Image formation

Monday March 21

Prof. Kristen Grauman
UT-Austin

Announcements

• Reminder: Pset 3 due March 30
• Midterms: pick up today

Recap: Features and filters

Transforming and describing images; textures, colors, edges

Recap: Grouping & fitting

Clustering, segmentation, fitting; what parts belong together?

Recap: Features and filters

Transforming and describing images; textures, colors, edges

Multiple views

Multi-view geometry, matching, invariant features, stereo vision

Multiple views

Multi-view geometry, matching, invariant features, stereo vision

Plan

• Today:
  – Local feature matching between views (wrap-up)
  – Image formation, geometry of a single view

• Wednesday: Multiple views and epipolar geometry
• Monday: Approaches for stereo correspondence
Previously

Local invariant features
- Detection of interest points
  - Harris corner detection
  - Scale invariant blob detection: LoG
- Description of local patches
  - SIFT: Histograms of oriented gradients
- Matching descriptors

Recall: Corners as distinctive interest points

- "edge": \( \lambda_1 \gg \lambda_2 \), \( \lambda_2 \gg \lambda_1 \)
- "corner": \( \lambda_1 \) and \( \lambda_2 \) are large, \( \lambda_1 \approx \lambda_2 \)
- "flat" region \( \lambda_1 \) and \( \lambda_2 \) are small;

One way to score the cornerness:

\[
f = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}
\]

Local features: main components

1) Detection: Identify the interest points

2) Description: Extract vector feature descriptor surrounding each interest point.

3) Matching: Determine correspondence between descriptors in two views

Kristen Grauman

"flat" region
\( \lambda_1 \) and \( \lambda_2 \) are small;

Harris Detector: Steps

Compute corner response \( f \)
Blob detection in 2D: scale selection

Laplacian-of-Gaussian = “blob” detector

\[
\nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}
\]

Filter scales

img1, img2, img3

Blob detection in 2D

We define the characteristic scale as the scale that produces peak of Laplacian response

Example

Original image at 1/4 the size

Scale invariant interest points

Interest points are local maxima in both position and scale.

Local features: main components

1) Detection: Identify the interest points

2) Description: Extract vector feature descriptor surrounding each interest point.

3) Matching: Determine correspondence between descriptors in two views

SIFT descriptor [Lowe 2004]

- Use histograms to bin pixels within sub-patches according to their orientation.

Why subpatches? Why does SIFT have some illumination invariance?

Making descriptor rotation invariant

- Rotate patch according to its dominant gradient orientation
- This puts the patches into a canonical orientation.

SIFT descriptor [Lowe 2004]

- Extraordinarily robust matching technique
  - Can handle changes in viewpoint
  - Up to about 60 degree out of plane rotation
  - Can handle significant changes in illumination
  - Sometimes even day vs. night (below)
  - Fast and efficient—can run in real time
  - Lots of code available

SIFT properties

- Invariant to
  - Scale
  - Rotation
- Partially invariant to
  - Illumination changes
  - Camera viewpoint
  - Occlusion, clutter

Local features: main components

1) Detection: Identify the interest points

2) Description: Extract vector feature descriptor surrounding each interest point.

3) Matching: Determine correspondence between descriptors in two views

Matching local features

Steve Seitz

Kristen Grauman
Matching local features

To generate **candidate matches**, find patches that have the most similar appearance (e.g., lowest SSD)

Simplest approach: compare them all, take the closest (or closest k, or within a thresholded distance)

Ambiguous matches

At what SSD value do we have a good match?
To add robustness to matching, can consider **ratio**: distance to best match / distance to second best match
If low, first match looks good.
If high, could be ambiguous match.

Matching SIFT Descriptors

- Nearest neighbor (Euclidean distance)
- Threshold ratio of nearest to 2nd nearest descriptor

Recap: robust feature-based alignment

- Extract features
- Compute **putative matches**
Recap: robust feature-based alignment

- Extract features
- Compute putative matches
- Loop:
  - Hypothesize transformation $T$ (small group of putative matches that are related by $T$)
  - Verify transformation (search for other matches consistent with $T$)

Source: L. Lazebnik

Applications of local invariant features

- Wide baseline stereo
- Motion tracking
- Panoramas
- Mobile robot navigation
- 3D reconstruction
- Recognition
- ...

Automatic mosaicing

Wide baseline stereo

http://www.cs.ubc.ca/~mbrown/autostitch/autostitch.html

[Image from T. Tuytelaars ECCV 2006 tutorial]
Recognition of specific objects, scenes

Plan

- Today:
  - Local feature matching between views (wrap-up)
  - Image formation, geometry of a single view
- Wednesday: Multiple views and epipolar geometry
- Monday: Approaches for stereo correspondence

Image formation

- How are objects in the world captured in an image?

Physical parameters of image formation

- Geometric
  - Type of projection
  - Camera pose
- Optical
  - Sensor’s lens type
  - Focal length, field of view, aperture
- Photometric
  - Type, direction, intensity of light reaching sensor
  - Surfaces’ reflectance properties

Image formation

- Let’s design a camera
  - Idea 1: put a piece of film in front of an object
  - Do we get a reasonable image?

Pinhole camera

- Add a barrier to block off most of the rays
  - This reduces blurring
  - The opening is known as the aperture
  - How does this transform the image?
Pinhole camera

- Pinhole camera is a simple model to approximate imaging process, perspective projection.

If we treat pinhole as a point, only one ray from any given point can enter the camera.

Camera obscura

- Reinerus Gemma-Frisius, observed an eclipse of the sun at Louvain on January 24, 1544, and later he used this illustration of the event in his book De Radio Astronomia et Geometrica, 1545. It is thought to be the first published illustration of a camera obscura...

Hammond, John H., The Camera Obscura, A Chronicle

In Latin, means ‘dark room’

Perspective effects

- Far away objects appear smaller
**Perspective effects**
- Parallel lines in the scene intersect in the image
- Converge in image on horizon line

**Projection properties**
- Many-to-one: any points along same ray map to same point in image
- Points \(\rightarrow\) points
- Lines \(\rightarrow\) lines (collinearity preserved)
- Distances and angles are **not** preserved

- Degenerate cases:
  - Line through focal point projects to a point.
  - Plane through focal point projects to line
  - Plane perpendicular to image plane projects to part of the image.

**Perspective and art**
- Use of correct perspective projection indicated in 1st century B.C. frescoes
- Skill resurfaces in Renaissance: artists develop systematic methods to determine perspective projection (around 1480-1515)

**Perspective projection equations**
- 3d world mapped to 2d projection in image plane

**Homogeneous coordinates**
- Is this a linear transformation? **no**—division by \(z\) is nonlinear
- Trick: add one more coordinate:

\[
(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}
\]

- Converting from homogeneous coordinates

\[
\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)
\]
**Perspective Projection Matrix**

- Projection is a matrix multiplication using homogeneous coordinates:

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1/f' & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
\Rightarrow \left( \frac{f'x}{z}, \frac{f'y}{z} \right)
\]

divide by the third coordinate to convert back to non-homogeneous coordinates

**Weak perspective**

- Approximation: treat magnification as constant
- Assumes scene depth \(<\) average distance to camera

\[
\begin{align*}
\text{Image plane} & : z = z_0 \\
\text{World points:} & \quad x' = \frac{f_x}{z} \quad y' = \frac{f_y}{z} \\
\text{Camera at constant distance} & \quad f = \text{focal length}
\end{align*}
\]

**Orthographic projection**

- Given camera at constant distance from scene
- World points projected along rays parallel to optical access

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z
\end{bmatrix}
\Rightarrow \left( \frac{x}{1}, \frac{y}{1} \right)
\]

**Physical parameters of image formation**

- Geometric
  - Type of projection
  - Camera pose
- Optical
  - Sensor’s lens type
  - focal length, field of view, aperture
- Photometric
  - Type, direction, intensity of light reaching sensor
  - Surfaces’ reflectance properties

**Pinhole size / aperture**

How does the size of the aperture affect the image we’d get?

Larger

Smaller

**Adding a lens**

- A lens focuses light onto the film
  - Rays passing through the center are not deviated
  - All parallel rays converge to one point on a plane located at the focal length f
Pinhole vs. lens

Cameras with lenses

- A lens focuses parallel rays onto a single focal point
- Gather more light, while keeping focus; make pinhole perspective projection practical

Thin lens equation

- How to relate distance of object from optical center (u) to the distance at which it will be in focus (v), given focal length f?

Thin lens equation

- Any object point satisfying this equation is in focus
Focus and depth of field

• Depth of field: distance between image planes where blur is tolerable.

Thin lens: scene points at distinct depths come in focus at different image planes. (Real camera lens systems have greater depth of field.)

Shapiro and Stockman

Focus and depth of field

• How does the aperture affect the depth of field?

A smaller aperture increases the range in which the object is approximately in focus.

Flower images from Wikipedia

Field of view

• Angular measure of portion of 3d space seen by the camera.


Field of view depends on focal length

• As \( f \) gets smaller, image becomes more wide angle
  - more world points project onto the finite image plane

• As \( f \) gets larger, image becomes more telescopic
  - smaller part of the world projects onto the finite image plane

Field of view depends on focal length

Size of field of view governed by size of the camera retina:

\[
\varphi = \tan^{-1}\left(\frac{d}{2f}\right)
\]

Smaller FOV = larger Focal Length

Slide from S. Seitz
Digital cameras

- Film → sensor array
- Often an array of charge coupled devices
- Each CCD is light sensitive diode that converts photons (light energy) to electrons

Summary

- Image formation affected by geometry, photometry, and optics.
- Projection equations express how world points mapped to 2d image.
  - Homogenous coordinates allow linear system for projection equations.
- Lenses make pinhole model practical.
- Parameters (focal length, aperture, lens diameter, …) affect image obtained.

Next time

- Geometry of two views