Brief Introduction to Geometry and Vision

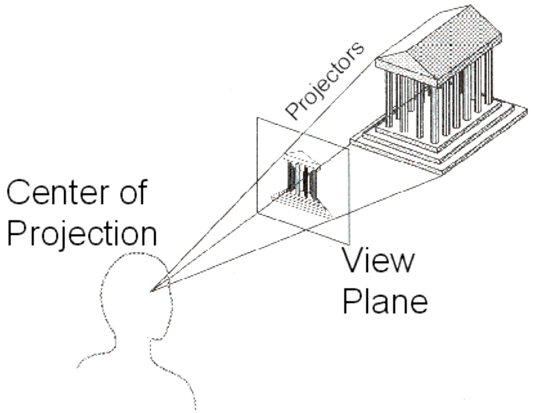
A.L. Yuille (UCLA)

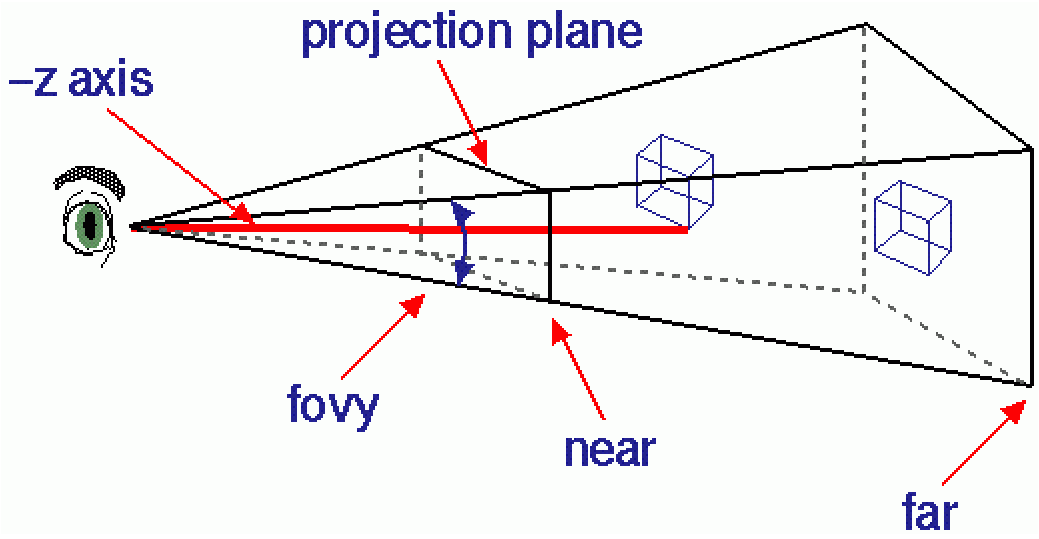
Plan of Talk

- Four Topics:
- (I) Basic Projection. Perspective. Vanishing Points.
- (II) Camera Calibration. Stereopsis. Essential Matrix. Fundamental Matrix.
- (III) Structure from Motion. Rigid. Extension to Non-Rigid.
- (IV) Geometric Priors. Manhattan World.

Geometry of Projection.

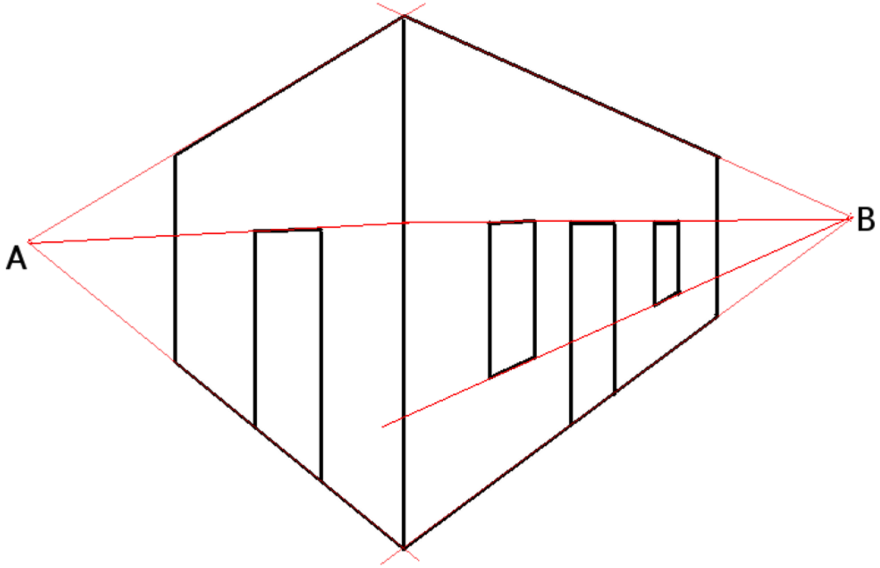
- Most analysis is based on the Pinhole camera model.
- Real cameras have lens (W. Freeman's lectures). See Szeliski's book for corrections to the pinhole camera model.





Properties of Perspective Projection

- Straight lines project to straight lines.
- Parallel lines in space project to lines which converge at a vanishing point.



Perspective 1

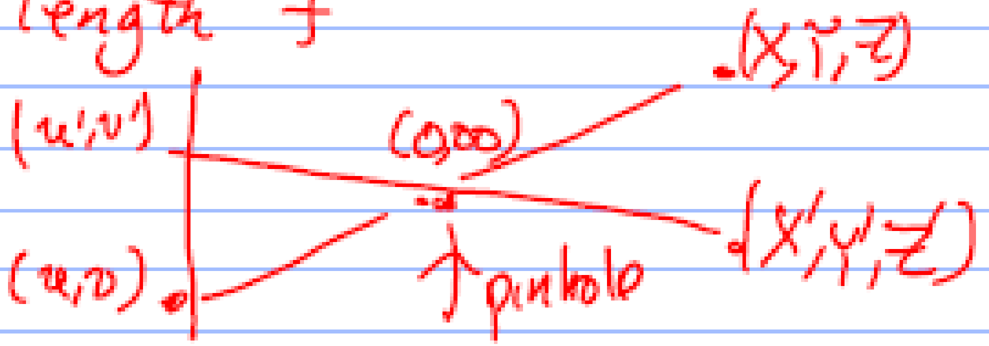
Note Title 7/24/2013

1) Perspective

$$u = \frac{fX}{Z}$$

$$v = \frac{fY}{Z}$$

3D position (X, Y, Z)
image coordinates (u, v)
focal length f



- Coordinate system based on camera.
- Origin $(0,0,0)$ at pinhole (lens)
- Camera is pointing in the Z direction

Perspective 2

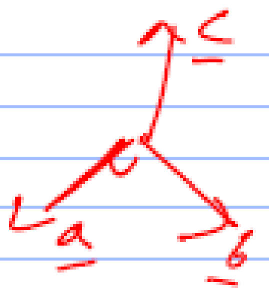
2) Camera Parameters

$\underline{a}, \underline{b}, \underline{c}$ right-angle triad

dot product

$\underline{a}.\underline{a} = \underline{b}.\underline{b} = \underline{c}.\underline{c} = 1$

$\underline{a}.\underline{b} = \underline{b}.\underline{c} = \underline{c}.\underline{a} = 0$



f · → $\underline{c}$ gaze direction

$\underline{a}, \underline{b}$ camera plane axes

$u\underline{a} + v\underline{b}$ (u, v)

$\underline{a} \wedge \underline{b} = \underline{c}$, $\underline{b} \wedge \underline{c} = \underline{a}$, $\underline{c} \wedge \underline{a} = \underline{b}$

× cross product

$\underline{o}$ camera position (of pinhole).

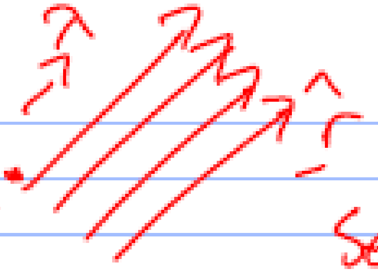
$$\boxed{\text{Camera Parameters: } \underline{a}, \underline{b}, \underline{c}, f, \underline{o}}$$

$$u = -f \frac{(\underline{r} - \underline{o}).\underline{a}}{\underline{r}.\underline{c}}, \quad v = -f \frac{(\underline{r} - \underline{o}).\underline{b}}{\underline{r}.\underline{c}}$$

More Complex Camera Models — e.g. u_0, v_0 center of image

Vanishing Points 1.

3) (i) Vanishing Points   Parallel Lines in Space



λ distance along line.

$\underline{r}_0 + \lambda \hat{\underline{r}}$ λ - length $\hat{\underline{r}}$ - direction of lines

Set $\underline{O} = 0$ pinhole at origin.

projects to line in image plane

$$u(\lambda) = -f\,\frac{(\underline{r}_0+\lambda\hat{\underline{r}})\cdot\underline{a}}{(\underline{r}_0+\lambda\hat{\underline{r}})\cdot\underline{c}}, \quad v(\lambda) = -f\,\frac{(\underline{r}_0+\lambda\hat{\underline{r}})\cdot\underline{b}}{(\underline{r}_0+\lambda\hat{\underline{r}})\cdot\underline{c}}$$

w.l.o.g. set $\underline{r}_0\cdot\underline{c}=0$. $\underline{r}_0 \to \underline{r}_0 - \frac{(\underline{r}_0\cdot\underline{c})}{(\hat{\underline{r}}\cdot\underline{c})}\hat{\underline{r}}$

$\hat{\underline{r}}\cdot\underline{c}=0$, only for lines perp. to direction of gaze $\underline{c}$.

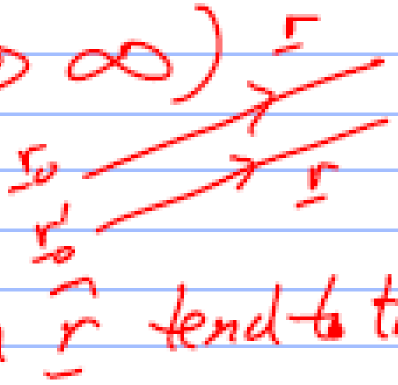
$$u(\lambda) = -f\,\frac{\hat{\underline{r}}\cdot\underline{a}}{\hat{\underline{r}}\cdot\underline{c}} - f\,\frac{\underline{r}_0\cdot\underline{a}}{\hat{\underline{r}}\cdot\underline{c}}\,\frac{1}{\lambda} \qquad v(\lambda) = -f\,\frac{\hat{\underline{r}}\cdot\underline{b}}{\hat{\underline{r}}\cdot\underline{c}} - f\,\frac{\underline{r}_0\cdot\underline{a}}{\hat{\underline{r}}\cdot\underline{c}}\,\frac{1}{\lambda}$$

λ distance along line in 3D. $1/\lambda$ is inverse distance

Vanishing points 2.

(4) Vanishing Point (ii)

As $\frac{1}{\lambda} \to 0$ (distance in space $\to \infty$)

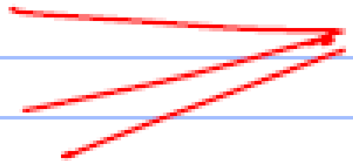


$$u(\lambda) \to -f \frac{\hat{\underline{r}}\cdot\underline{a}}{\hat{\underline{r}}\cdot\underline{c}}$$

$$v(\lambda) \to -f \frac{\hat{\underline{r}}\cdot\underline{b}}{\hat{\underline{r}}\cdot\underline{c}}$$

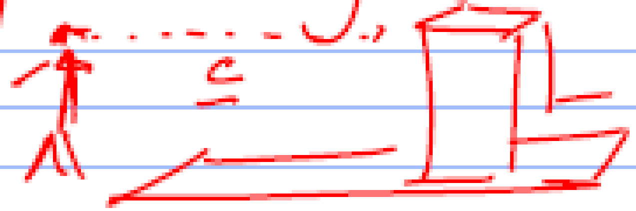
independent of $\underline{r}_0$

So all lines in direction $\hat{\underline{r}}$ tend to the same vanishing point.



vanishing points may be at infinity

eg



$\hat{\underline{r}}\cdot\underline{c} = 0$

How many vanishing points? see later.

Linear approximations: Weak Orthographic

- Perspective projection can often be approximated by scaled orthographic projection (e.g., if Z is constant).
- This is a linear operation.
- Parallel lines project to parallel lines (vanishing points at infinity).
- This is often a good approximation which is easy to use.
- Maths of weak orthographic projection.

Linear Projection 1

(5) Linear Projection Models

Can often approximate perspective projection by linear models - simplifies analysis.

(eg. if relative depth change is small)

$$(u, v) = \begin{pmatrix} K_1 & K_2 & K_3 \\ H_1 & H_2 & H_3 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

affine cameras, many variants.

weak orthographic $\underline{K} \cdot \underline{H} = 0 \quad |\underline{K}| = |\underline{H}|$

$\underline{K} = K_1, K_2, K_3$

$\underline{H} = H_1, H_2, H_3$

simplest case

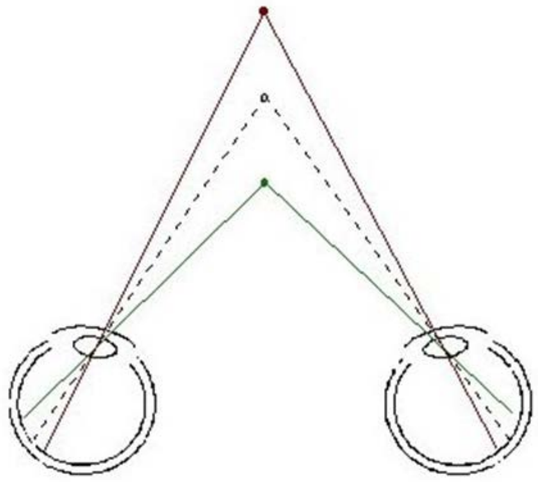
$u = s X$

$v = s Y$

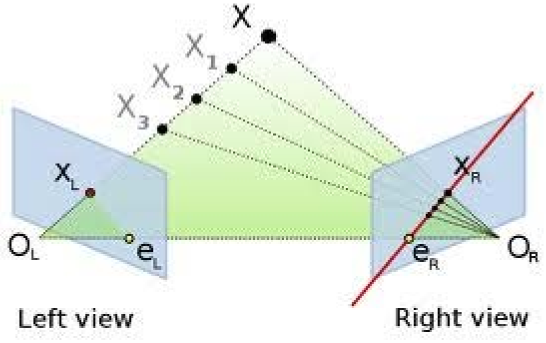
s - scale factor.

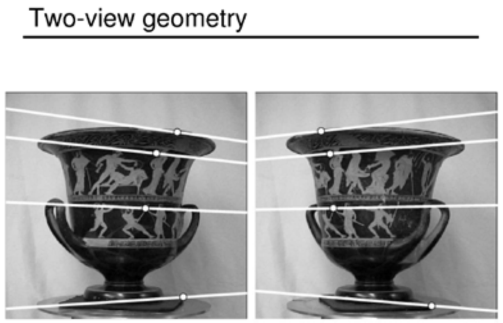
Two Cameras. Binocular Stereo.

- Binocular stereo.
- Estimate depth from two eyes/cameras by triangulation.
- Requires solving the correspondence problem between points in the two images.
- Correspondence problem is helped by the epipolar line constraint.
- Camera calibration needed.



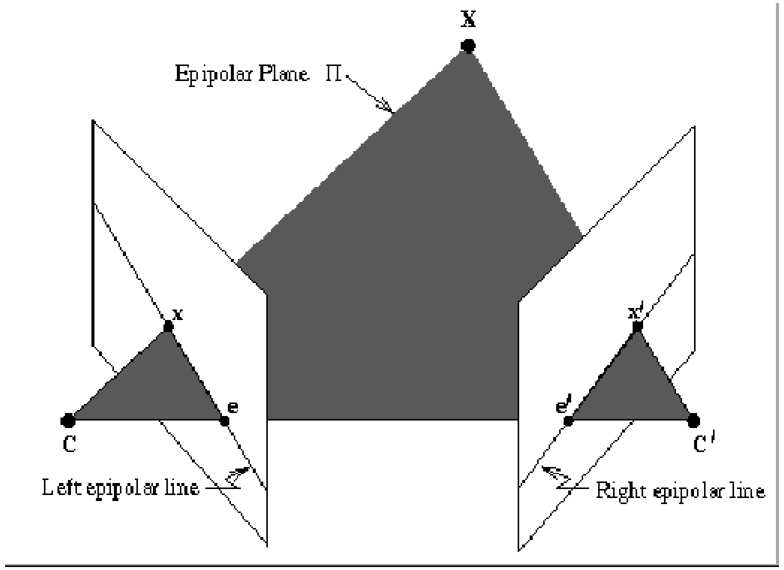
Epipolar Lines:





- Points on one epipolar line can only be matched tc . corresponding epipolar line.
- Epipolar lines depend on the camera parameters.
- If both cameras are parallel, then epipolar lines are horizontal.
- Geometric demonstration

  of epipolar line constraint.



Stereo Algorithms can exploit epipolar line constraints

Simplest model: estimate the disparity d at each point (convert to depth by geometry).

Matching unambiguous, despite epipolar line constraints.

Regularize by smoothness (ordering),

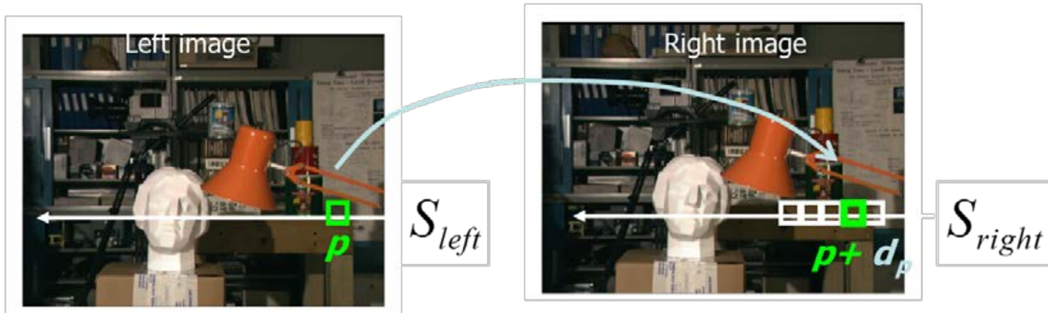
E.g., Marr and Poggio 1978. Arbib and Dev. 1977.

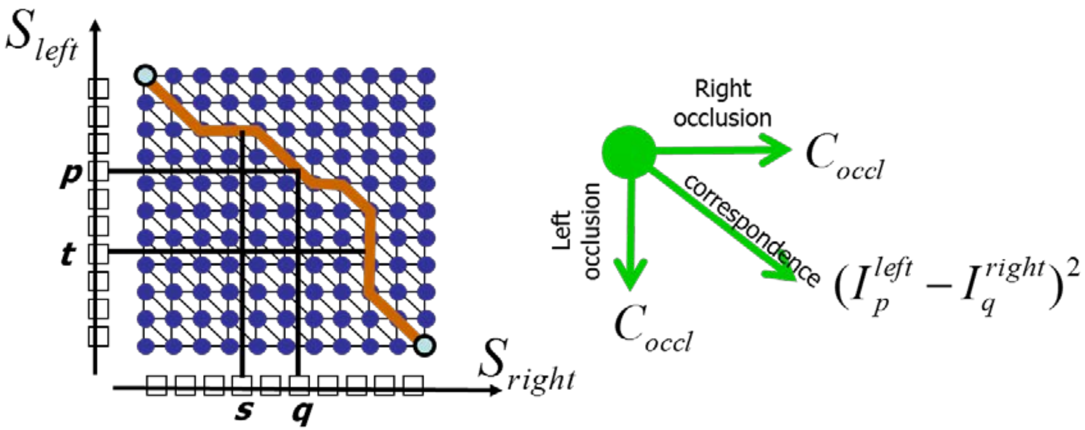
Simple Energy function: (Boykov)

$$E(d_1, d_2, \ldots, d_n) = \sum_{p \in S_{left}} (I_p^{left} - I_{p+d_p}^{right})^2 + \sum_{p \in S_{left}} (d_p - d_{p+1})^2$$

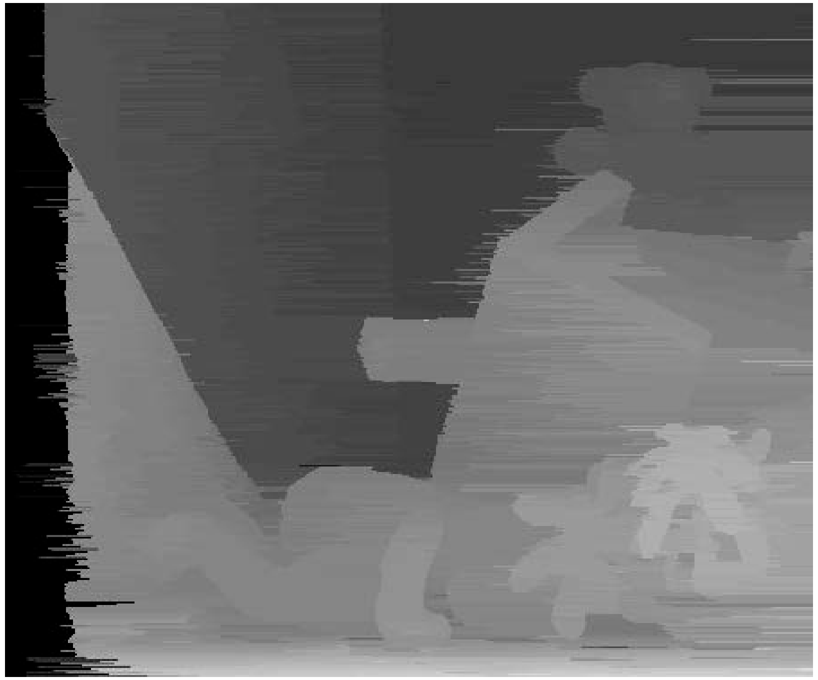
Exploit Epipolar Line constraint

- The epipolar line constraint reduces correspondence to a one-dimensional problem.
- Dynamic programming can be applied.



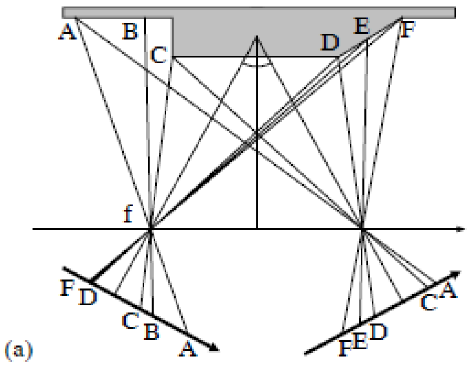


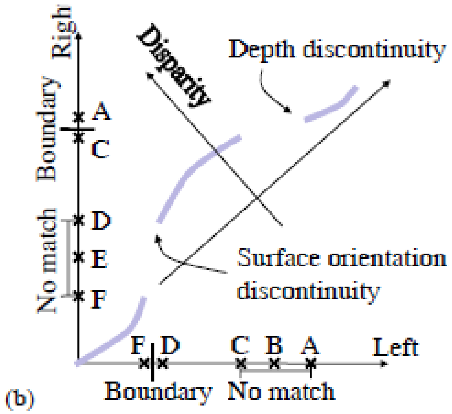
Results using Dynamic Programming



Half-Occlusions.

- Da Vinci's stereopsis.
- Points are half-occluded: visible to one eye/camera but not to the other.





- This gives cues for the detection of boundaries.
- Geiger, Ladendorf, Yuille, 1992, Belhumeur and Mumford 1992.
- H. Ishikawa and D. Geiger. 1998 (across epipolar lines).

Camera Callibration

- Essential Matrix (Longuet-Higgins 1981). Fundamental Matrix (Q.T. Luong and O. D. Faugeras 1992, Hartley 1992).
- More calibration (Z. Zhang 2000).
- More reading on geometry:
- R. Hartley and A. Zisserman. Multiple View Geometry in computer vision. 2003.
- Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. An Invitation to 3-D Vision. 2004.

Essential Matrix 1

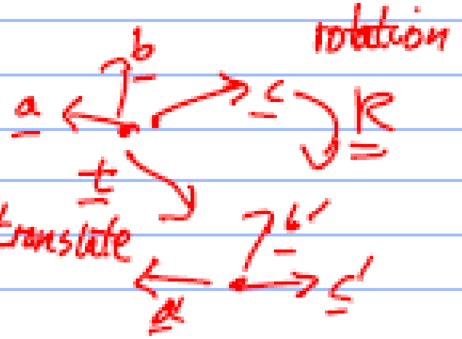
(6) Essential Matrix   Fundamental Matrix (World Coordinates)

$\underline{X} = (X_1, X_2, X_3)$ $\underline{X}' = (X_1', X_2', X_3')$ same 3D point
different coordinate system.

Analysis based on camera coordinates.

$$\begin{pmatrix}u\\v\end{pmatrix} = \frac{f}{X_3}\begin{pmatrix}X_1\\X_2\end{pmatrix}, \qquad \begin{pmatrix}u'\\v'\end{pmatrix} = \frac{f}{X_3'}\begin{pmatrix}X_1'\\X_2'\end{pmatrix}$$

$$\underline{u} \triangleq \begin{pmatrix}u\\v\\1\end{pmatrix} = \frac{f}{X_3}\begin{pmatrix}X_1\\X_2\\X_3\end{pmatrix}, \quad \underline{u}' \triangleq \begin{pmatrix}u'\\v'\\1\end{pmatrix} = \frac{f}{X_3'}\begin{pmatrix}X_1'\\X_2'\\X_3'\end{pmatrix}$$



$$\underline{X} = \underline{R}\,\underline{X}' + \underline{t}$$

$\underline{R}$ – relative rotation of cameras
$\underline{t}$ – translation

Essential Matrix 2

(7) Essential Matrix $\quad \underline{\underline{E}} = \underline{\underline{R}}\,[t]_\times$

← rotation

→ matrix representation of cross product with $\underline{t}$

$[t]_\times$ is matrix with components $\sum_k \epsilon_{ijk} t_k$

ϵ_{ijk} anti-symmetric tensor

Claim $\quad \boxed{\underline{u}'\,\underline{\underline{E}}\,\underline{u} = 0}$

imposes constraints on correspondence. | Epipolar Line

Proof $\quad \underline{u}'^T \underline{\underline{E}}\,\underline{u} \propto \underline{X}'^T \underline{\underline{E}}\,\underline{X}$

T - transpose

$$\underline{X}'^T \underline{\underline{E}}\,\underline{X} = (\underline{X} - \underline{t})^T \underline{\underline{R}}^T \underline{\underline{R}}\,[t]_\times \underline{X} = (\underline{X} - \underline{t})^T [t]_\times \underline{X} = 0$$

$$\sum_{ijk} (X_i - t_i)\,\epsilon_{ijk} t_k X_j = \underbrace{\sum_{ijk} \epsilon_{ijk} X_i X_j t_k}_{0} - \underbrace{\sum_{ijk} \epsilon_{ijk} t_i X_j t_k}_{0}$$

Note: $\epsilon_{ijk} = -\epsilon_{jik}$, $\quad \epsilon_{ijk} = -\epsilon_{ikj}$ $\quad \epsilon_{123} = 1$

Structure from Motion: Rigid

- Linear Projection. 3D structure can be estimated by linear algebra (Singular Value Decomposition).
- Camera parameters can also be estimated.
- This estimation is up to an ambiguity.

- Main paper:
- C. Tomasi and T. Kanade. 1991.
- But see also: L.L. Kontsevich, M.L. Kontsevich, A. Kh. Shen.
- “Two Algorithms fro Reconstruction Shapes}. Avtometriya. 1987.

Structure from Motion: Rigid.

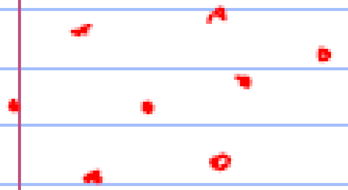
- Linear projection.
- Set of images is rank 3.
- R. Basri and S. Ullman. Recognition by Linear Combinations of Models. 1991.
- Maths of SVD.

Structure from Motion: Rigid 1

8) Structure from Motion: Rigid   Linear Projection

Set of 3D points $\underline{X}^{\mu} = (X_1^{\mu}, X_2^{\mu}, X_3^{\mu})$ (X_1, X_2, X_3) position in 3D.

$\mu = 1$ to N.



M views $\begin{pmatrix}\underline{K}^t \\ \underline{H}^t\end{pmatrix} = \begin{pmatrix} K_1^t & K_2^t & K_3^t \\ H_1^t & H_2^t & H_3^t \end{pmatrix}$

Measurements in Image

$$u^{\mu,t} = \underline{K}^t \cdot \underline{X}^{\mu} = \sum_{i=1}^{3} K_i^t X_i^{\mu}$$

$$v^{\mu,t} = \underline{H}^t \cdot \underline{X}^{\mu} = \sum_{i=1}^{3} H_i^t X_i^{\mu}$$

Note: $u^{\mu,t}$ & $v^{\mu,t}$ rank 3 $(\sum_{i=1}^{3})$ Set of image points of an object lie in a 3D space (Basri & Ullman).

Structure from Motion: Rigid 2

1) Estimate shape $(\underline{X}^{\mu})$ and viewpoint $(\underline{K}^{t}, \underline{H}^{t})$.

Assume noise: Gaussian (enables solution by linear algebra).

$$u^{\mu,t} = \sum_{i=1}^{3} K^{t}_{i} X^{\mu}_{i} + n_1^{\mu,t}$$

$$v^{\mu,t} = \sum_{i=1}^{3} H^{t}_{i} X^{\mu}_{i} + n_2^{\mu,t}$$

n_1, n_2 zero mean additive Gaussian noise.

Task: Minimize

$$\underset{\text{w.r.t. } K,H,X}{E[K,H,X]} = \sum_{\mu,t}\Big(u^{\mu,t} - \sum_{i=1}^{3} K^{t}_{i} X^{\mu}_{i}\Big)^2 + \sum_{\mu,t}\Big(v^{\mu,t} - \sum_{i=1}^{3} H^{t}_{i} X^{\mu}_{i}\Big)^2$$

simplify consider

$$\sum_{\mu,t}\Big(u^{\mu,t} - \sum_{i=1}^{3} K^{t}_{i} X^{\mu}_{i}\Big)^2$$

Structure from Motion: 3

(10) Solution by Linear Algebra (SVD) up to ambiguity.

$$E[K,X] = \sum_{\mu,t} \left(u^{\mu,t} - \sum_{i=1}^{3} K_i^t X_i^\mu \right)^2$$

Bilinear Problem.
If K known, solution of X is linear.
If X known, solution of K is linear.

Global minimum can be found by Singular Value Decomposition (SVD)

Matrix $\underline{U}$ - components $u^{\mu,t}$

Express $\underline{U} = \underline{E}\,\underline{D}\,\underline{F}^T$

$$\underline{D} = \begin{pmatrix} d_1 & 0 & & 0 \\ 0 & d_2 & & \\ & & d_3 & \\ & & & \ddots \end{pmatrix}$$
diagonal

CHECK $\underline{E}\,\underline{E}^T = \underline{I} = \underline{F}\,\underline{F}^T$

columns of $\underline{E}$ and $\underline{F}$ correspond to eigenvectors of $\underline{U}\,\underline{U}^T$ and $\underline{U}^T\underline{U}$ respectively.

Structure from Motion: Rigid 4

(10) Solution by Liner Algebra (SVD) up to ambiguity.

$E[K,X] = \sum_{\mu,t} \left(u^{\mu,t} - \sum_{i=1}^{3} K_i^t X_i^\mu \right)^2$

Bilinear Problem.
If K known, solution of X is linear.
If X known, solution of K is linear.

Global minimum can be found by Singular Value Decomposition (SVD)

Matrix $\underline{U}$ – components $u^{\mu,t}$

Express $\underline{U} = \underline{E}\,\underline{D}\,\underline{F}^T$

$\underline{D} = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & \\ & & d_3 \\ & & & \ddots \end{pmatrix}$ diagonal

CHECK $\underline{E}\,\underline{E}^T = \underline{I} = \underline{F}\,\underline{F}^T$

column's of $\underline{E}$ and $\underline{F}$ correspond to eigenvectors of $\underline{U}\,\underline{U}^T$ and $\underline{U}^T\underline{U}$ respectively.

Structure from Motion: 5

(12) Solution

Let $e_k(t)_{k=1,2,3}$ $f_k(\mu)$ be first three columns of $\underline{U}$ and $\underline{V}$

Then solutions are of form: $\underline{U} = (e_1(t)\ e_2(t) \ldots)$

$\underline{V} = (f_1(\mu)\ f_2(\mu) \ldots)$

$$K_i^t = \sum_{k=1}^{3} P_{ik}\, e_k(t)$$

$$X_i^\mu = \sum_{k=1}^{3} Q_{ik}\, f_k(\mu)$$

where $\underline{P}\,\underline{Q}^T = \underline{I}_3$

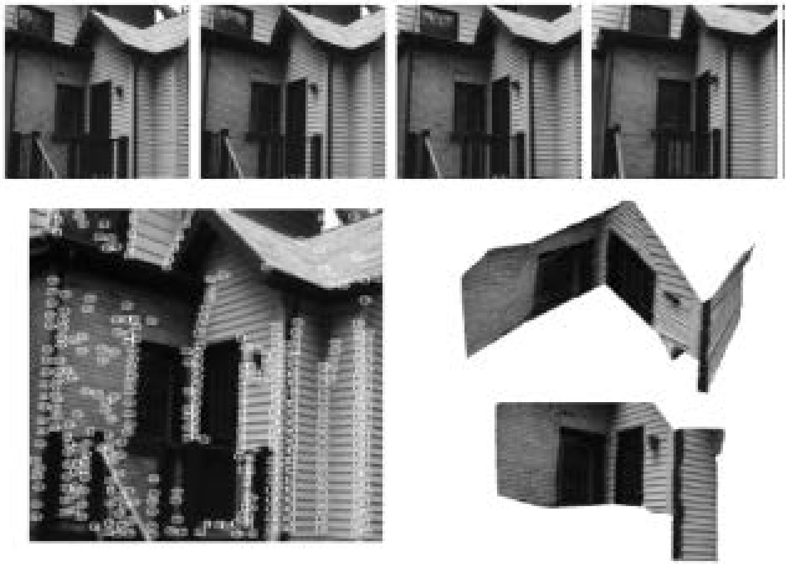
Same ambiguity

$\underline{P} \to \underline{P}\,\underline{A}$ $\quad \underline{Q}^T \to \underline{A}^{-1}\underline{Q}$

Structure From Motion: Rigid.

SVD Results: From Tomasi and Kanade. 1991

But see: Kontsevich et al.



Extension to Non-Rigid Motion

- This approach can be extended to a special class of non-rigid motion.
- The object can be expressed as a linear sum of basis functions. The sum varies over time.
- C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. CVPR. 2000.
- Theory clarified by:
- Y. Dai, H. Li, and M. He. A Simple Prior-free Method for Non Rigid Structure from Motion Factorization, in CVPR 2012 (ORAL). IEEE CVPR Best Paper Award-2012. (Code available).
- http://users.cecs.anu.edu.au/~hongdong/

Non-rigid motion

(12) Relax Rigidity: Basis Function Models (Bregler et al)

$u^{\mu,t} = \underline{K}^t \cdot \underline{X}^{\mu,t} = \sum_{i=1}^{3} K^t_i X^{\mu,t}_i$

$v^{\mu,t} = \underline{H}^t \cdot \underline{X}^{\mu,t} = \sum_{i=1}^{3} H^t_i X^{\mu,t}_i$

relax rigidity $X^{\mu}_i \rightarrow X^{\mu,t}_i$

Too relaxed

Problem is ill-posed.

Now suppose that the object $\underline{X}^{\mu,t}$ can be expressed as a linear combination of bases:

$$\underline{X}^{\mu,t} = \sum_{a=1}^{A} \beta^t_a \underline{B}^{\mu}_a$$

bases for object

coefficients, depend on t.

$$u^{\mu,t} = \sum_{i=1}^{3} \sum_{a=1}^{A} K^t_i \beta^t_a B^{\mu}_{i,a}$$

$$v^{\mu,t} = \sum_{i=1}^{3} \sum_{a=1}^{A} H^t_i \beta^t_a B^{\mu}_{i,a}$$

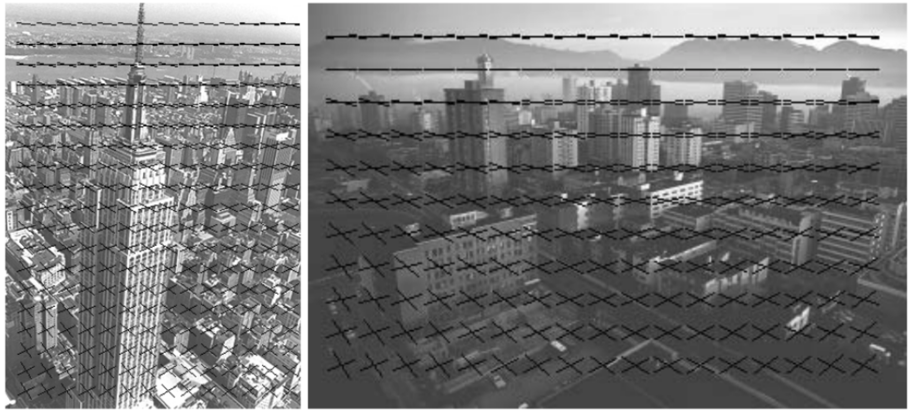
SVD - twice

More Complex - ambiguity

Best Paper CVPR 2012.

Manhattan World

- Many scenes, particularly man-made scenes, have a natural three-dimensional coordinate systems caused by the structure of the world.



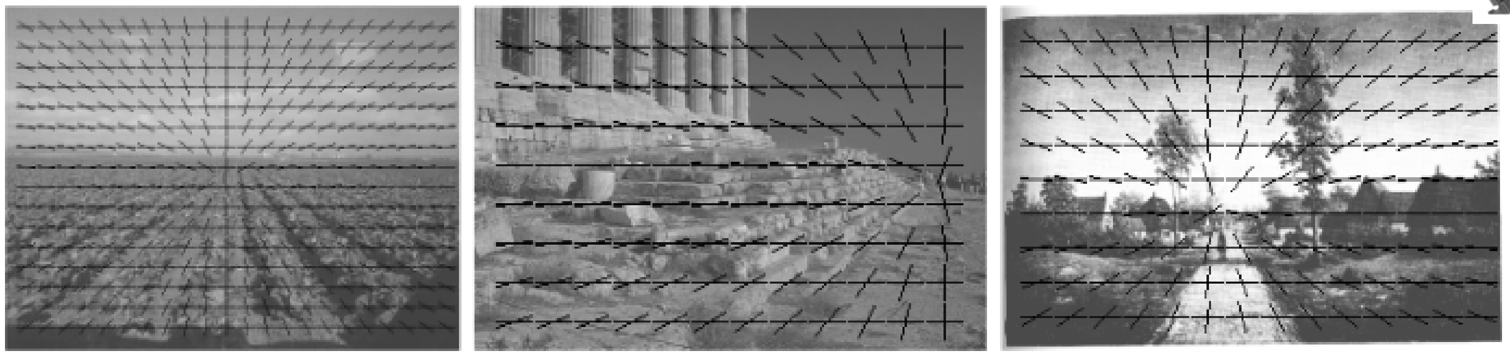
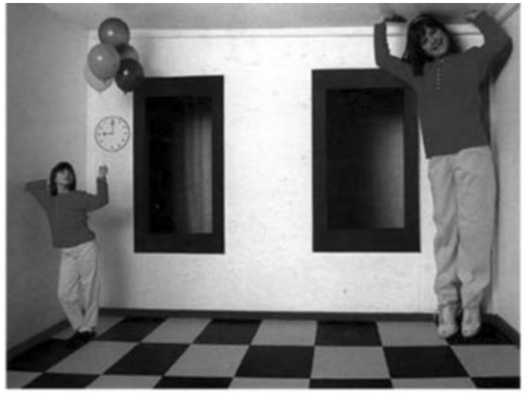
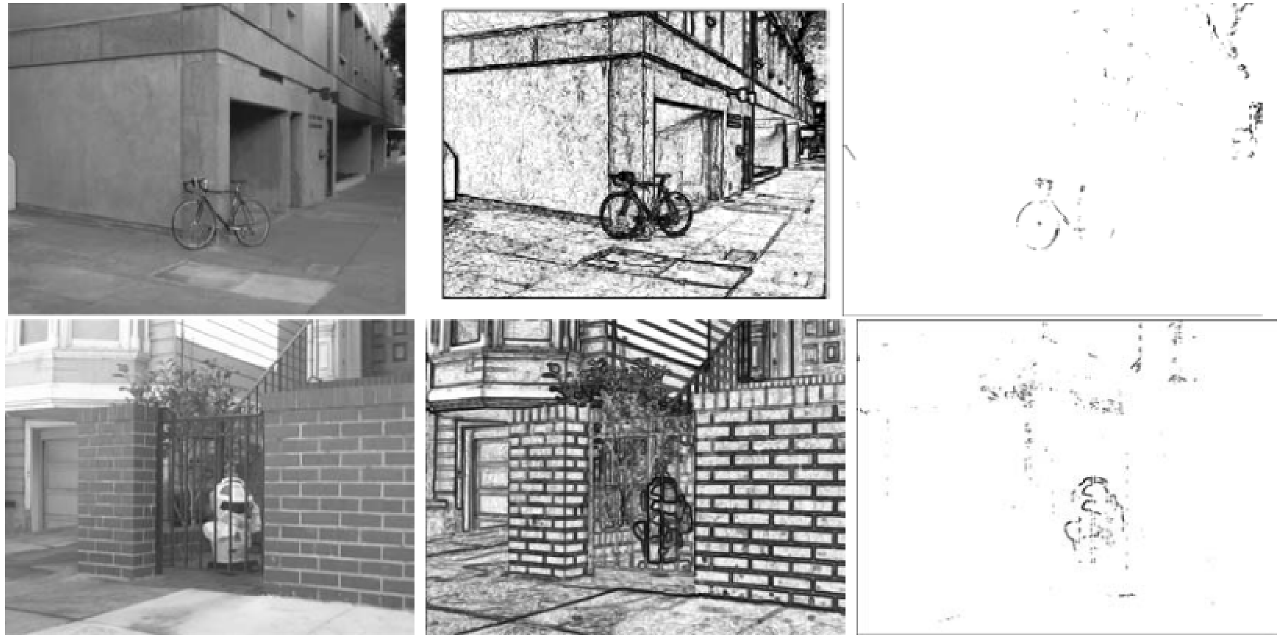


Figure 5: Results on an American mid-west broccoli field, the ruins of the

Manhattan World: 2

- Back to Ames Room:
- Non-Manhattan Edges:





Manhattan World 1

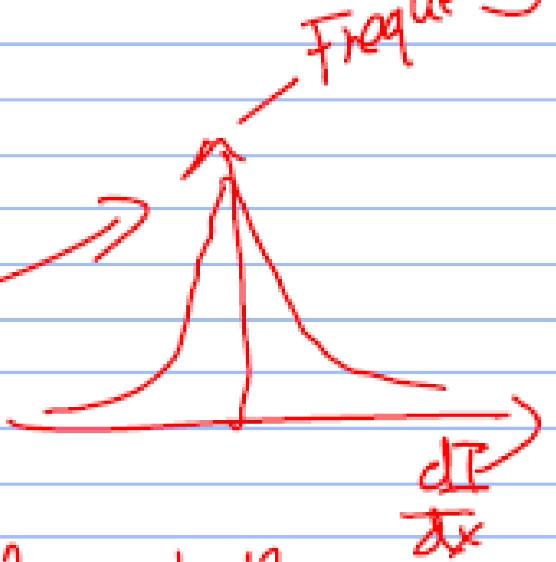
(13) Manhattan World.

Exploit regularities in the geometric structure of scenes.

Statistics of derivative operators in images (Simoncelli, Zhu, ..)

Statistical justification of piecewise smoothness.

Histogram of $\frac{dI}{dx}$ peaks at $\frac{dI}{dx} = 0$



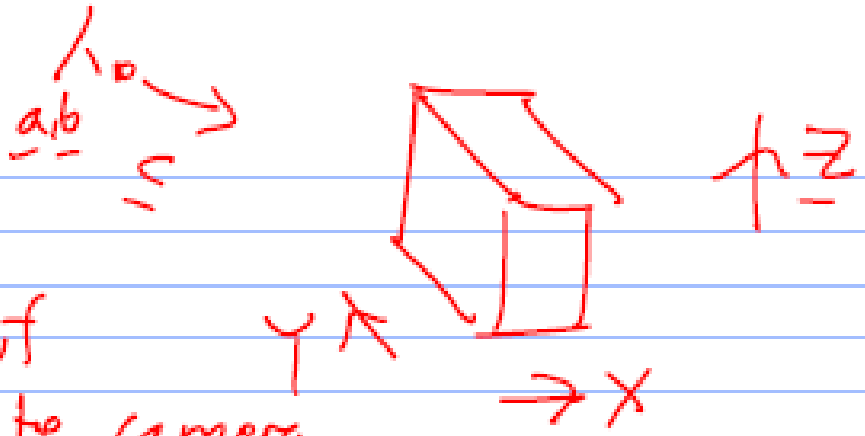
But there are also regularities in the directions of edges, depending on the orientation of the camera relative to the geometric structure of the scene.

Manhattan World 2

(24) Manhattan World

· distribution of edges in the image depends on camera parameters $(\underline{a}, \underline{b}, \underline{c}), f$

· use structure of scene to callibrate camera.
— estimate $(\underline{a}, \underline{b}, \underline{c}), f$.



Assume : A pixel $\underline{u}$ image is

(i)	An edge in the X direction	$m_{\underline{u}} = 1$
(ii)	" " " " Y "	$m_{\underline{u}} = 2$
(iii)	" " " " Z "	$m_{\underline{u}} = 3$
(iv)	An unaligned edge	$m_{\underline{u}} = 4$
(v)	Not an edge	$m_{\underline{u}} = 5$

The m's are hidden variables,

Manhattan World 3

15)

Generative Model for the image gradient $\underline{\nabla I}$ at $\underline{u}$

$P(\underline{\nabla I}(\underline{u}) \mid m_{\underline{u}}, \underline{\Psi})$ $\qquad \underline{\Psi} = (\underline{a}, \underline{b}, \underline{c}, f)$

camera parameters.

If $m_{\underline{u}} = 5$ (no edge)

$P(\underline{\nabla I}(\underline{u}) \mid m_{\underline{u}} = 5, \underline{\Psi}) \simeq 0$, unless $|\underline{\nabla I}(\underline{u})|$ is small

if $m_{\underline{u}} = 1$ (edge in X direction)

$P(\underline{\nabla I}(\underline{u}) \mid m_{\underline{u}} = 1, \underline{\Psi}) \simeq 0$, unless $|\underline{\nabla I}(\underline{u})|$ is large

<u>and</u> $\underline{\nabla I}(\underline{u})$ points in direction of projection of X-line.

Manhattan World 4

(16) Full Model.

Specify prior $P(m_{\underline{u}})$

don't care about $m_{\underline{u}}$, so sum it out. (Freeman's talk).

$$P(\underline{\nabla I}(\underline{u}) \mid \underline{\Psi}) = \sum_{m=1}^{5} P(\underline{\nabla I}(\underline{u}) \mid \underline{\Psi}, m_{\underline{u}}) P(m_{\underline{u}})$$

Assume Independence → for Full Image.

$$P(\{\underline{\nabla I}(\underline{u})\} \mid \underline{\Psi}) = \prod_{\underline{u}} P(\underline{\nabla I}(\underline{u}) \mid \underline{\Psi})$$

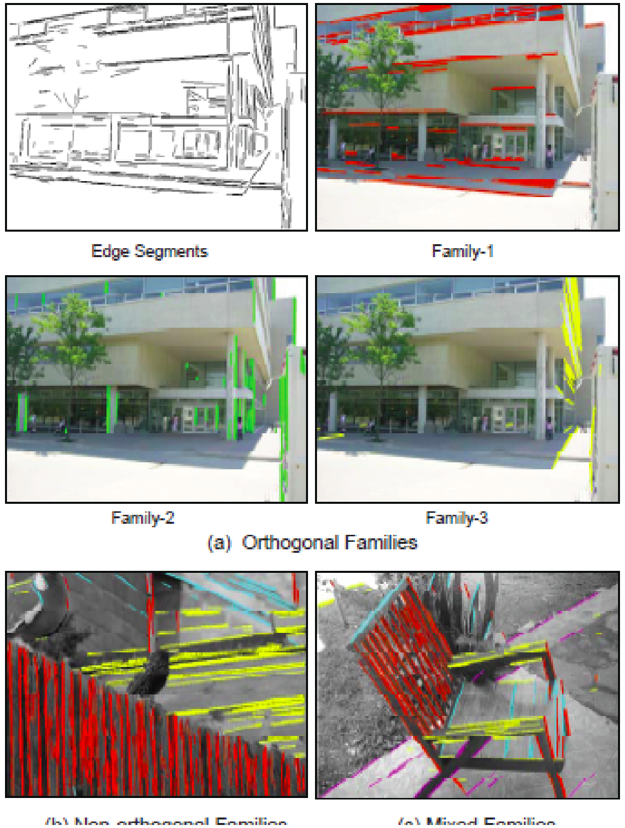
Estimate $\hat{\underline{\Psi}} = \arg\max_{\underline{\Psi}} P(\{\underline{\nabla I}(\underline{u})\} \mid \underline{\Psi})$.

Very Simple → Works fairly well

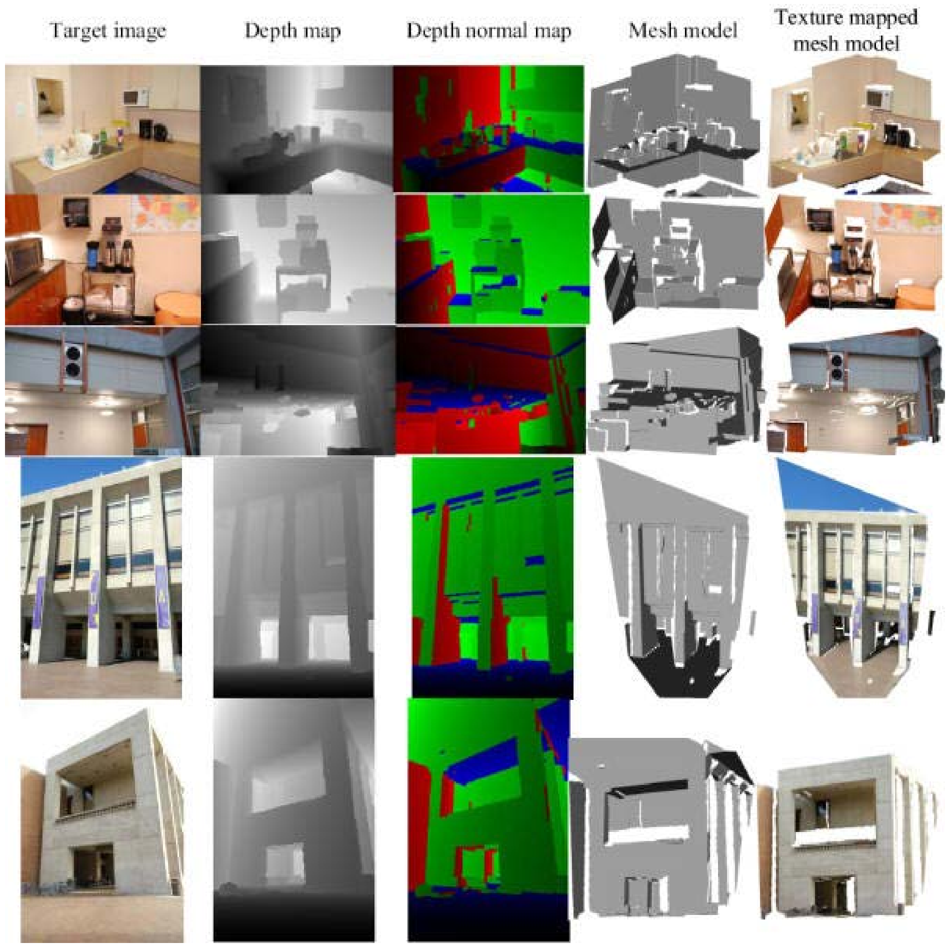
More Sophisticated Models

Beyond Manhattan:

- There are parallel lines in the scene, but they are not orthogonal.



Manhattan World



- Manhattan World stereo:
- Piecewise planar surfaces with dominant
- Directions.
- Instead of assuming
- Piecewise smoothness.
- Video available:
- http://grail.cs.washington.edu/projects/manhattan/

- Y. Fuukawa, B. Curless, S. Seitz, and R. Szeliski. 2009.

Manhattan World Grammar

- Website: http://www.youtube.com/watch?v=s0mhpKFv36g