



스테레오 비전(Stereo Vision)

2022학년도 제 1학기

<http://giplab.kyonggi.ac.kr>

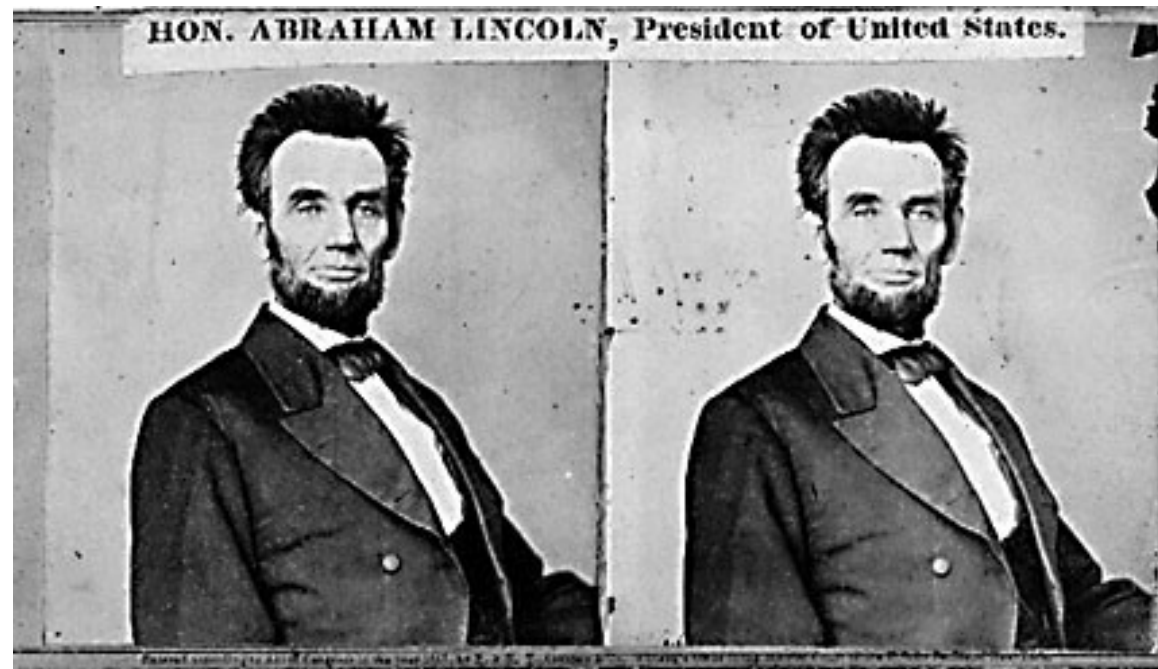


Introduction(1)

- ❖ Stereo matching 은 두 개 이상의 이미지가 주어졌을 때, 같은 장면(scene)이나 객체(object)를 계산하여 형상으로 표현하는 것
 - ❖ 2D의 위치를 3D의 깊이로 변환하여 두 픽셀을 매칭시켜 3D를 추정하는 과정
- ❖ 본 강좌에서는
 - ❖ 어떻게 더 완벽한 3D모델을 만드는지를 설명.
 - ❖ Stereo matching의 기초적인 원리를 설명.
 - ❖ 영상에서 epipolar line을 일치하도록 수정 (Rectification) 하는 방법을 설명.

Introduction(2)

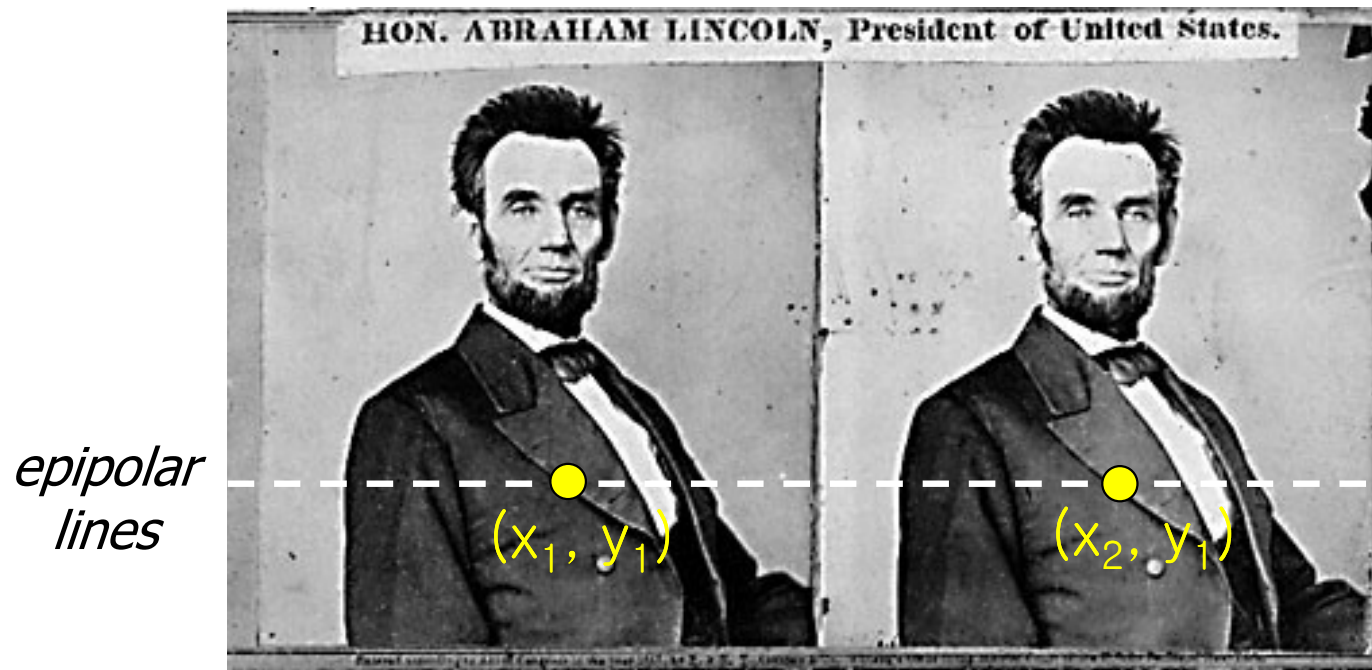
Stereo



- ❖ Given two images from different viewpoints
 - ❖ How can we compute the **depth of each point** in the image?
 - ❖ Based on *how much each pixel moves* between the two images

Introduction(3)

Epipolar geometry

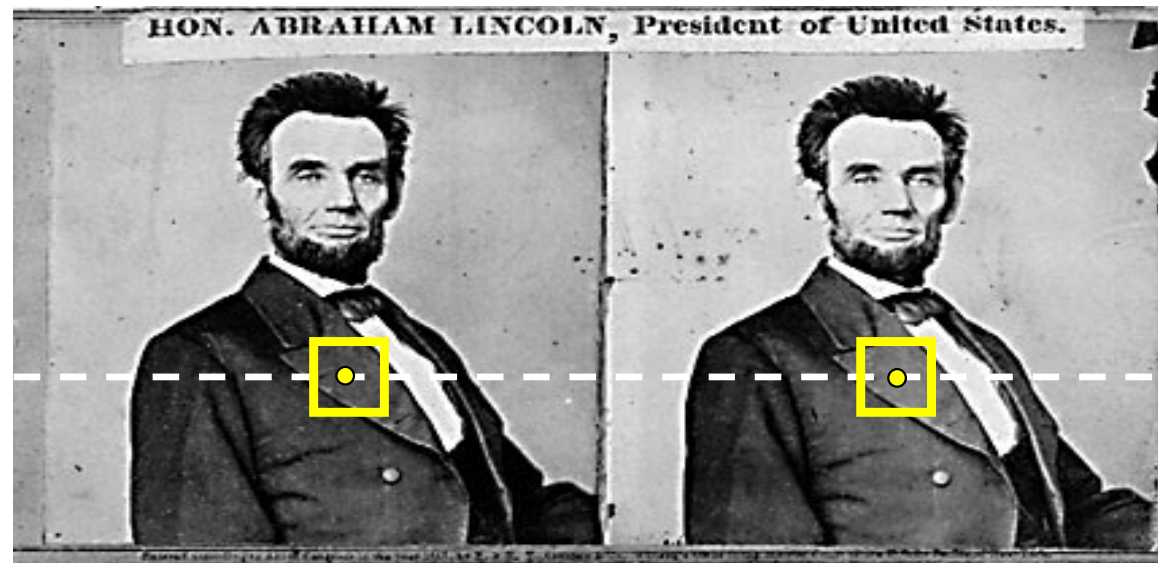


Two images captured by a purely horizontal translating camera
(*rectified stereo pair*)

$$x_2 - x_1 = \text{the } \textit{disparity} \text{ of pixel } (x_1, y_1)$$

Introduction(4)

Basic Stereo Algorithm



For each epipolar line

For each pixel in the left image

- compare with every pixel on same epipolar line in right image
- pick pixel with minimum match cost

Improvement: match *windows*

Introduction(5)

❖ (a)와 (b)로 (c) depth map를 만듦



(a)



(b)



(c)

❖ (d)와 (e)로 (f) 3D model를 만듦



(d)



(e)



(f)

Introduction(6)

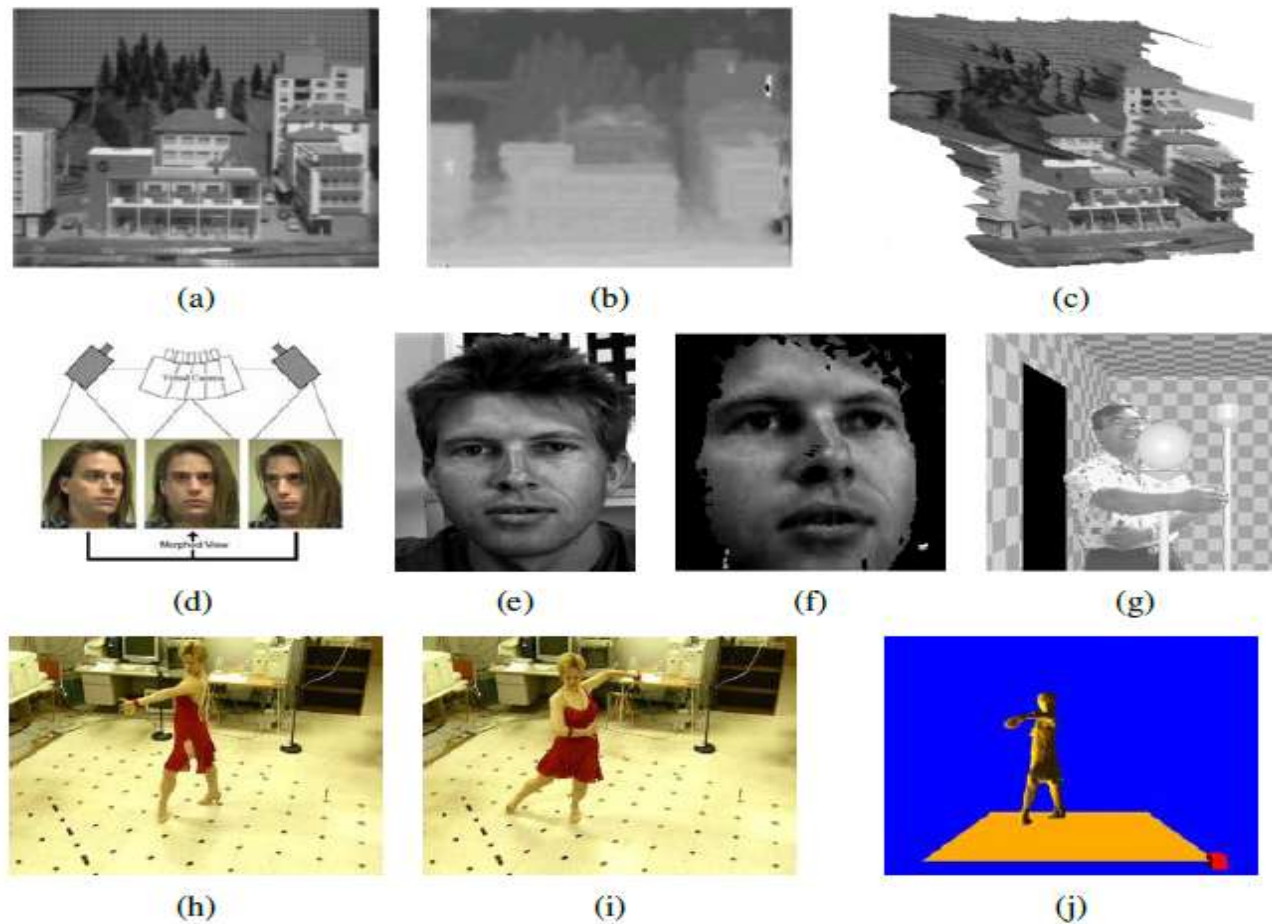
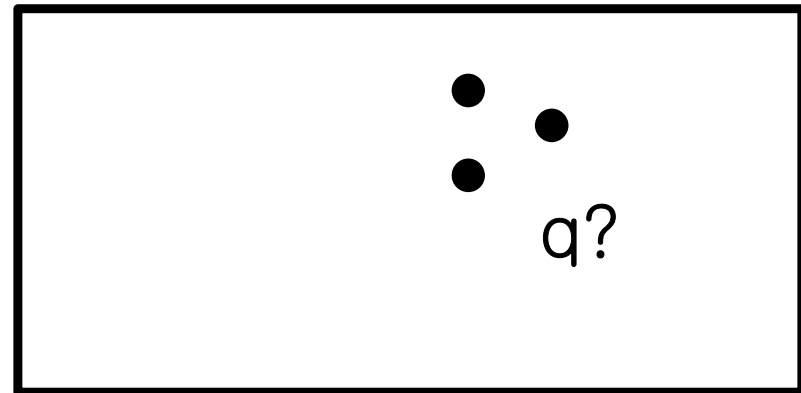
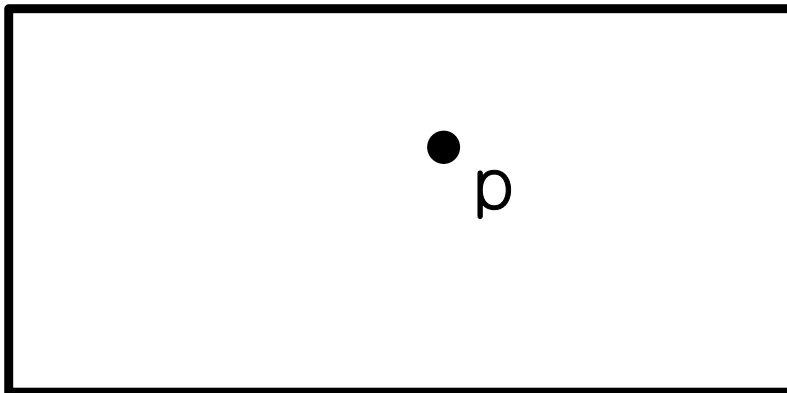


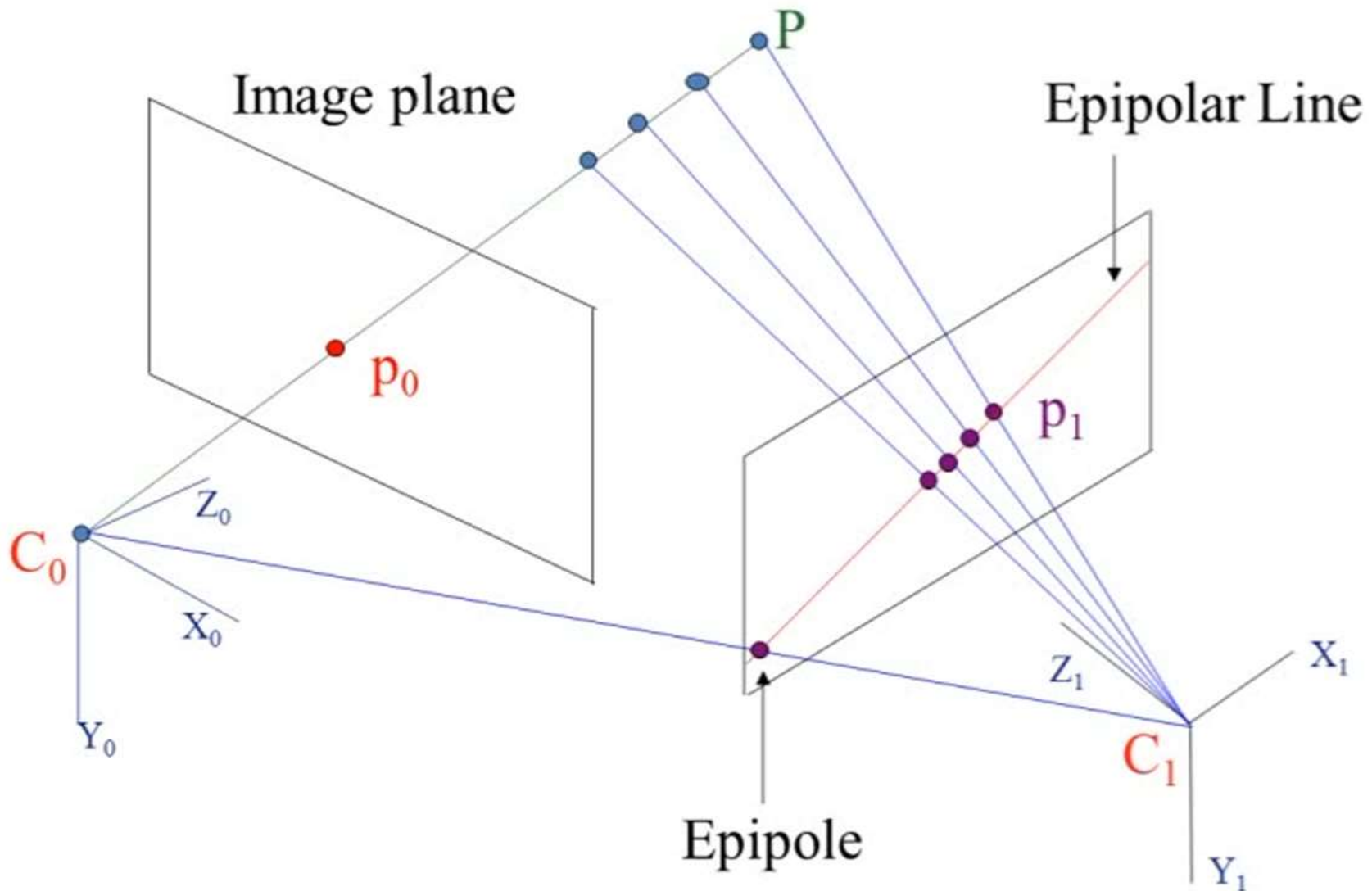
Figure 11.2 Applications of stereo vision: (a) input image, (b) computed depth map, and (c) new view generation from multi-view stereo (Matthies, Kanade, and Szeliski 1989) © 1989 Springer; (d) view morphing between two images (Seitz and Dyer 1996) © 1996 ACM; (e–f) 3D face modeling (images courtesy of Frédéric Devernay); (g) *z-keying* live and computer-generated imagery (Kanade, Yoshida, Oda *et al.* 1996) © 1996 IEEE; (h–j) building 3D surface models from multiple video streams in Virtualized Reality (Kanade, Rander, and Narayanan 1997).

1.1 Epipolar geometry(1)

- ❖ 서로 다른 view에서 같은 장면을 찍은 이미지가 있다.
- ❖ 한 영상에 p 라는 점이 주어졌을 때, 다른 영상에서 이에 대응하는 점 q 의 위치는 어디일까?



Epipolar geometry(2)



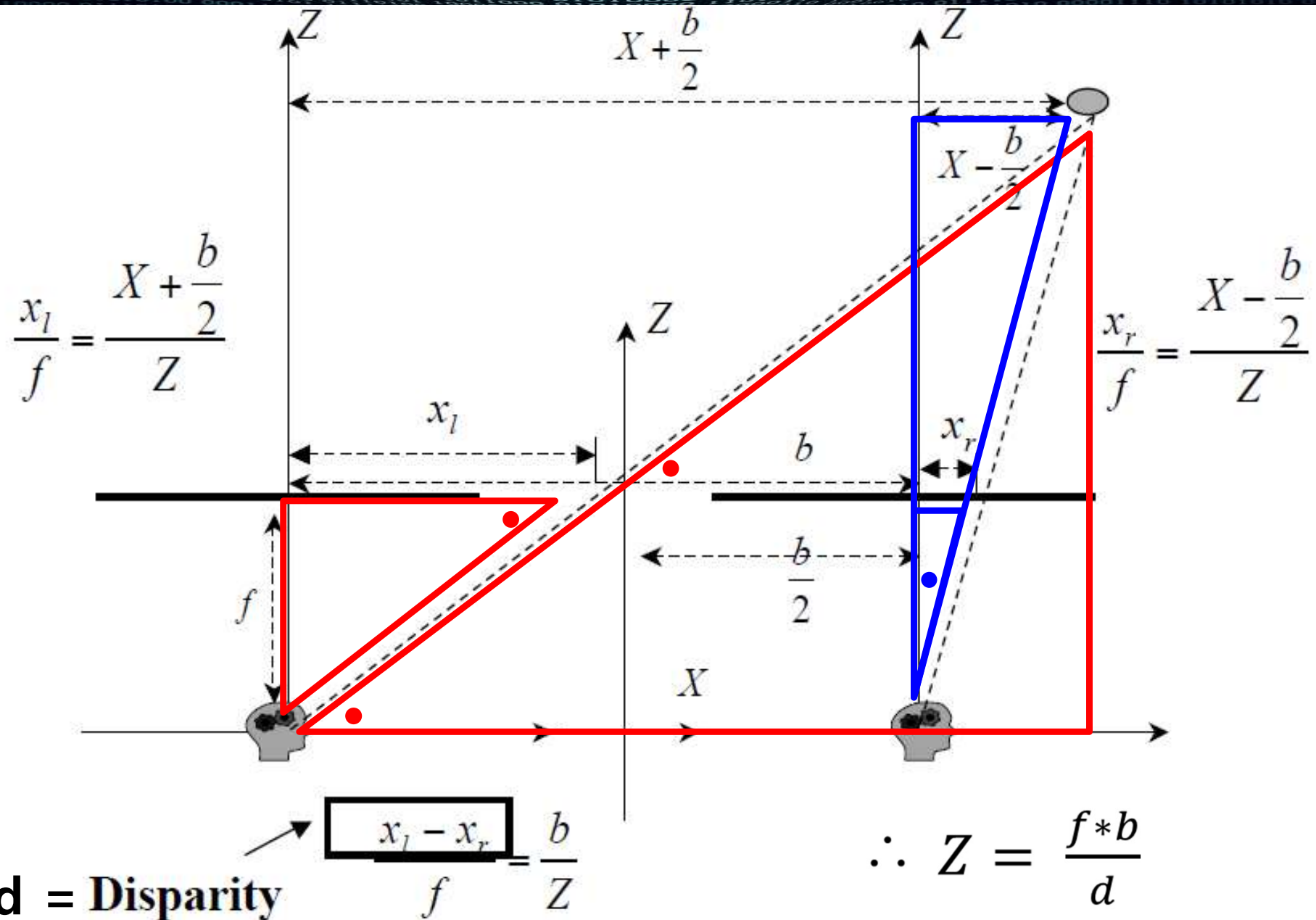
Epipolar geometry(3)

- ❖ Epipolar constraint
 - ❖ 한 영상에서 영상좌표 p_0 로부터 대응되는 다른 영상의 영상좌표 p_1 을 유일하게 결정할 수는 없지만, p_1 을 지나는 직선인 epipolar line은 유일하게 결정할 수 있다.
 - ❖ 단, 스케일까지는 결정할 수 없다.
- ❖ Triangulation : 삼각측량법
 - ❖ 두 영상 평면 사이의 기하학적 관계가 주어지고, 두 영상 평면상의 매칭 쌍, p, q 가 주어지면 이로부터 원래 공간의 3D공간 좌표 P 를 결정할 수 있다. (깊이(Depth)를 구할 때 쓰일 수 있음)

Triangulation(삼각측량법)을 사용한 Depth 구하기

- ❖ Triangulation 에서 주요한 두 가지 문제
 - ❖ 초점거리 f 와 기준선 b 를 알아야 함
 - ❖ 사전 지식이나 카메라 보정을 사용
 - ❖ (x_r, y_r) 에 대응하는 점 (x_l, y_l) 을 찾아야 함
⇒ Correspondence problem

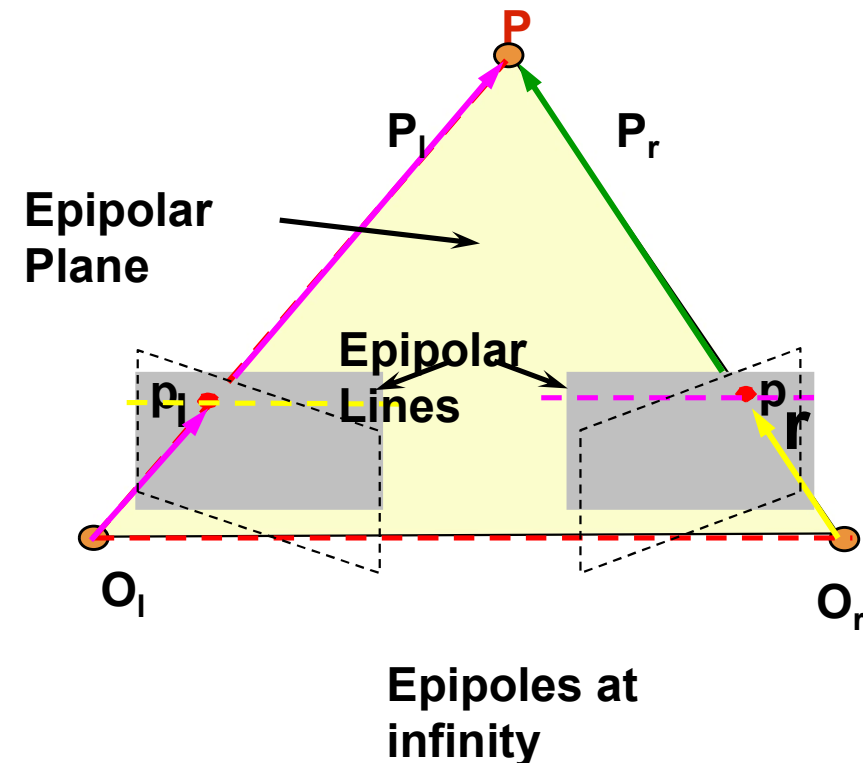
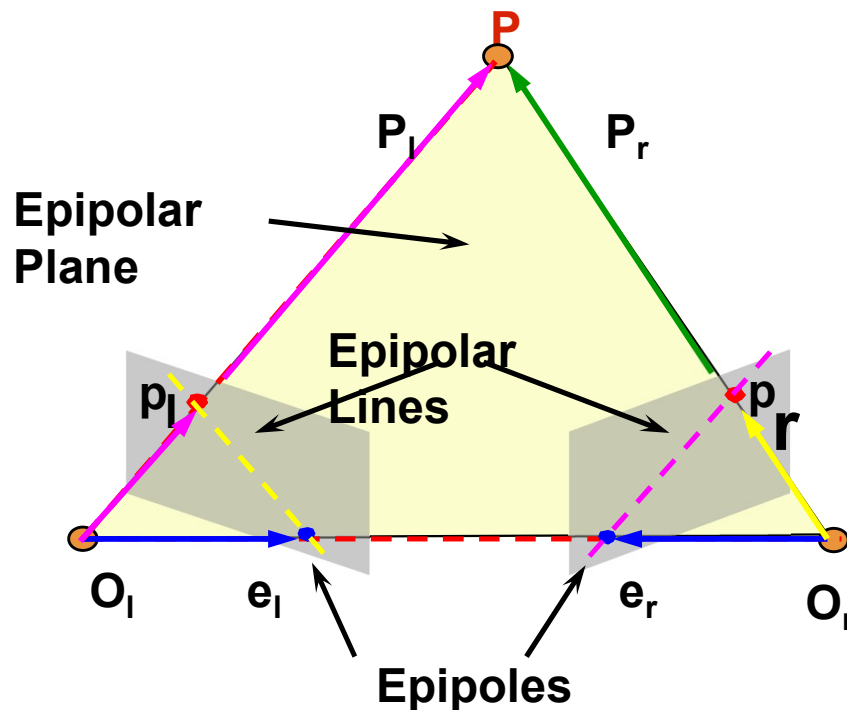
Triangulation(삼각측량법)을 사용한 Depth 구하기



1.1.1 Rectification(1)

- ❖ Epipolar line 을 평행으로 맞춰주는 것
- ❖ 이를 통해 두 이미지의 대응점이 같은 라인(epipolar line)에 위치하게 됨

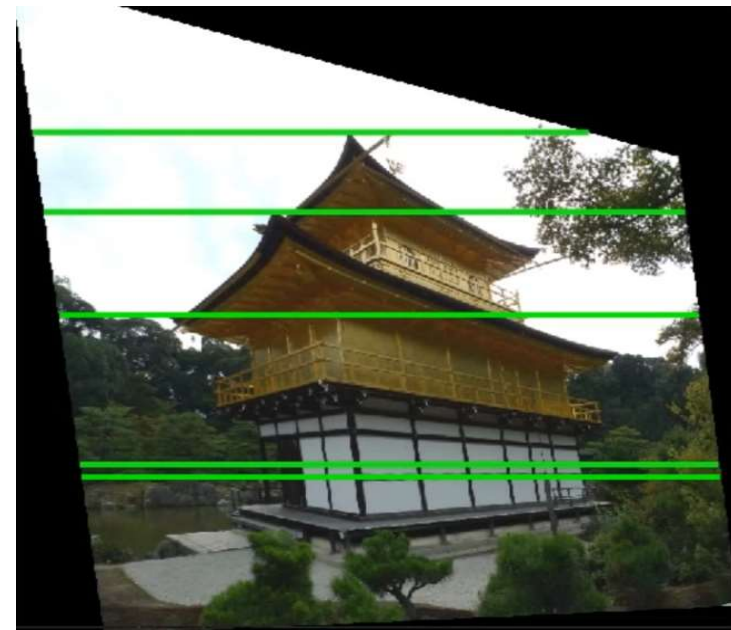
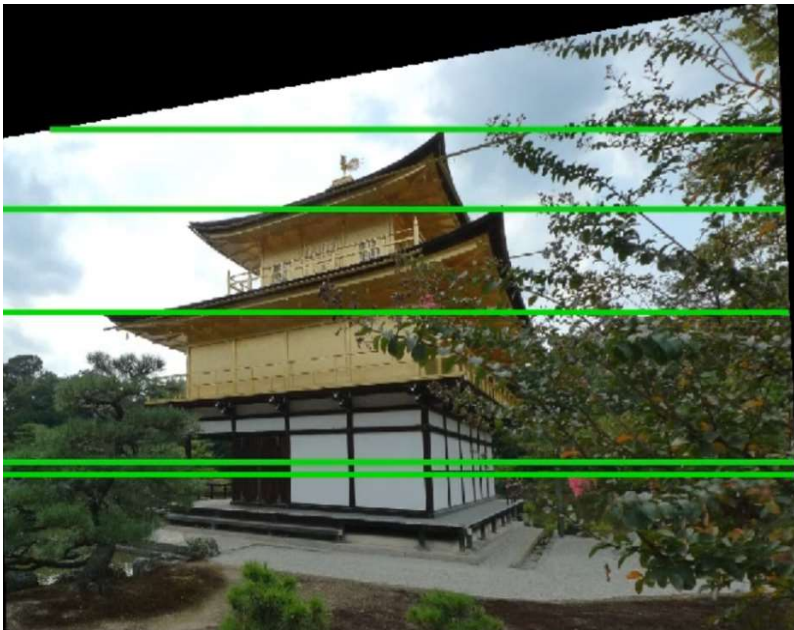
Rectified Images



Rectification(2)

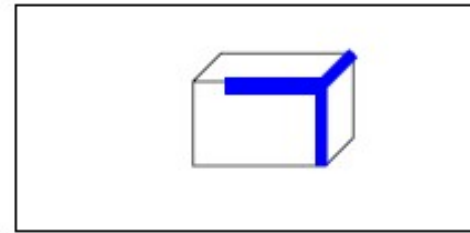
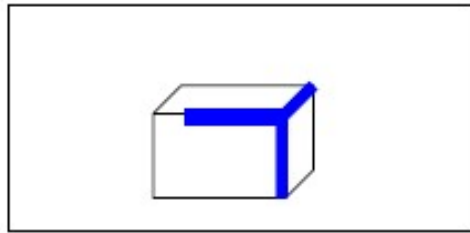


Rectification(3)

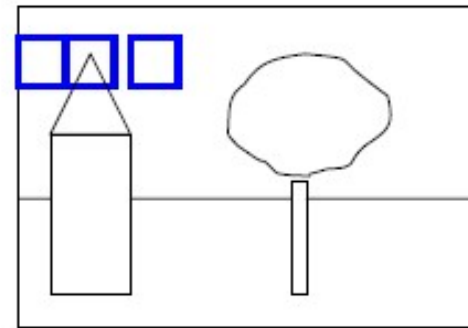
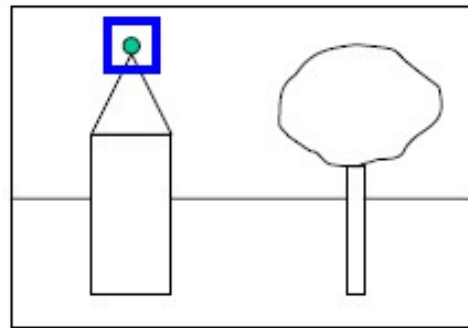


1.2 Stereo Correspondence

- ❖ Sparse correspondence
 - ❖ 특정 부분이나 코너부분 비교

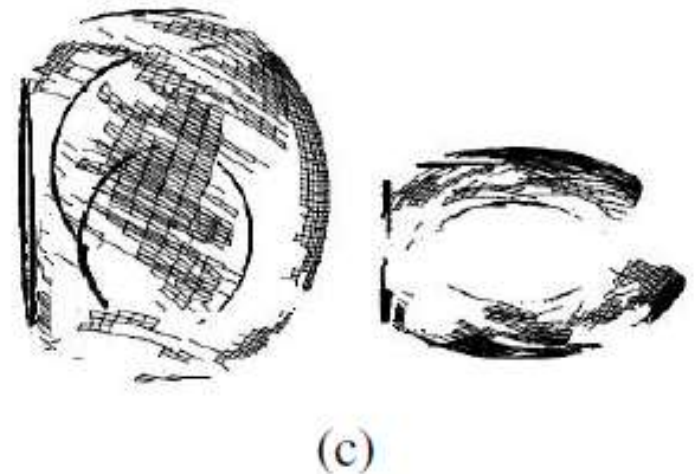
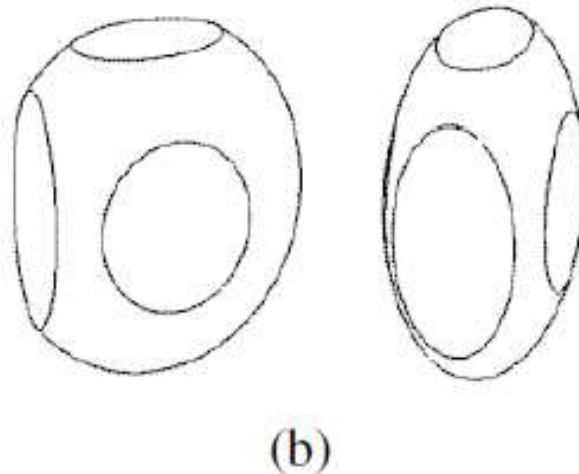
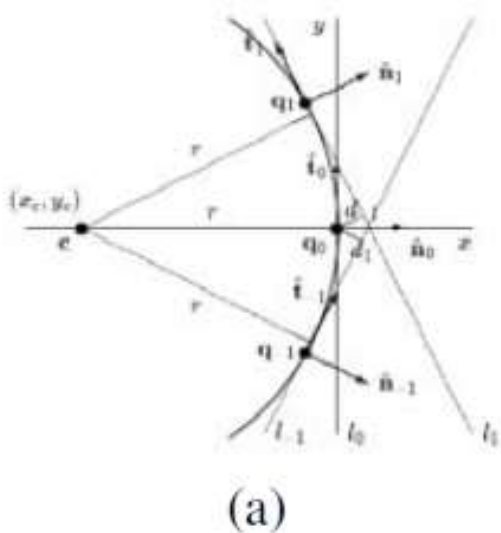


- ❖ Dense correspondence
 - ❖ 작은 부분(윈도우) 상호 상관도 또는 SSD 사용



1.2 Sparse Correspondence(희소 대응)

- ❖ 3D curves and profiles(3d 곡선과 옆모습)
 - ❖ (a) : epipolar plane에 맞는 원호
 - ❖ (b) : 절단된 면과 타원체의 합성 예
 - ❖ (c) : 경사와 평면도로 부터 부분적으로 재구성된 메쉬



Sparse correspondence(희소 대응)

- ❖ (d) : 실제 캔 이미지
- ❖ (e) : 엣지 추출
- ❖ (f) : 부분적으로 복원된 옆모습 곡선
- ❖ (g) : 부분적으로 재구성된 표면 메쉬



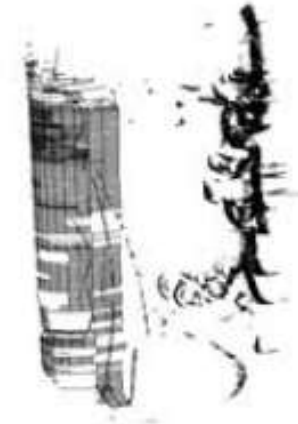
(d)



(e)



(f)



(g)

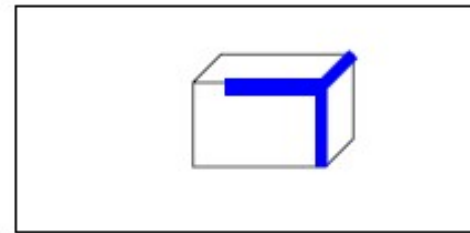
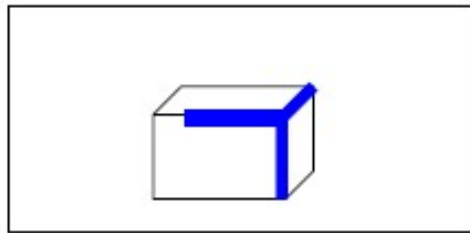
1.3 Dense correspondence(조밀한 대응)

- ❖ 조밀한 대응 알고리즘
 - ❖ 일치하는 부분의 비용을 계산
 - ❖ 비용을 통합
 - ❖ 차이를 계산하고 최적화
 - ❖ 차이를 정제
- ❖ Sum-of-Squared-Differences(SSD) algorithm를 사용하여 물체를 분리 할 수 있다.
 - ❖ Local algorithms : 화소 근처의 정보만 이용
 - ❖ Global algorithms : 영상 전체에 대하여 연산

Stereo Correspondence

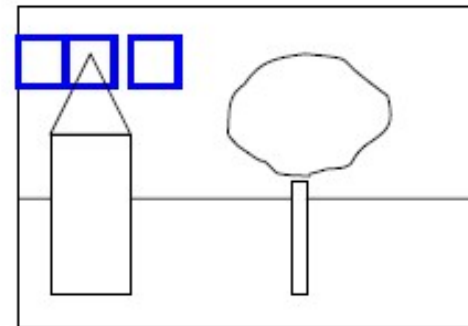
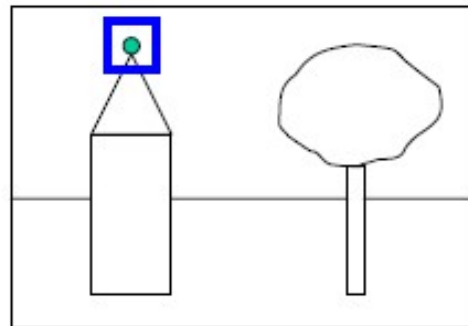
❖ Sparse correspondence

- ❖ 특정 부분이나 코너부분 비교



❖ Dense correspondence

- ❖ 작은 부분(윈도우) 상호 상관도 또는 SSD 사용



Stereo matching 알고리즘

- ❖ Local method
 - ❖ Block matching
 - ❖ Feature matching
- ❖ Global method
 - ❖ Graph-cut
 - ❖ Energy Minimization

1.4 Local method(1)

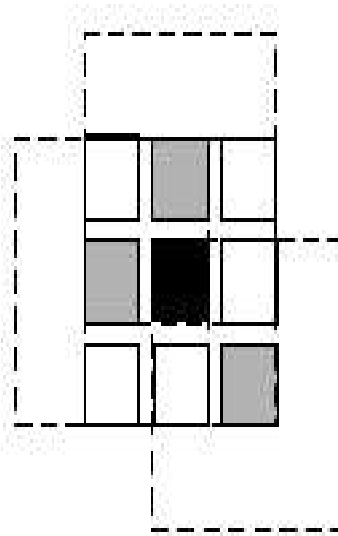


Figure 11.8 Shiftable window (Scharstein and Szeliski 2002) © 2002 Springer. The effect of trying all 3×3 shifted windows around the black pixel is the same as taking the minimum matching score across all *centered* (non-shifted) windows in the same neighborhood. (For clarity, only three of the neighboring shifted windows are shown here.)

Local method(2)

- ❖ (a) : 평가점들이 선택된 원본이미지
- ❖ (b) : windows 변수
- ❖ (c) : 적응적 가중치
- ❖ (d) : 분할 기반

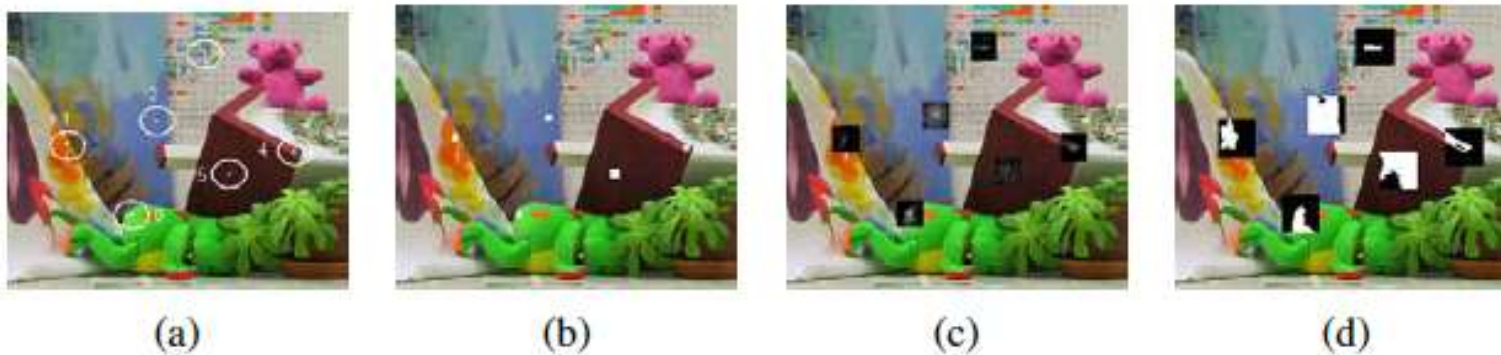


Figure 11.9 Aggregation window sizes and weights adapted to image content (Tombari, Mattoccia, Di Stefano *et al.* 2008) © 2008 IEEE: (a) original image with selected evaluation points; (b) variable windows (Veksler 2003); (c) adaptive weights (Yoon and Kweon 2006); (d) segmentation-based (Tombari, Mattoccia, and Di Stefano 2007). Notice how the adaptive weights and segmentation-based techniques adapt their support to similarly colored pixels.

Local method(3)

- ❖ Sub-pixel estimation and uncertainty (서브픽셀 추정 및 불확실성)
 - ❖ (a) : 입력 이미지
 - ❖ (b) : 추정된 깊이맵(파란색이 가까움)
 - ❖ (c) : 추정된 신뢰도(빨간색이 높음), textured areas 높은 신뢰

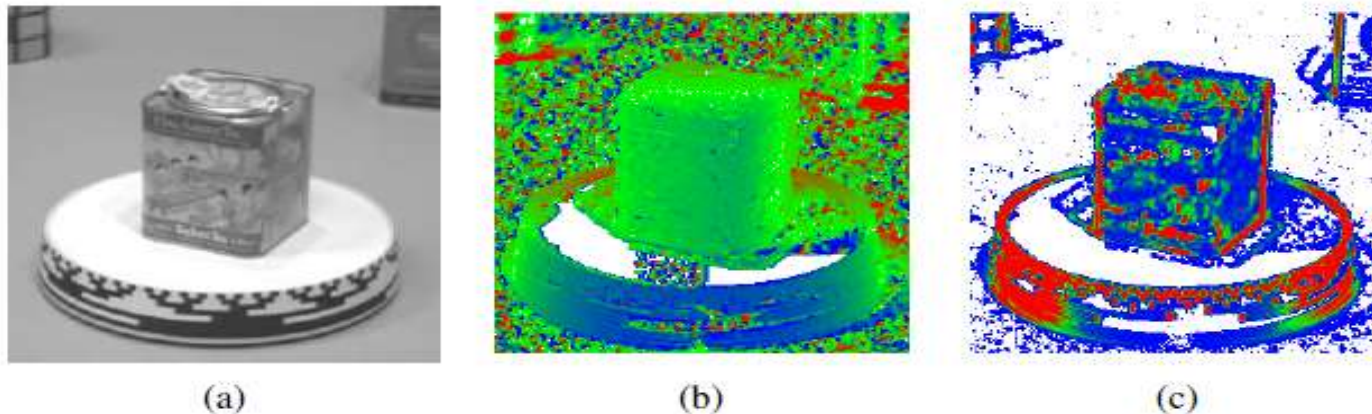


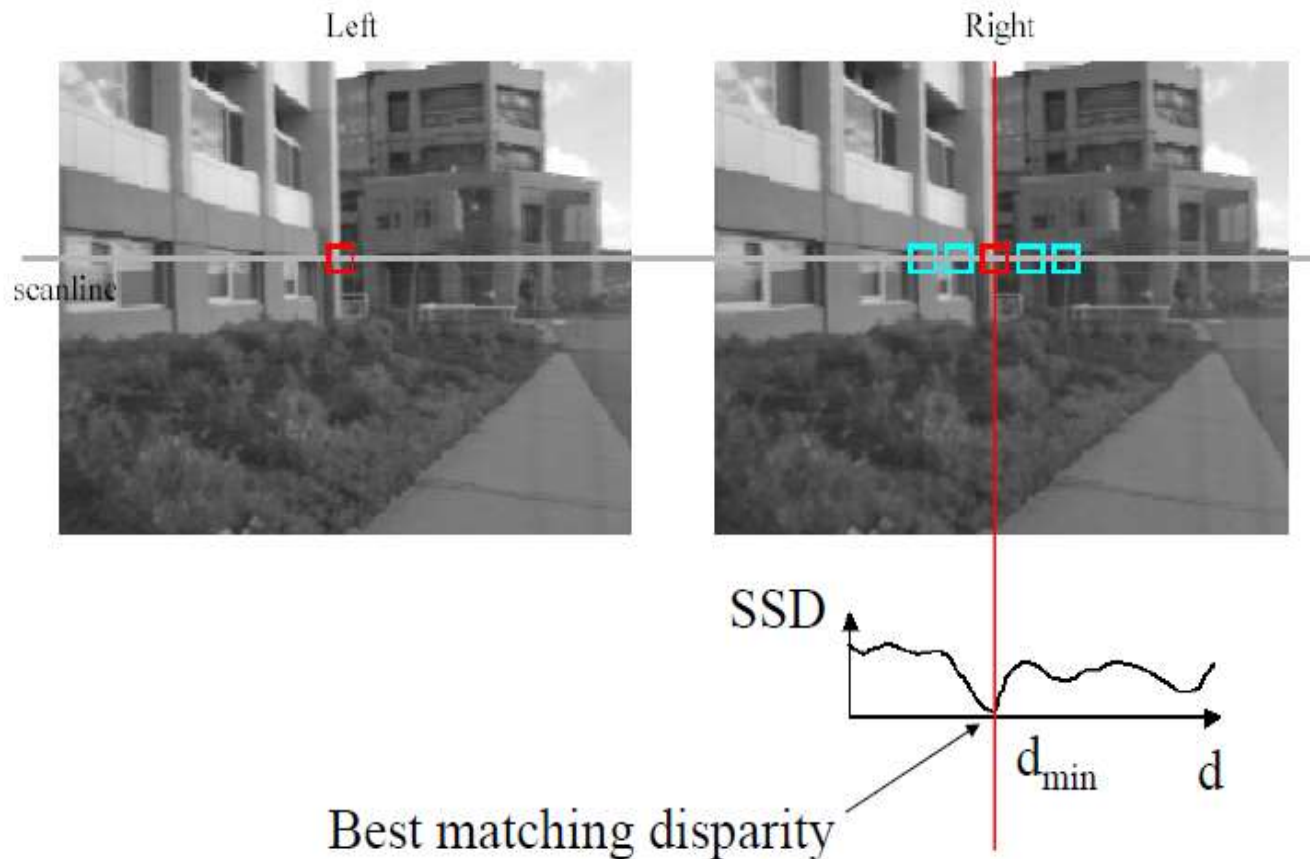
Figure 11.10 Uncertainty in stereo depth estimation (Szeliski 1991b): (a) input image; (b) estimated depth map (blue is closer); (c) estimated confidence (red is higher). As you can see, more textured areas have higher confidence.

Local method(4)

- ❖ Application : Stereo-based head tracking
 - ❖ 실시간 스테레오 알고리즘은 컴퓨터나 게임 시스템의 상호 작용하여 사용자의 위치를 추적하기 위한 것이다.
 - ❖ 초기 : 머리추적장치 및 스테레오 안경 사용, 현재는 대체됨
 - ❖ 헤드 트래킹은 실시간으로 변형 될 수 있어, "가상 거울"을 구성하는데 사용 될 수 있다.
 - ❖ 이 외에도 시선 보정에 사용됨
ex) 화상회의에서 일반적으로 모니터 상단에 카메라가 배치되므로 사용자의 시선이 아래로 떨어져 다른 참가자를 찾고 있는 것처럼 나타남.

Local method(5)

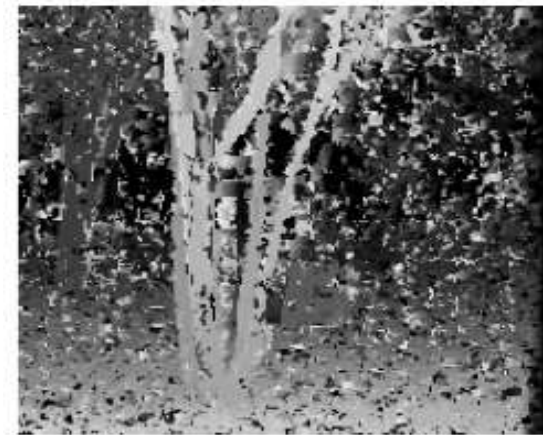
- ❖ SSD(Sum of Squared Differences)
 - ❖ 제곱 차의 합을 이용함



❖ Problems with window size

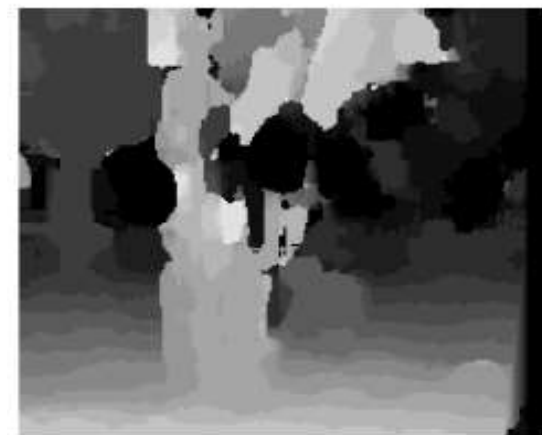


Input stereo pair



$W = 3$

- ❖ Window size 작으면...
 - ❖ 좋은 정밀도, 자세하다.
 - ❖ 노이즈에 민감하다
- ❖ Window size 크면...
 - ❖ 노이즈에 강하다
 - ❖ 축약적이고 자세하지 못하다.



$W = 20$

1.5 Global optimization(1)

- ❖ Cooperative algorithms
 - ❖ 인간의 스테레오 비전의 계산 모델에서 영감
 - ❖ 전역 최적화 알고리즘과 유사
 - ❖ 반복적으로 비선형 연산을 사용하여 변이 추정을 업데이트
- ❖ Coarse-to-fine and incremental warping
 - ❖ 대부분의 알고리즘은 모든 경우의 수를 열거하고 그 중 최고를 선택
 - ❖ 때때로, 빠른 접근을 위해 (infinitesimal) optical flow computation 사용
 - ❖ 이미지는 연속적으로 변형하고 만족할 때까지 점진적으로 갱신 추정

Global optimization(2)

- ❖ Dynamic programming
 - ❖ Global optimization 알고리즘은 동적 프로그래밍을 기반
 - ❖ 처음에는 스테레오 기반에서 스테레오 비전을 위해 사용됨
 - ❖ 최근에는 스캔라인 매칭 문제에서 사용됨

Global optimization(3)

❖ Segmentation-based techniques

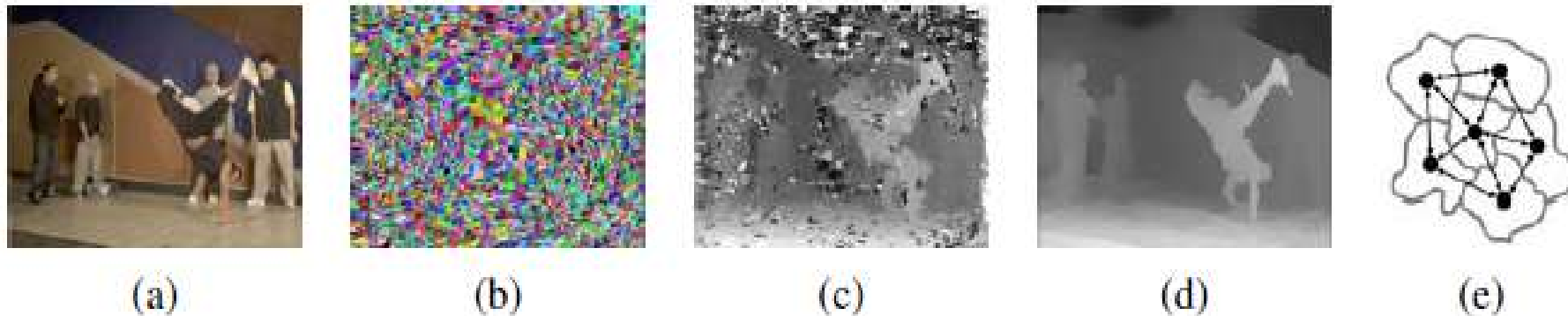


Figure 11.12 Segmentation-based stereo matching (Zitnick, Kang, Uyttendaele *et al.* 2004)
© 2004 ACM: (a) input color image; (b) color-based segmentation; (c) initial disparity estimates; (d) final piecewise-smoothed disparities; (e) MRF neighborhood defined over the segments in the disparity space distribution (Zitnick and Kang 2007) © 2007 Springer.

Global optimization(4)

❖ Segmentation-based techniques



Figure 11.13 Stereo matching with adaptive over-segmentation and matting (Taguchi, Wilburn, and Zitnick 2008) © 2008 IEEE: (a) segment boundaries are refined during the optimization, leading to more accurate results (e.g., the thin green leaf in the bottom row); (b) alpha mattes are extracted at segment boundaries, which leads to visually better compositing results (middle column).

Global optimization(5)

- ❖ Application: Z-keying and background replacement
- ❖ 실시간 스테레오 매칭 어플리케이션
- ❖ 과거 스튜디오에서 제한
- ❖ 현재는 일반 PC에서 화상 회의 어플리케이션을 통해 사용 가능하도록 만들고 있다.



(g)

Global method(1)

❖ Energy-Minimization

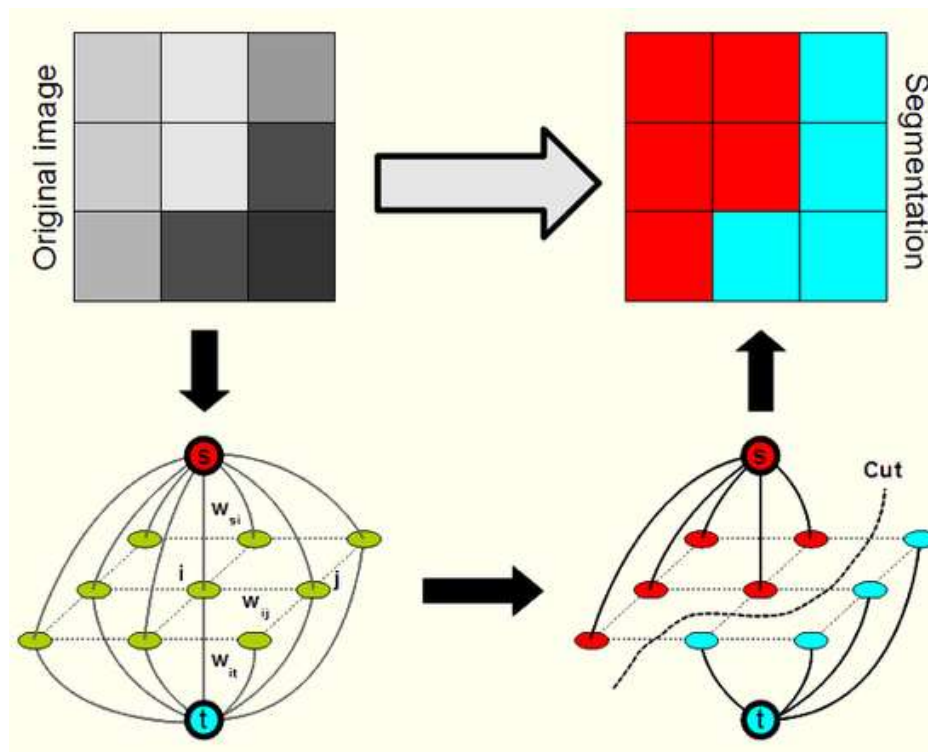
❖ 일반적인 2D문제에 대한 하

❖ Graph-Cut

$$E_{\text{total}}(\mathbf{d}) = E_{\text{data}}(\mathbf{d}) + \lambda E_{\text{smoothness}}(\mathbf{d})$$

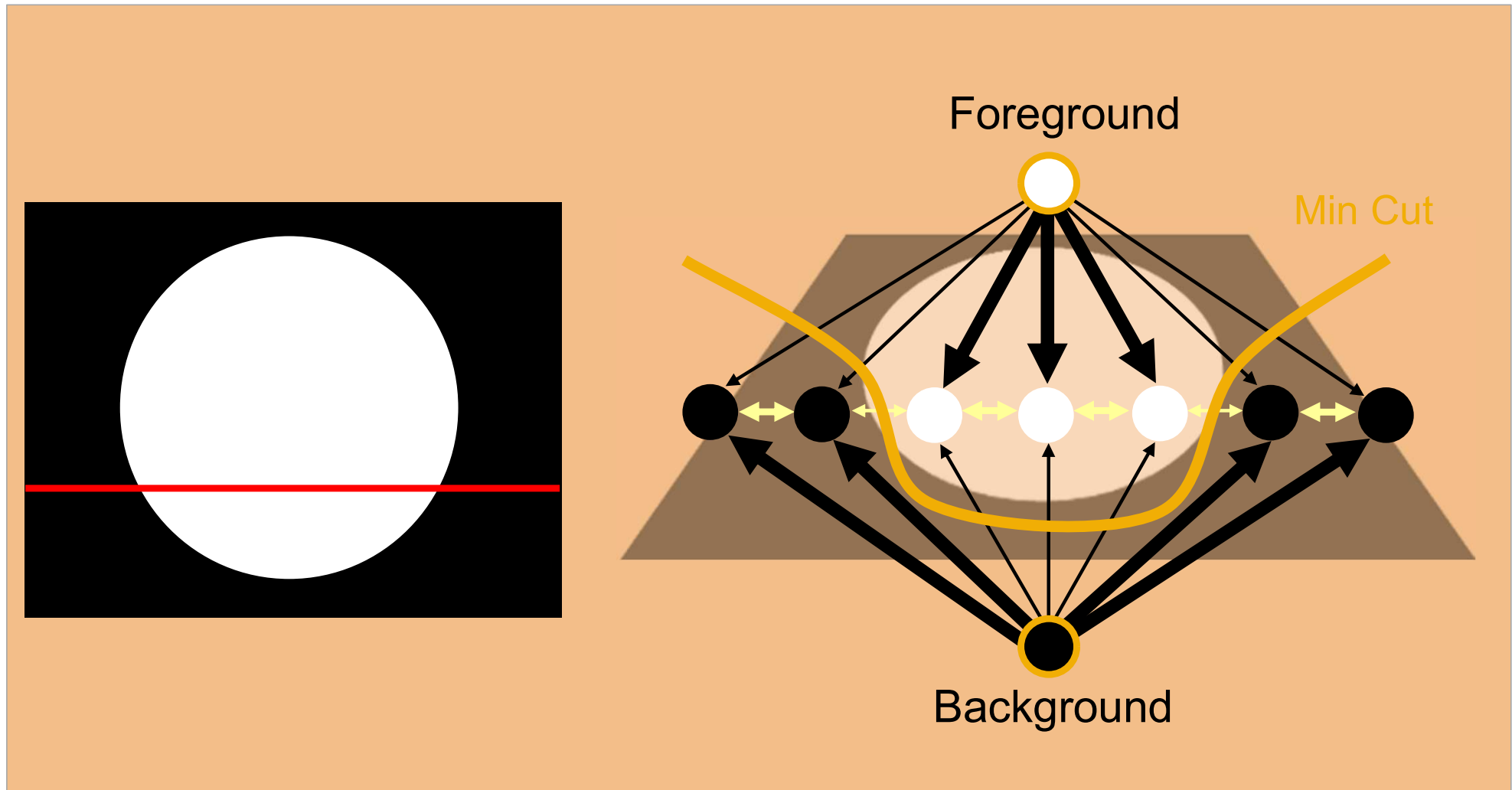
$$E_{\text{data}}(\mathbf{d}) = \sum_{x,y} f_{x,y}(d_{x,y})$$

$$E_{\text{smoothness}}(\mathbf{d}) = \sum_{x,y} \rho(d_{x,y} - d_{x-1,y}) + \sum_{x,y} \rho(d_{x,y} - d_{x,y-1})$$



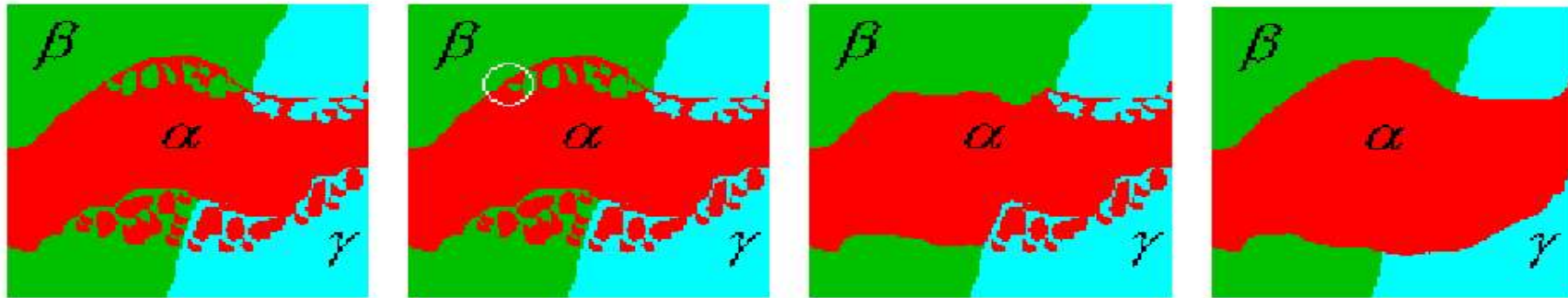
Global method(2)

❖ Graph-Cut



Global method(3)

❖ Graph-Cut based Segmentation



(a) initial labeling

(b) standard move

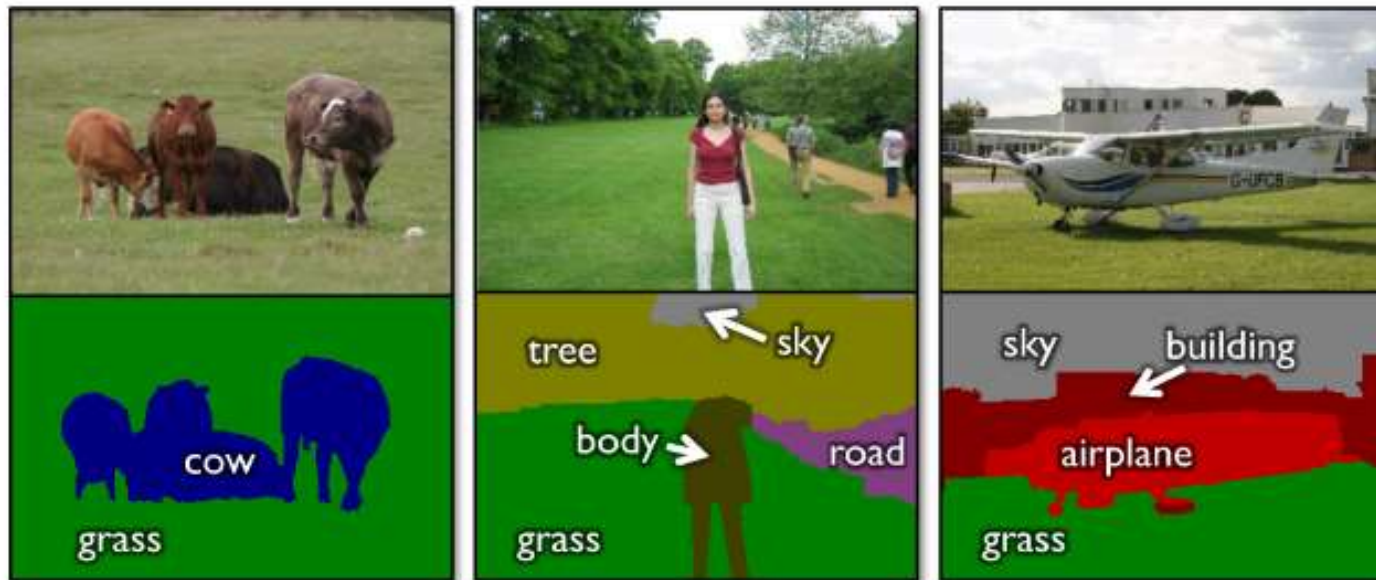
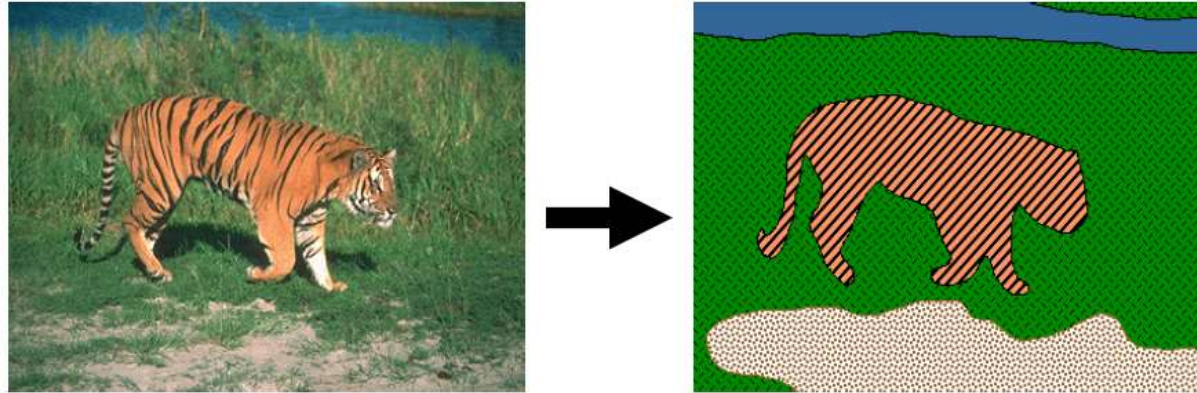
(c) α - β -swap

(d) α -expansion

- ❖ α - β -swap : α , β 의 일부 픽셀을 더 유사한 β , α 로 바꾸는 과정
- ❖ α -expansion : 표시되지 않은 부분을 α 의 서브셋으로

Global method(4)

❖ Graph-Cut



object classes	building	grass	tree	cow	sheep	sky	airplane	water	face	car
bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat

Local method vs Global method

- ❖ Local method
 - ❖ 계산 속도가 상대적으로 빠르다.
 - ❖ 하지만 디테일이 떨어진다.
- ❖ Global method
 - ❖ 계산 속도가 상대적으로 느리다.
 - ❖ 구하는 속도가 오래 걸린다.
 - ❖ 따라서 Dynamic Programming 를 이용
 - ❖ Local method보다 디테일하다.

Multi-view stereo(1)

❖ Z-keying 를 사용한 Background replacement



Multi-view stereo(2)

❖ Volumetric and 3D surface reconstruction

❖ 멀티 뷰 스테레오의 목적은 카메라의 관점에서 촬영한 이미지로부터 완벽한 3D 객체 모델을 재구성하는 것이다.

❖ (a) : 댄서 주변을 멀티카메라로 찍고, 계산하여 나타냄

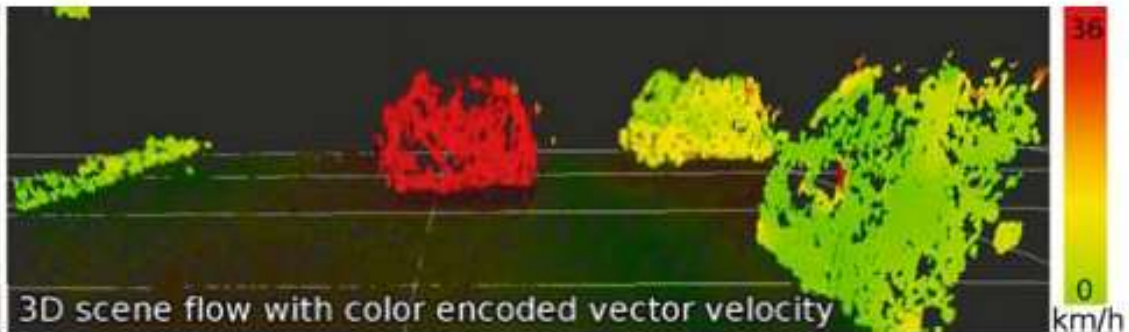
❖ (b) : 스테레오 카메라로부터 얻어지는 차트 찍고 계



(a)



Left camera image



3D scene flow with color encoded vector velocity

(b)

Multi-view stereo(3)

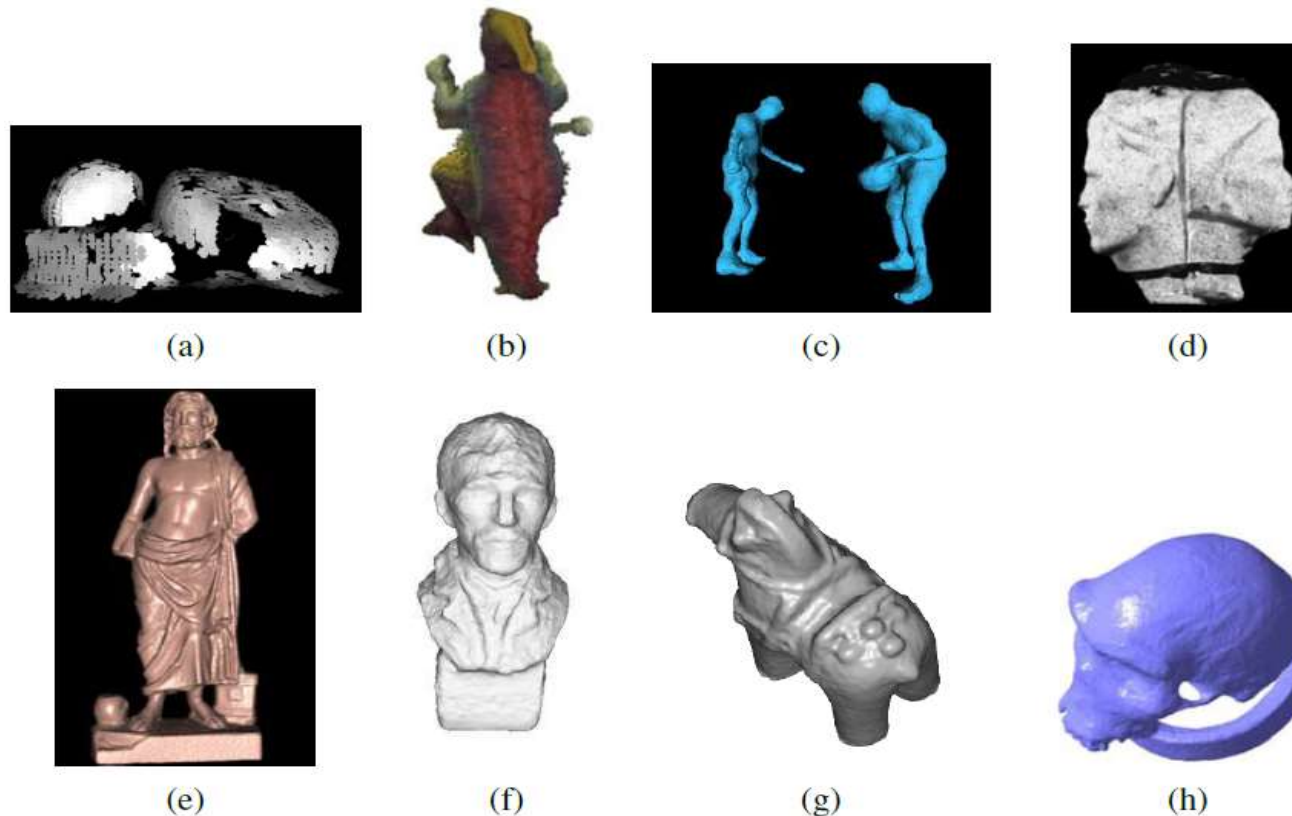


Figure 11.19 Multi-view stereo algorithms: (a) surface-based stereo (Fua and Leclerc 1995); (b) voxel coloring (Seitz and Dyer 1999) © 1999 Springer; (c) depth map merging (Narayanan, Rander, and Kanade 1998); (d) level set evolution (Faugeras and Keriven 1998) © 1998 IEEE; (e) silhouette and stereo fusion (Hernandez and Schmitt 2004) © 2004 Elsevier; (f) multi-view image matching (Pons, Keriven, and Faugeras 2005) © 2005 IEEE; (g) volumetric graph cut (Vogiatzis, Torr, and Cipolla 2005) © 2005 IEEE; (h) carved visual hulls (Furukawa and Ponce 2009) © 2009 Springer.

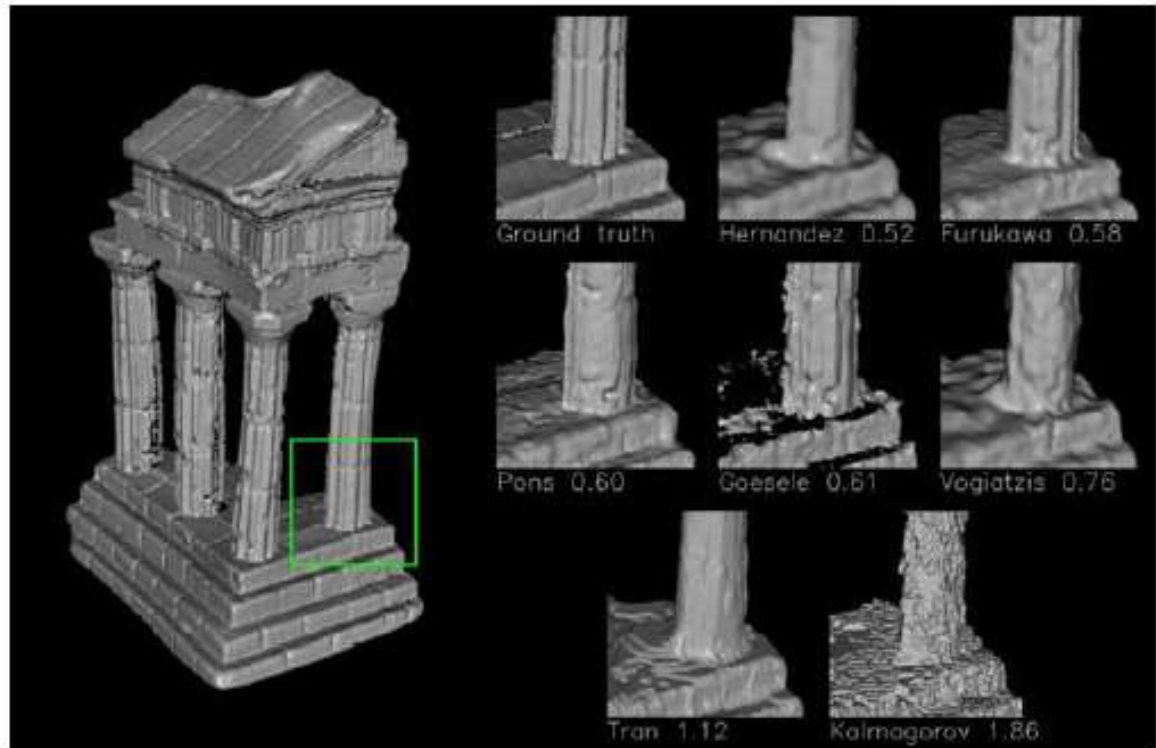
Multi-view stereo(4)

❖ (b) voxel coloring 방법



Multi-view stereo(5)

- ❖ Shape from silhouettes(실루엣에서 모양)
 - ❖ 대부분 전경과 배경 분할을 하면, 초기화하거나 정확한 3D 모델 또는 멀티 뷰 스테레오를 할 수 있는 좋은 방법이다.

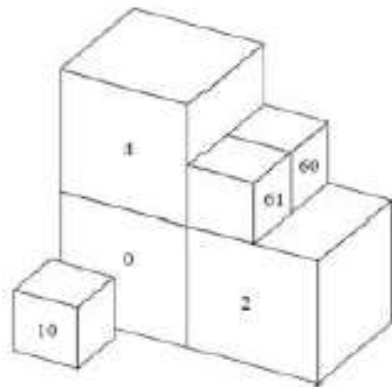


7가지 평가 알고리즘에 대한 결과 (숫자는 각각 기법들의 정확도)

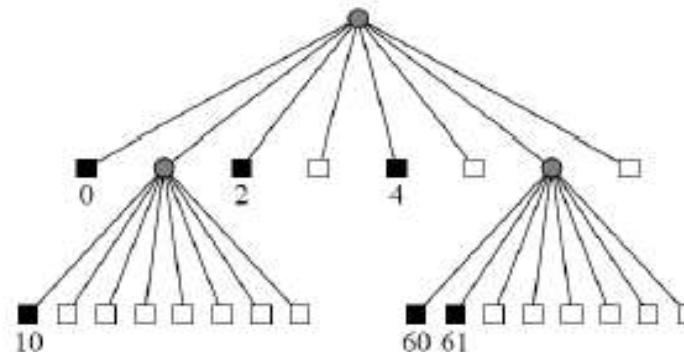
Multi-view stereo(6)

❖ Shape from silhouettes(실루엣에서 모양)

❖ 실루엣에서 다양한 복원



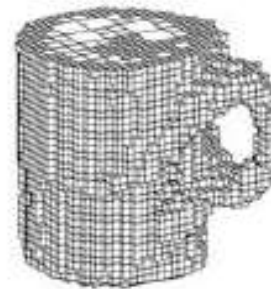
(a)



(b)



(c)



(d)

(a) : 팔진트리 표현

(b) : 팔진트리구조

(c) : 턴테이블 객체
입력 이미지

(d) : 3차원
Volumetric
octree모델

Multi-view stereo 보충자료

- ❖ 평가 알고리즘 사이트

- ❖ <http://vision.middlebury.edu/mview/eval/>

- ❖ Multi-view stereo

- ❖ <http://www.youtube.com/watch?v=oyiaAYrFrLM>