

Camera and Image Motion

Computational Photography

CSE 273

Lecture 7

Announcements

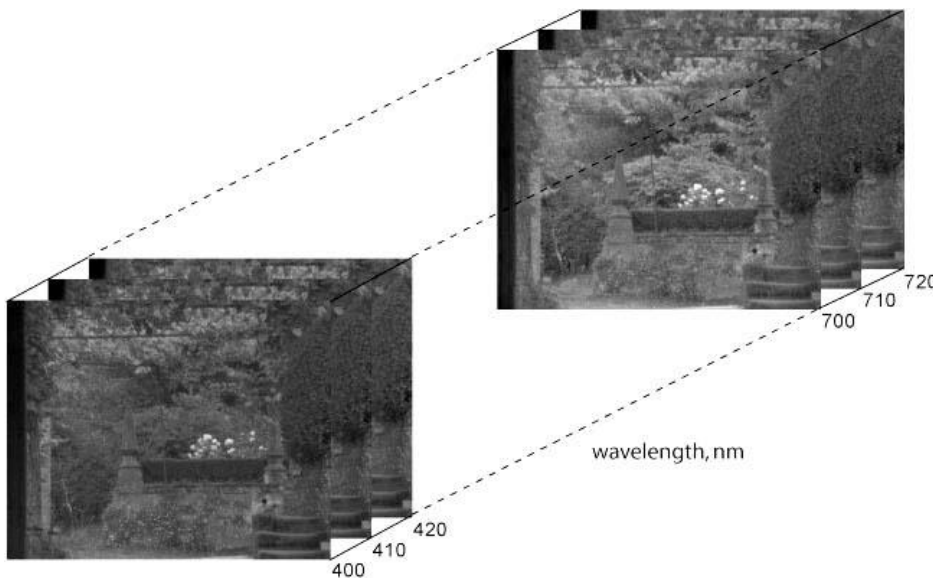
- Assignment 3 is due today, 11:59 PM
- Assignment 4 will be released today
 - Due Feb 7, 11:59 PM

Burst photography

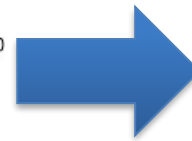
- Acquisition of several images by a single camera in quick succession

Burst photography

- Hyperspectral image (HSI) set to RGB image



Hyperspectral image



RGB image

Burst photography

- Set of standard dynamic range (SDR) images to a high dynamic range (HDR) image
- Super resolution



Camera and image motion

- Burst photography
 - Super resolution
- Video
 - Mosaic construction from video (or a set of images)
 - Video stabilization

Camera and image motion

- Camera motion: rotation only (i.e., camera center does not move)
- Image motion: 2D projective transformation

Camera motion

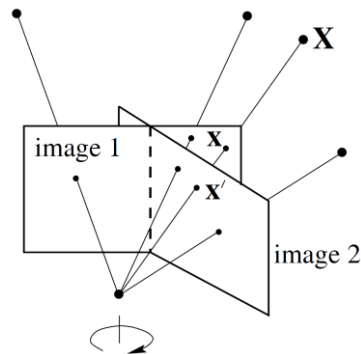
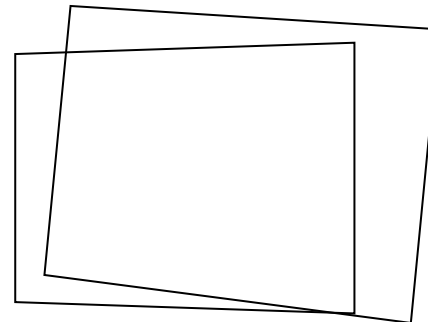
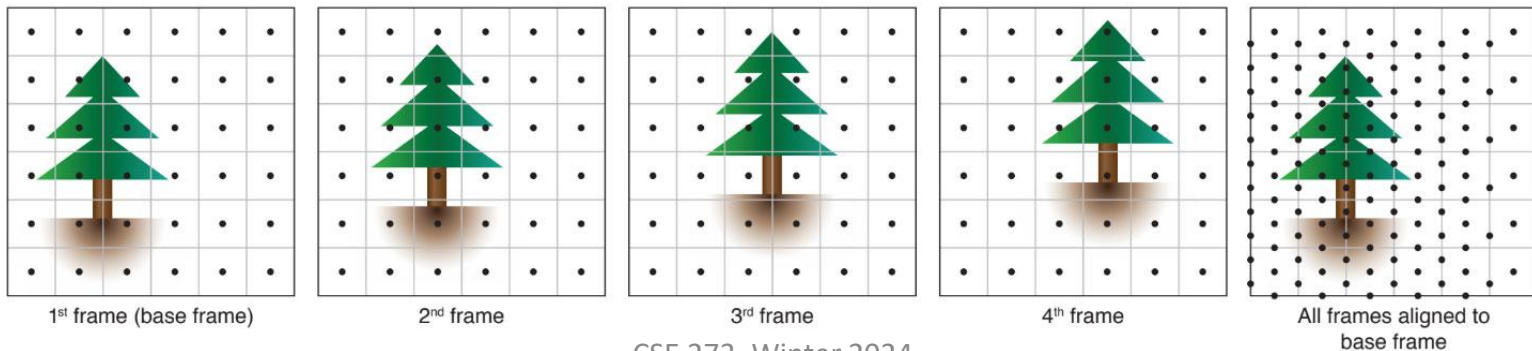


Image motion



Super resolution

- Camera motion (e.g., hand shake) causes pixel-level differences over each image in the burst
- Requires subpixel alignment of all images
- Can be used to compute a super resolution image



Super resolution

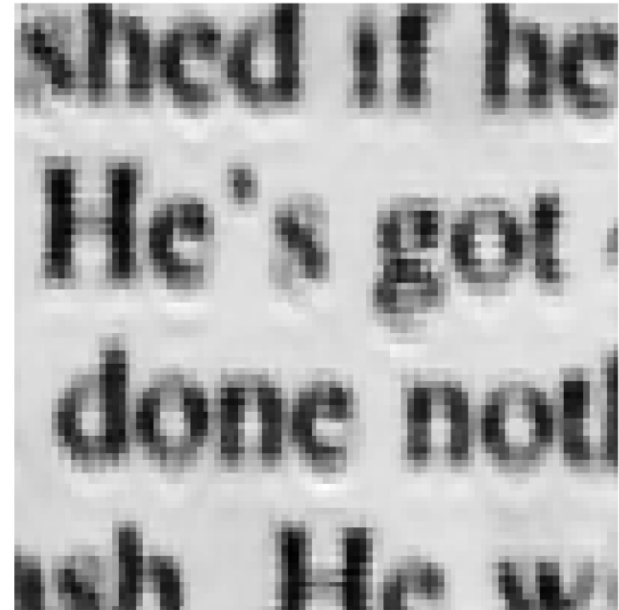
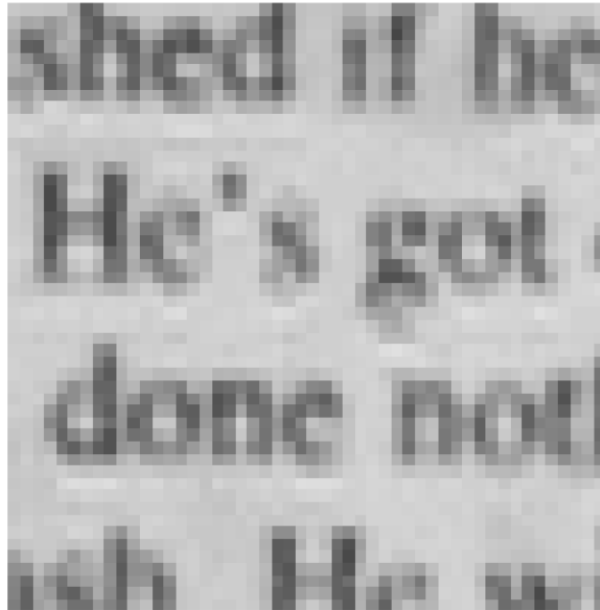


Figure 10. (Left) A small section of one of 15 low-resolution input images. (Right) An estimate of the texture at 2.0 times higher resolution using the MAP estimator.

D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom.
CVPR 1998.

2D projective transformation

Linear transformation of points in homogeneous coordinates

$$\mathbf{x}' = \mathbf{H}\mathbf{x}$$
$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix}$$
$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} xh_{11} + yh_{12} + wh_{13} \\ xh_{21} + yh_{22} + wh_{23} \\ xh_{31} + yh_{32} + wh_{33} \end{bmatrix}$$

Nonlinear transformation of points in inhomogeneous coordinates

$$\tilde{x} = \frac{x'}{w'} = \frac{xh_{11} + yh_{12} + wh_{13}}{xh_{31} + yh_{32} + wh_{33}} \text{ and}$$
$$\tilde{y} = \frac{y'}{w'} = \frac{xh_{21} + yh_{22} + wh_{23}}{xh_{31} + yh_{32} + wh_{33}}$$

Do not forget this is a nonlinear transformation

Two view geometry

- Rotation about the same camera center
 - Second camera translation vector

The translation $\mathbf{t}'$ of the second camera (or view) rotating about the same camera center $\tilde{\mathbf{C}}$ of the first camera (or view) is given by

$$\begin{aligned}\tilde{\mathbf{C}} &= -\mathbf{R}^\top \mathbf{t} = -\mathbf{R}'^\top \mathbf{t}' \\ \mathbf{R}' \mathbf{R}^\top \mathbf{t} &= \mathbf{t}'\end{aligned}$$

where

$\mathbf{R}$	First camera rotation matrix
$\mathbf{t}$	First camera translation vector
$\mathbf{R}'$	Second camera rotation matrix
$\mathbf{t}'$	Second camera translation vector

Camera projection

Project 3D point in world coordinate frame $\mathbf{X}$ under camera projection matrix $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$

$$\mathbf{x} = \mathbf{P}\mathbf{X}$$

$$\mathbf{x} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\mathbf{X}$$

$$\mathbf{K}^{-1}\mathbf{x} = [\mathbf{R} \mid \mathbf{t}]\mathbf{X}$$

$$\hat{\mathbf{x}} = \hat{\mathbf{P}}\mathbf{X}$$

Camera
calibration
matrix



where

Camera projection matrix

$$\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}] = \mathbf{K}\hat{\mathbf{P}}$$

Normalized camera projection matrix

$$\hat{\mathbf{P}} = [\mathbf{R} \mid \mathbf{t}] = \mathbf{K}^{-1}\mathbf{P}$$

Image point in pixel coordinates

$$\mathbf{x} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\mathbf{X} = \mathbf{K}\hat{\mathbf{x}}$$

Image point in normalized coordinates

$$\hat{\mathbf{x}} = [\mathbf{R} \mid \mathbf{t}]\mathbf{X} = \mathbf{K}^{-1}\mathbf{x}$$

Rotation about the same camera center

Calibrated cameras

Given two cameras, the first camera and the second camera, rotating about the same camera center $\tilde{\mathbf{C}}$, a point $\hat{\mathbf{x}}$ in normalized coordinates in the first image maps to a point $\hat{\mathbf{x}}'$ in normalized coordinates in the second image as follows.

$$\mathbf{x} = \mathbf{K}\mathbf{R}[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}$$

$$\mathbf{K}^{-1}\mathbf{x} = \mathbf{R}[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}$$

$$\hat{\mathbf{x}} = \mathbf{R}[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}, \text{ where } \hat{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{x}$$

$$\mathbf{R}^\top \hat{\mathbf{x}} = [\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}$$

$$\mathbf{x}' = \mathbf{K}'\mathbf{R}'[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}$$

$$\mathbf{K}'^{-1}\mathbf{x}' = \mathbf{R}'[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}$$

$$\hat{\mathbf{x}}' = \mathbf{R}'[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}, \text{ where } \hat{\mathbf{x}}' = \mathbf{K}'^{-1}\mathbf{x}'$$

$$\hat{\mathbf{x}}' = \mathbf{R}'\mathbf{R}^\top \hat{\mathbf{x}}$$

where

$\mathbf{K}$	First camera calibration matrix
$\mathbf{R}$	First camera rotation matrix
$\mathbf{t}$	First camera translation vector
$\mathbf{K}'$	Second camera calibration matrix
$\mathbf{R}'$	Second camera rotation matrix
$\mathbf{t}'$	Second camera translation vector

Rotation about the same camera center

Calibrated cameras

- Case of unrotated first camera

If $R = I$ (i.e., the first camera is unrotated), then

$$\begin{aligned}\hat{\mathbf{x}} &= [I \mid -\tilde{\mathbf{C}}]\mathbf{X} & \hat{\mathbf{x}}' &= R'[I \mid -\tilde{\mathbf{C}}]\mathbf{X} \\ \hat{\mathbf{x}}' &= R'\hat{\mathbf{x}}\end{aligned}$$

which is often written as $\hat{\mathbf{x}}' = R\hat{\mathbf{x}}$, where R is the rotation of the second camera
Mapping in pixel coordinates

$$\begin{aligned}\hat{\mathbf{x}}' &= R\hat{\mathbf{x}}, \text{ where } \hat{\mathbf{x}} = K^{-1}\mathbf{x} \text{ and } \hat{\mathbf{x}}' = K'^{-1}\mathbf{x}' \\ K'^{-1}\mathbf{x}' &= RK^{-1}\mathbf{x} \\ \mathbf{x}' &= K'RK^{-1}\mathbf{x} \\ \mathbf{x}' &= H\mathbf{x}, \text{ where } H = K'RK^{-1}\end{aligned}$$

Rotation about the same camera center

Uncalibrated cameras

Given two cameras, the first camera and the second camera, rotating about the same camera center $\tilde{\mathbf{C}}$, points $\mathbf{x}_i$ in pixel coordinates in the first image map to points $\mathbf{x}_i$ in pixel coordinates in the second image as follows.

$$\begin{aligned}\mathbf{x}_i &= \mathbf{K}\mathbf{R}[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}_i \forall i & \mathbf{x}'_i &= \mathbf{K}'\mathbf{R}'[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}_i \forall i \\ (\mathbf{K}\mathbf{R})^{-1}\mathbf{x}_i &= [\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}_i \forall i & \mathbf{x}'_i &= \mathbf{K}'\mathbf{R}'(\mathbf{K}\mathbf{R})^{-1}\mathbf{x}_i \forall i \\ \mathbf{x}'_i &= \mathbf{H}\mathbf{x}_i \forall i, \text{ where } \mathbf{H} = \mathbf{K}'\mathbf{R}'(\mathbf{K}\mathbf{R})^{-1}\end{aligned}$$

Rotation about the same camera center

Uncalibrated cameras

- Case of unrotated first camera

If $R = I$ (i.e., the first camera is unrotated), then

$$\begin{aligned} \mathbf{x}_i &= K[I \mid -\tilde{C}]\mathbf{X}_i \forall i & \mathbf{x}'_i &= K'R'[I \mid -\tilde{C}]\mathbf{X}_i \forall i \\ K^{-1}\mathbf{x}_i &= [I \mid -\tilde{C}]\mathbf{X}_i \forall i & \mathbf{x}'_i &= K'R'K^{-1}\mathbf{x}_i \forall i \\ & & \mathbf{x}'_i &= H\mathbf{x}_i \forall i, \text{ where } H = K'R'K^{-1} \end{aligned}$$

which is often written as $H = K'RK^{-1}$, where R is the rotation of the second camera

Note if H is determined from point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ in pixel coordinates, then

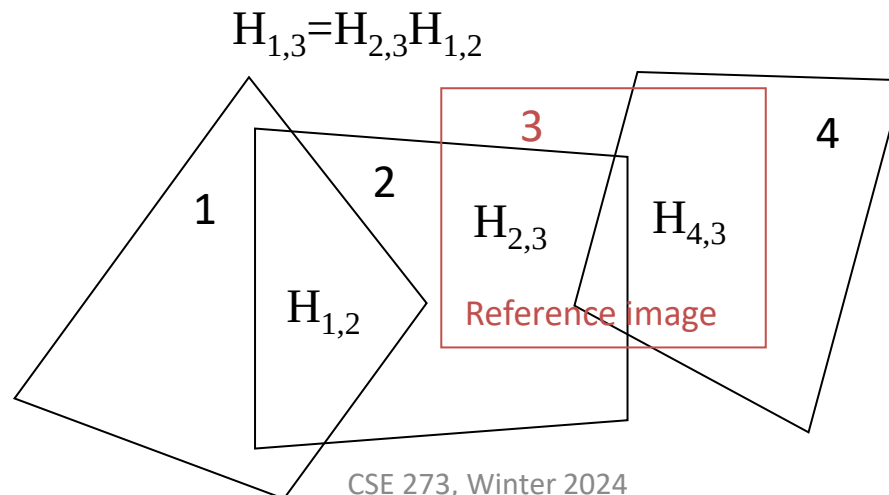
$$H = K'RK^{-1}$$

$$R = K'^{-1}HK$$

In general, the resulting R will not be in $SO(3)$. If the camera calibration matrices are known, then determine R directly from $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ in normalized coordinates.

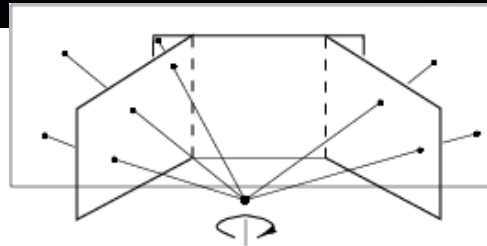
Mosaic construction from images

- Select one of the images as a reference image *ref*
- For the remaining images
 - Calculate or estimate the 2D projective transformation $H_{n,ref}$ from image *n* to *ref*
 - Compose 2D projective transformations
 - Transform image under $H_{n,ref}$



Camera and image motion

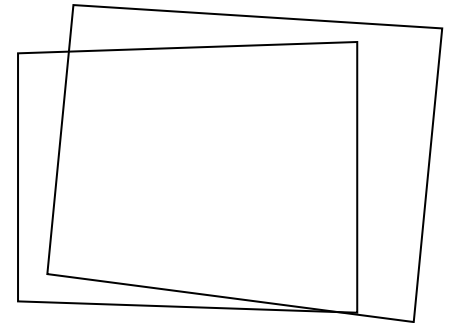
- Mosaic construction from rotating camera
 - Planar mosaic (map images to plane)



Camera and image motion

- Camera motion measurements may be inaccurate or imprecise
 - Initial estimate of image motion
 - Guided feature matching
- Camera motion may be unknown
 - Estimate camera and/or image motion from images directly
 - Direct methods
 - Feature based methods

Image motion



2D projective transform estimation

- Direct methods
 - Estimate directly from spatial derivatives and difference of image intensities using all pixels
 - B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. International Joint Conference on Artificial Intelligence, 1981.
- Feature based methods
 - Estimate from a set of feature correspondences
 - E.A. Maxwell. The methods of plane projective geometry based on the use of general homogeneous coordinates. Cambridge University Press, 1946.

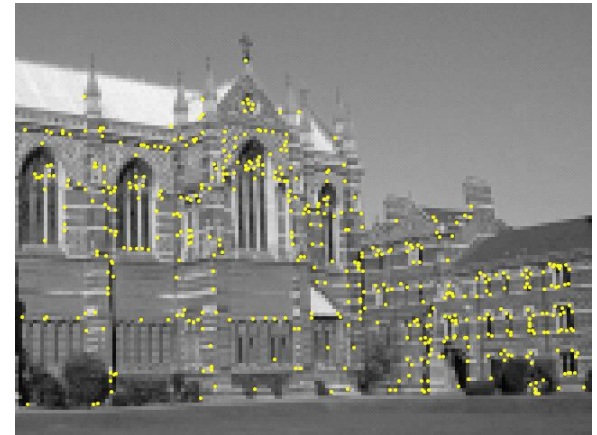
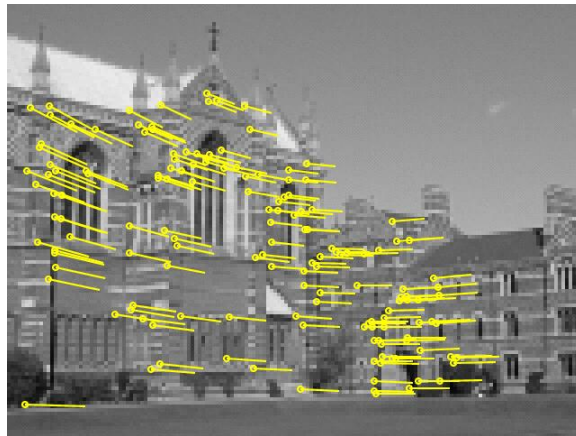
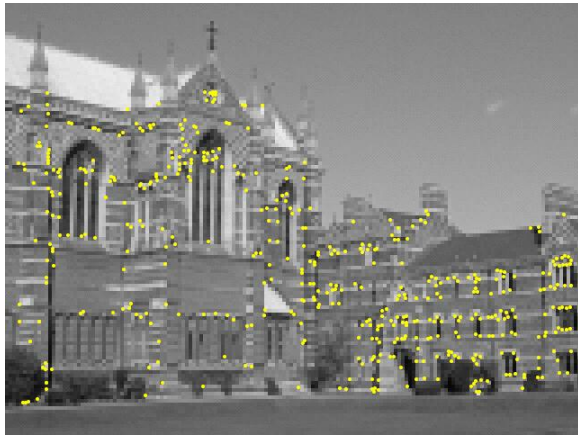
Direct methods

- Given two images J and I , estimate the transform from J to I
- Assumptions
 - Intensity constancy
$$J(x, y) = I(x + u(x, y), y + v(x, y))$$
 - Small displacement (u, v)
$$I_x u + I_y v + I_t = 0$$
- Requires pyramidal implementation
 - Coarse-to-fine iterative estimation refinement
 - J.R. Bergen et al. Hierarchical model-based motion estimation. ECCV 1992.



Feature based methods

- Given a set of feature correspondences between two images J and I , estimate the transform from J to I

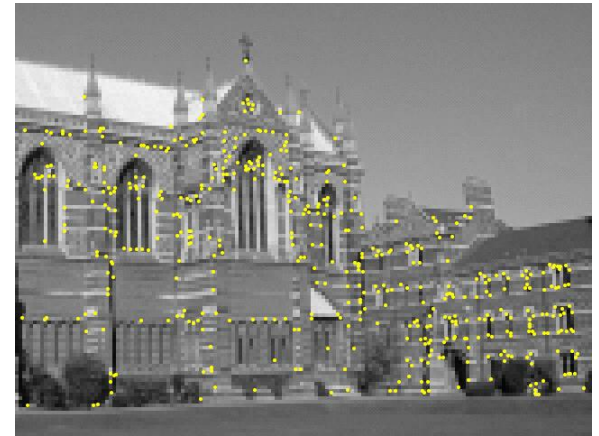
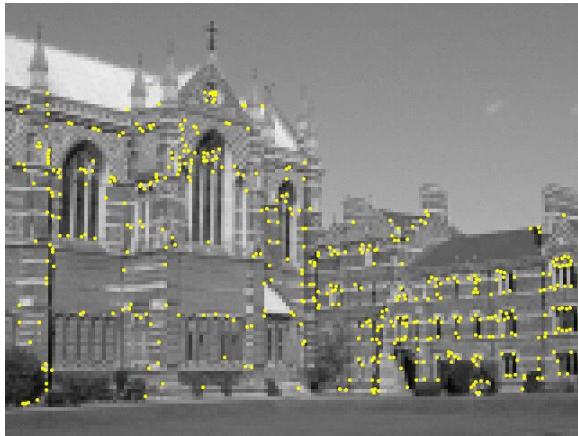


Feature based methods

- Establish feature correspondences
 - Feature detection
 - Feature descriptors
 - Putative feature correspondences
 - Feature matching (images) or tracking (video)
 - Outlier rejection
 - Random sample consensus (RANSAC) or M-estimator sample consensus (MSAC)
 - Minimal solution of model (e.g., 2D projective transformation)
- Model estimation
 - Linear and/or nonlinear solution

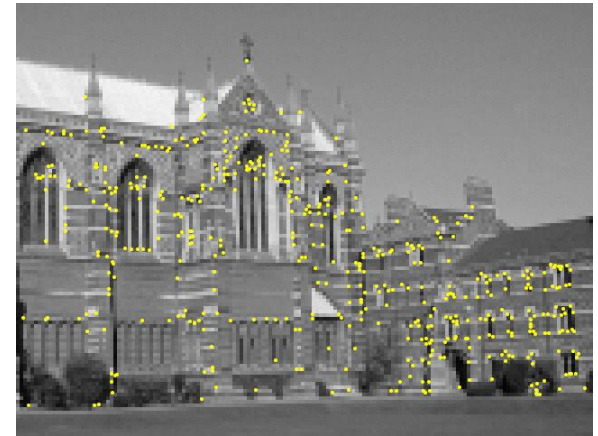
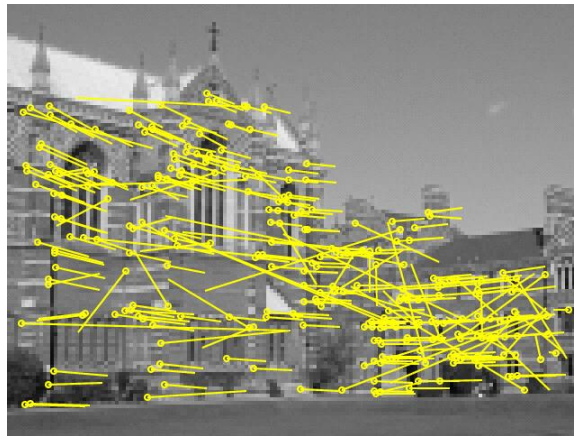
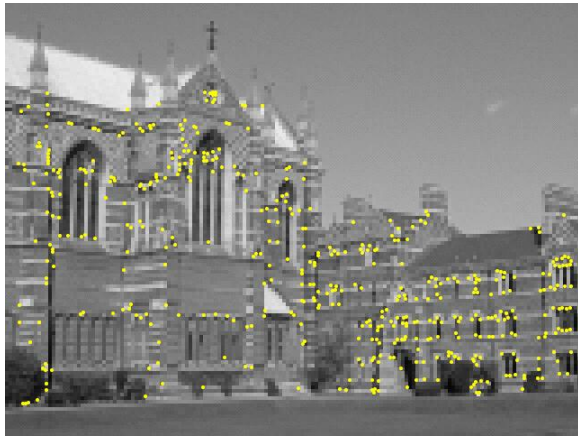
Feature based methods

- Feature detection



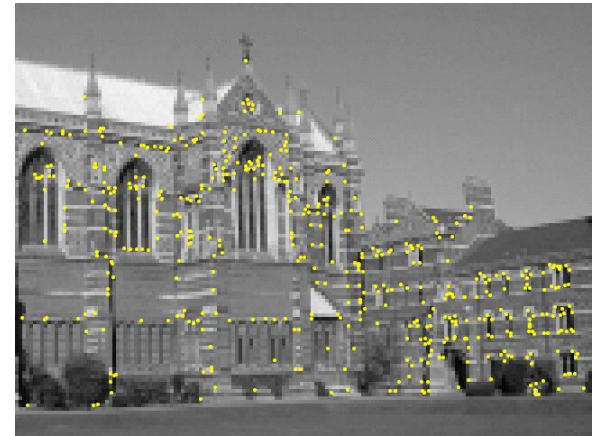
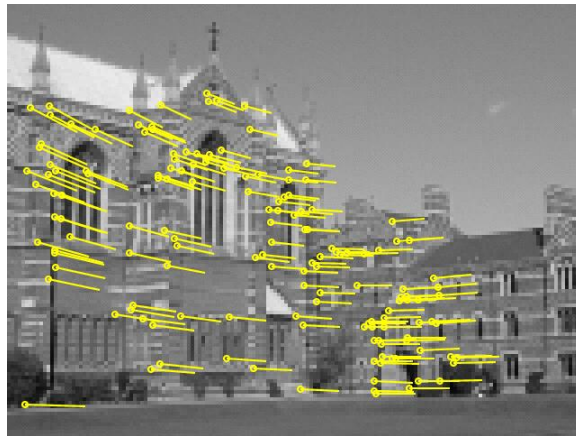
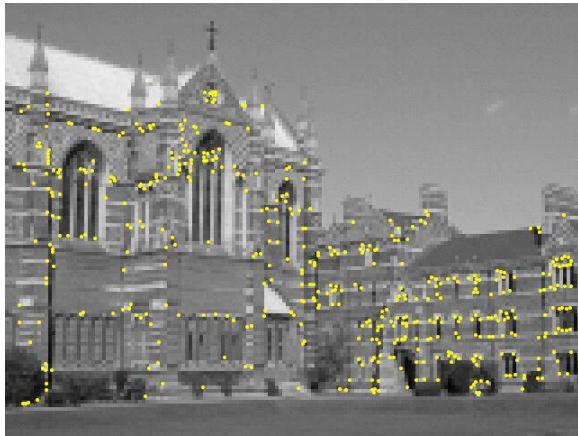
Feature based methods

- Putative feature correspondences



Feature based methods

- Outlier rejection

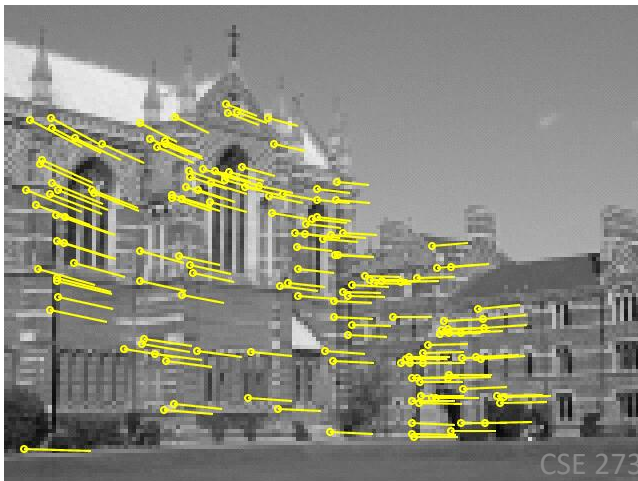


Outlier rejection

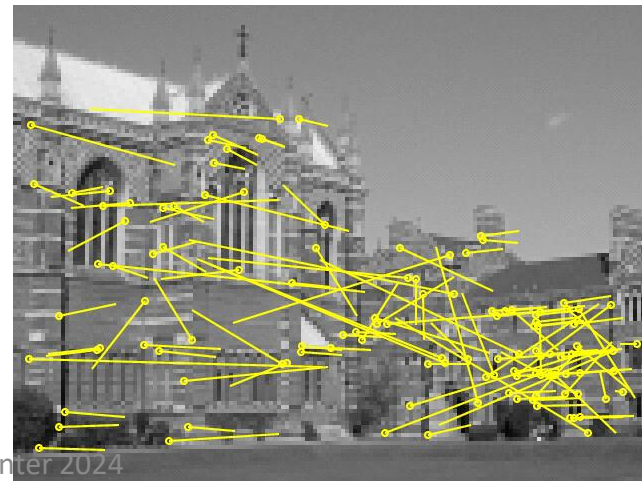
Putative feature correspondences



Inliers

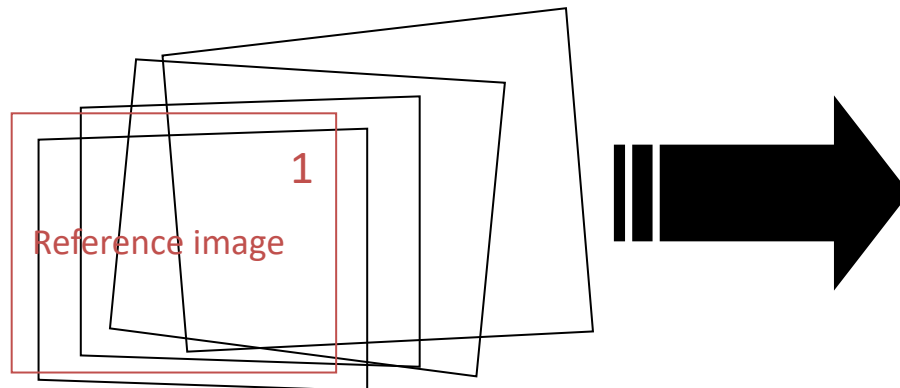


Outliers



Mosaic construction from video

- Set image 1 as the reference image
- For the remaining images
 - Sequentially estimate the 2D projective transformation $H_{n,n-1}$ from the current image n to the previous image $n-1$
 - Compose 2D projective transformations $H_{n,1} = H_{n-1,1}H_{n,n-1}$
 - Transform the image under $H_{n,1}$



Camera and image motion

- Mosaic construction from video

Misalignment

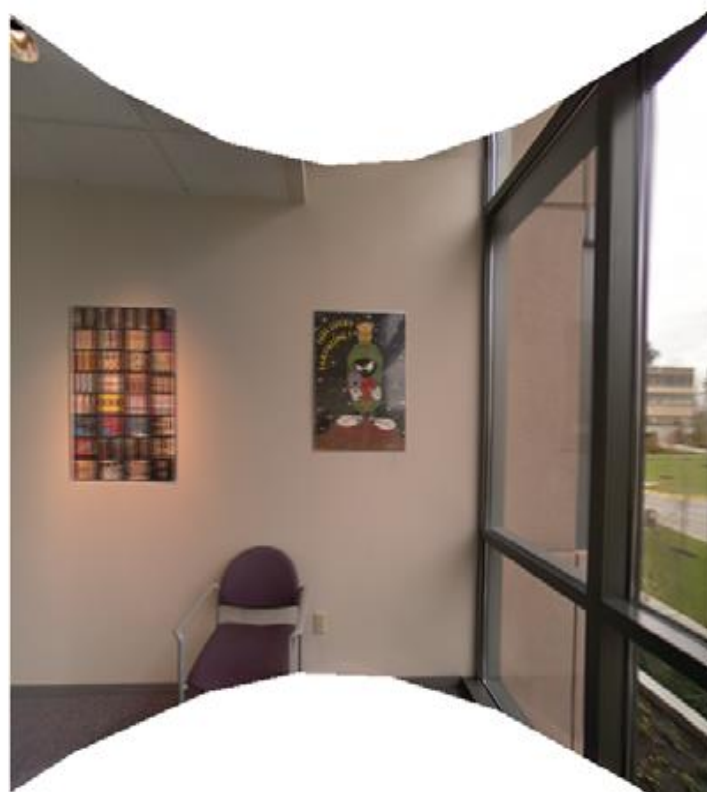
Composition of transformations accumulates error of each transformation estimate



Optimization

- Minimization of all errors
 - Simultaneous adjustment of all parameters
 - Cannot be performed sequentially
- Direct
 - H.-Y. Shum and R. Szeliski. Construction and refinement of panoramic mosaics with global and local alignment. ICCV 1998.
- Feature based
 - 2D block adjustment

Optimization to align all images



Three images

Camera and image motion

- Mosaic construction from rotating camera
 - Cylindrical mosaic
 - Map images to inside of cylinder
 - Spherical mosaic
 - Map images to inside of sphere



Camera and image motion

- Mosaic construction from rotating camera
 - Cylindrical mosaic
 - Map images to inside of cylinder, then unroll cylinder



Camera and image motion

- Mosaic construction from rotating camera
 - Spherical mosaic (map images to inside of sphere)



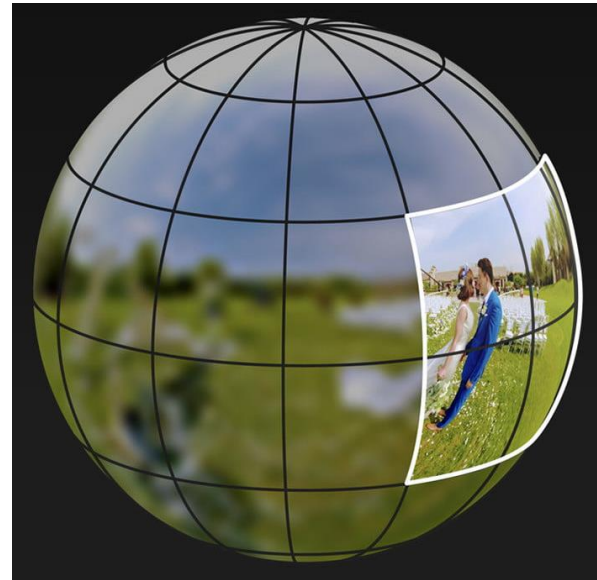
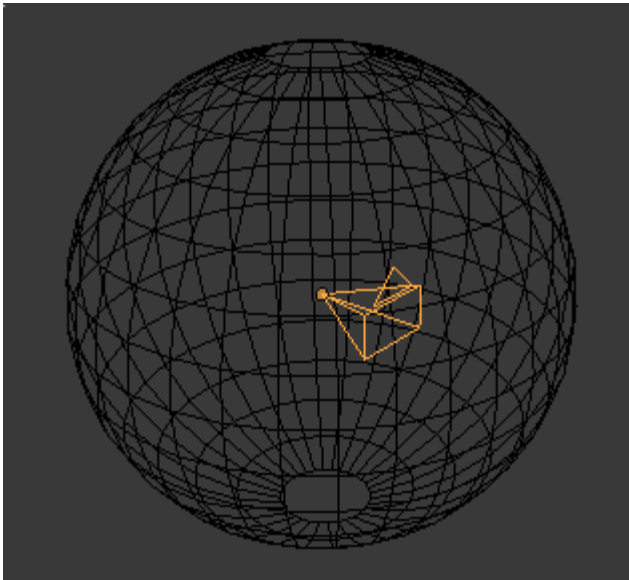
Camera and image motion

- Mosaic construction from rotating camera
 - Spherical mosaic (map images to inside of sphere)



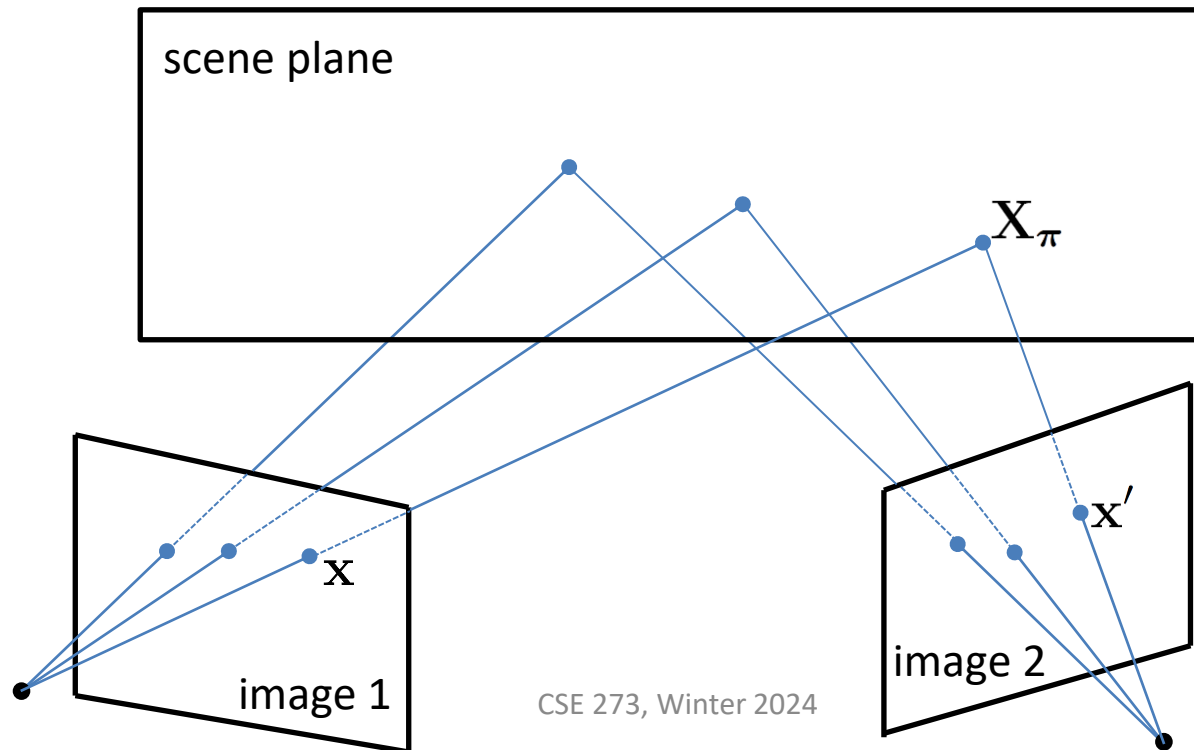
Camera and image motion

- Mosaic construction from rotating camera
 - Enables image synthesis from virtual camera with same camera center as physical camera



Camera and image motion

- Camera motion: arbitrary, but imaging a scene plane
- Image motion: 2D projective transformation



Imaging a plane

- Calibrated cameras

The cameras $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$ and $\mathbf{P}' = \mathbf{K}'[\mathbf{R}' \mid \mathbf{t}']$ image the points $\mathbf{X}_{\pi,i}$ on the plane $\boldsymbol{\pi} = (\mathbf{n}^\top, d)^\top$ as

$$\mathbf{x}_i = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\mathbf{X}_{\pi,i} \forall i \text{ and } \mathbf{x}'_i = \mathbf{K}'[\mathbf{R}' \mid \mathbf{t}']\mathbf{X}_{\pi,i} \forall i$$

2D projective transformation

$$\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i \forall i, \text{ where } \mathbf{H} = \mathbf{K}'\mathbf{R}' \left(\mathbf{I} - (\mathbf{R}'^\top \mathbf{t}' - \mathbf{R}^\top \mathbf{t}) \left(\frac{\mathbf{n}}{d - \mathbf{t}^\top \mathbf{R}\mathbf{n}} \right)^\top \right) \mathbf{R}^\top \mathbf{K}^{-1}$$

Imaging a plane

- Uncalibrated cameras

The points $\mathbf{X}_{\pi,i}$ on the plane π may be written as

$$\mathbf{X}_{\pi,i} = \mathbf{M}\mathbf{x}_{\pi,i} \forall i, \text{ where } \mathbf{M} \text{ is the null space of } \pi^\top$$

Projection

$$\mathbf{x}_i = \mathbf{P}\mathbf{X}_{\pi,i} \forall i$$

$$\mathbf{x}_i = \mathbf{P}\mathbf{M}\mathbf{x}_{\pi,i} \forall i$$

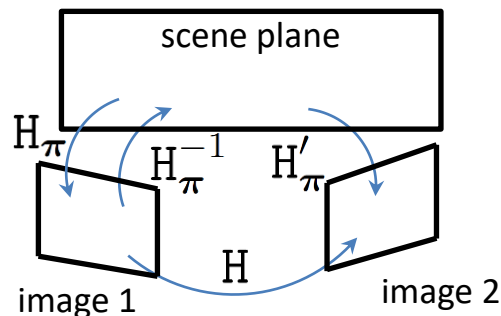
$$\mathbf{x}_i = \mathbf{H}_\pi \mathbf{x}_{\pi,i} \forall i, \text{ where } \mathbf{H}_\pi = \mathbf{P}\mathbf{M}$$

$$\mathbf{x}'_i = \mathbf{P}'\mathbf{X}_{\pi,i} \forall i$$

$$\mathbf{x}'_i = \mathbf{P}'\mathbf{M}\mathbf{x}_{\pi,i} \forall i$$

$$\mathbf{x}'_i = \mathbf{H}'_\pi \mathbf{x}_{\pi,i} \forall i, \text{ where } \mathbf{H}'_\pi = \mathbf{P}'\mathbf{M}$$

2D projective transformation



$$\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i \forall i, \text{ where } \mathbf{H} = \mathbf{H}'_\pi \mathbf{H}_\pi^{-1}$$

$$\mathbf{H} = \mathbf{P}'\mathbf{M}(\mathbf{P}\mathbf{M})^{-1}$$

Two view geometry

- Same model for rotation about the same camera center and imaging a plane
 - 2D projective transformation

3×3 homogeneous (i.e., defined up to nonzero scale) matrix (8 degrees of freedom)

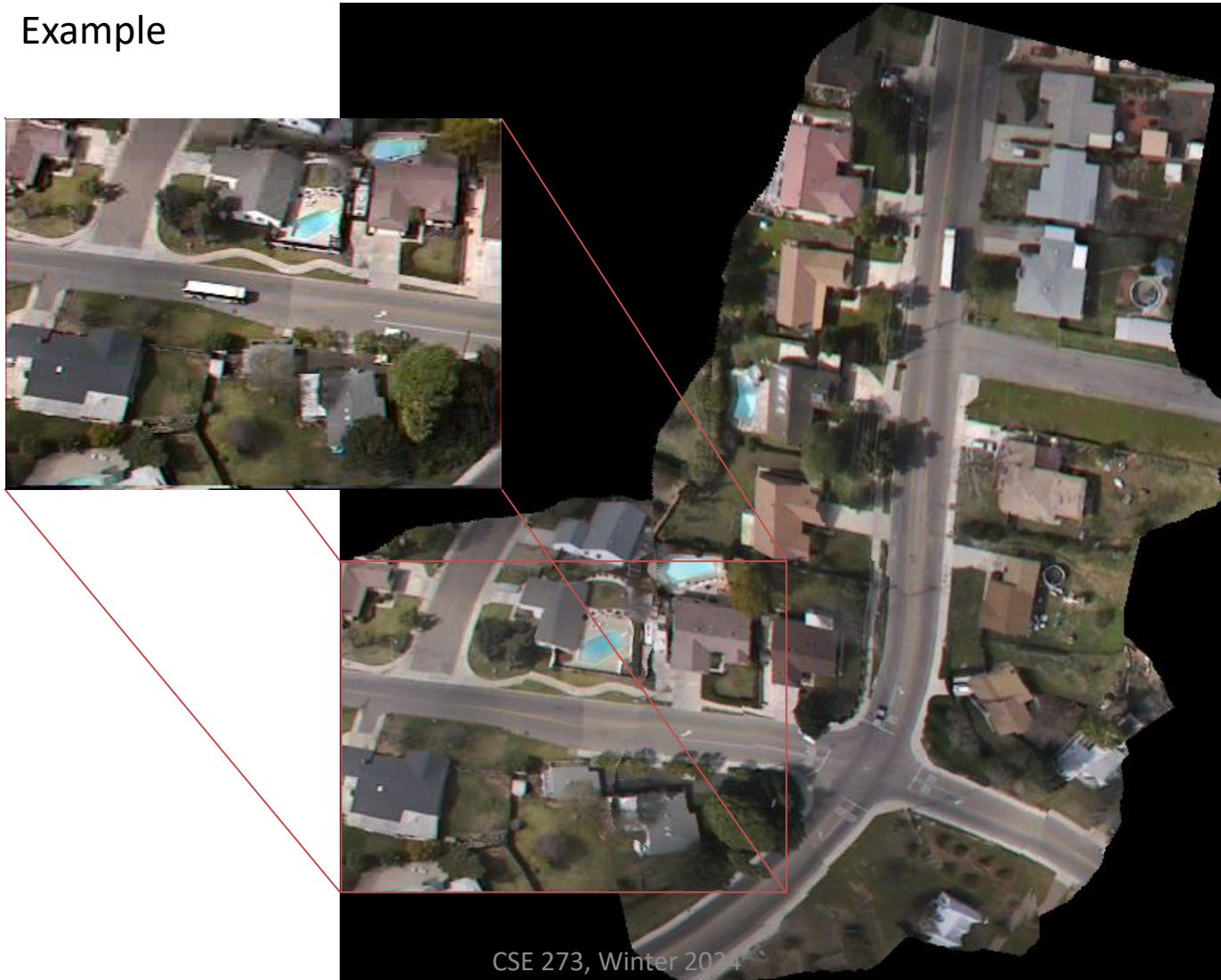
$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_1 \\ a_{21} & a_{22} & t_2 \\ v_1 & v_2 & u \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix}$$

$$\mathbf{x}' = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^\top & u \end{bmatrix} \mathbf{x}, \text{ where } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \mathbf{t} = (t_1, t_2)^\top, \text{ and } \mathbf{v} = (v_1, v_2)^\top$$

$$\mathbf{x}' = \mathbf{H}_P \mathbf{x}, \text{ where } \mathbf{H}_P = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^\top & u \end{bmatrix}$$

Mosaic construction from video

Example



Mosaic construction

Misalignment

Composition of transformations accumulates error of each transformation estimate

Follow with optimization to align all video frames



Complete mosaic with blending



Limitation

- 2D projective transformation is only valid for rotating camera or planar scene
- Nonplanar scene and translating camera requires 3D model



Issue: dynamic scene

- Stationary objects are aligned, moving objects are not aligned
- Burst photography
 - Hyperspectral image to RGB image
 - High dynamic range imaging
 - Super resolution
- Camera and image motion
 - Mosaic construction from video (or a set of images)

High dynamic range imaging of dynamic scene

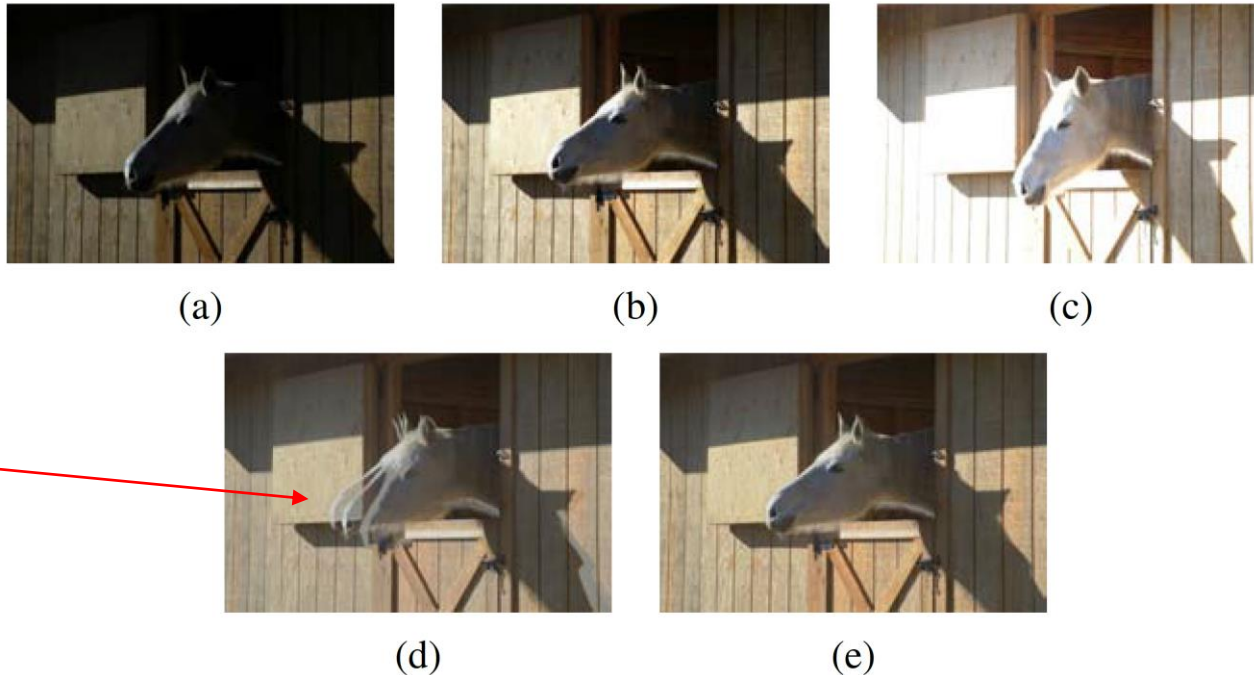


Figure 10.16 Merging multiple exposures to create a high dynamic range composite (Kang, Uyttendaele, Winder *et al.* 2003): (a–c) three different exposures; (d) merging the exposures using classic algorithms (note the ghosting due to the horse’s head movement); (e) merging the exposures with motion compensation.

Constructing mosaic of dynamic scene



(a)



(b)



(c)



(d)

Figure 9.14 Final composites computed by a variety of algorithms (Szeliski 2006a): (a) average, (b) median, (c) feathered average, (d) p -norm $p = 10$, (e) Voronoi, (f) weighted ROD vertex cover with feathering, (g) graph cut seams with Poisson blending and (h) with pyramid blending.

Constructing mosaic of dynamic scene



(e)



(f)



(g)



(h)

Figure 9.14 Final composites computed by a variety of algorithms (Szeliski 2006a): (a) average, (b) median, (c) feathered average, (d) p -norm $p = 10$, (e) Voronoi, (f) weighted ROD vertex cover with feathering, (g) graph cut seams with Poisson blending and (h) with pyramid blending.

Video stabilization

- Estimate frame to frame motion (e.g., similarity transformation)
 - Parameterize motion (e.g., scale, rotation, translation)
 - Over time window centered at current frame
 - Compose transformations such that current frame is reference frame
 - Low pass filter (i.e., smooth) parameters
 - Apply transformation comprised of filtered parameters

Camera and image motion

- Video stabilization



Original video

Stabilized video

Next lecture

- Camera arrays and light field photography