ASIFT:

A New Framework for Fully Affine Invariant Image Comparison

Guoshen Yu,
Ecole Polytechnique, France

Jean-Michel Morel
ENS Cachan, France

For more information,

Google ASIFT

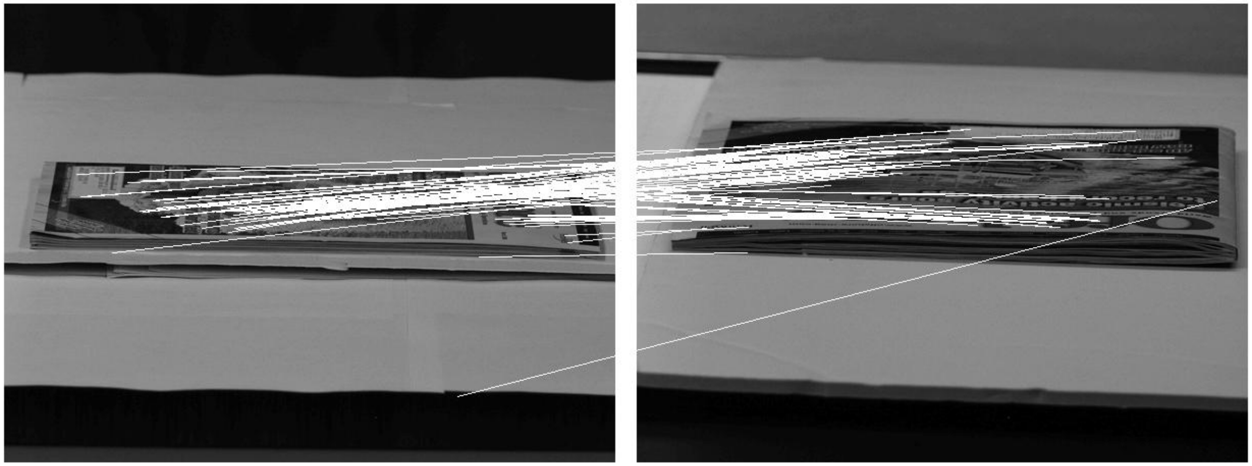
Do they contain the same objects?

Challenge: a very large transition tilt.

The transition tilt measures the amount of the viewpoint change and will be defined later.

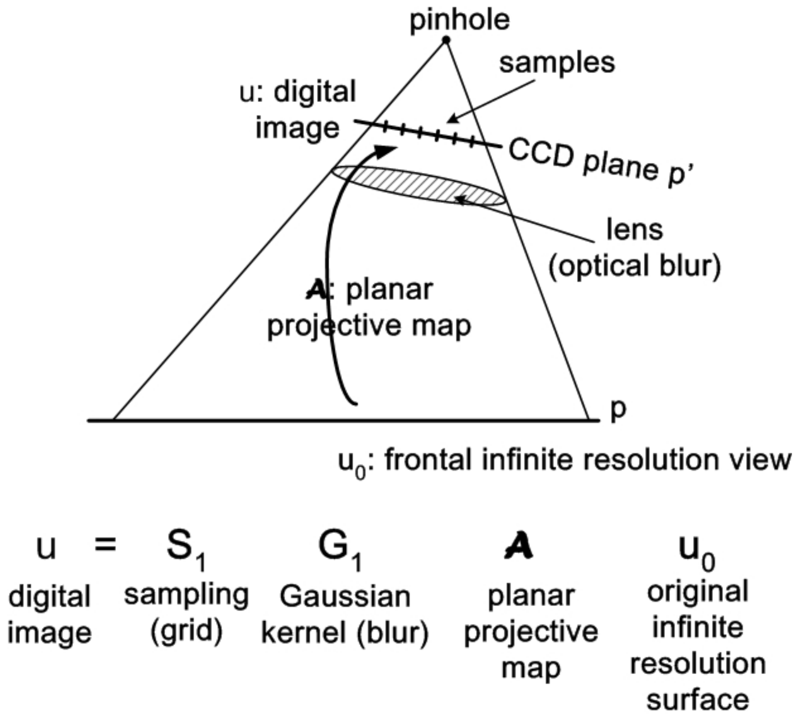
The new state of the art:

It is by now possible to recognize a solid object in a digital image, no matter what the angle and the distance, up to limits that only depend on resolution.



90 correct matches, 4 outliers. The matches were obtained by the *Affine-SIFT* method (ASIFT), a variant of the SIFT method.

Camera Model



The projective camera model $u = \mathbf{S}_1\mathbf{G}_1\mathbf{A}\mathbf{u}_0$.

- $\mathbf{A}$ is a planar projective transform (homography) .

- $\mathbf{G}_1$ is an optical (Gaussian) kernel.

- $\mathbf{S}_1$ is the CCD sampling. Shannon condition satisfied: $u = \mathbf{S}_1\mathbf{G}_1\mathbf{A}\mathbf{u}_0 \longrightarrow \mathbf{u} = \mathbf{G}_1\mathbf{A}\mathbf{u}_0$.

Affine Simplification

If the object's shape is locally smooth, local deformations in a single view can be approximated by several different local affine transforms.

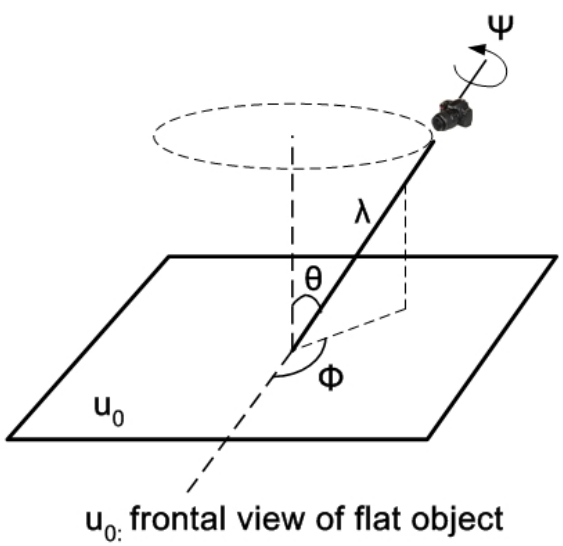
Affine transforms map rectangles to parallelograms.

Geometric Interpretation of the Six Affine Parameters

$$u = \mathbf{S}_1 \mathbf{G}_1 \mathbf{A} \mathbf{u}_0.$$

$\mathbf{A}$ is an affine map: $\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{H}_\lambda \mathbf{R}_1(\psi) \mathbf{T}_t \mathbf{R}_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$



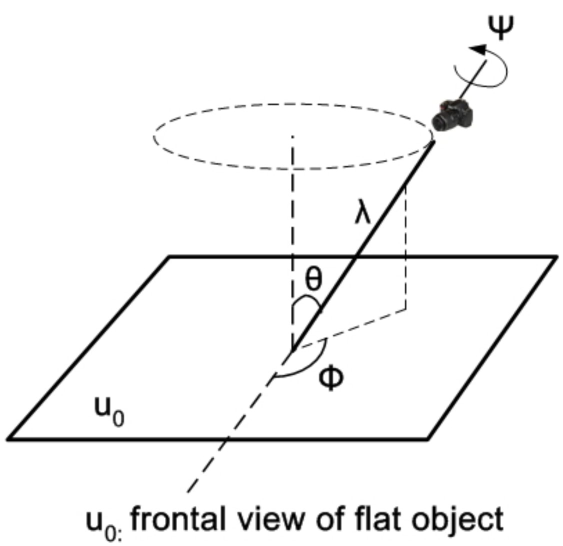
- ϕ: *longitude* angle between optical axis and a fixed vertical plane.
- $\theta = \arccos(1/t)$: *latitude* angle between optical axis and the normal to the image plane. **Tilt** $t > 1 \leftrightarrow \theta \in [0°, 90°]$.
- ψ: rotation angle of camera around optical axis.
- λ: *zoom* parameter.
- $\mathcal{T} = (e, f)^T$: translation, not presented here.

Geometric Interpretation of the Six Affine Parameters

$$u = \mathbf{S}_1 \mathbf{G}_1 \mathbf{A} \mathbf{u}_0.$$

$$\mathbf{A} \text{ is an affine map: } \begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{H}_\lambda \mathbf{R}_1(\psi) \mathbf{T}_t \mathbf{R}_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$



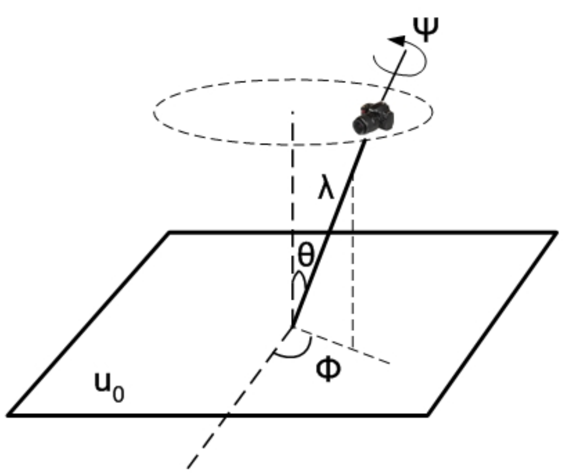
- ϕ: ***longitude*** **angle between optical axis and a fixed vertical plane.**

Geometric Interpretation of the Six Affine Parameters

$$u = \mathbf{S}_1 \mathbf{G}_1 \mathbf{A} \mathbf{u}_0.$$

A is an affine map: $\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{H}_\lambda \mathbf{R}_1(\psi) \mathbf{T}_t \mathbf{R}_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$



$u_{0:}$ frontal view of flat object

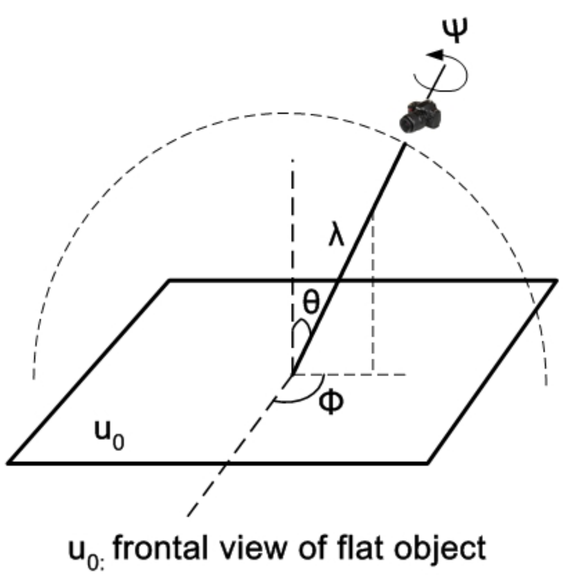
- ϕ: ***longitude*** **angle between optical axis and a fixed vertical plane.**

Geometric Interpretation of the Six Affine Parameters

$$u = \mathbf{S}_1\mathbf{G}_1\mathbf{A}\mathbf{u}_0.$$

$\mathbf{A}$ is an affine map: $\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{H}_\lambda \mathbf{R}_1(\psi)\mathbf{T}_t\mathbf{R}_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$



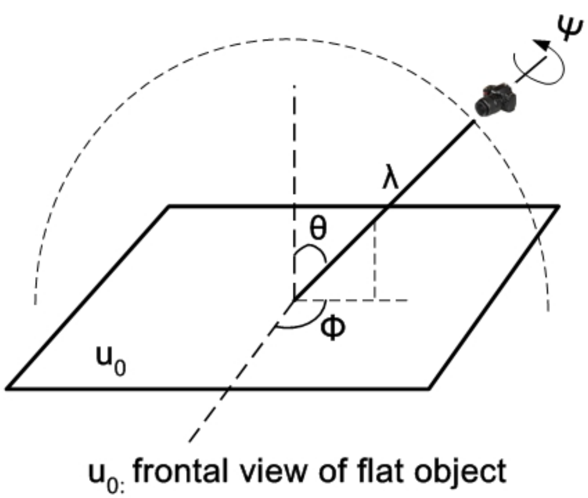
- $\theta = \arccos(1/t)$: ***latitude* angle between optical axis and the normal to the image plane. Tilt** $t > 1 \leftrightarrow \theta \in [0°, 90°]$.

Geometric Interpretation of the Six Affine Parameters

$$u = \mathbf{S}_1 \mathbf{G}_1 \mathbf{A} \mathbf{u}_0.$$

$$\mathbf{A} \text{ is an affine map: } \begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{H}_\lambda \mathbf{R}_1(\psi) \mathbf{T}_t \mathbf{R}_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$



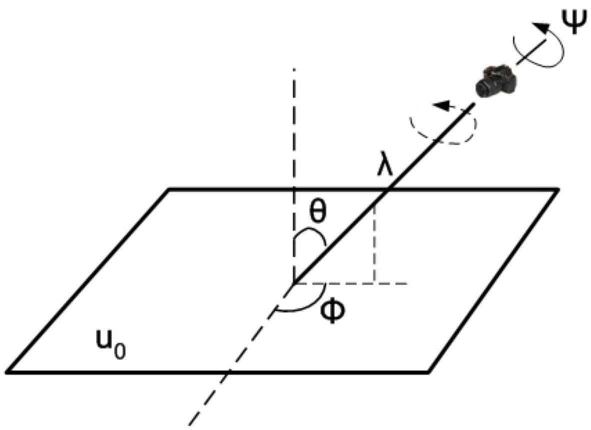
- $\theta = \arccos(1/t)$: ***latitude* angle between optical axis and the normal to the image plane. Tilt** $t > 1 \leftrightarrow \theta \in [0°, 90°]$.

Geometric Interpretation of the Six Affine Parameters

$$u = \mathbf{S}_1 \mathbf{G}_1 \mathbf{A} \mathbf{u}_0.$$

$\mathbf{A}$ is an affine map: $\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{H}_\lambda \mathbf{R}_1(\psi) \mathbf{T}_t \mathbf{R}_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$



u0: frontal view of flat object

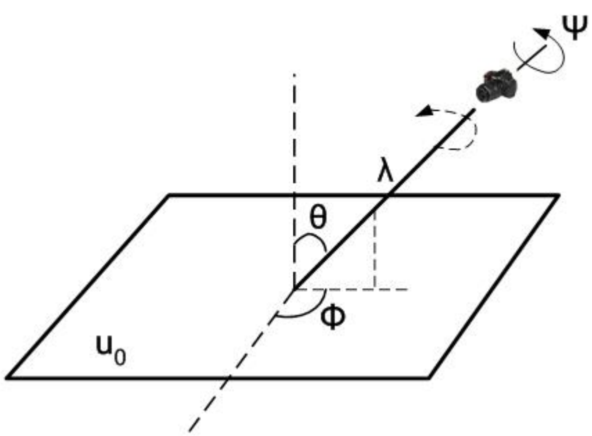
- ψ: **rotation angle of camera around optical axis.**

Geometric Interpretation of the Six Affine Parameters

$$u = \mathbf{S}_1 \mathbf{G}_1 \mathbf{A} \mathbf{u}_0.$$

A is an affine map: $\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{H}_\lambda \mathbf{R}_1(\psi) \mathbf{T}_t \mathbf{R}_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$



$u_{0:}$ frontal view of flat object

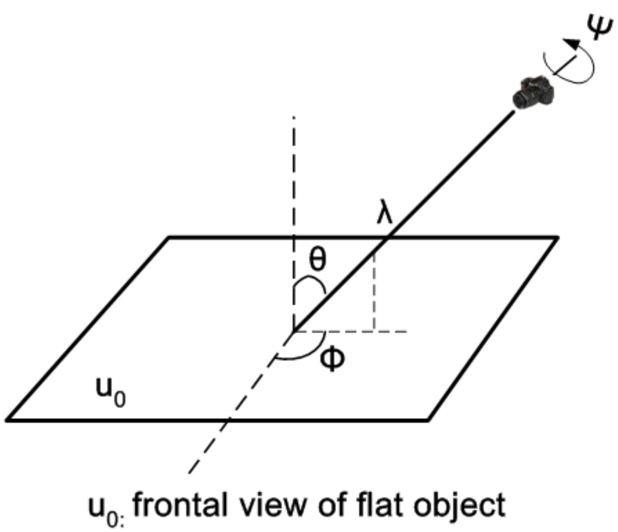
- ψ: **rotation angle of camera around optical axis.**

Geometric Interpretation of the Six Affine Parameters

$$u = \mathbf{S}_1 \mathbf{G}_1 \mathbf{A} \mathbf{u}_0.$$

A is an affine map: $\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{H}_\lambda \mathbf{R}_1(\psi) \mathbf{T}_t \mathbf{R}_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$



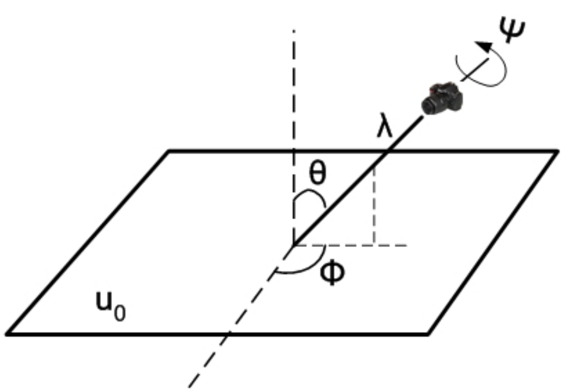
- λ: ***zoom* parameter.**

Geometric Interpretation of the Six Affine Parameters

$$u = \mathbf{S}_1 \mathbf{G}_1 \mathbf{A} \mathbf{u}_0.$$

$$\mathbf{A} \text{ is an affine map: } \begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$$

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \mathbf{H}_\lambda \mathbf{R}_1(\psi) \mathbf{T}_t \mathbf{R}_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

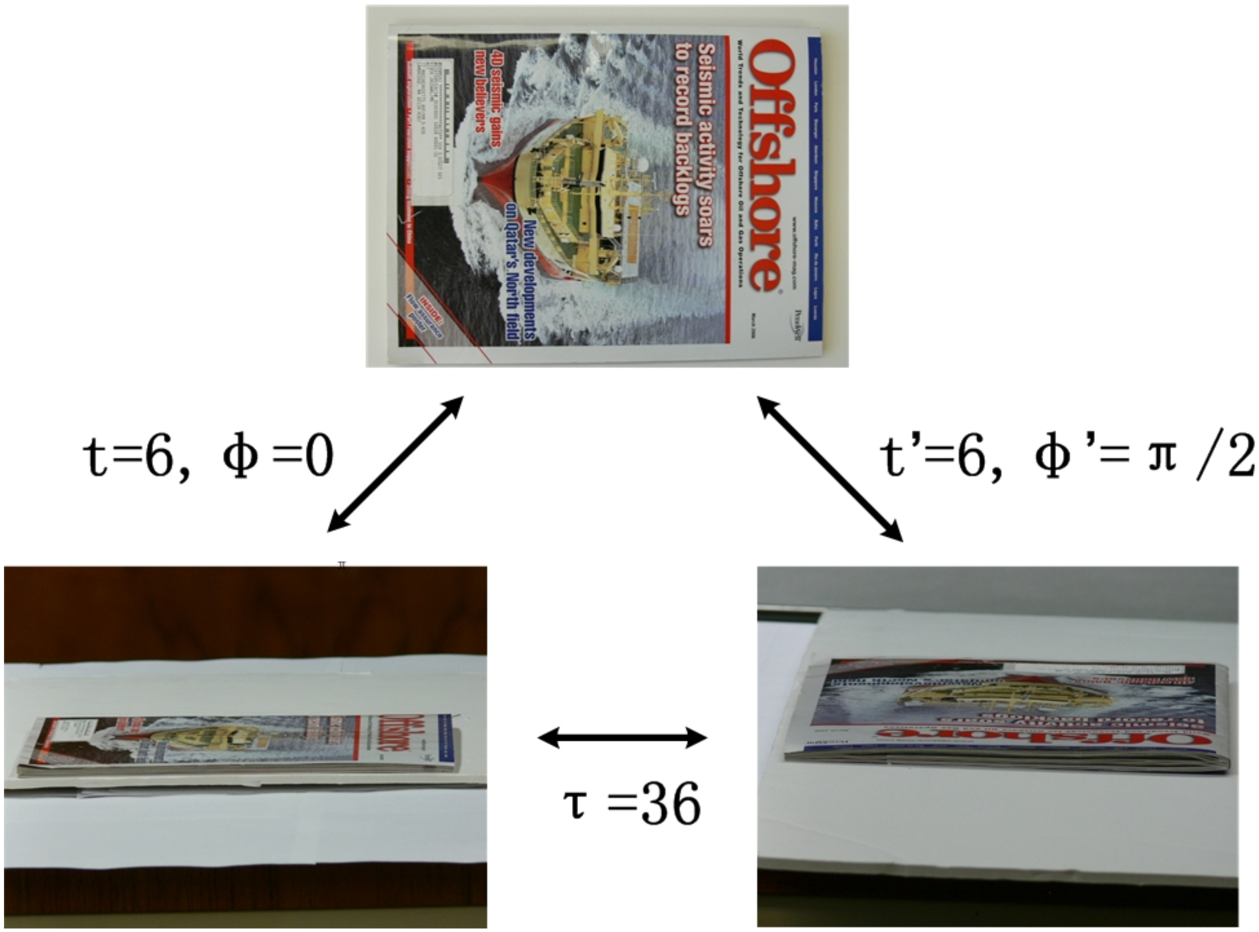


u_0: frontal view of flat object

- λ: ***zoom* parameter.**

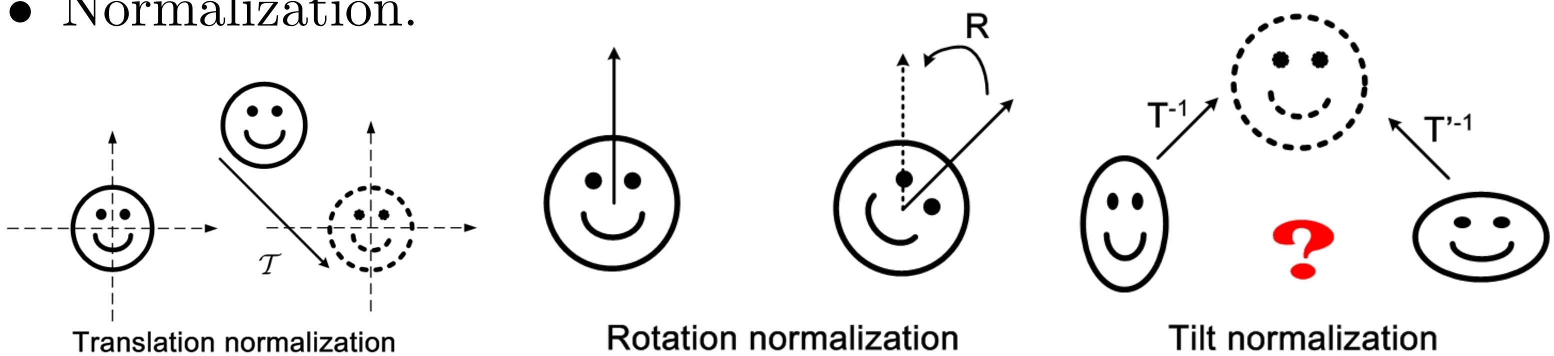
High Transition Tilts

Both compared images are usually slanted views. The *transition tilt* quantifies the tilt between two such images.



Affine Invariance: Simulation v.s. Normalization

- Simulation.
  - all 6 parameters impossible, e.g. 10^6.
- Normalization.

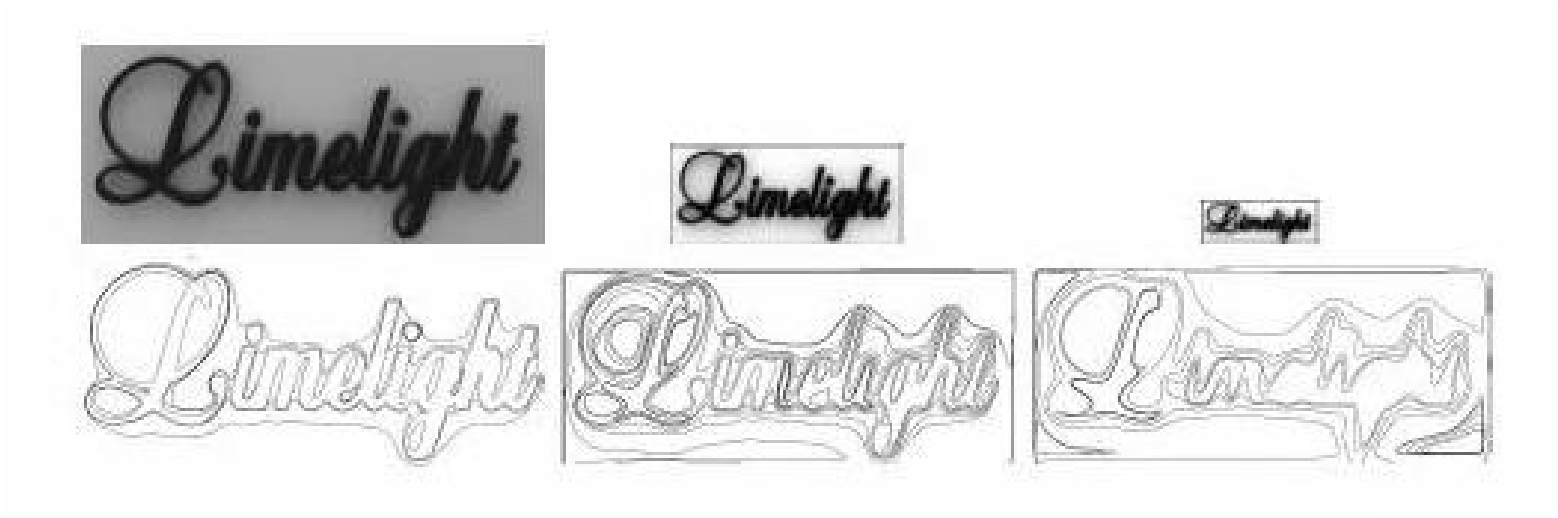


$$\mathbf{u} = \mathbf{G}_1\mathbf{A}\mathbf{u}_0, \ \ \mathbf{v} = \mathbf{G}_1\mathbf{u}_0 \Rightarrow \mathbf{u} = \mathbf{A}\mathbf{v} \ ?$$

Non-commutation: in general $\mathbf{G}_1\mathbf{A}u_0 \neq \mathbf{A}\mathbf{G}_1u_0$

- Translation $\mathcal{T}$ and rotation $\mathbf{R}$ can be normalized.
  Strong commutation with blur $\Rightarrow$ normalization possible.
- Zoom $\mathbf{H}_\lambda$ and tilt $\mathbf{T}$ cannot be normalized *stricto sensu*.
  Weak commutation with blur $\Rightarrow$ simulation necessary.
  $\mathbf{H}_\lambda\mathbf{G}_1 = \mathbf{G}_{1/\lambda}\mathbf{H}_\lambda \Rightarrow \mathbf{H}_\lambda\mathbf{v} \neq \mathbf{u}$

Affine Invariance: Simulation v.s. Normalization

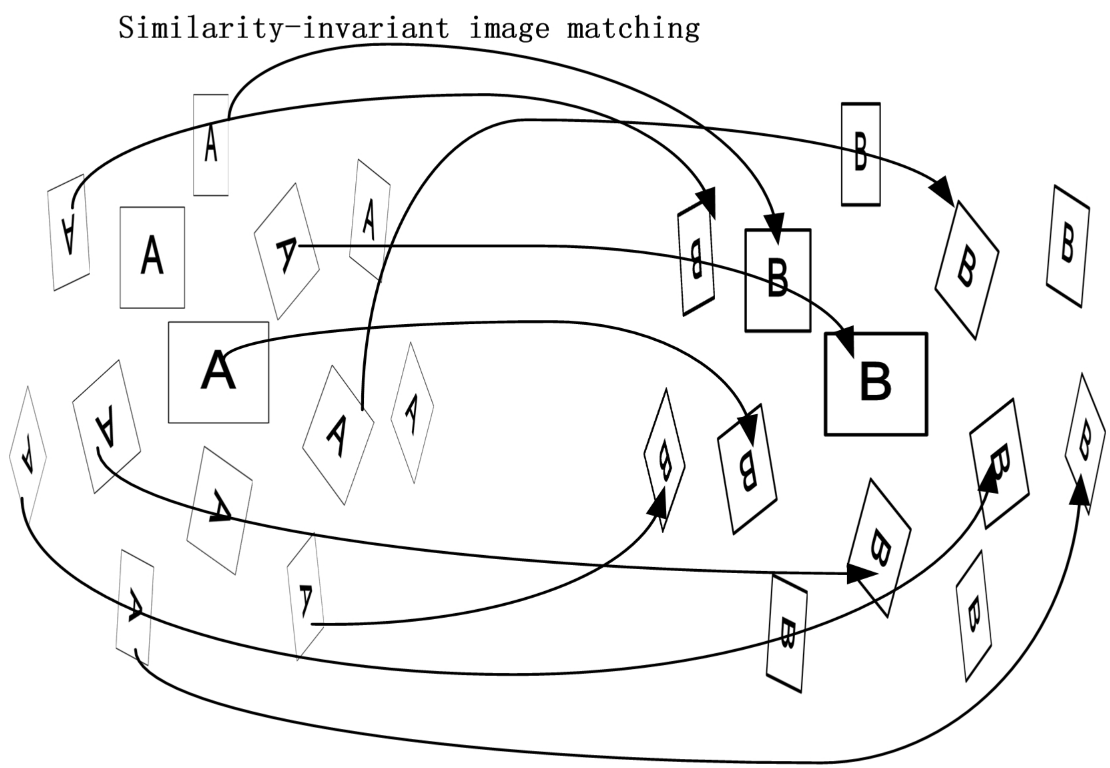


State-of-the-art

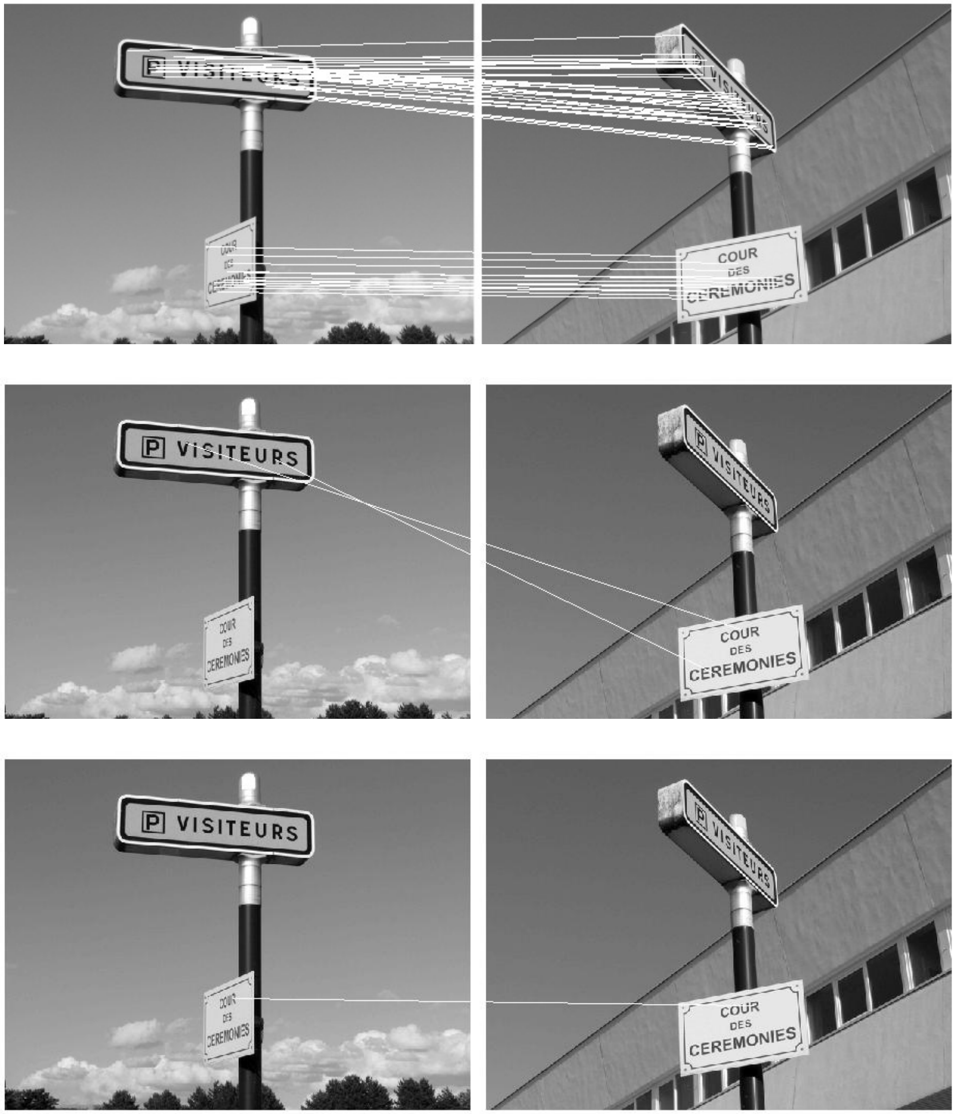
- SIFT (Scale-Invariant Feature Transform) [Lowe 99, 04]:
  - Rotation and translation are **normalized**.
  - Zoom is **simulated** in the scale space.
  - No treatment on latitude and longitude: modest robustness $\tau_{\max} < 2$.
- MSER (Maximally Stable Extremal Region) [Matas et al. 02] and LLD (Level Line Descriptor) [Musé et al. 06]
  - Attempt to **normalize** all the parameters.
  - Weakness: limited affine invariance $\tau_{\max} < 10$, not scale invariant, small number of features.
- Other methods: Harris-Affine, Hessian-Affine [Mikolajczyk and Schmid 04]

Affine-SIFT (ASIFT) Overview

- Simulate latitude, longitude to achieve full affine invariance.
- Simulated images are compared by a rotation-, translation- and zoom-invariant algorithm, e.g., SIFT. (SIFT normalizes translation and rotation and simulates zoom.)



Experiments: Image Matching



Road signs.

Transition tilt: $\tau \approx 2.6$.

Number of correct matches:

ASIFT (top)—50;

SIFT (middle)—0;

MSER (bottom)—1.

Experiments: Image Matching

Statue of Liberty. Transition tilt $\tau \in [1.3, \infty)$.

Number of correct matches: ASIFT (left)—22, SIFT (right)—1, MSER (not shown)—0.

Consistency of ASIFT: reduction from ASIFT to SIFT

Theorem 1 *Let* $u = \mathrm{G}_1 A \mathcal{T}_1 u_0$ *and* $v = \mathrm{G}_1 B \mathcal{T}_2 u_0$ *be two images obtained from an infinite resolution image* u_0*. Then ASIFT, applied with a dense set of tilts and longitudes, simulates two views of* u *and* v *that are obtained from each other by a translation, a rotation, and a camera zoom. As a consequence, these images match by the SIFT algorithm.*

Proof:

ASIFT works because *geometric* tilts can be reverted by *simulated* tilts in the orthogonal direction. The price to pay is a zoom-out.

Parameter Sampling Range

- Longitude angle $\phi \in [0, \pi)$.
- Tilt $t_{\max} = 4\sqrt{2}$ ($\theta \approx 80°$).

t = 3.8 ($\theta = 74.7°$), 116 correct ASIFT matches.

Parameter Sampling Range

- Longitude angle $\phi \in [0, \pi)$.
- Tilt $t_{\max} = 4\sqrt{2}$ ($\theta \approx 80°$).

t = 5.6 ($\theta = 79.7°$), 26 correct ASIFT matches.

Parameter Sampling Range

- Longitude angle $\phi \in [0, \pi)$.
- Tilt $t_{\max} = 4\sqrt{2}$ ($\theta \approx 80°$).

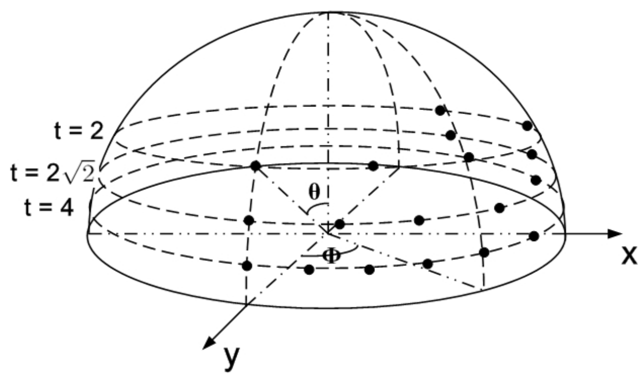
t = 8 ($\theta = 82.8°$), 0 ASIFT match.

Parameter Sampling

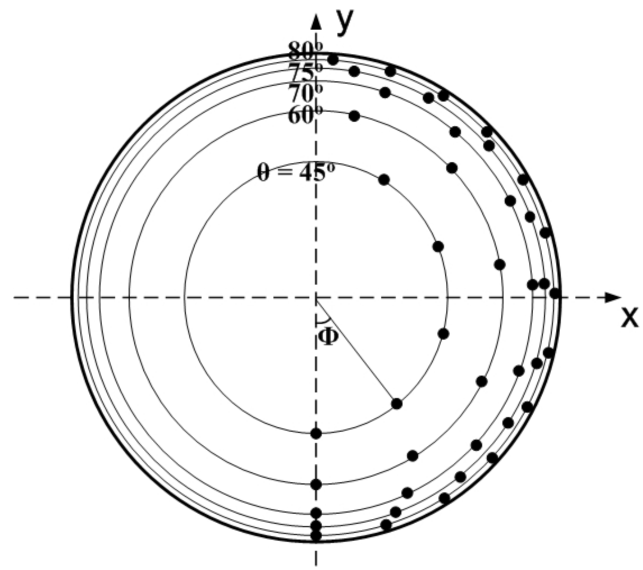
- Geometric sampling of t: $\triangle t = t_{k+1}/t_k = \sqrt{2}$.

t	1	$\sqrt{2}$	2	$2\sqrt{2}$	4	$4\sqrt{2}$
θ	$0°$	$45°$	$60°$	$69.3°$	$75.5°$	$79.8°$

- Arithmetical sampling of ϕ: $\triangle\phi = \phi_{k+1} - \phi_k = \triangle\phi = \frac{72°}{t}$.



Perspective view



View from the zenith

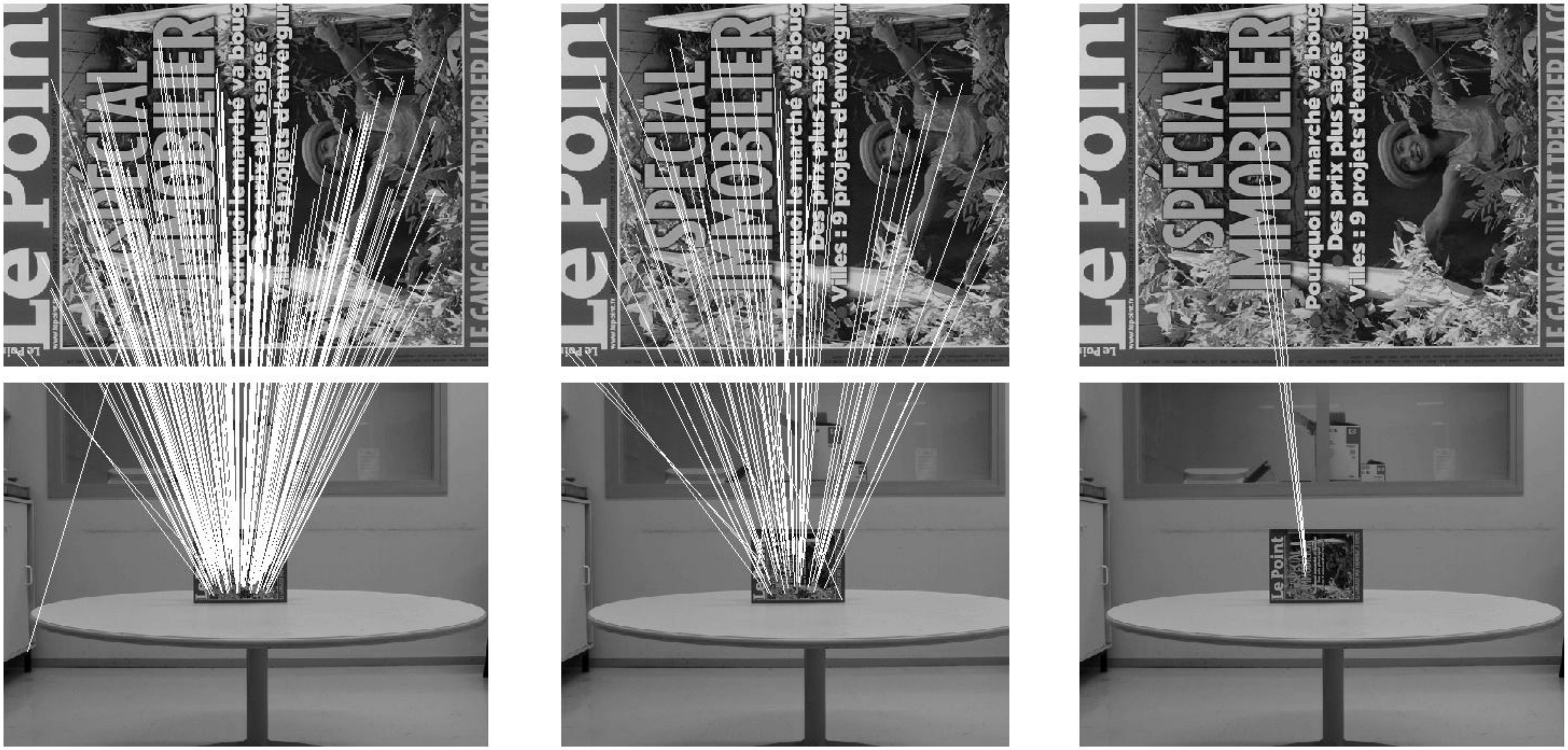
Acceleration: Multi-resolution ASIFT

1. ASIFT on low-resolution images ($r \times r$ sub-sampled) .
2. ASIFT on high-resolution images obtained with the M best affine transforms (only in case of success in (1)).

ASIFT Complexity

- Complexity proportional to (area of query) $\times$ (searched area).
- Image area proportional to number of simulated tilts.
  - $t = 1,\ \sqrt{2},\ 2,\ 2\sqrt{2},\ 4,\ 4\sqrt{2}$.
  - Number of rotations for tilt t is about $2.5t$.
  - At tilt t, image area $\sim 1/t$.
- Simulated area on one side: $\frac{1+5\times 2.5}{9} \approx 1.5$ times original image.
- ASIFT complexity: $(1.5)^2 \times$SIFT $= 2.25\times$SIFT, $\tau_{max} = 32$.

Experiments: Image Matching

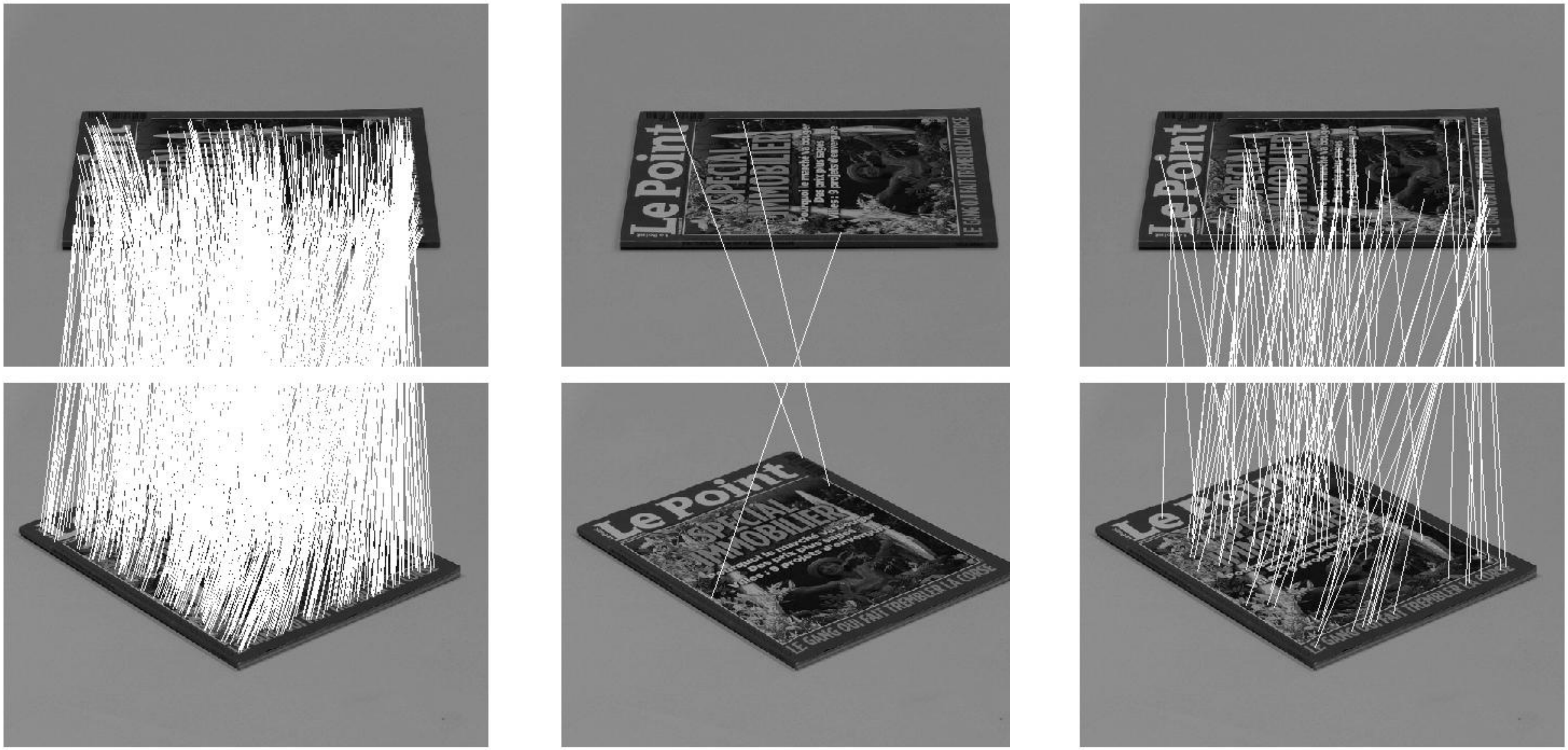


Zoom change. Number of correct matches: ASIFT (left)—222; SIFT (middle)—87; MSER (right)—4.

Experiments: Image Matching

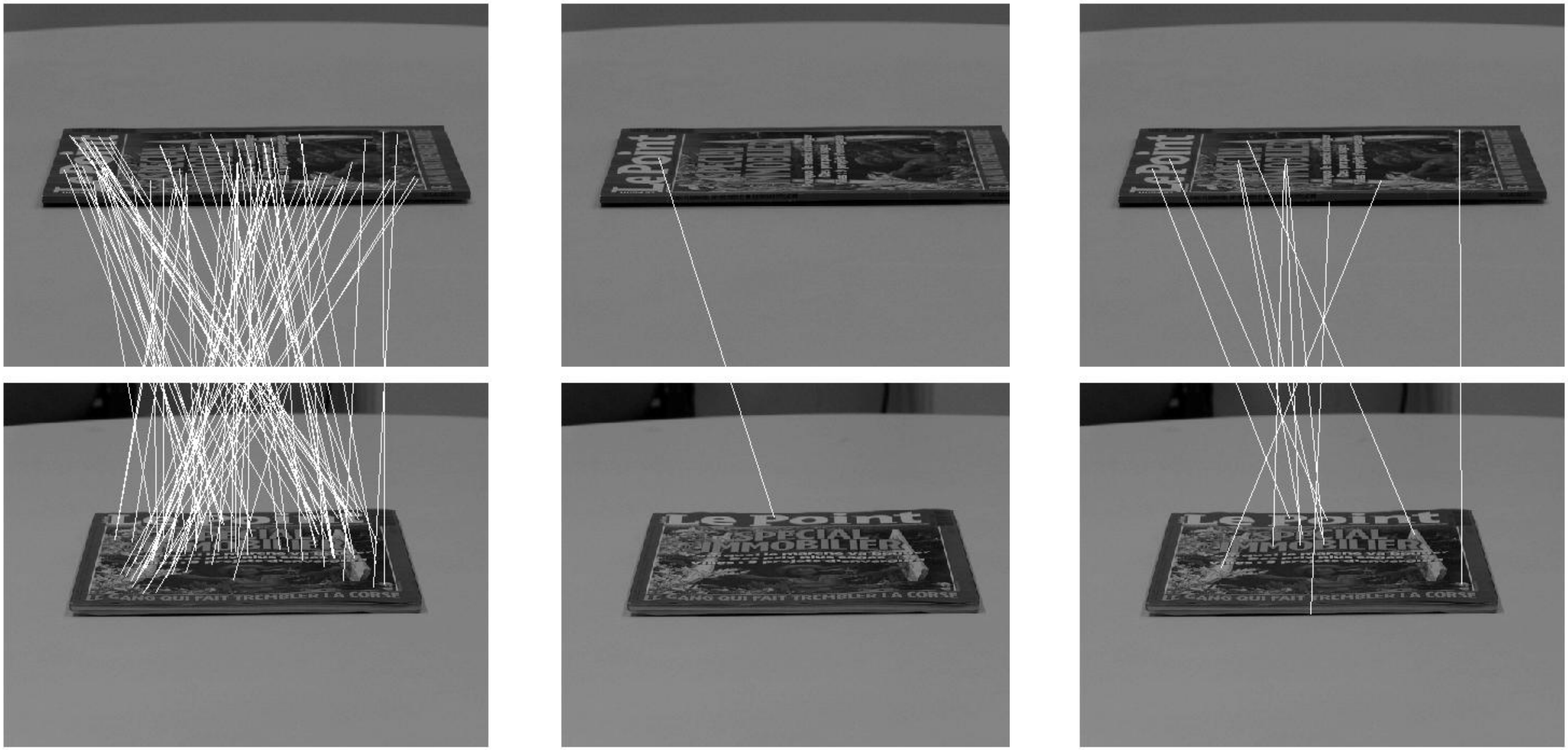
Frontal v.s. -80° angle, zoom $\times 10$: absolute tilt $t = 5.8$. Number of correct matches: ASIFT (left)—75; SIFT (middle)—1; MSER (right)—2.

Experiments: Image Matching



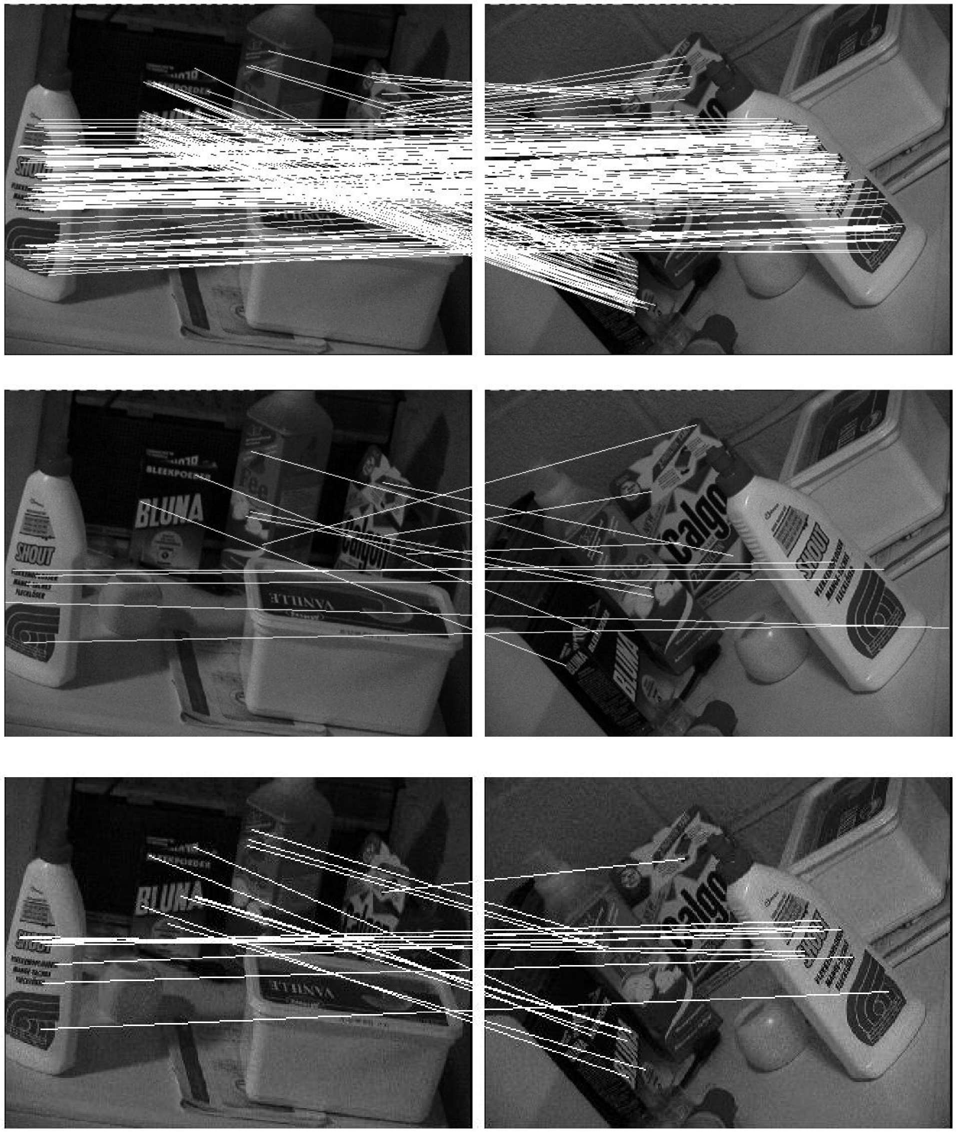
Correspondences between the magazine images taken with absolute tilts $t_1 = t_2 = 2$ with longitude angles $\phi_1 = 0^\circ$ and $\phi_2 = 50^\circ$, transition tilt $\tau = 3$. Number of correct matches: ASIFT (left)—881; SIFT (middle)—2; MSER (right)—87.

Experiments: Image Matching



Correspondences between the magazine images taken with absolute tilts $t_1 = t_2 = 4$ with longitude angles $\phi_1 = 0^\circ$ and $\phi_2 = 90^\circ$, transition tilt $\tau = 16$. Number of correct matches: ASIFT (left)—88; SIFT (middle)—1; MSER (right)—9.

Experiments: Image Matching



Images proposed by Matas et al.

Number of correct matches:

ASIFT (top)—254;

SIFT (middle)—10;

MSER (bottom)—22.

Experiments: Image Matching

Parkings.

Transition tilt: $\tau \approx 15$.

Number of correct matches:

ASIFT (top)—78;

SIFT (middle)—0;

MSER (bottom)—0.

Experiments: Image Matching

Ecole Polytechnique.

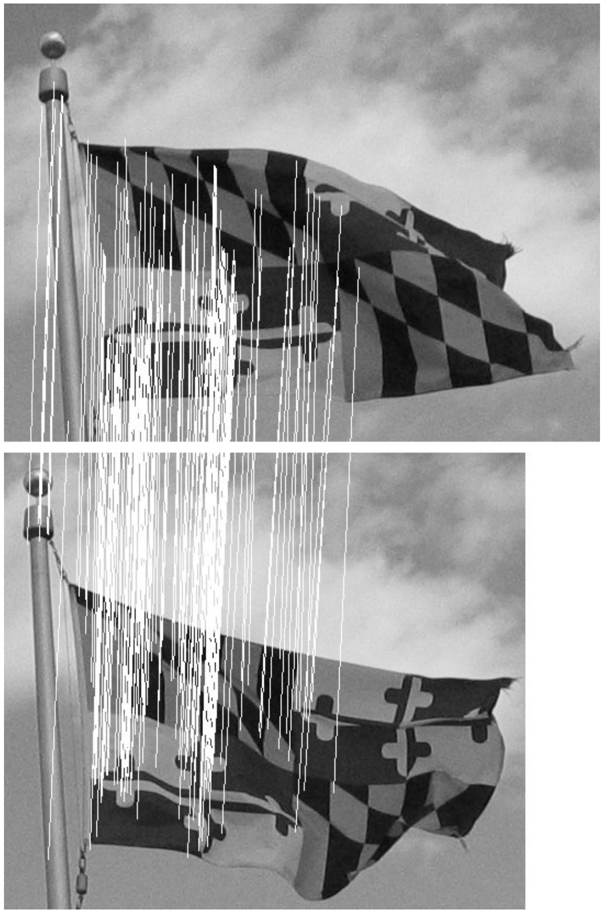
Transition tilt $\tau = 2.4$. Number of correct matches: ASIFT (left)—103; SIFT (middle)—13; MSER (right)—4.

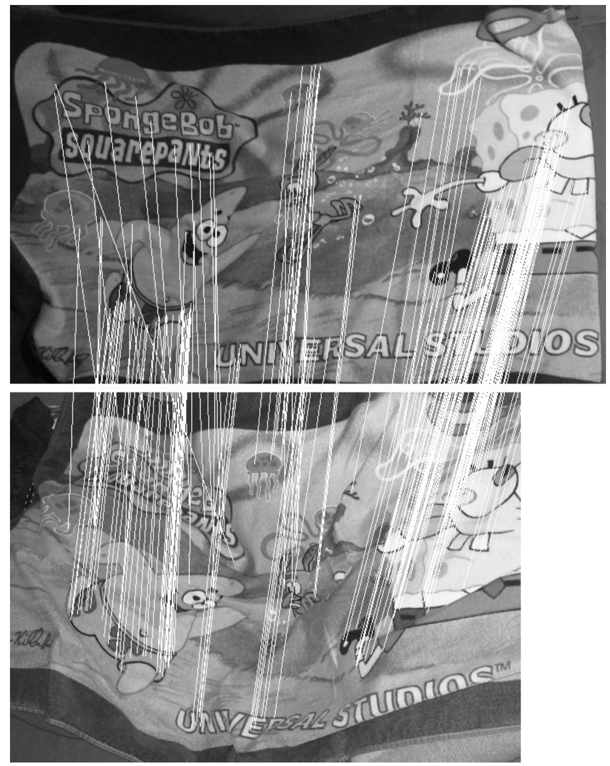
Experiments: Image Matching

Pentagon. Transition tilt $\tau \approx 2.5$.

Number of correct matches: ASIFT (left)—378, SIFT (middle)—6, MSER(right)—17.

Experiments: Image Matching





Left: flag. ASIFT (shown)—141, SIFT—31, MSER—2.

Right: SpongeBob. ASIFT (shown)—370, SIFT—75, MSER—4.

Experiments: Object Tracking

References:

- J.M. Morel and G.Yu, ASIFT: A New Framework for Fully Affine Invariant Image Comparison, SIAM Journal on Imaging Sciences, vol.2, issue 2, 2009.
- G. Yu and J.M. Morel, A Fully Affine Invariant Image Comparison Method, IEEE ICASSP, Taipei, 2009.
- J.M. Morel and G.Yu, On the consistency of the SIFT Method, Preprint, CMLA 2008-26, Sept 2008.

Website and Online Demo: try ASIFT with your images!

- http://www.cmap.polytechnique.fr/~yu/research/ASIFT/demo.html

For more information,

Google ASIFT