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

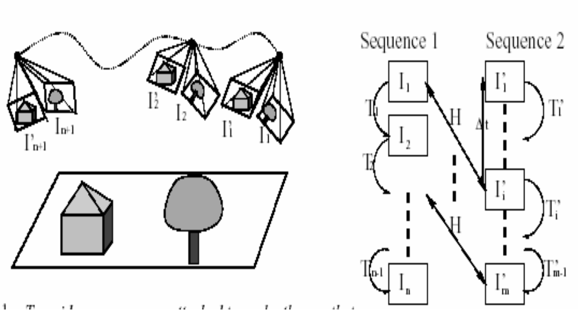
- Main approaches
  - Directly match image intensities
  - Match extracted local image features
  - Motion based (trajectory)
- The similarity between images
- Can two sequences without any spatial overlap be aligned?

Homography

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

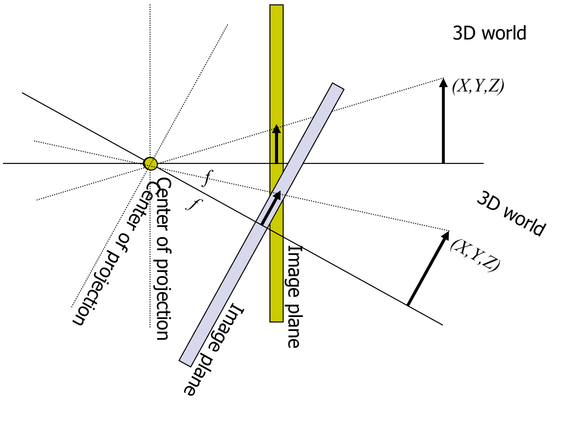
$$p' \cong Hp$$

Problem formulation



Recover the H (spatial alignment) and Δt (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 Δt 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}'.

$$p'_{i+1} \cong T_i p'_i \qquad p_{i+1} \cong T_i p_i$$

$$p'_i \cong H p_i \qquad p'_{i+1} \cong H p_{i+1}$$

- We can derive

$$T_i' \cong H T_i H^{-1} \text{ or } T_i' = s_i H T_i H^{-1}$$

- Therefore

$$eig(T_i') = s_i eig(T_i)$$

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

- s_i can be estimated by least squares minimization.

$$s_i H T_i - T_i' H = 0$$

2D Scenes

Rearrange H as $\vec{h} = [H_{11} H_{12} H_{13} H_{21} H_{22} H_{23} H_{31} H_{32} H_{33}]^t$

$$M_i \vec{h} = \vec{0}$$

Where M_i is a 9×9 matrix defined by T_i, T_i' and s_i

The constraints from all the transformation $T_1, \ldots T_n$ and T_1', … T_n' can be combined into a single set of linear equations: $A\vec{h} = \vec{0}$

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

$\vec{h}$ may be given by the eigenvector of A^tA 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 $p_{i+1}{}^t F_i p_i = 0$.

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

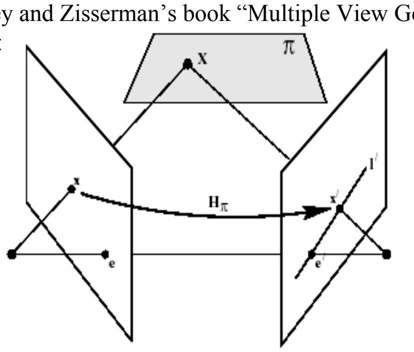
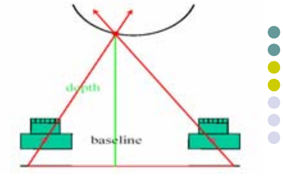
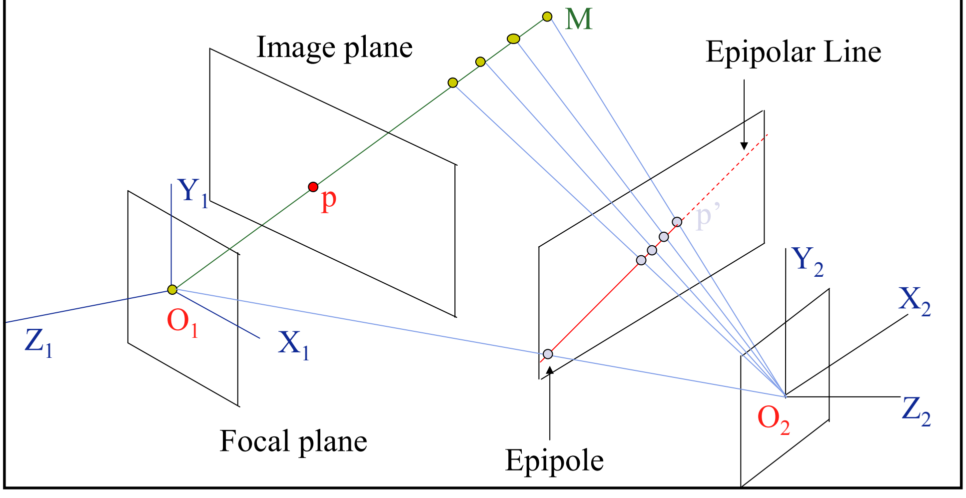


Fig. 9.5. *A point* x *in one image is transferred via the plane* π *to a matching point* x' *in the second image. The epipolar line through* x' *is obtained by joining* x' *to the epipole* e'*. In symbols one may write* $x' = H_\pi x$ *and* $l' = [e']_\times x' = [e']_\times H_\pi x = Fx$ *where* $F = [e']_\times H_\pi$ *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

$$\begin{bmatrix} e_i^t & \vec{0}^t & (e_i')_x e_i^t \\ \vec{0}^t & e_i^t & (e_i')_y e_i^t \end{bmatrix}_{2\times 9} \vec{h} = 0$$

  - 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

$$sim(T_i, T'_{i+\Delta t}) = \frac{eig(T_i)^t \, eig(T'_{i+\Delta t})}{\|eig(T_i)\| \cdot \|eig(T'_{i+\Delta t})\|}$$

- Search Δt to maximize the function:

$$SIM(\Delta t) = \sum_i sim(T_i, T'_{i+\Delta t})^2$$

- Handle sequences of different frame rates
  - PAL(25 frames/sec) vs. NTSC(30 frames/sec)

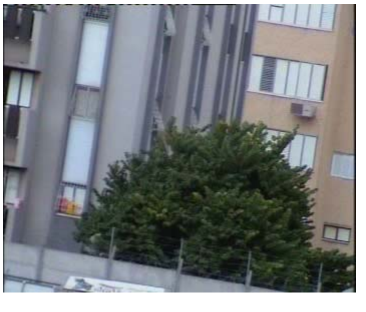
$$SIM(\Delta t) = \sum_i sim(T_{5i}^{5(i+1)}, T_{6i+\Delta t}^{6(i+1)+\Delta t})^2$$

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

Applied Transformation	Recovered Transformation	Max Residual Misalignment
Horizontal shift of 352 pixels	Horizontal shift of 351.6 pixels	0.7 pixels
Zoom factor = 2	Zoom factor = 1.9992	0.4 pixels
Zoom factor = 4	Zoom factor = 4.0048	0.4 pixels
Rotation by 180°	Rotation by 180.00°	0.01 pixels

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 Δt is constant
- Relative camera motion is constant