Alignment of Non-Overlapping Sequences
Y. Caspi and M. Irani

Presented by
Yinghua Hu
Goal

- Align temporally and spatially two non-overlapping image sequences

Example sequence 1
Example sequence 2

Aligned Sequence
Background

- Main approaches
  - Directly match image intensities
  - Match extracted local image features
  - Motion based (trajectory)
- The similarity between images
- Can two sequences without any spatial overlap be aligned?

Homography

- For the corresponding feature points $p$ and $p'$ between temporally corresponding frames $I_1$ and $I_1'$, there is an inter-camera homography $H$ which leads to

\[ p' \cong Hp \]
Problem formulation

Recover the $H$ (spatial alignment) and $\Delta t$ (temporal alignment) directly from $T_1, \ldots, T_n$ and $T_1', \ldots, T_m'$.

Perspective Projection
(Origin at image center)
Recover Spatial Alignment

- Assume $\Delta t$ is known and recover $H$.
- Two cases:
  - Scene that is planar or distant from the cameras (2D scenes).
    - homographies
  - Non-planar scene (3D scenes).
    - fundamental matrix

2D Scenes

- $P$ be a 3D point in the scene, denote by $p_i$ and $p_i'$ its image coordinates in frame $I_i$ and $I_i'$, and by $p_{i+1}$ and $p_{i+1}'$ its image coordinates in frame $I_{i+1}$ and $I_{i+1}'$.

\[
p'_{i+1} \equiv T_ip_i' \\
p_{i+1} \equiv T_ip_i \\
p_i' \equiv Hp_i \\
p_{i+1}' \equiv Hp_{i+1}
\]

- We can derive

\[
T_i' \equiv HT_iH^{-1} \text{ or } T_i' = s_iHT_iH^{-1}
\]

- Therefore

\[
eig(T_i') = s_i\eig(T_i)
\]

For $3 \times 3$ matrix $A$, $\eig(A) = [\lambda_1, \lambda_2, \lambda_3]^t$

- $s_i$ can be estimated by least squares minimization.

\[
s_iHT_iT_i'H = 0
\]
2D Scenes

Rearrange $H$ as

$$\tilde{h} = [H_{11}H_{12}H_{13}H_{21}H_{22}H_{23}H_{31}H_{32}H_{33}]^T$$

$M$\tilde{h} = \tilde{0}$

Where $M_i$ is a $9 \times 9$ matrix defined by $T_i$, $T'_i$ and $s_i$

The constraints from all the transformation $T_1, \ldots T_n$ and $T'_1, \ldots T'_n$ can be combined into a single set of linear equations:

$$Ah = 0$$

Where

$$A = \begin{bmatrix}
M_1 \\
\vdots \\
M_n
\end{bmatrix}$$

$\tilde{h}$ may be given by the eigenvector of $A^T A$ corresponding to the smallest eigenvalue.

3D scenes

The input to algorithm is two sequences of fundamental matrices. For corresponding image points $p_i, p_{i+1}$ in successive frames $I_i$ and $I_{i+1}$, there is $p_{i+1}^t F p_i = 0$.

Each fundamental matrix can be decomposed into a epipole+homography (Figure from Hartley and Zisserman’s book “Multiple View Geometry in Computer Vision”):

![Diagram showing epipolar geometry](image)

Fig. 9.5. A point $x$ in one image is transferred via the plane $\pi$ to a matching point $x'$ in the second image. The epipolar line through $x' \pi$ is obtained by joining $x'$ to the epipole $e'$. In symbols one may write $x' = H_{e'} x$ and $x' = [e'] \cdot H_{e'} x = F x$ where $F = [e'] \cdot H_{e'}$ is the fundamental matrix.
**Stereo Constraints**

**Epipolar Geometry**

3D scenes

- Homography-based constraints
  - Impose plane consistency
    - “Plane+Parallax” approach (requires a visible planar surface)
    - “Threading” method (less restrictive)
  - Epipole-based constraints
    - No issue of plane consistency
    - Every pair of corresponding epipoles imposes two linear constraints on $H$

\[
\begin{bmatrix}
  e_i' & 0' & (e_i')_x e_i' \\
  0' & e_i' & (e_i')_y e_i' \\
\end{bmatrix} \vec{h} = 0
\]

- Can be combined with homography-based constraints or used alone for solving $H$.
- If used alone, four pairs of corresponding epipoles are sufficient.
Recover Temporal Synchronization

- Similarity measure
  \[ \text{sim}(T_i, T'_i) = \frac{\text{eig}(T_i)' \text{eig}(T'_i)}{\|\text{eig}(T_i)\| \cdot \|\text{eig}(T'_i)\|} \]

- Search \( \Delta t \) to maximize the function:
  \[ SIM(\Delta t) = \sum \text{sim}(T_i, T'_{i, \Delta t})^2 \]

- Handle sequences of different frame rates
  - PAL(25 frames/sec) vs. NTSC(30 frames/sec)
  \[ SIM(\Delta t) = \sum \text{sim}(T_{5(i+1)}, T'_{6(i+1), \Delta t})^2 \]

Applications

- Alignment of non-overlapping sequences
  - Can be used to generate wide-screen movies from two narrow field-of-view movies
- Alignment of sequences obtained at different zooms
  - Two sequences display different features
  - Can be used in surveillance
- Multi-Sensor Alignment
  - Two sequences display different features
Non-overlapping sequence 1

Non-overlapping sequence 2
Alignment result

Sequence in different zooms
Alignment results for different zooms

Sequence from visible-light camera
Sequence from infra-red camera

Fused sequence from different sensors
Verify the accuracy

- A real video sequence is warped with known manipulations
- Calculate the recovered homography $H$
- Quantitative results

<table>
<thead>
<tr>
<th>Applied Transformation</th>
<th>Recovered Transformation</th>
<th>Max Residual Misalignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal shift of 352 pixels</td>
<td>Horizontal shift of 351.6 pixels</td>
<td>0.7 pixels</td>
</tr>
<tr>
<td>Zoom factor = 2</td>
<td>Zoom factor = 1.9992</td>
<td>0.4 pixels</td>
</tr>
<tr>
<td>Zoom factor = 4</td>
<td>Zoom factor = 4.0048</td>
<td>0.4 pixels</td>
</tr>
<tr>
<td>Rotation by $180^\circ$</td>
<td>Rotation by $180,00^\circ$</td>
<td>0.01 pixels</td>
</tr>
</tbody>
</table>

Main contribution

- Align two sequences without “coherent appearance”.
- Alignment is based on “coherent temporal behavior”.

Limitations

- Two cameras share the same center of projection
- Time shift $\Delta t$ is constant
- Relative camera motion is constant