2D Matching

• Problem
  1. need to match images to maps or models
  2. need to match images to images

• Applications
  1. land use inventory matches images to maps
  2. object recognition matches image to model
  3. determining blood flow from before and after X-rays

• Methods to Study
  1. recognition by alignment
  2. pose clustering
  3. geometric hashing
  4. local focus feature matching
  5. relational matching
  6. interpretation tree
  7. discrete relaxation
Some Tools to Use

- Algebra of 2D affine transformations
  1. scaling
  2. rotation
  3. translation
  4. shear, reflection
- Least-Squares fitting to obtain best transformation parameters
- nonlinear warping
- Some algorithms of general use
  1. graph matching
  2. pose clustering
  3. discrete relaxation
  4. interpretation tree
Registration of 2D Data

A mapping between 2D spaces $M$ and $I$. $M$ may be a model and $I$ and image, but in general any 2D spaces are possible.

$$
M[x, y] = I[g(x, y), h(x, y)]
$$
$$
I[r, c] = M[g^{-1}(r, c), h^{-1}(r, c)]
$$

Definition Image registration is the process by which points of two images from similar viewpoints of essentially the same scene are geometrically transformed so that corresponding feature points of the two images have the same coordinates after transformation.
Registering Image ${}^{1}I_{t1}$ to ${}^{2}I_{t2}$

Matching control point pairs are:

| 288 210 31 160 | 232 288 95 205 | 195 372 161 229 | 269 314 112 159 |
| 203 424 199 209 | 230 336 130 196 | 284 401 180 124 | 327 428 198 69 |
| 284 299 100 146 | 337 231 45 101 | 369 223 38 64 |

The Transformation Matrix is:

$$
\begin{bmatrix}
-0.0414 & 0.773 & -119 \\
-1.120 & -0.213 & 526 \\
0.0 & 0.0 & 1.0
\end{bmatrix}
$$

Images of same scene and best affine mapping from the left image into the right image using 11 control points. $[x,y]$ coordinates for the left image with $x$ increasing downward and $y$ increasing to the right; $[u,v]$ coordinates for the right image with $u$ increasing downward and $v$ toward the right. The 11 clusters of coordinates directly below the images are the matching control points $x,y,u,v$. 
Scaling (Affine Transformations)

\[
\begin{bmatrix}
x' \\
y'
\end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} cx \\ cy \end{bmatrix} = c \begin{bmatrix} x \\ y \end{bmatrix}
\] (2)

\[
\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} c_x & 0 \\ 0 & c_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_xx \\ c_yy \end{bmatrix}
\] (3)

Scaling both coordinates of a 2D vector by scale factor 2.
Rotation (Affine Transformations)

\[
R_\theta([x, y]) = R_\theta(x[1, 0] + y[0, 1])
= xR_\theta([1, 0]) + yR_\theta([0, 1])
= x[\cos \theta, \sin \theta] + y[-\sin \theta, \cos \theta]
= [x\cos \theta - y\sin \theta, x\sin \theta + y\cos \theta]
\]

\[
\begin{bmatrix}
x' \\
y'
\end{bmatrix}
= \begin{bmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{bmatrix}
\begin{bmatrix}
x \\
y
\end{bmatrix}
= \begin{bmatrix}
x\cos \theta - y\sin \theta \\
x\sin \theta + y\cos \theta
\end{bmatrix}
\quad (4)
\]

Rotation of any 2D point in terms of rotation of the basis vectors.
Orthogonal and Orthonormal Transforms

2 Definition A set of vectors is said to be **orthogonal** if all pairs of vectors in the set are perpendicular; or equivalently, have scalar product of zero.

3 Definition A set of vectors is said to be **orthonormal** if it is an orthogonal set and if all the vectors have unit length.

- orthogonal transformations preserve angles
- orthonormal transformations preserve distances as well
- rotations and translations preserve both
- rigid transformation is combined rotation and translation
Translation Requires Homogeneous Coords

4 Definition The homogeneous coordinates of a 2D point \( \mathbf{P} = [x, y]^t \) are \([sx, sy, s]^t\), where \( s \) is a scale factor, commonly 1.0.

Matrix Representation of a 2D Translation

\[
\begin{bmatrix}
  x' \\
  y' \\
  1
\end{bmatrix} =
\begin{bmatrix}
  1 & 0 & x_0 \\
  0 & 1 & y_0 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  x \\
  y \\
  1
\end{bmatrix} =
\begin{bmatrix}
  x + x_0 \\
  y + y_0 \\
  1
\end{bmatrix}
\] (5)
Rotation, Scaling and Translation

Image \( I \) from a square-pixel camera looking vertically down on a workbench \( W \): feature points \( (x_i, y_i) \) in image coordinates need to be rotated, scaled, and translated to obtain workbench coordinates \( (x_w, y_w) \).
Modeling Shear

Map the basis vectors $[1, 0], [0, 1]$ onto $[1, e_v], [0, 1]$ and/or Map the basis vectors $[1, 0], [0, 1]$ onto $[1, 0], [e_u, 1]$.

Figure 1: (Left) v-axis shear and (right) u-axis shear.

\[
\begin{bmatrix}
x \\ y \\ 1
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 \\ e_u & 1 & 0 \\ 0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
u \\ v \\ 1
\end{bmatrix} \quad (6)
\]

\[
\begin{bmatrix}
x \\ y \\ 1
\end{bmatrix} = \begin{bmatrix}
1 & e_v & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
u \\ v \\ 1
\end{bmatrix} \quad (7)
\]
General Affine Transformations

1. We already covered *scaling*, *rotation*, *translation*, *shearing*, now we add *reflection*.

2. *Reflection* about the **u-axis** maps $[1,0]$, $[0,1]$ onto $[1,0]$, $[0,-1]$ respectively.

3. *Reflection* about the **v-axis** maps $[1,0]$, $[0,1]$ onto $[-1,0]$, $[0,1]$ respectively.

4. All operations *scaling*, *rotation*, *translation*, *shearing*, *reflection* (in 2D) have inverses of the same type (e.g. can be undone).

5. **Any combination** of *scaling*, *rotation*, *translation*, *shearing*, *reflection* is modeled by matrix multiplication—with the combined form:

   \[
   \begin{bmatrix}
   x \\
   y \\
   1
   \end{bmatrix}
   =
   \begin{bmatrix}
   a_{11} & a_{12} & a_{13} \\
   a_{21} & a_{22} & a_{23} \\
   0 & 0 & 1
   \end{bmatrix}
   \begin{bmatrix}
   u \\
   v \\
   1
   \end{bmatrix}
   \]  

   (8)

6. Ideally, these 6 parameters can be determined by matching (registering) 3 points in 2 images. In practice, a least squares solution from matching many more points gives a more robust registration formula—SEE LATER SLIDE.
Solving for an RST Mapping

5 Definition Control points are clearly distinguishable and easily measured points used to establish known correspondences between different coordinate spaces.

Assume two corresponding control points $^iP_j$ and $^wP_j$

1. determine rotation $\theta$

   (a) direction of the vector $P_1P_2$ in $I$ is determined as $\theta_i = \arctan((^i y_2 - ^i y_1)/(^i x_2 - ^i x_1))$

   (b) direction of the vector in $W$ is determined as $\theta_w = \arctan((^w y_2 - ^w y_1)/(^w x_2 - ^w x_1))$

   (c) $\theta = \theta_w - \theta_i$.

2. $\theta$ is determined: all $\sin$ and $\cos$ elements are known

3. solve 3 equations and 3 unknowns for $s$ and $x_0$, $y_0$.

\[
\begin{bmatrix}
  x_w \\
  y_w \\
  1
\end{bmatrix} = \begin{bmatrix}
  1 & 0 & x_0 \\
  0 & 1 & y_0 \\
  0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
  s & 0 & 0 \\
  0 & s & 0 \\
  0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
  \cos \theta & -\sin \theta & 0 \\
  \sin \theta & \cos \theta & 0 \\
  0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
  x_i \\
  y_i \\
  1
\end{bmatrix}
\]

\[x_w = x_i s \cos \theta - y_i s \sin \theta + x_0 \]  \hspace{1cm} (9)

\[y_w = x_i s \sin \theta + y_i s \cos \theta + y_0 \]  \hspace{1cm} (10)
Extracting a SubImage by Sampling

Distorted face of Andrew Jackson extracted from a $20 bill by defining an affine mapping with shear.

\[
\begin{bmatrix}
  x \\
  y \\
  1
\end{bmatrix}
= \begin{bmatrix}
x_0 \\
y_0
\end{bmatrix}
+ \frac{r}{n} \left( \begin{bmatrix}
x_1 \\
y_1
\end{bmatrix}
- \begin{bmatrix}
x_0 \\
y_0
\end{bmatrix} \right)
+ \frac{c}{m} \left( \begin{bmatrix}
x_2 \\
y_2
\end{bmatrix}
- \begin{bmatrix}
x_0 \\
y_0
\end{bmatrix} \right)
\]

(12)

\[
\begin{bmatrix}
x \\
y \\
1
\end{bmatrix}
= \begin{bmatrix}
(x_1 - x_0)/n \\
(y_1 - y_0)/n
\end{bmatrix}
\begin{bmatrix}
x_0 \\
y_0
\end{bmatrix}
+ \begin{bmatrix}
(x_2 - x_0)/m \\
(y_2 - y_0)/m
\end{bmatrix}
\begin{bmatrix}
x_0 \\
y_0
\end{bmatrix}
\begin{bmatrix}
r \\
c
\end{bmatrix}
\]

[r, c] subscripts the output image; [x, y] subscripts the input image. The output image has n rows and m columns. The mapping computes, for each [r, c] of output where to access the intensity in the input [x, y].
Extracting a SubImage by Sampling 2

Figure 2: (Left) Image of a scene containing signage used in Chapter 8 and (right) a new image cut out of the original using a sampling transformation.
Object Recognition by Alignment

1. Can recognize an object if we can register object model features to image features—match corners, holes, boundaries, etc.

2. Recognition also localizes the object (determines object pose $\equiv$ position and orientation).

3. Generic Recognition by Allignment
   (a) Automatically hypothesize the match of a few salient features.
   (b) Derive a mapping based on the matching features.
   (c) Verify or refute this match using other object features.
   (d) If verified, then recognition and localization accomplished; else try another hypothesis.
Object Recognition by Alignment

(Left) Model object and (right) three holes detected in an image.

**Model Point Locations and Interpoint Distances** (†coordinates are for the center of holes).

<table>
<thead>
<tr>
<th>point</th>
<th>coordinates †</th>
<th>to A</th>
<th>to B</th>
<th>to C</th>
<th>to D</th>
<th>to E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(8,17)</td>
<td>0</td>
<td>12</td>
<td>15</td>
<td>37</td>
<td>21</td>
</tr>
<tr>
<td>B</td>
<td>(16,26)</td>
<td>12</td>
<td>0</td>
<td>12</td>
<td>30</td>
<td>26</td>
</tr>
<tr>
<td>C</td>
<td>(23,16)</td>
<td>15</td>
<td>12</td>
<td>0</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>D</td>
<td>(45,20)</td>
<td>37</td>
<td>30</td>
<td>22</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>E</td>
<td>(22,1)</td>
<td>21</td>
<td>26</td>
<td>15</td>
<td>30</td>
<td>0</td>
</tr>
</tbody>
</table>

**Image Point Locations and Interpoint Distances** (†coordinates are for the center of holes).

<table>
<thead>
<tr>
<th>point</th>
<th>coordinates †</th>
<th>to H1</th>
<th>to H2</th>
<th>to H3</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>(31,9)</td>
<td>0</td>
<td>21</td>
<td>26</td>
</tr>
<tr>
<td>H2</td>
<td>(10,12)</td>
<td>21</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>H3</td>
<td>(10,24)</td>
<td>26</td>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>
Automatic Control Point Matches

- control points match by feature type
- pairs of matching control points must be at same distance
- can compute RST from one correctly matching pair

1. compute $\theta$ from the difference in heading
   The direction of the vector from $A$ to $B$ in the model is $\theta_1 = \arctan(9.0/8.0) = 0.844$ and the heading of the corresponding vector from $H_2$ to $H_3$ in the image is $\theta_2 = \arctan(12.0/0.0) = \pi/2 = 1.571$. The rotation is thus $\theta = 0.727$ radians.

2. distance matrices show that scale $s = 1$

3. compute translation $[u_0, v_0]$

$$
\begin{bmatrix}
  u \\
  v \\
  1
\end{bmatrix}
= \begin{bmatrix}
  10 \\
  12 \\
  1
\end{bmatrix}
= \begin{bmatrix}
  \cos \theta & -\sin \theta & u_0 \\
  \sin \theta & \cos \theta & v_0 \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  8 \\
  17 \\
  1
\end{bmatrix}
$$

The two resulting linear equations readily produce $u_0 = 15.3$ and $v_0 = -5.95$. 
Generalized Hough Transform

Assume affine transformation from model into image space. Compute transforms with significant supporting evidence.

1. process image to obtain items of evidence $E_j$
2. for each item of evidence $E_j$,
   vote for registration parameters $\alpha_k$ that it supports
3. significant clusters in $\alpha$-space yield registration of model to image data
A Best Affine Mapping

A general affine transformation from 2D to 2D requires six parameters and can be computed from only 3 matching pairs of points \( ([x_j, y_j], [u_j, v_j])_{j=1,3} \).

\[
\begin{bmatrix}
  u \\
  v \\
  1
\end{bmatrix}
= \begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  x \\
  y \\
  1
\end{bmatrix}
\] (14)

A Least-Squares approach will be more robust.

\[
\varepsilon(a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}) = \sum_{j=1}^{n} \left( (a_{11}x_j + a_{12}y_j + a_{13} - u_j)^2 + (a_{21}x_j + a_{22}y_j + a_{23} - v_j)^2 \right) \] (15)

Taking the six partial derivatives of the error function with respect to each of the six variables and setting this expression to zero gives us the six equations represented in matrix form:

\[
\begin{bmatrix}
  \sum x_j^2 & \sum x_j y_j & \sum x_j & 0 & 0 & 0 \\
  \sum x_j y_j & \sum y_j^2 & \sum y_j & 0 & 0 & 0 \\
  \sum x_j & \sum y_j & \sum 1 & 0 & 0 & 0 \\
  0 & 0 & 0 & \sum x_j^2 & \sum x_j y_j & \sum x_j \\
  0 & 0 & 0 & \sum y_j & \sum y_j^2 & \sum y_j \\
  0 & 0 & 0 & \sum x_j & \sum y_j & \sum 1
\end{bmatrix}
\begin{bmatrix}
  a_{11} \\
  a_{12} \\
  a_{13} \\
  a_{21} \\
  a_{22} \\
  a_{23}
\end{bmatrix}
= \begin{bmatrix}
  \Sigma u_j x_j \\
  \Sigma u_j y_j \\
  \Sigma u_j \\
  \Sigma v_j x_j \\
  \Sigma v_j y_j \\
  \Sigma v_j
\end{bmatrix}
\] (16)
Best 2D Affine Mapping

Matching control point pairs are:
288 210 31 160  232 288 95 205  195 372 161 229  269 314 112 159
203 424 199 209  230 336 130 196  284 401 180 124  327 428 198  69
284 299 100 146  337 231 45 101  369 223 38  64

The Transformation Matrix is:

\[
\begin{bmatrix}
-0.0414 & 0.773 & -119 \\
-1.120 & -0.213 & 526 \\
0.0 & 0.0 & 1.0
\end{bmatrix}
\]

Residuals (in pixels) for 22 equations are as follows:

0.18 -0.68 -1.22 0.47 -0.77 0.06 0.34 -0.51 1.09 0.04 0.96
1.51 -1.04 -0.81 0.05 0.27 0.13 -1.12 0.39 -1.04 -0.12 1.81

Images of same scene and best affine mapping from the left image into the right image using 11 control points. [x,y] coordinates for the left image with x increasing downward and y increasing to the right; [u,v] coordinates for the right image with u increasing downward and v toward the right. The 11 clusters of coordinates directly below the images are the matching control points x,y,u,v.