3D Reconstruction for Rendering

Topic in Image-Based Modeling and Rendering

CSE291 J00

Lecture 6

This lecture

- [DTM96] Paul E. Debevec, Camillo J. Taylor, Jitendra Malik, "Modeling and Rendering Architecture from Photographs: A hybrid geometry and image-based approach". *Technical report UCB/CSD-96-893, University of California at Berkeley, 1996*.
- [TK94] Camillo J. Taylor and David J. Kriegman "Structure and Motion from Line Segments in Multiple Images". *IEEE Trans. Pattern Anal. Machine Intell. 17(11) November 1995.*

Introduction

- Presents an approach for modeling and rendering existing architectural scenes from sparse sets of still photographs.
- Geometry-based methods: modeling program is used to constructing model.
  - Drawbacks: labor-intensive, difficult to verify and most of all: unrealistic.
- Image-based: creating model directly from photographs.
  - Relies on *stereo algorithms,* has a lot constrains on the input.

Hybrid approach

- Combine the strengths of both geometry-based and image-based methods.

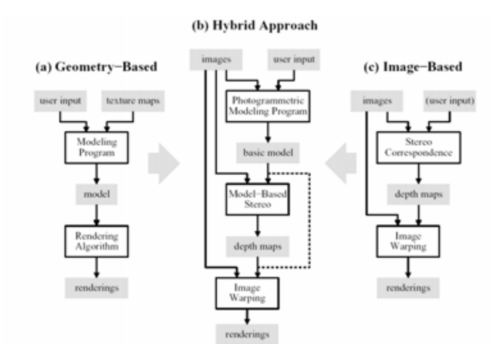
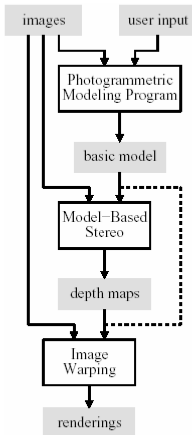


Figure 1: *Schematic of how our hybrid approach combines geometry-based and image-based approaches to modeling and rendering architecture from photographs.*

Src [DTM96]

Hybrid approach



- Photogrammetric modeling: geometric model of the architecture is recovered interactively.
- View-dependent texture mapping: create novel view.
- Model-based stereo: additional geometric detail can be recovered.

Example

Original photograph with marked edges | Recovered model | Model edges projected onto photograph | Synthetic rendering

Src [DTM96]

Photogrammetric modeling

- Scene represented as a constrained hierarchical model of parametric primitives (*block*).
- Relationships between blocks are represented by a rotation matrix and a translation vector.

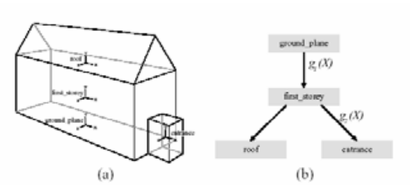
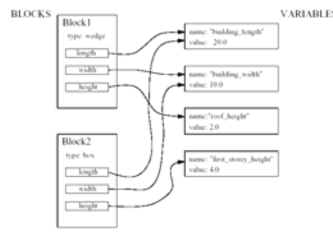


Figure 4: **(a)** *A geometric model of a simple building.* **(b)** *The model's hierarchical representation. The nodes in the tree represent parametric primitives (called* **blocks***) while the links contain the spatial relationships between the blocks.*

Source [DTM96]



Representation of block parameters as symbol references. A single variable can be referenced by the model in multiple places, allowing constraints of symmetry to be embedded in the model.

Advantages.

- Advantages of modeling the scene with blocks:
  - Most architecture can be readily decomposed into a set of blocks.
  - Blocks implicitly model common architectural constraints.
  - Convenient to manipulate block primitives.
  - Surface of the scene is readily obtained from the blocks.
  - Reduce the number of parameters need to recover.

Reconstruction

- Based on the reconstruction of infinite straight lines. [TK94]
- Model parameters and camera positions are computed by minimizing objective function O.

$$O = \sum_{j=1}^{m} \sum_{i=1}^{n} Error\left(F(p_i, q_j), u_{ij}\right)$$

Disparity between the **actual** and **expected** image measurements

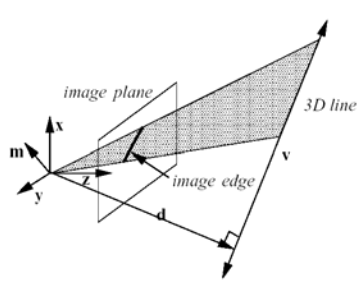
Camera position of image *i*.

Structure of feature *j*.

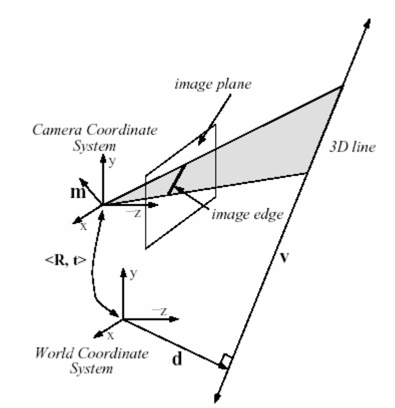
Feature *i* in image *j*.

Geometry of Straight Lines

- A straight line is represented by a tuple <**v**,**d**>.
- The line and the origin define a plane whose normal is vector **m**
- The edge in the image will be: $m_x x + m_Y y + m_z = 0$



Projection Function *F*



$$\begin{aligned} {}^c\hat{\mathbf{v}} &= {}^c_w\mathbf{R}\,{}^w\hat{\mathbf{v}} \\ {}^c\mathbf{d} &= {}^c_w\mathbf{R}\,({}^w\mathbf{d} - {}^w\mathbf{t}_c + ({}^w\mathbf{t}_c \cdot {}^w\hat{\mathbf{v}})\,{}^w\hat{\mathbf{v}}) \\ {}^c\mathbf{m} &= {}^c\hat{\mathbf{v}} \times {}^c\mathbf{d} \\ &= {}^c_w\mathbf{R}\,\{{}^w\hat{\mathbf{v}} \times ({}^w\mathbf{d} - {}^w\mathbf{t}_c)\} \\ {}^c\hat{\mathbf{m}} &= {}^c\mathbf{m} / \|{}^c\mathbf{m}\| \end{aligned}$$

$$F(R,t,v,d) \rightarrow m$$

The Error Function

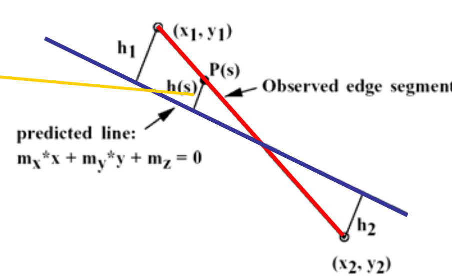
- Error function represents disparity between the **actual** and **expected** image measurements.

$$h(s) = h_1 + s\frac{h_2 - h_1}{l}$$

$$l = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$h_1 = \frac{m_x x_1 + m_y y_1 + m_z}{\sqrt{m_x^2 + m_y^2}}$$

$$h_2 = \frac{m_x x_2 + m_y y_2 + m_z}{\sqrt{m_x^2 + m_y^2}}$$



The Error Function <cont.>

- The total error between the observed edge segment and the predicted edge as:

$$Error = \int_0^1 h^2(s)ds = \frac{l}{3}(h_1^2 + h_1 h_2 + h_2^2)$$
$$= \mathbf{m}^T (A^T B A)\mathbf{m}$$

$$A = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{bmatrix}$$
$$B = \frac{l}{3(m_x^2 + m_y^2)} \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

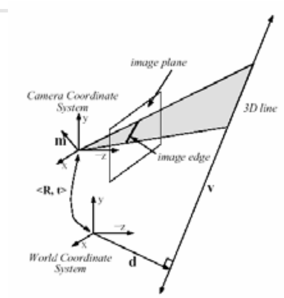
This is a projection function of <$\mathbf{R}_j,\mathbf{t}_j,\mathbf{v}_i,\mathbf{d}_i$>

Recovery Algorithm - Overview

- Minimizing object function *O* with respect to <$\mathbf{R}_j,\mathbf{t}_j,\mathbf{v}_i, \mathbf{d}_i$>
  - Estimates for camera rotations R_j.
  - Estimates for camera translations t_j and the parameters of the model <v_i, d_i>.
  - Non-linear minimization over the entire parameter space.

Estimate R_j

- Initial estimates for R_j can be entered manually.
- Can be re-estimated after v_i are estimated.



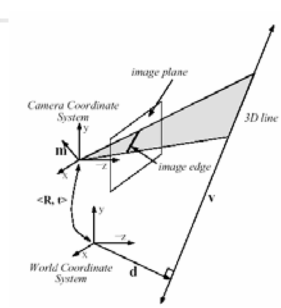
Estimate $\mathbf{v}_i$

- Based on constraints

$$^{c}m^{T}\left({}^{c}_{w}R\,{}^{w}v\right)=0$$

- Use **m'** – measured normal to the plane passing through camera center and the observed edge

$$^{c}m' = \begin{bmatrix} x_1 \\ y_1 \\ -1 \end{bmatrix} \times \begin{bmatrix} x_2 \\ y_2 \\ -1 \end{bmatrix}$$



Estimate $\mathbf{v}_i$ <cont.>

$$ {}^c m^T \left({}^c_w R \, {}^w v \right) = 0 \longrightarrow C_1 = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(m'^T_{ij} R_j v_i \right)^2 = \sum_{i=1}^{n} C_{A_i} $$

- $\mathbf{v}_i$ is determined by minimizing each C_{A_i} with respect to $\mathbf{v}_i$.
  - *2n degrees of freedom*.
- Improve the estimates of $\mathbf{R}_j$, $\mathbf{v}_i$ by minimizing C_1 with respect to (R_j, v_i).
  - *2n+3(m-1) degrees of freedom*.

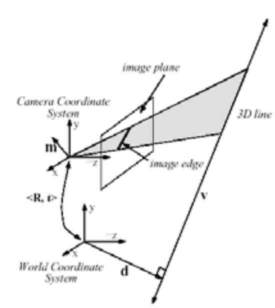
Estimates $\mathbf{d}_i$ and $\mathbf{t}_j$

- Again, from constraint:

$$ {}^c m^T \left({}^c_w R^w \left({}^w d - {}^w t_c \right) \right) = 0 $$

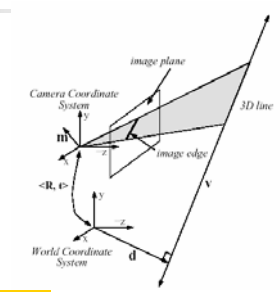
- $\mathbf{d}_i$ and $\mathbf{t}_j$ are estimated to minimize the objective function:

$$ C_2 = \sum_{j=1}^{m} \sum_{i=1}^{n} \left(m'^T_{ij} R_j \left(d_i - t_j \right) \right)^2 $$



Estimates $\mathbf{d}_i$ and $\mathbf{t}_j$

- Note that $\mathbf{v}_i$ is orthogonal to $\mathbf{d}_i$.



$$v_i v_i^x = 0$$
$$v_i v_i^y = 0$$
$$v_i^x v_i^y = 0$$

$$d_i = \alpha v_i^x + \beta v_i^y$$

$$C_2 = \sum_{j=1}^{m} \sum_{i=1}^{n} \left(m'^T_{ij} R_j \left(\alpha v_i^x + \beta \hat{v}_i^y - t_x \hat{x} - t_y \hat{y} - t_z \hat{z} \right) \right)^2$$

- Then, closed form linear least squares can be applied to obtain estimates for these parameters.

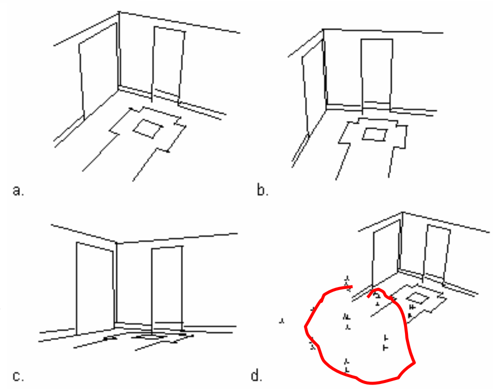
Minimizing the Objective Function.

$$O = \sum_{j=1}^{m} \sum_{i=1}^{n} Error\left(F(p_i, q_j), u_{ij} \right)$$

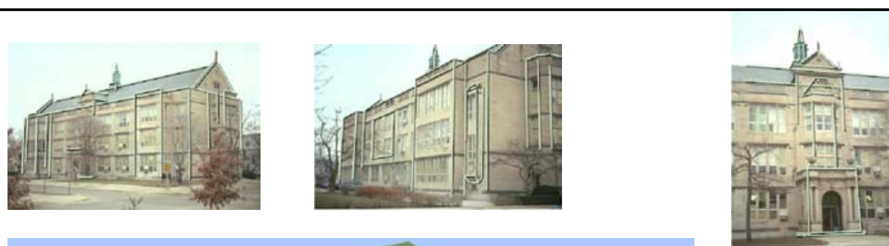
- Using variant of the Newton-Raphson method.
- Optimization requires fewer than ten iterations.
- Edge of the recovered models typically conform to the original photographs to within a pixel.

Six images taken from a sequence of 24 taken from a section of Yale Center for Systems Science office complex

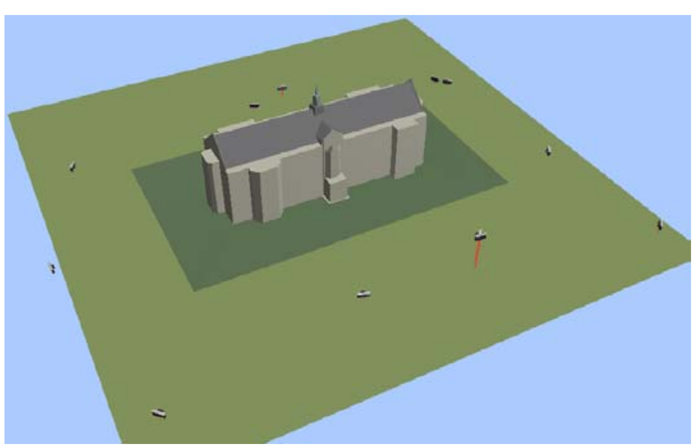
Source [TK96]



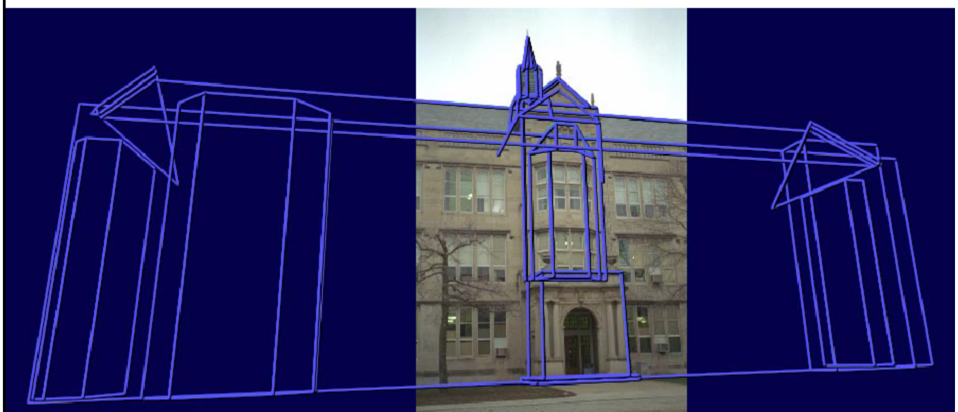
*A set of views of the reconstruction of the scene. The small coordinate axes in figure **d** represent the reconstructed camera positions.*



Above: *Three of twelve photographs used to reconstruct the entire exterior of University High School in Urbana, Illinois. The superimposed lines indicate the edges the user has marked.*



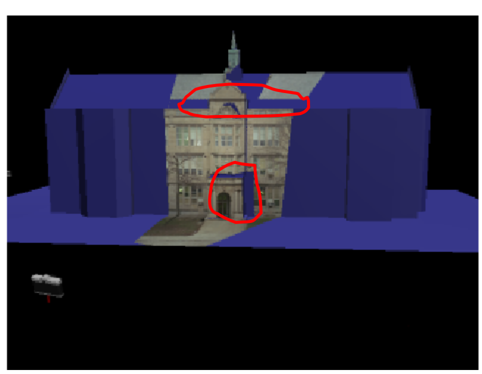
Left: *The high school model, reconstructed from twelve photographs. Aerial view showing the recovered camera positions.*



The edges of the reconstructed model, projected through the recovered camera positions and overlaid on the corresponding images. The recovered model conforms to the photographs to within a pixel in all twelve images, indicating that the building has been accurately reconstructed.

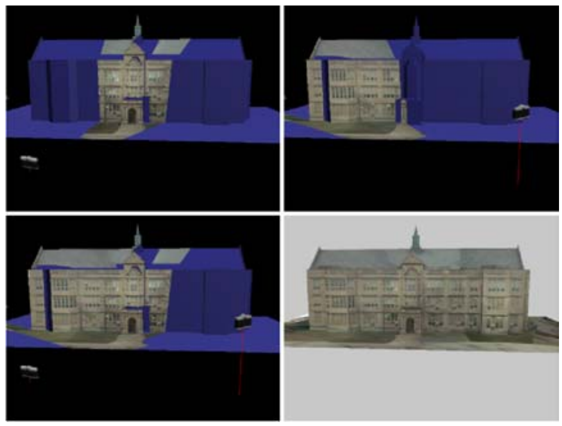
View-Dependent Texture-Mapping

- Projecting the original photographs onto the model based on viewing frustum.
- Single-image projection: image-space shadow map.



Composition of multiple images

- Need multiple images in order to render the entire model.



Compositing multiple images.

- Use the one with viewing angle closest to that of rendering view.
- Projected image weights are computed at every pixel of every projected rendering.
- In fact, a single weight is used across a flat surface.

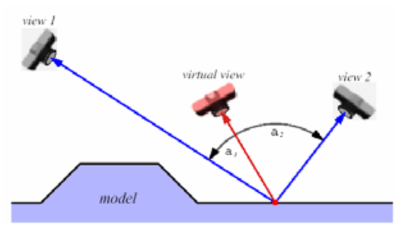
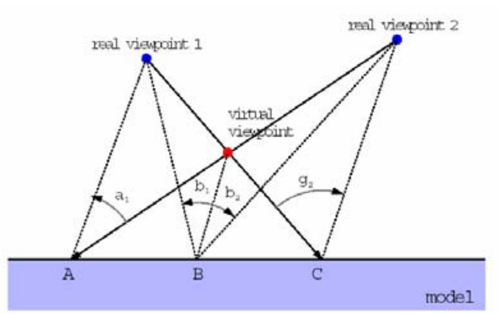


Figure 12: *The weighting function used in view-dependent texture mapping. The pixel in the virtual view corresponding to the point on the model is assigned a weighted average of the corresponding pixels in actual views 1 and 2. The weights w_1 and w_2 are inversely inversely proportional to the magnitude of angles α_1 and α_2. Alternately, more sophisticated weighting functions based on expected foreshortening and image resampling can be used.*

Src [DTM96]

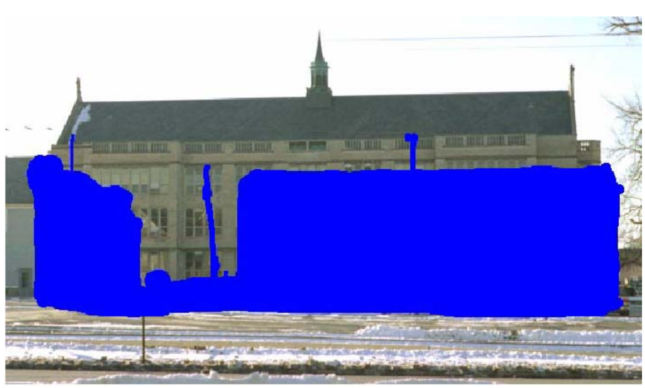
Blending images

- For pixel *B*, both views are equally good.
- One solution would be to use pixel values from view 1 to the left of *B* and pixel values from view 2 to the right of *B*.
- A better solution is to blend pixels from view 1 and view 2 in the area between *A* and *C* according to their relative fitness values.



Removal of obstructions.

- Unwanted objects in the original photograph may be projected onto the surface of the model.
- "Mask out" these objects with a reserved color.



A photograph in which the user has masked out two interposed buildings that obscure the architecture of interest. These masked regions will consequently not be used in the rendering process.

Filling a holes.

- Unfilled pixels on the periphery of the hole are successively filled with the average values of the neighboring pixels.
- After each iteration, the region of unfilled pixels is eroded by a width of one pixel.

University High School fly-around, with trees.

University High School fly-around, without trees. For this sequence, the trees were masked out of the twelve original photographs of the building. The masked sections were then filled in automatically by the image composition algorithm.

Model-Based Stereo

- Some geometric detail is not captured in the model.
- Measures how the actual scene deviates from the approximate model.
- The model serves to place the images into a common frame of reference.

Traditional stereo

- Given 2 images (*key* and *offset*) model-based stereo computes the associated depth map for the key image by determining corresponding points in the key and offset images.
- When the key and offset images are taken from relatively far apart, corresponding pixel neighborhood can be foreshortened very differently.

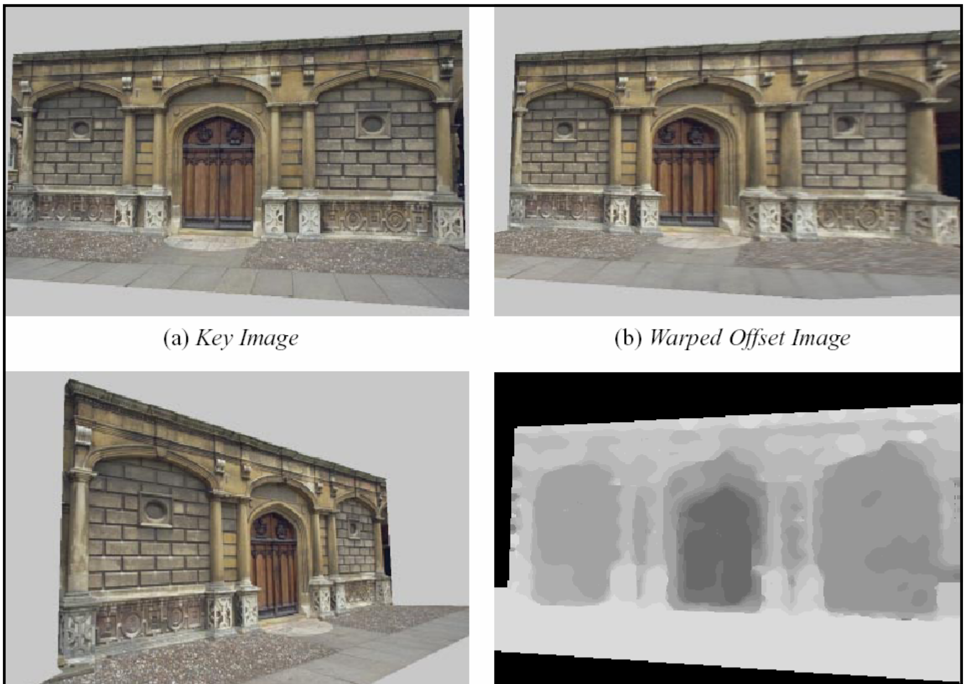
Warp offset

- Project the offset image onto the model and view it from the position of the key image. (warped offset).
- Pixel neighborhoods are compared between the key and warp offset images.

Src [DTM96]

Advantage of warp offset.

- Reduction of differences in foreshortening.
- Any point in the scene which lies on the approximate model will have zero disparity between the key image and the warped offset image.
- Easy to convert to a depth map for the key image.
- Less sensitive to noise in image measurements.



(a) *Key Image* (b) *Warped Offset Image*

(c) *Offset Image* (d) *Computed Disparity Map*

Epipolar Geometry

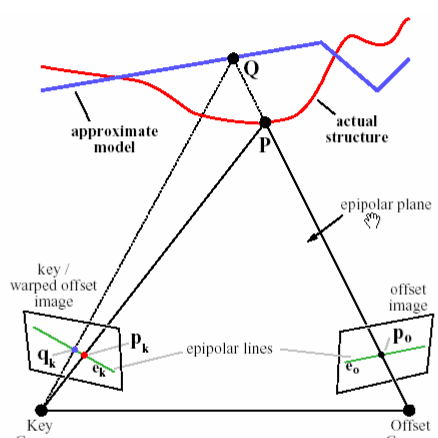
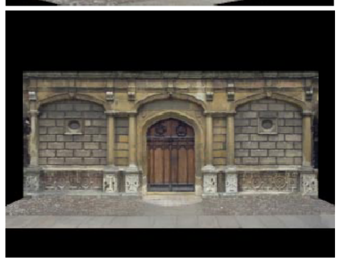


Figure 15: *Epipolar geometry for model-based stereo.*

Src [DTM96]

Result of model-based stereo.

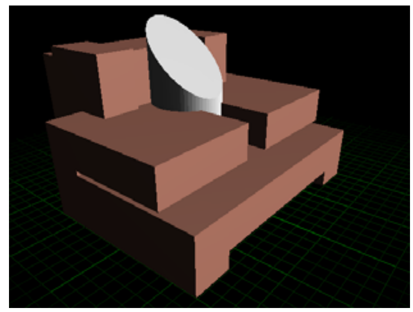


Novel views of the scene generated from four original photographs. The depth is computed from model-based stereo and the frames are made by compositing image-based renderings with view-dependent texture-mapping.

Result – The Campanile

- **The Campanile**, a short film shown at the SIGGRAPH'97 Electronic Theatre, used Façade to construct a photorealistic model of Berkeley's clock tower and the surrounding campus

Result



San Francisco Museum of Modern Art by Yizhou Yu

Improvement of Façade.

3-D Model Reconstruction of theTaj Mahal using Façade.
George Borshukov

Commercial use.

Seven photographs of a street scene in downtown Palo Alto, northern California. Modeling time 3 hours.

Source: http://www.canoma.com/movies.html

Past, present and future …

- Commercial products: Canoma, REALVIZ® ImageModeler, PhotoModeler Pro 4.5
- Manex entertainment: The Matrix.
- Using architectural conventions and practical considerations to enhance the reconstruction model.
- Reconstruct from single, uncalibrated image.
- And more…