

다중 비디오 영상을 이용한 실시간 가상공간 영상 재구성 시스템

*최혁승, **한태우, **이주호, **양현승
*삼성전자 디지털미디어연구소
**한국과학기술원 전자전산학과
e-mail : hyok.choi@samsung.com

Real-time Virtual Volumetric Scene Reconstruction System from Multiple Video Streaming

*Hyok S. Choi, ** Taewoo Han, **Juho Lee, **Hyun S. Yang
*Digital Media R&D Center, Samsung Electronics Co., Ltd.
**Computer Science Division, Department of EECS, KAIST

요 약

근래의 컴퓨터 그래픽스 분야의 중요 목표 중 하나는 동적으로 변화하는 3 차원 가상 공간을 재현해 내는 것이다. 일반적으로 공간정보 취득 기술은 특수한 하드웨어나 환경을 전제로 하는 능동적인 방법과, 특정 환경을 전제하지 않으나 상대적으로 계산 복잡도가 높은 수동적 방법으로 나뉜다. 이 논문에서는 수동적 알고리즘의 계산 복잡도를 개선하여, 특수한 환경이나 물리적인 전제 없이 비교적 간단한 하드웨어를 이용하여 정보 취득 후 되도록 짧은 시간(latency)내에 가상 영상을 재구성하는 시스템을 설계 구현한다.

1. 서론

컴퓨터 비전 분야에서 2 차원의 영상들로 3 차원 정보 모델을 가진 영상을 재구성(virtual volumetric scene reconstruction)하는 문제는 오래되었지만 아직 어려운 문제 중 하나이다. 일반적으로 이러한 문제는 기하학적인 모델과 재료적 특성, 조명 등의 고려를 기반으로 합성하는 방법으로 접근할 수 있다.[1]

근래 연구가 활발한, 동적 변화 영상의 가상 공간 재현 연구는 크게 능동적인 방법과 수동적 방법으로 분류된다. 수동적인 방법들은 디지털 카메라 등의 수동적인 장치만을 이용하여 취득하는 것이며, 능동적인 방법은 특수한 광을 이용한 특수한 장치나 레이저를 이용한 3 차원 거리 스캐너 등을 이용한 방법이다. 일반적으로 능동적인 방법은 고품질의 결과를 만들 수 있으나, 대개 매우 많은 비용이 요구되거나 정보취득(acquisition)에 매우 시간이 많이 걸리고, 물리적인 변화(특수 광원, 도색 등)가 필수적인 경우가 많아 응용에 일반적으로 한계가 있다. 수동적인 방법은 이러한 전제를 요구하지 않는 장점이 있으나, 일반적으로 계

산복잡도가 매우 높아 즉각적으로 결과를 얻기가 힘들며, 결과 품질에 개선 여지가 높다.[1]

본 논문에서는 기존의 수동적 방법의 계산복잡도를 개선한 알고리즘을 제시하며, 다수의 동기식 실시간 비디오 영상 획득, 조명환경에 상대적으로 강건한 윤곽 추출을 이용하여, 특수한 환경이나 물리적인 제약이 없는 실시간 영상재구성시스템을 설계 구현하였다.

2. 수동적 가상 영상 재구성 기술

기존의 수동적 가상영상재구성기술(passive virtual volumetric reconstruction)은 다음과 같이 분류된다.

가. 공간단위색차법(voxel coloring)

공간단위색차법은 색상의 일관성(color consistency)에 의존한다.[2-4] 시야가 가려지지 않은 공간상의 동일한 지점에 대해 서로 다른 카메라에 영상(projection)된 지점의 영상(image)의 색상이 동일한 색상이라면 그 지점은 존재하는 것으로 간주하고, 다른 색상의 경우 허공으로 간주하는 방법이다. 개념상 간단하지만, 특정 지점에 대한 시선의 가려짐(visibility)에 대한 고려, 물

체의 외각부분에 대한 색상 존재성(color consistency)의 임계치 문제, 카메라의 성능이나 조명, 외란에 대한 취약성 등이 과제이다.

나. 입체시각법(stereo vision)

입체시각법은 한 쌍의 2 차원 영상을 이용하여 거리정보를 획득하는 방법이다. 많은 연구가 진행되었으나, 두 영상 사이의 동일 지점에 대한 강건한(robust) 위치 파악이 매우 어려운 과제이다.[5-6] 일반적으로 서로 상응하는 기본선(baseline)을 따라 표준화된 상호관계값(normalized cross-correlation)을 이용하며, 대부분 부동 거리(disparity range)를 일정 값 이하로 제한하게 되며, 재구성된 결과가 최대 2.5 차원인 한계가 있다.

다. 영상기반렌더링(image-based rendering)

전통적인 모델링(modeling) 및 렌더링(rendering) 기술들에 대해서 현실적인 대안으로 제시되어 왔던 영상기반렌더링(image-based rendering)의 최대의 장점은 결과물의 현실성이다.[8-10] 전통적인 모델링 과정 없이, 영상으로부터 바로 결과영상을 합성해내는 이러한 방법은 상황의 기하적, 영상적 복잡도와 상관없이 일정 수행 시간으로 결과를 내기도 한다. 하지만 결과물의 깊이(depth)가 상대적으로 매우 적어 2 차원에 가깝다.

라. 시각체법(visual hull)

시각체(visual hull)는 실제 물체에 대해서 모든 시계(vi7ew)로부터 같은 외각선(silhouette)을 만들어 내는 최대의 모양(maximal shape)으로 정의된다.[10] 그 물체의 시계로부터 얻어지는 외각의 볼록체(convex hull)와 유사하나, 홀(hole)이 존재할 수 있는 차이가 있다. 참조시계(reference view)의 집합의 크기가 무한으로 갈수록, 추정 시각체(inferred visual hull)는 원래의 시각체(visual hull)에 근사하지만, 얻어지는 시각체(visual hull)는 오목한 부분 때문에, 원래의 물체와 동일한 결과를 만들 수는 없다.[11-12] [그림 1]

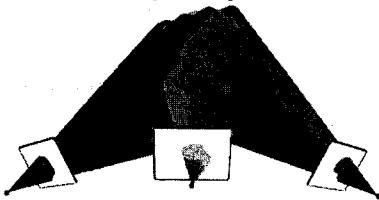


그림 1 : 시각체(visual hull)

3. 시스템 알고리즘(system algorithms)

가. 설계목표

초기 입체시각법(stereo vision)을 이용한 시스템 설계 및 검증을 통해, 최종 구현 시스템의 설계 목표를 다음과 같이 설정하였다. 첫째, 획득을 위하여 특수한 환경 변화를 요구하지 않아야 한다. 즉, 그리드(grid)등의 조작된 조명, 레이저(laser) 등의 능동적 장치를 전제로 하지 않아야 한다. 둘째, 처리를 위하여 특수한 하드웨어를 요구하지 않아야 한다. 즉, 다수의 DSP,

다수의 프로세서나 워크스테이션(workstation)을 이용한 분산환경(distributed computing) 등을 전제로 하지 않아야 한다. 셋째, 되도록 효율적으로 구현되어 영상 취득 후, 결과대기시간(latency)이 적어야 한다. 궁극적으로는 실시간(real-time)으로 처리 가능할 수 있어야 한다. 넷째, 되도록 실제에 가까운 재구성 결과 영상을 획득할 수 있어야 한다.

따라서, 본 연구에서는 상기 조건에 부합하도록 수동적 획득 방법(passive acquisition method) 중 시각체법(visual hull)을 토대로, 일반적인 단위공간(voxel space) 계산 알고리즘 대신, 광선들의 교차(intersection) 단위 알고리즘으로 계산복잡도를 대폭 개선한 시각체샘플링(visual hull sampling)을 설계하고, 결과 영상은 영상기반 렌더링(image-based rendering)을 이용하도록 하였다.

나. 시각체 샘플링(visual hull sampling)

시각체를 통한 물체 복원 방식은 각 카메라에서 얻은 외각선 영상에 기반하여, 투영중심(COP: Center Of Projection)과 외각선(silhouette) 영상으로 이루어지는 다면체(polyhedron)들을 각각의 카메라에 대해 계산한 후, 다면체들의 교집합을 구함으로써 물체의 시각체(visual hull)를 얻는다.[그림 1]

본 논문에서는 임의의 시점에 카메라가 있다고 가정하고 일정간격으로 카메라의 COP 에서 물체방향으로 광선(ray)을 쏘서 거리영상 통해 원하는 시각체 정보만을 얻는다. 즉 시각체 샘플링(visual hull sampling)으로 반드시 필요한 정보만 획득한다.[그림 2]

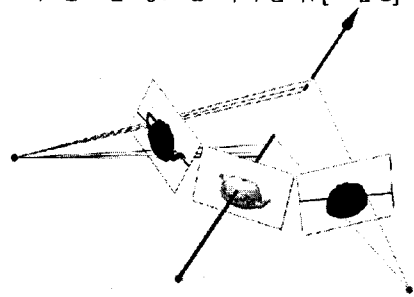


그림 2 : 시각체 샘플링

촬영에 사용된 각 카메라의 교정(calibration) 정보와 촬영 영상이 주어졌을 때, 가상 카메라에서의 거리 영상을 구하는 알고리즘은 다음과 같다. :

```

procedure VisualHullSampling(var camInfo : cameraInfo[], img:image[]);
var
    numCameras : integer; {카메라의 개수 = 동시 촬영된 영상의 수}
    pixelSize : integer; {영상의 픽셀의 총 개수}
    virRangelImage : IntervalSet[]; {가상 거리 영상 결과}
    silEdges : SilhouetteEdgeSet[]; {실루엣 에지 집합}
    i, j : integer; {인덱스}
begin
    for i:=1 to numCameras do {각 영상에 대해 외각선영상 구함}
        silEdges[i] := computeSilhouette(img[i]);
    end (for)
    for j:=1 to pixelSize do
        virRangelImage[j] := <[0, infinite]>; {가상 거리영상 초기화}
    end (for)
    for j:=1 to pixelSize do

```

```

for i:=1 to numCameras do
  Line3D ray := computeRay(camInfo[i], j); {광선 구하기}
  Line2D line := projectRay(ray, camInfo[i]); {2D 투영}
  IntervalSet set2 := Intersect(ray, silEdges[i]); {교점}
  IntervalSet set3 := unprojectLines(set2, ray, camInfo[i]);
  VirRangelImage[j].IntersectWith(set3); {교집합}
end (for)
end (for)
end {procedure}
    
```

가상 거리 영상의 각 픽셀에 대해 주어진 카메라 영상에 의해 물체가 있을 수 있는 구간을 결정하는 것은 5 단계의 과정을 거친다. 먼저 카메라의 교점 정보를 바탕으로 3 차원 광선을 구하고, 두번째로 그 광선을 해당하는 카메라의 영상 평면에 투영한다. 세번째 단계에서는 영상에서의 실루엣과 영상 평면에 투영된 직선의 교점을 구하고, 네번째 단계에서는 교점들로 이루어진 물체가 있을 수 있는 구간을 3 차원으로 되돌린다. 마지막으로 이렇게 하여 구해진 3 차원 상에서의 구간들과 현 픽셀에서의 가상 거리 영상의 구간들의 교집합을 구하게 된다.

다. 시각체 샘플링 알고리즘의 계산 복잡도 개선

카메라의 광선과 다른 영상평면 사이의 관계를 생각해 보면, 가상 카메라로부터 오는 여러 개의 3 차원 광선(ray)들이 다른 카메라로부터 얻어지는 영상에서 똑같은 2 차원 직선에 투영된다는 것을 알 수 있다. [그림 3]은 다른 영상평면에서 같은 2 차원 직선에 투영되는 가상 카메라 광선들의 예를 보여준다.

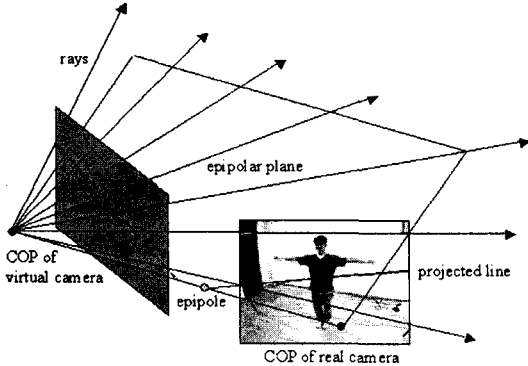


그림 3 : 가상카메라 광선들의 영상평면으로의 투영

따라서, 에피폴라 라인(epipolar line)과 외각선과의 교점 구간을 찾는 과정이 중복되어 발생되므로, 선분 캐싱(caching)을 사용할 수 있다. 먼저 캐시 인덱스(cache index)를 계산하여 캐시에 있는 교점 구간이 이미 계산되어 있는지를 검사하고, 이미 계산되어 있지 않은 경우에만 교점 구간을 새로 계산하여 캐시에 저장한다. 인덱스로는 [그림 4]와 같이 직선과 영상의 경계(boundary)의 교점 중에서 에피폴라(epipole)와 가장 멀리 있는 점을 사용한다. 구현시 영상의 경계선과의 2 개의

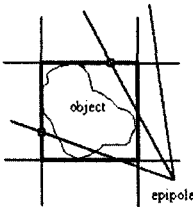


그림 4 : 캐시인덱스

교점과 에피폴라의 거리 계산은 부동 소수형 연산으로 복잡하기 때문에 영상 평면에서 에피폴라의 위치 정보를 이용하여 2 개의 교점 중에서 인덱스로 사용할 교점을 찾아내는 것이 계산비용을 줄일 수 있다.

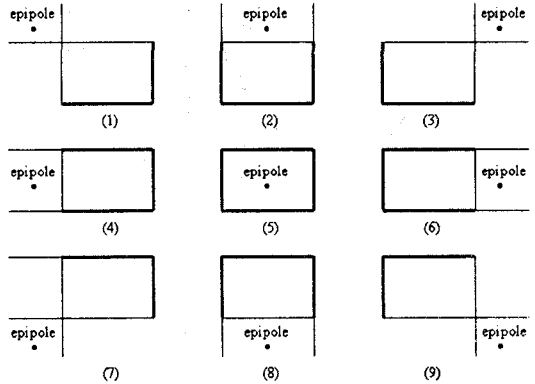
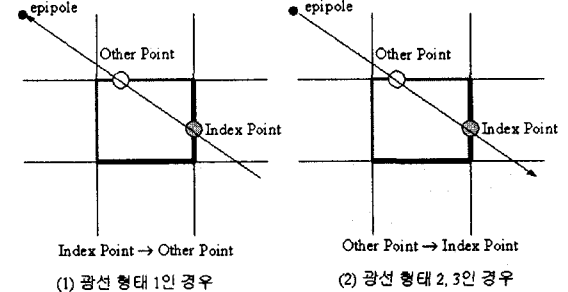


그림 5 : 영상평면에서 에피폴라(epipole)의 위치

에피폴라의 위치는 [그림 5]와 같이 9 가지 경우로 분류될 수 있는데, 굵은 선과 만나는 점을 캐시의 인덱스로 사용한다. (5)번의 경우에는 에피폴라가 영상의 내부에 있기 때문에 영상 경계선과의 교점 중에서 에피폴라의 거리가 작은 점을 인덱스로 사용한다.



Index Point → Other Point (1) 광선 형태 1인 경우
Other Point → Index Point (2) 광선 형태 2, 3인 경우

그림 6 : 에피폴라 위치 (1)인 경우 검색 방향 및 구간

영상 평면 내에서 2 차원 교점 구간을 빠르게 찾기 위해서는 직선을 따라가면서 교점을 찾는 과정에서 방향과 구간을 결정하여야 한다. 이는 에피폴라의 위치와 광선 형태로 결정된다. [그림 6]은 에피폴라 위치 (1)인 경우의 검색의 방향과 구간의 예를 보여준다. 에피폴라 위치 (1)일 때, 광선 형태가 1 번인 경우 인덱스 점에서 에피폴라를 향해 다른 점까지이고, 광선 형태가 2, 3 번인 경우는 다른 점에서 인덱스 점까지이다.

또한 속도 향상을 위하여 영상 경계선 대신 실루엣으로 추출된 부분의 직사각형 형태의 경계선을 설정하여 사용한다. 이렇게 할 경우 직선을 따라가면서 교점을 구하는 것의 범위도 줄고, 캐시 인덱스와 관계되어 있는 에피폴라의 위치도 [그림 5]의 (5)번의 경우가 적게 나오기 때문에 계산량이 줄어든다.

4. 시스템 구현 및 결과

구현된 시스템 구성은 [그림 7]과 같다.

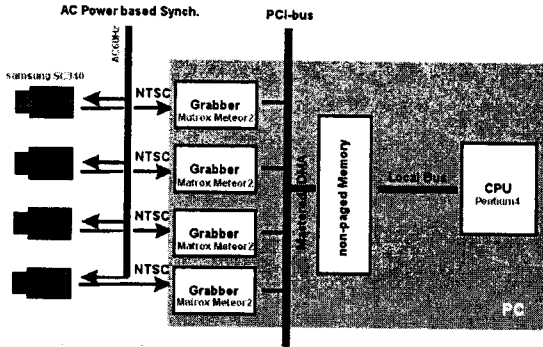


그림 7 : 영상획득 및 처리 시스템 구성도(4camera)

본 시스템은 컬러 NTSC 출력의 일반 CCD 카메라 및 5mm~15mm 렌즈 4 set 를 사용하며, AC 전원(60Hz)을 동기화원(synchronization source)으로 설계하였다. 단 하나의 일반 PC 에 4 개의 영상획득 카드를 non-paged memory 에 4 채널 master-dma 로 구현되었다. 실험환경은 창이 있는 일반 사무실이며, 반사인덱스(reflection index, r)을 이용하여 강건한 외각선 획득 알고리즘을 구현하여 사용하였다.[13]

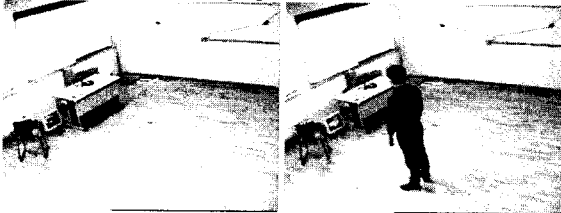


그림 8 : 배경영상

그림 9 : 입력영상

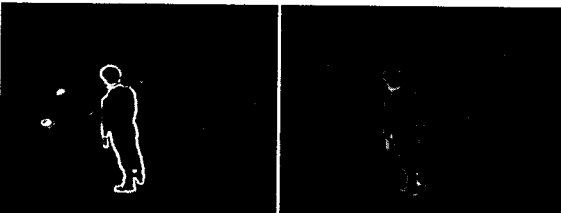


그림 10 : 외각선 추출영상 그림 11 : 구분(segment)영상

구현된 시스템은 4 개의 320x240 24 비트 컬러영상을 실시간 입력으로 가상 공간 영상을 초당 12 프레임으로 재구성한다. [그림 11] [그림 12]

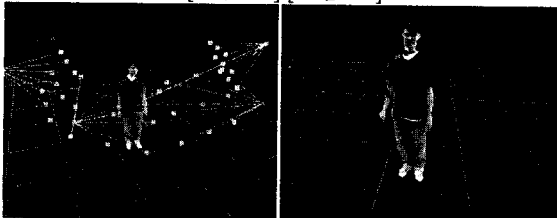


그림 12 : 시각체샘플링결과 그림 13 : 회전 및 확대 결과

5. 결론

본 논문에서는 수동 영상 재구성 알고리즘인 시각체법(visual hull)을 개선하여, 임의 시점의 거리영상을

광선 투영과 광선 교차 알고리즘을 이용한 시각체샘플링(visual hull sampling)과 캐싱(caching) 알고리즘, 가상의 카메라의 광선 방향과 가장 가까운 영상들의 색상정보를 사용하는 알고리즘을 구현하였다. 구현된 시스템은 다수의 실시간 영상입력으로 임의 가상 카메라 영상을 특별한 장치나 하드웨어 없이, 실시간에 가깝게 재구성한다.

본 시스템은 실시간 3 차원 동적 모델링 외에도, 향후 실시간 원격 가상존재(realtime remote virtual presence) 시스템 구성, 스포츠, 예술공연등의 이벤트에 대한 임의 뷰(view)에 대한 가상 3 차원 영상 복원 등 다양한 분야에 응용이 가능할 것이다.

참고문헌

- [1] G. Slabaugh, W. Bruce Culbertson, Thomas Malzbender, and Ron Shafer, "A Survey of Methods for Volumetric Scene Reconstruction from Photographs", International Workshop on Volume Graphics 2001, pp. 81-100, Stony Brook, New York, June 21-22, 2001.
- [2] K. N. Kutulakos and S. M. Seitz, "What Do N Photographs Tell Us about 3D Shape?", TR680, Computer Science Dept., Univ. of Rochester, January 1998.
- [3] W. B. Culbertson, T. Malzbender, and G. Slabaugh, "Generalized Voxel Coloring", Proceedings of the ICCV Workshop, Vision Algorithms Theory and Practice, Springer-Verlag Lecture Notes in Computer Science 1883, pp.100-115, September 1999.
- [4] P. Eisert, E. Steinbach, and B. Girod, "Multi-Hypothesis, Volumetric Reconstruction of 3-D Objects from Multiple Calibrated Camera Views", Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pp.3509-3512, 1999.
- [5] Y. Yang, A. Yuille, and J. Lu, "Local, Global, and Multilevel Stereo Matching", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 274-279, 1993.
- [6] Q. Chen and G. Medioni, "A Volumetric Stereo Matching Method: Application to Image-Based Modeling", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.29-34, June 1999.
- [7] S. E. Chen and L. Williams, "View Interpolation for Image Synthesis", pp. 279-288, SIGGRAPH 93.
- [8] P. Debevec, C. Taylor, and J. Malik, "Modeling and Rendering Architecture from Photographs", pp. 11-20, SIGGRAPH 96.
- [9] M. Levoy and P. Hanrahan, "Light Field Rendering", pp. 31-42, SIGGRAPH 96.
- [10] A. Laurentini, "The Visual Hull Concept for Silhouette-based Image Understanding", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.16, No. 2, February. 1994.
- [11] J. Koenderink, "Solid Shape", The MIT Press, 1990.
- [12] E. Boyer and M. Berger, "3D Surface Reconstruction Using Occluding Contours", IJCV 22, pp. 219-233, 1997.
- [13] Naoya Ohta, "A Statistical Approach to Background Subtraction for Surveillance Systems", Proceedings of ICCV, 2001.