedings of 3D Digital Imaging and Modeling, pp. 145-152.

[5] Yushuang Liu, Heidrich, W. "Interactive 3D model acquisition and registration", Computer Graphics and Applications, 2003. Proceedings. 11th Pacific Conference, pages 115-122, 2003

[6] T. Jaeggli, T.P. Koninckx, L. Van Gool, "Online 3D Acquisition and Model Integration, IEEE International Workshop on Projector-Camera Systems - ICCV03, cdrom proc, October 12, 2003, Nice, France.

[7] V. Popescu, G. Bahmutov, M. Mudure, E. Sacks, "The ModelCamera," Elsevier Journal of Graphical Models, January 2006.

[8] R. Benjema and F. Schmitt, "Fast global registration of 3D sampled surface using a multi-z-buffer technique, " Image and Vision Computing, 17, 113-123, 1999.

[9] R. Bergevin, M. Soucy, H. Gagnon, and D. Laurendeau, "Toward a general multi-view registration technique," IEEE Trans. on Pattern Analysis and Machine Intelligence, 18 (5), 1996.

[10] Soon-Yong Park and Murali Subbarao, "An Accurate and Fast Point-to-Plane Registration Technique," Pattern Recognition Letter, Vol. 24, No. 16, pp. 2967-2976, Dec 2003