ASIFT:

A New Framework for Fully Affine Invariant Image Comparison

Guoshen Yu,
Ecole Polytechnique, France

Jean-Michel Morel
ENS Cachan, France

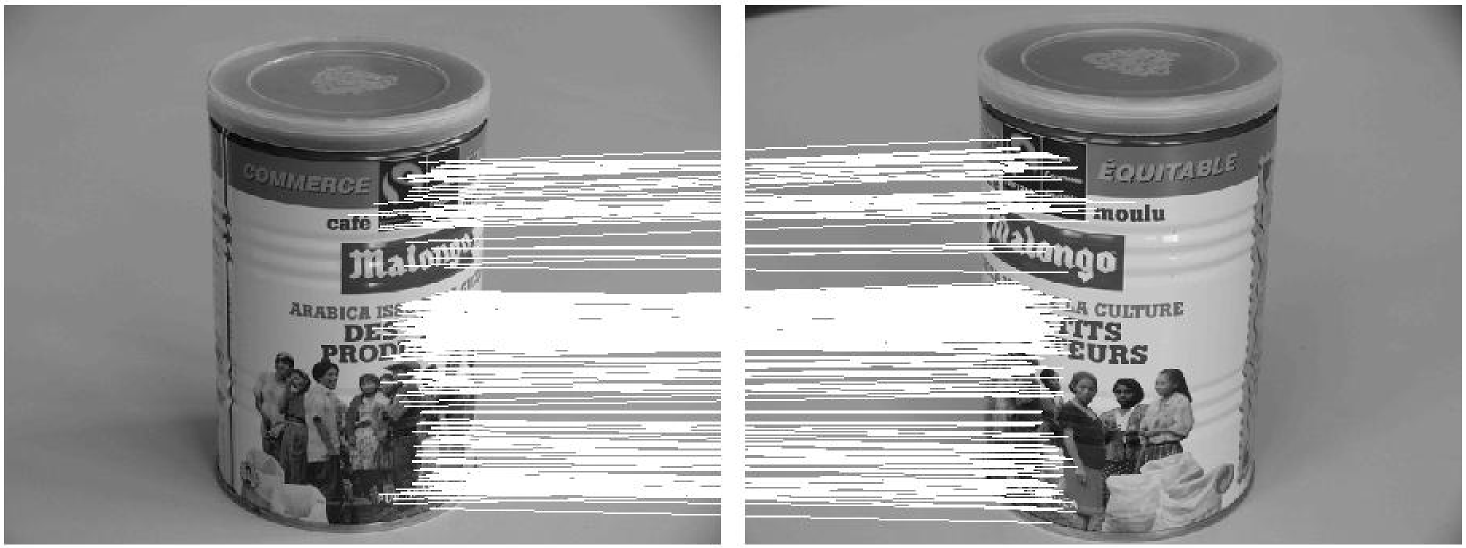
Proposed method: ASIFT.

State of the art: SIFT.

State of the art: MSER.

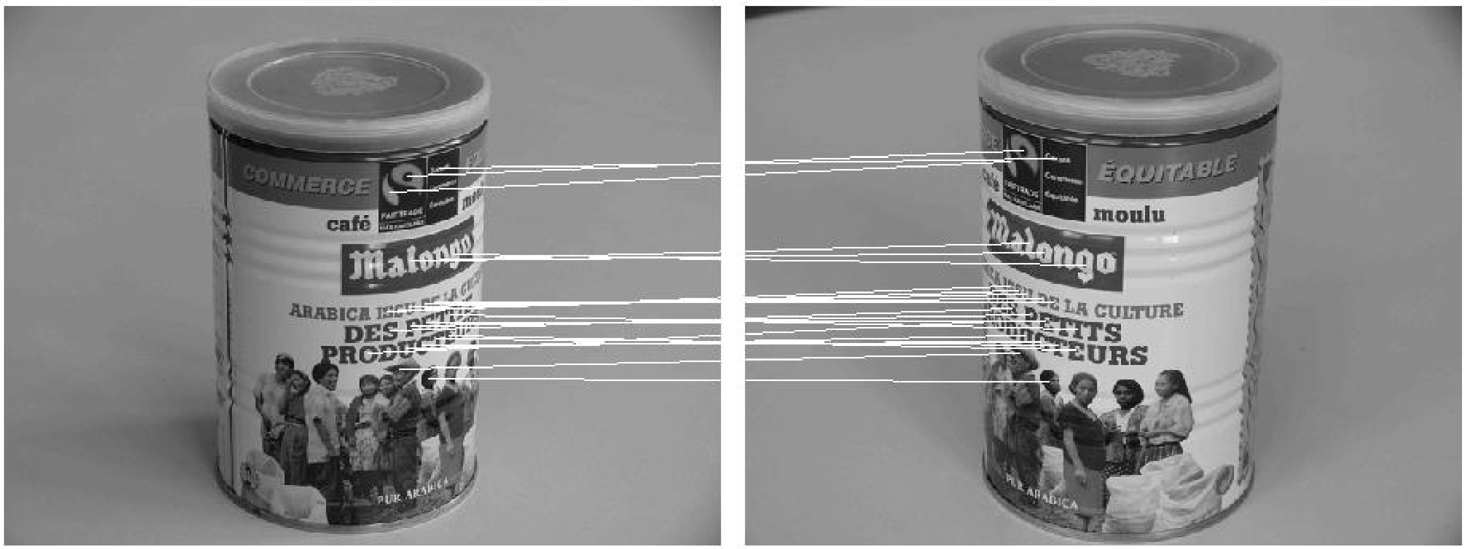
State of the art: Hessian Affine.

State of the art: Harris Affine.

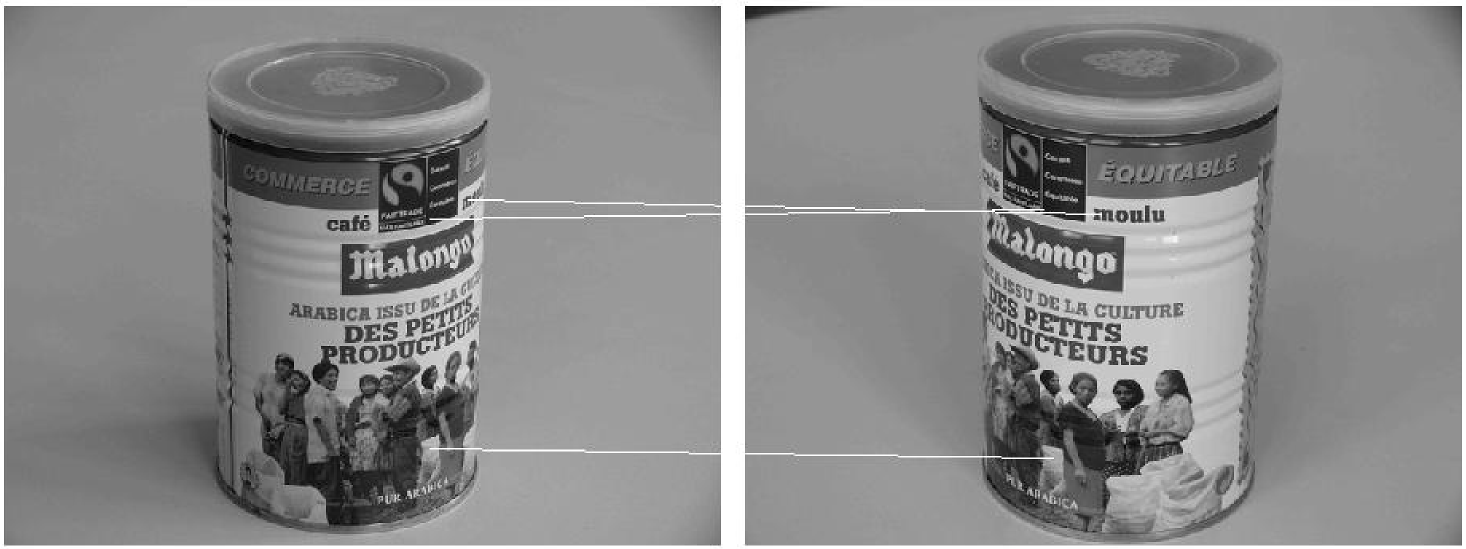


Proposed method: ASIFT.

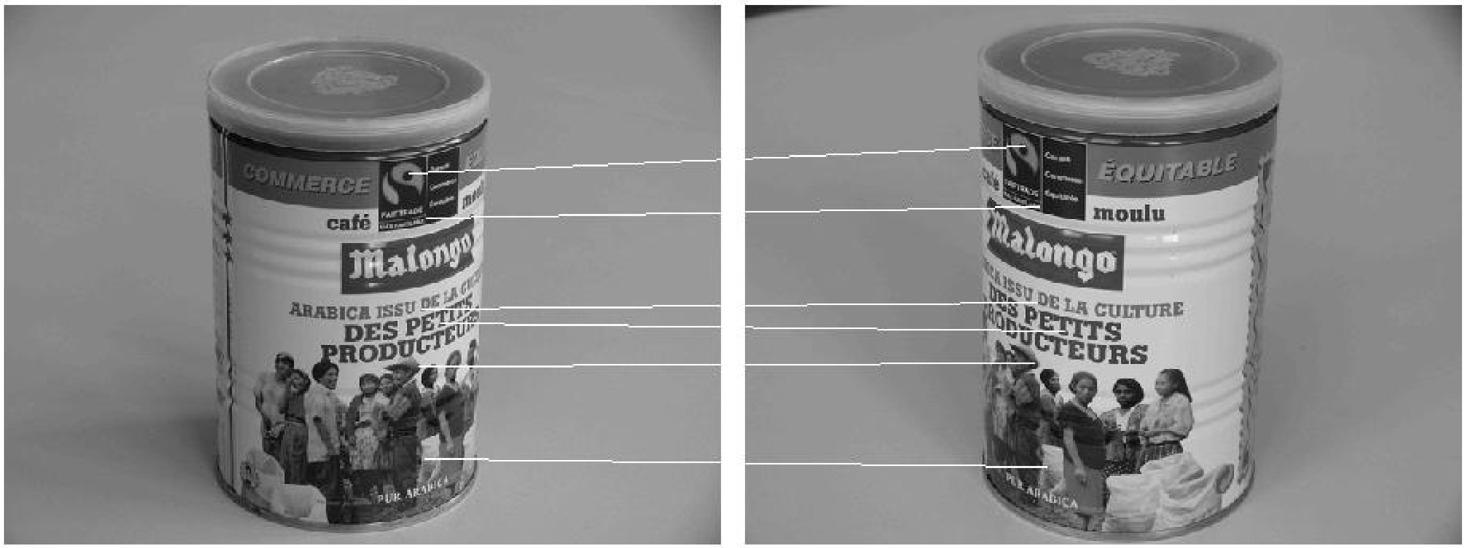
State of the art: SIFT.



State of the art: MSER.



State of the art: Hessian Affine.



State of the art: Harris Affine.

Proposed method: ASIFT.

State of the art: SIFT.

State of the art: MSER.

State of the art: Hessian Affine.

State of the art: Harris Affine.

Proposed method: ASIFT.

State of the art: SIFT.

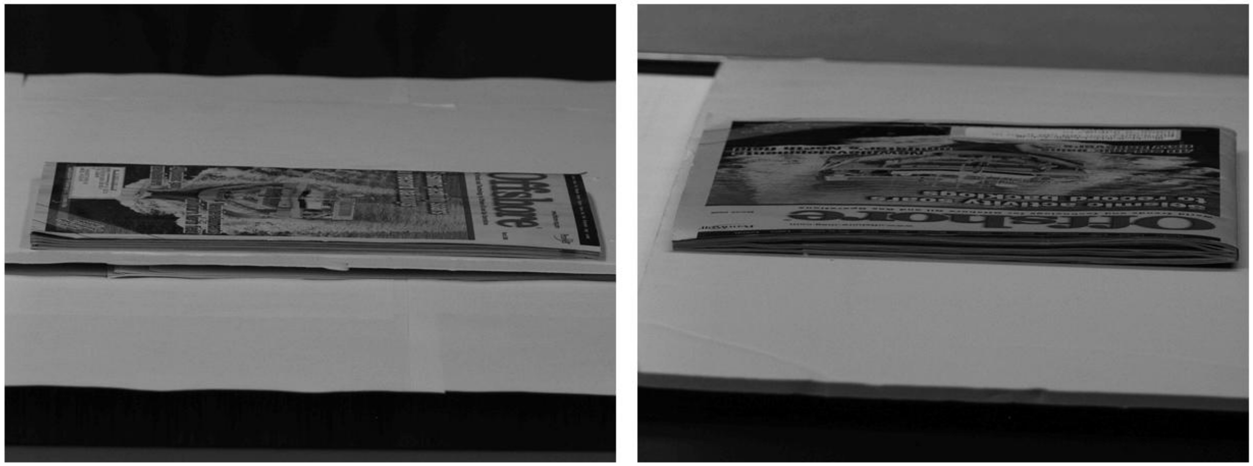
State of the art: MSER.

State of the art: Hessian Affine.

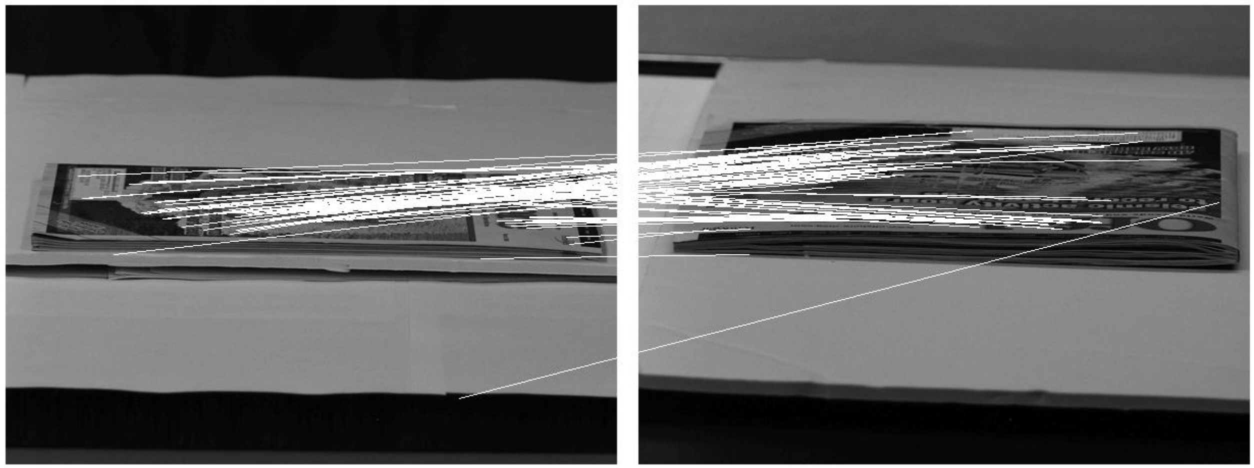
State of the art: Harris Affine.

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.

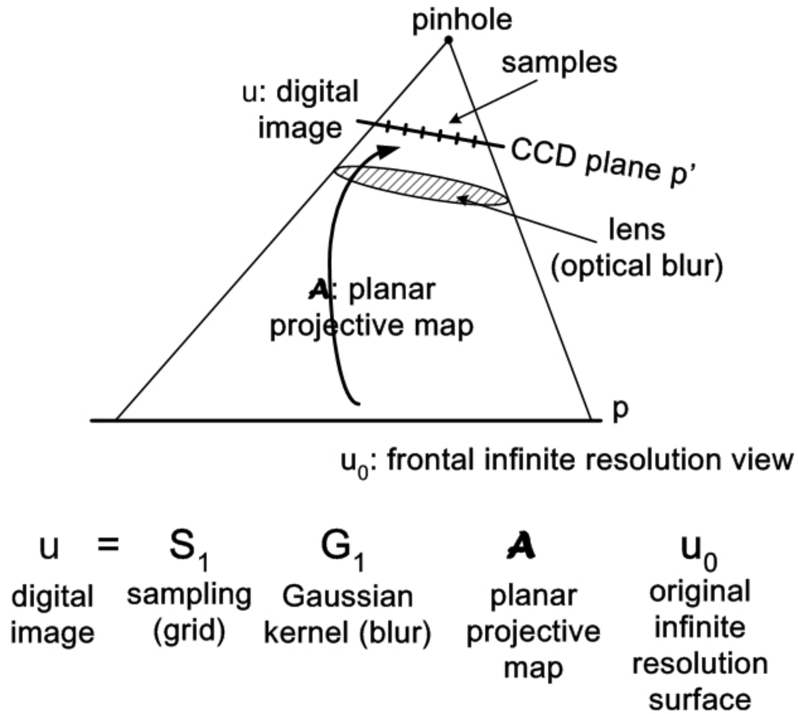


In this pair: A very large transition tilt (extreme angle).
The transition tilt will be defined later.



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 anti-aliasing gaussian filter.

- $\mathbf{S}_1$ 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.

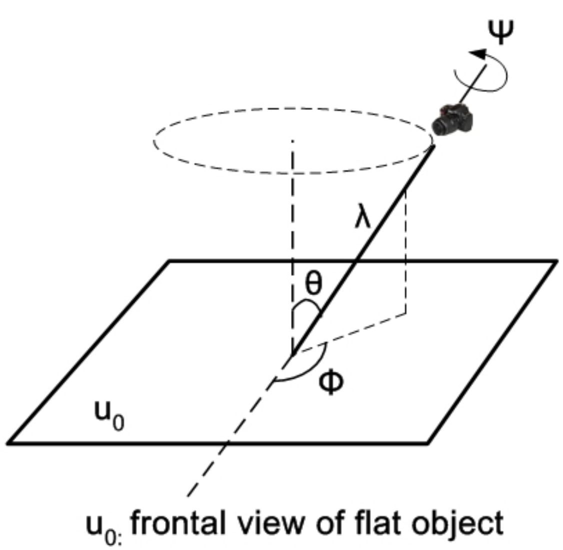
Affine transforms map rectangles to parallelograms.

Geometric Interpretation of the Six Affine Parameters

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

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

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



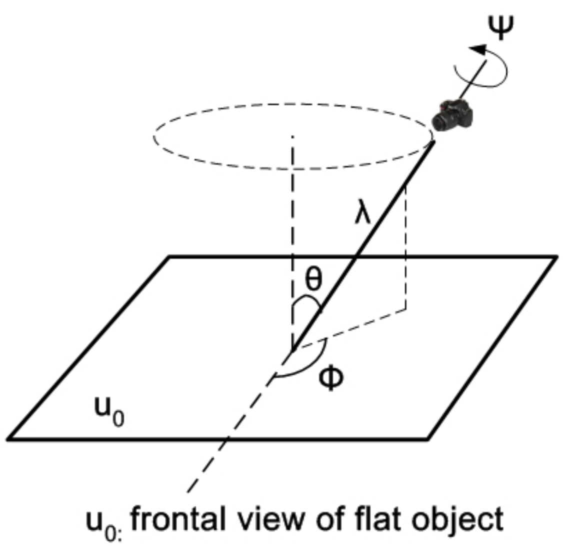
- ϕ: *longitude* angle between optical axis and a fixed vertical plane.
- $\theta = \arccos(1/t)$: *latitude* angle between optical axis and the normal to the image plane. **Tilt** $t > 1 \leftrightarrow \theta \in [0°, 90°]$.
- ψ: rotation angle of camera around optical axis.
- λ: *zoom* parameter.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

Geometric Interpretation of the Six Affine Parameters

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

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

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



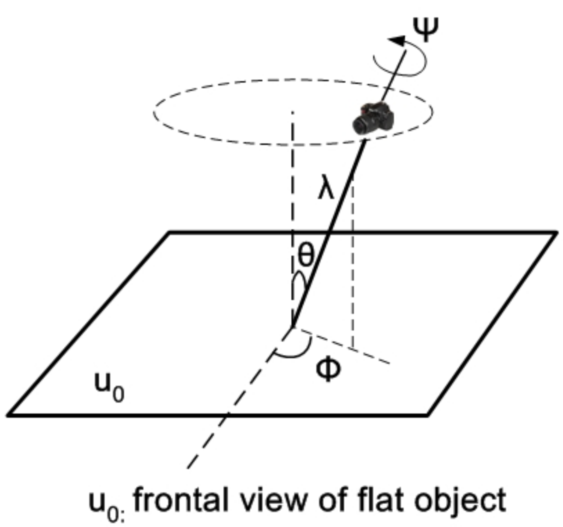
- ϕ: ***longitude*** **angle between optical axis and a fixed vertical plane.**

Geometric Interpretation of the Six Affine Parameters

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

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

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



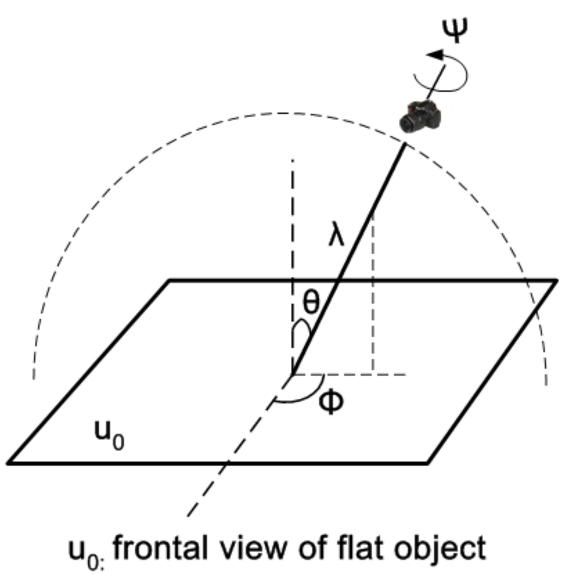
- ϕ: ***longitude* angle between optical axis and a fixed vertical plane.**

Geometric Interpretation of the Six Affine Parameters

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

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

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



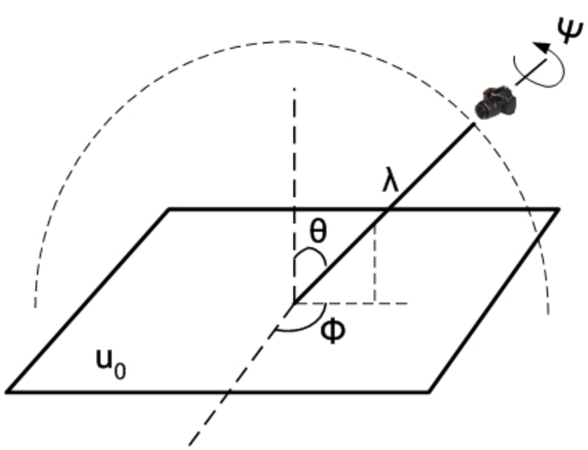
- $\theta = \arccos(1/t)$: ***latitude* angle between optical axis and the normal to the image plane. Tilt** $t > 1 \leftrightarrow \theta \in [0^\circ, 90^\circ]$.

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: $\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$

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



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

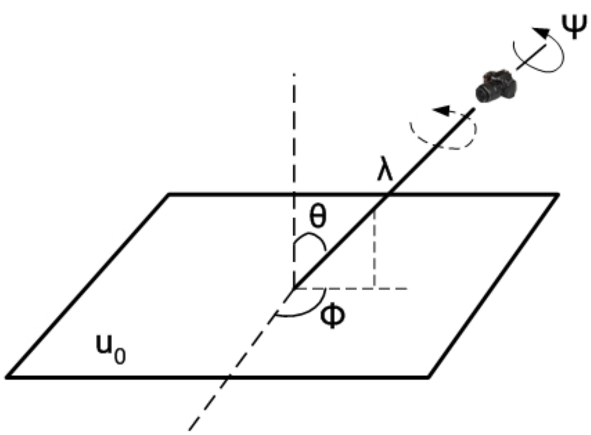
- $\theta = \arccos(1/t)$: ***latitude* angle between optical axis and the normal to the image plane. Tilt** $t > 1 \leftrightarrow \theta \in [0°, 90°]$.

Geometric Interpretation of the Six Affine Parameters

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

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

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



u_0: frontal view of flat object

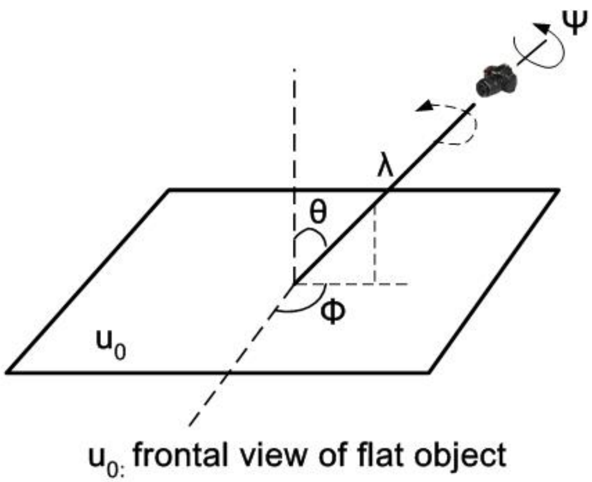
- ψ: **rotation angle of camera around optical axis.**

Geometric Interpretation of the Six Affine Parameters

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

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

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



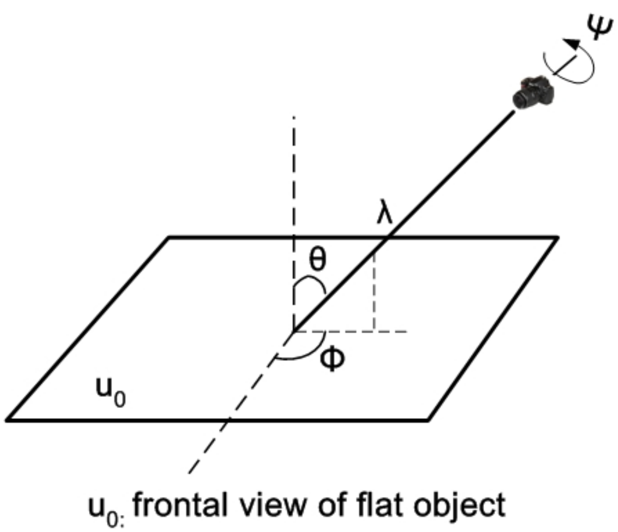
- ψ: **rotation angle of camera around optical axis.**

Geometric Interpretation of the Six Affine Parameters

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

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

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



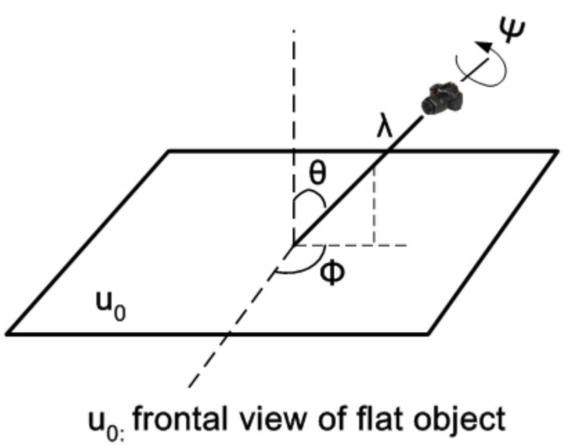
- λ: ***zoom* parameter.**

Geometric Interpretation of the Six Affine Parameters

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

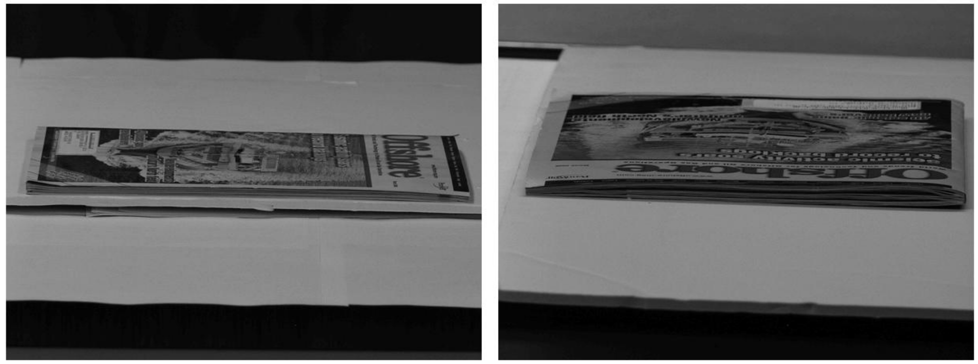
A is an affine map: $\begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$

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



- λ: ***zoom* parameter.**

Transition Tilts

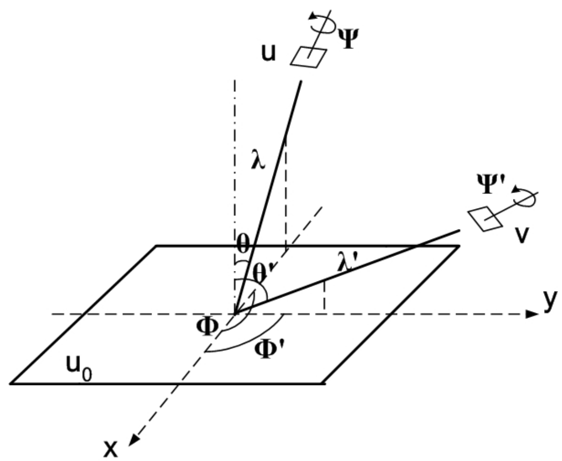


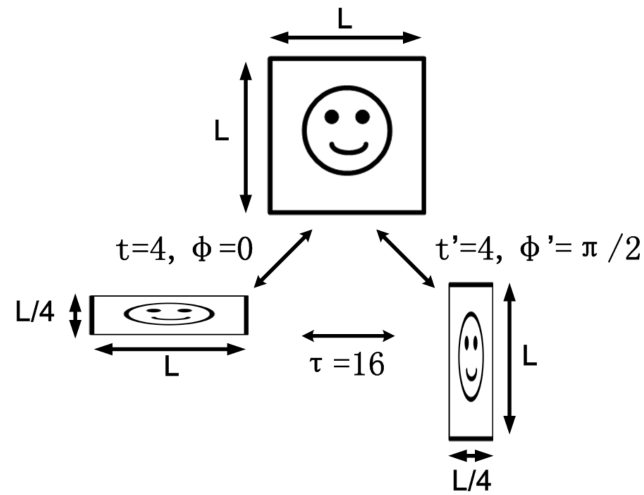
Both compared images are usually slanted views. The *transition tilt* quantifies the tilt between two such images.

Definition Consider two views of a planar image, $u_1(x, y) = u(A(x, y))$ and $u_2(x, y) = u(B(x, y))$ where A and B are two linear maps such that BA^{-1} is not a similarity. We call *transition* tilt $\tau(u_1, u_2)$ and *transition* rotation $\phi(u_1, u_2)$ the unique parameters such that

$$BA^{-1} = H_\lambda R_1(\psi) T_\tau R_2(\phi). \quad (1)$$

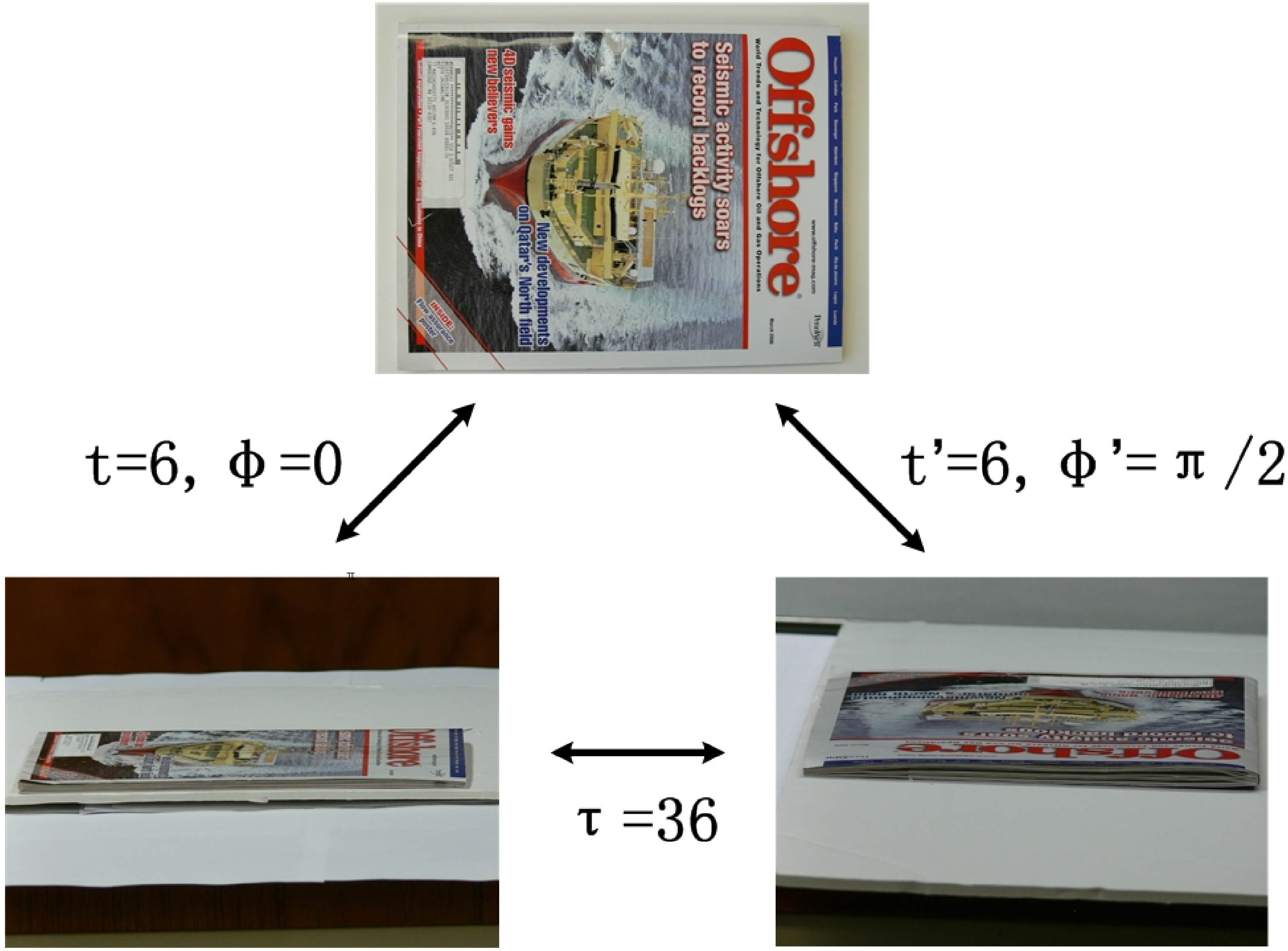
Properties of Transition Tilts





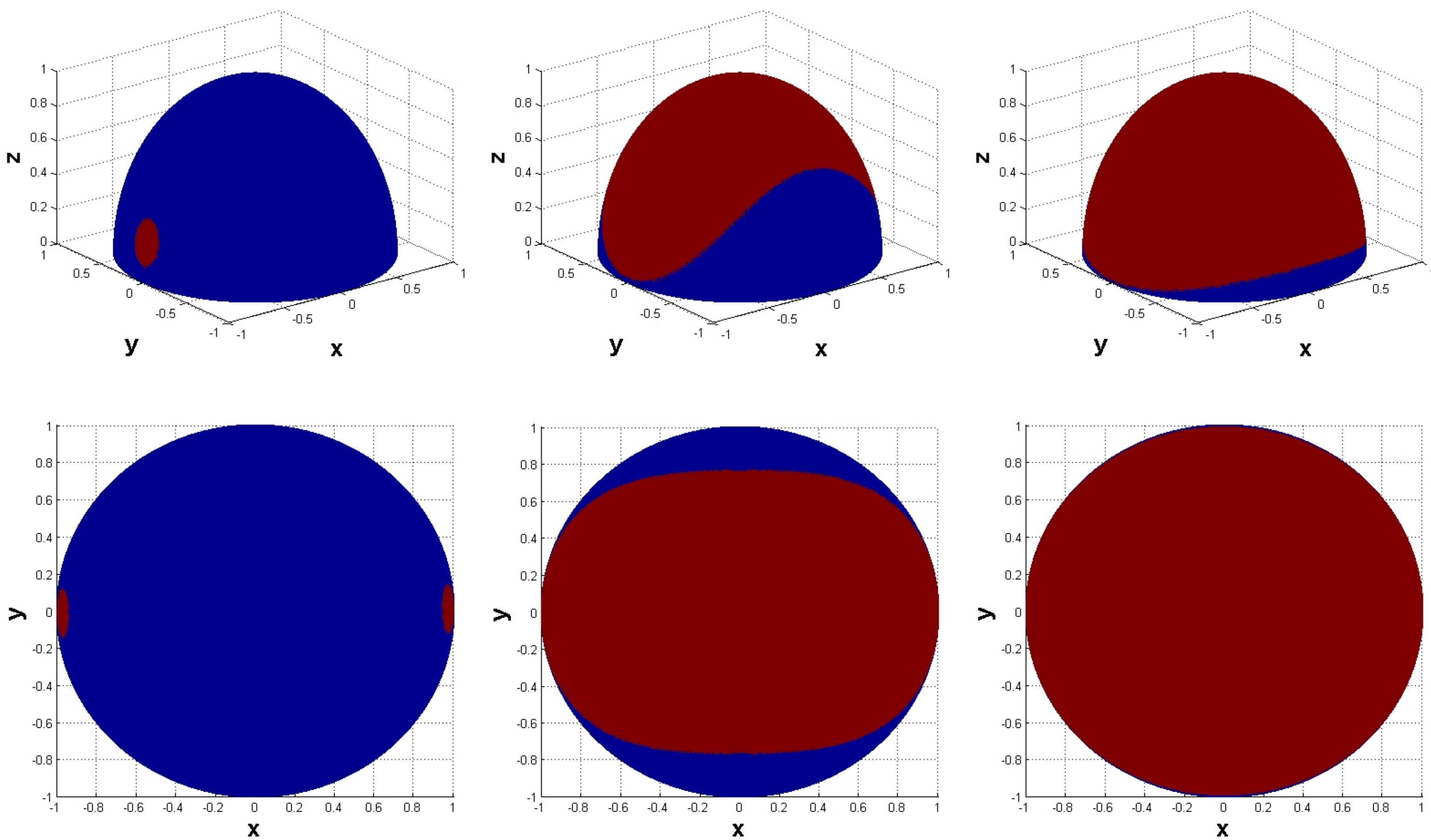
- The transition tilt is symmetric, i.e., $\tau(u_1, u_2) = \tau(u_2, u_1)$;
- The transition tilt only depends on the absolute tilts and on the longitude angle difference: $\tau(u_1, u_2) = \tau(t, t', \phi - \phi')$;
- One has $t'/t \le \tau \le t't$, assuming $t' = \max(t', t)$;
- The transition tilt is equal to the absolute tilt: $\tau = t'$, if the other image is in frontal view ($t = 1$).

High Transition Tilts



$$\tau = 36 \Rightarrow \theta = \mathbf{88.41^\circ}$$

High Transition Tilts

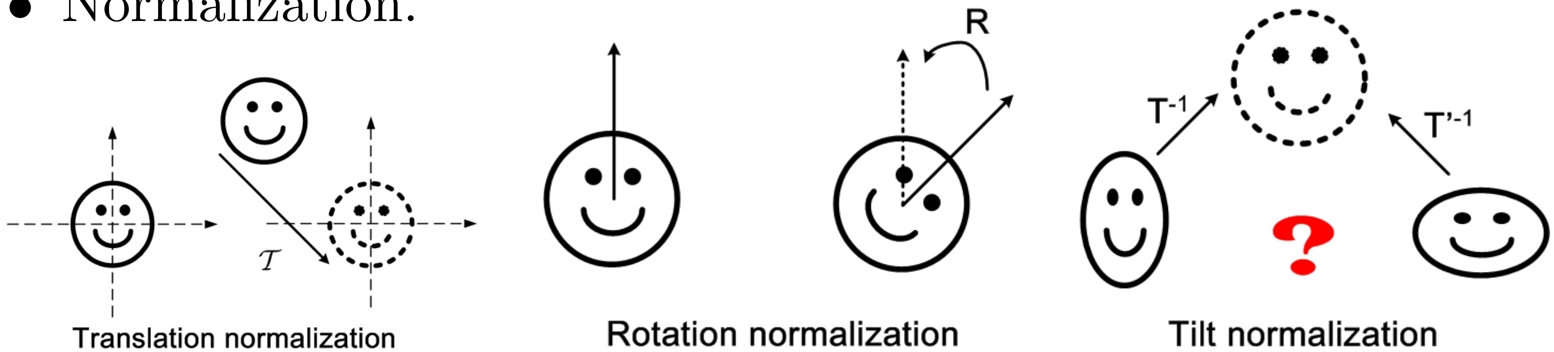


$\tau < 2$ (SIFT) $\qquad \tau < 10$ (MSER) $\qquad \tau < 40$ (ASIFT)

$$\theta = 80^\circ$$

Affine Invariance: Simulation v.s. Normalization

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

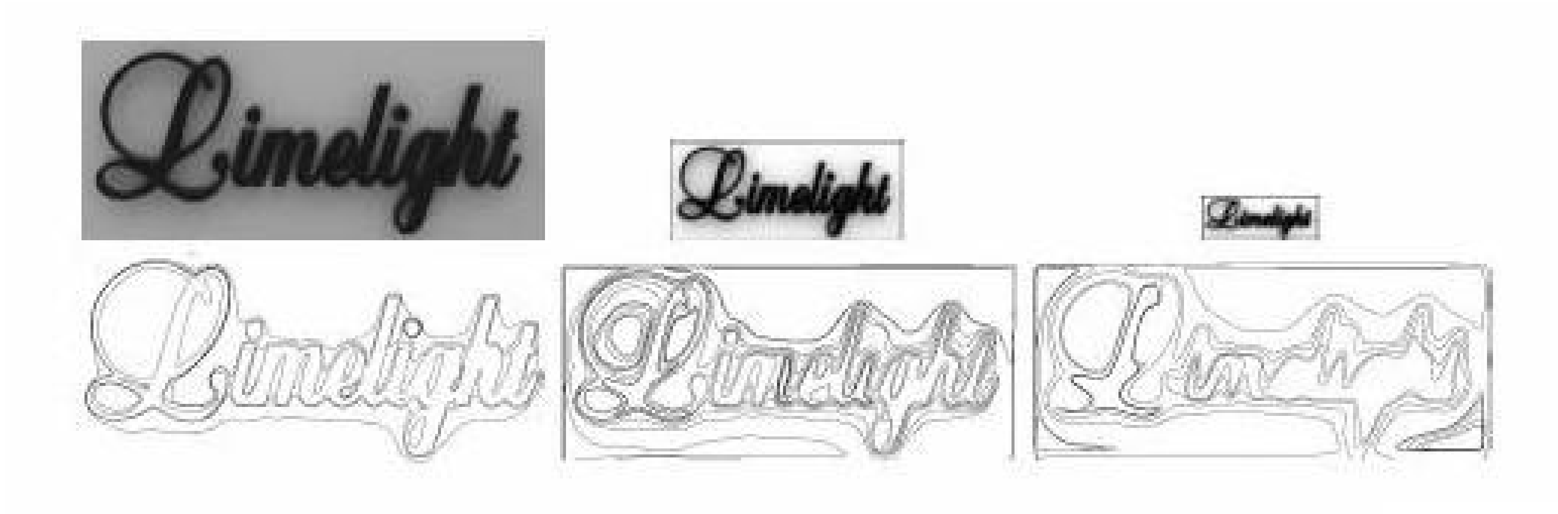


$$\mathbf{u} = \mathbf{G}_1\mathbf{A}\mathbf{u}_0, \;\; \mathbf{v} = \mathbf{G}_1\mathbf{u}_0 \Rightarrow \mathbf{u} = \mathbf{A}\mathbf{v} \; ?$$

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

- 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]

State-of-the-art

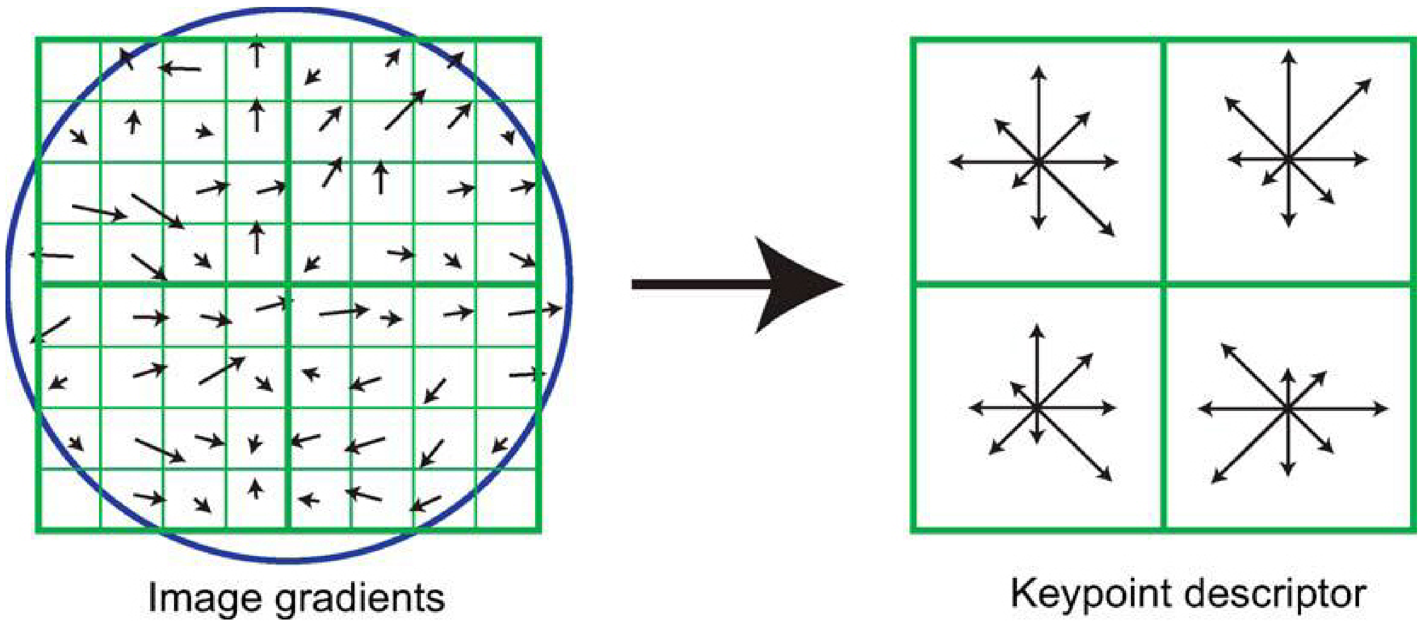
- Other methods: [Baumberg, 00; Tuytelaars and Van Gool, 00, 04; Mikolajczyk and Schmid, 02, 04, 05; Schaffalitzky and Zisserman, 02; Brown and Lowe, 02, S. Belongie, J. Malik, and J. Puzicha, 02, Kadir, Zisserman, Brady, 04, Ke and Sukthankar, 04]
- Evaluations: [Mikolajczyk and Schmid 03, 05, K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, 05]
  - SIFT-based descriptors perform best.
  - MSER outperforms other affine invariant detectors such as Hessian Affine and Harris Affine.

SIFT: Scale Invariant Features Transform

- the initial digital image is $\mathbf{S}_1\mathbf{G}_1\mathbf{A}\mathbf{u}_0$, $\mathbf{A}$ is any similarity, $\mathbf{u}_0$ is the underlying infinite resolution planar image;
- at all scales $\sigma > 0$, the SIFT method computes $\mathbf{u}(\sigma, \cdot) = \mathbf{G}_\sigma\mathbf{G}_1\mathbf{A}\mathbf{u}_0$ and "key points" $(\sigma, \mathbf{x})$, namely scale and space extrema of $\Delta\mathbf{u}(\sigma, \cdot)$;
- the blurred $\mathbf{u}(\sigma, \cdot)$ image is sampled around each key point at a pace proportional to $\sqrt{1+\sigma^2}$;
- directions of the sampling axes are fixed by a dominant direction of $\nabla\mathbf{u}(\sigma, \cdot)$ in a σ-neighborhood of the key point;
- this yields rotation, translation and scale invariant samples: the 4 parameters of $\mathbf{A}$ have been eliminated!;
- the final SIFT descriptor keeps only orientations of the gradient to gain invariance w.r. light conditions.

SIFT Feature Points

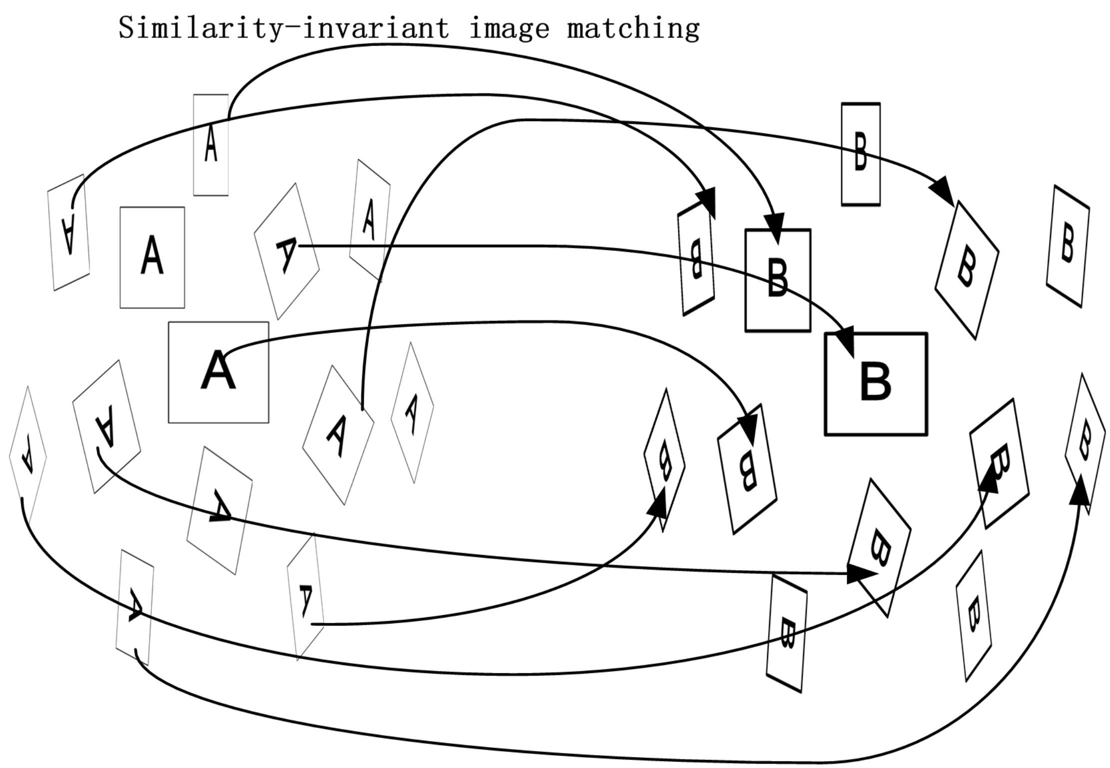
SIFT: Scale Invariant Features Transform



Each key-point is associated *a square image patch whose size is proportional to the scale and whose side direction is given by the assigned direction*. Example of a 2×2 descriptor array of orientation histograms (right) computed from an 8×8 set of samples (left). The orientation histograms are quantized into 8 directions and the length of each arrow corresponds to the magnitude of the histogram entry.

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.)



Inverting Tilts

Definition Given $t > 1$, the tilt factor, define

- the **geometric** tilt : $T_t^x u_0(x, y) := u_0(tx, y)$.
  In the y direction, $T_t^y u_0(x, y) := u_0(x, ty)$.
- the **simulated** tilt (taking into account camera blur): $\mathbb{T}_t^x v := T_t^x \mathrm{G}^x_{\sqrt{t^2-1}} *_x v$.
  In the y direction, $\mathbb{T}_t^y v := T_t^y \mathrm{G}^y_{\sqrt{t^2-1}} *_y v$.
- **Main Formula**
  For $t \geq 1$, $\mathbb{T}_t^y \mathrm{G}_1 T_t^x = \mathrm{G}_1 H_t$.
  Geometric tilts in x are reversed by simulated tilts in y up to a zoom-out scale change.

ASIFT Algorithm

1. Apply a dense set of rotations to both images u and v.
2. Apply in continuation a dense set of *simulated* tilts $\mathbb{T}_t^x$ to all rotated images.
3. Perform a SIFT comparison of all pairs of resulting images.

Notice that by the relation

$$\mathbb{T}_t^x R\left(\frac{\pi}{2}\right) = R\left(\frac{\pi}{2}\right) \mathbb{T}_t^y, \tag{1}$$

ASIFT simulates tilts in the y direction, up to a rotation.

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 *by cameras at infinity with arbitrary position and focal lengths. 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 that ASIFT works

$$BA^{-1} = H_\lambda R_1 T_t^x R_2.$$

Compare: $u = \mathbf{G}_1 u_0, \quad v = \mathbf{G}_1 R_1 T_t^x R_2 H_\lambda u_0.$

Applying R_1^{-1} to v yields $v \to v' = \mathbf{G}_1 T_t^x R_2 H_\lambda u_0.$

Then revert T_t^x by applying the simulated tilt in the y direction to v':

$\mathbb{T}^y := T_t^y \mathbf{G}^y_{\sqrt{t^2-1}} *_y$. Indeed (main formula):

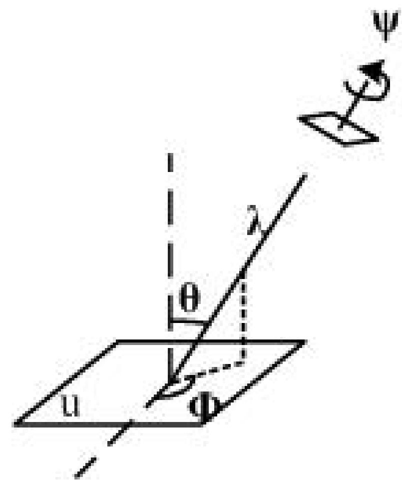
$$\mathbb{T}_t^y \mathbf{G}_1 T_t^x = \mathbf{G}_1 H_t.$$

Thus by application of $\mathbb{T}^y$ to v' we get

$$v' \to \mathbf{G}_1 H_t R_2 H_\lambda u_0 = \mathbf{G}_1 H_{t\lambda} R_2 u_0,$$

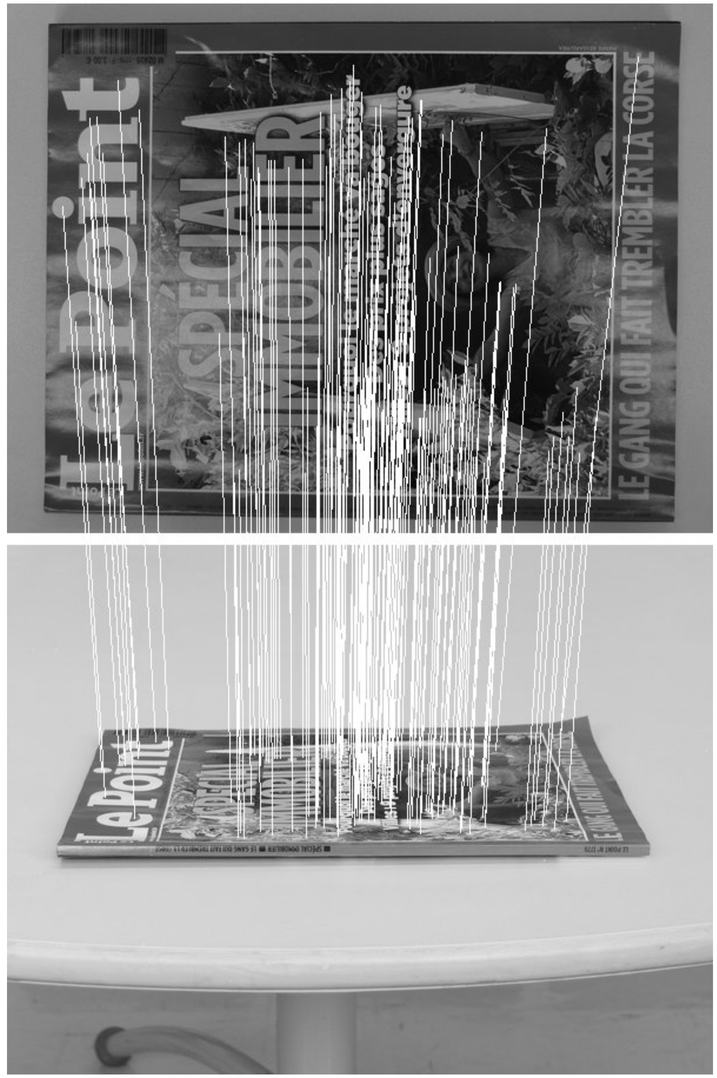
which is SIFT equivalent to u.

Parameter Sampling Range



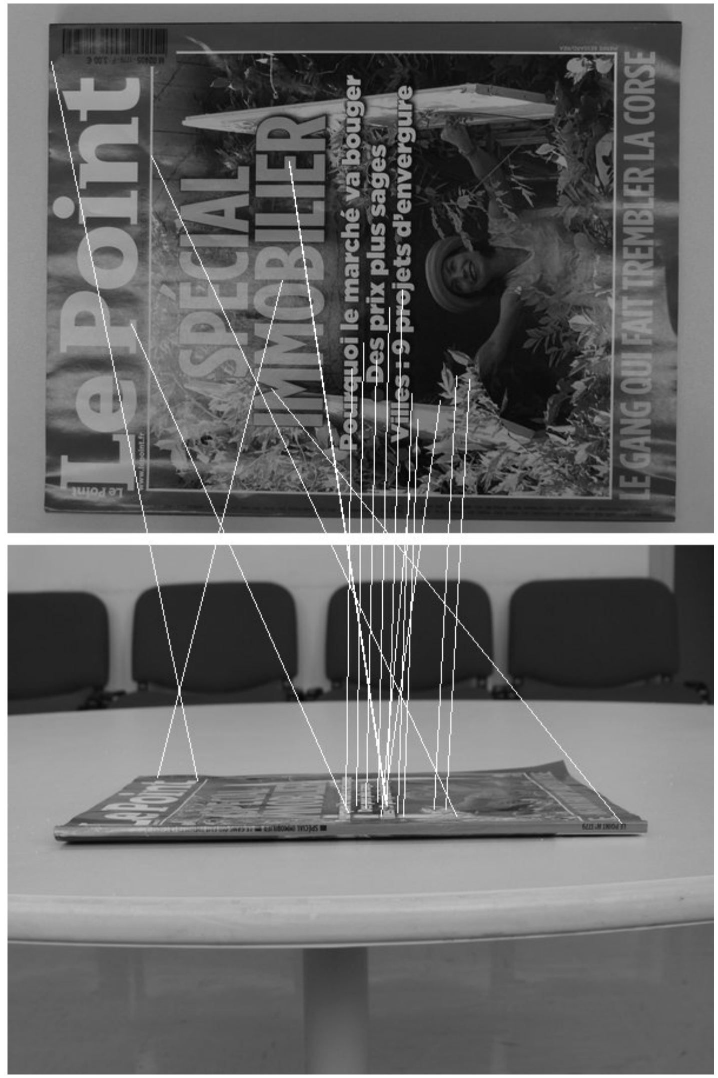
- Longitude angle $\phi \in [0, \pi)$.
  - $\mathbf{R}_1(\psi)\mathbf{T}_t\mathbf{R}_2(\phi + \pi) = \mathbf{R}_1(\psi + \pi)\mathbf{T}_t\mathbf{R}_2(\phi)$.
- Tilt $t = 1/\cos\theta \in [1, t_{\max}]$.
  - Physical limitation: planar and Lambertian.
  - $t_{\max} = 4\sqrt{2}$ obtained experimentally.
  - The resulting $\tau_{\max} = 32$.

Parameter Sampling Range



t = 3 ($\theta = 70.5°$), 151 correct ASIFT matches.

Parameter Sampling Range



t = 5.2 ($\theta = 78.9°$), 12 correct ASIFT matches.

Parameter Sampling Range

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

Parameter Sampling Range

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

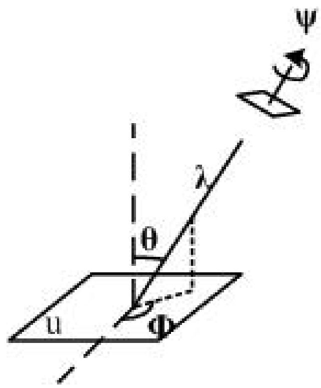
Parameter Sampling Range

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

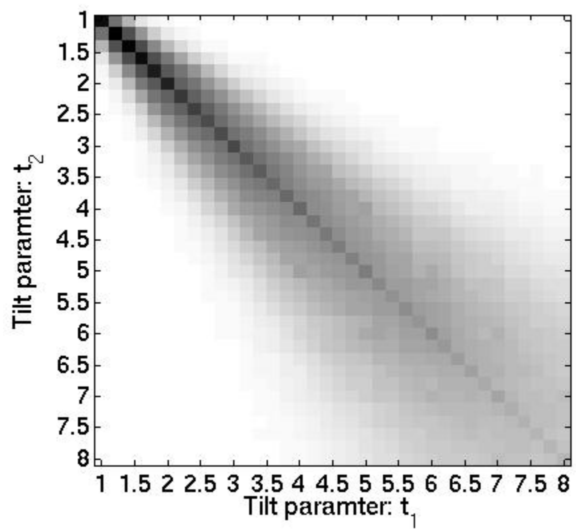
Parameter Sampling Range

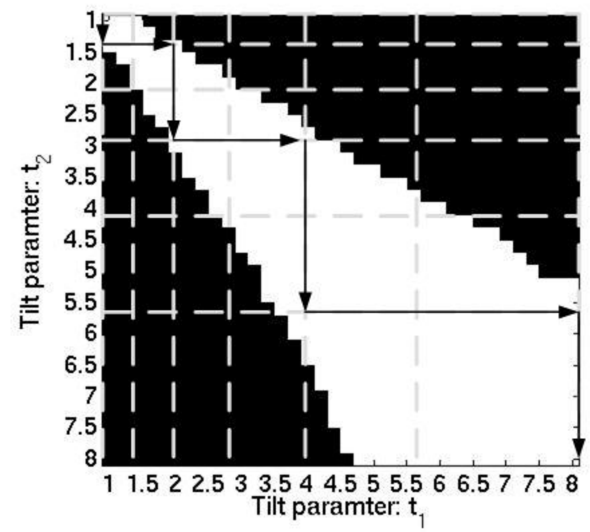
t = 8 ($\theta = 82.8°$), 0 ASIFT match.

Parameter Sampling Step: $\triangle t$



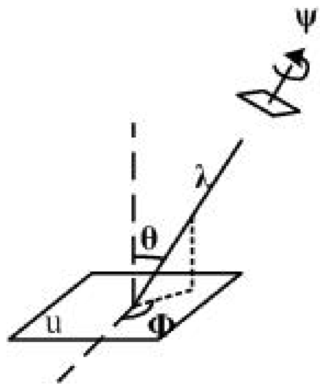
- $t = 1/\cos\theta$, θ is the latitude angle.
- θ sampled with higher precision when $\theta \to 90°$.
- Geometric sampling of t: $\triangle t = t_{k+1}/t_k$.
- $\triangle t = \sqrt{2}$ is obtained experimentally: compare $\mathbf{u} = \mathbf{T}_{t_1}\mathbf{u}_0$ and $\mathbf{v} = \mathbf{T}_{t_2}\mathbf{u}_0$ with SIFT.



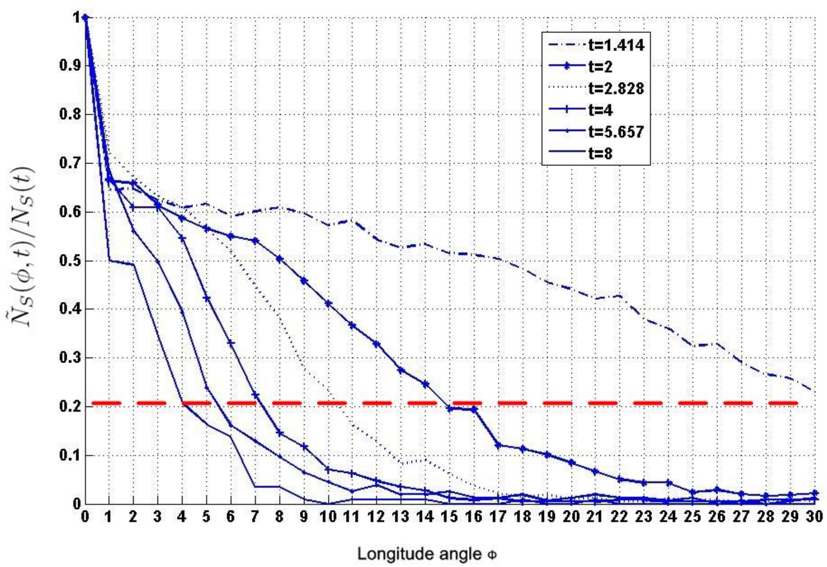


t	1	$\sqrt{2}$	2	$2\sqrt{2}$	4	$4\sqrt{2}$
θ	$0°$	$45°$	$60°$	$69.3°$	$75.5°$	$79.8°$

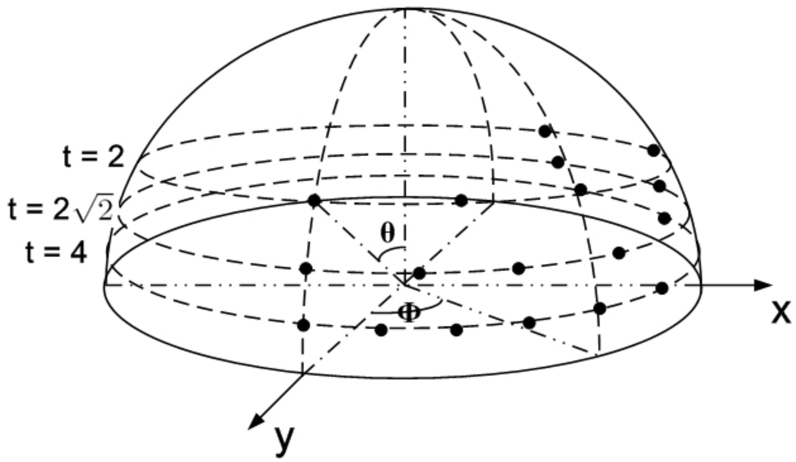
Parameter Sampling Step: $\triangle\phi$



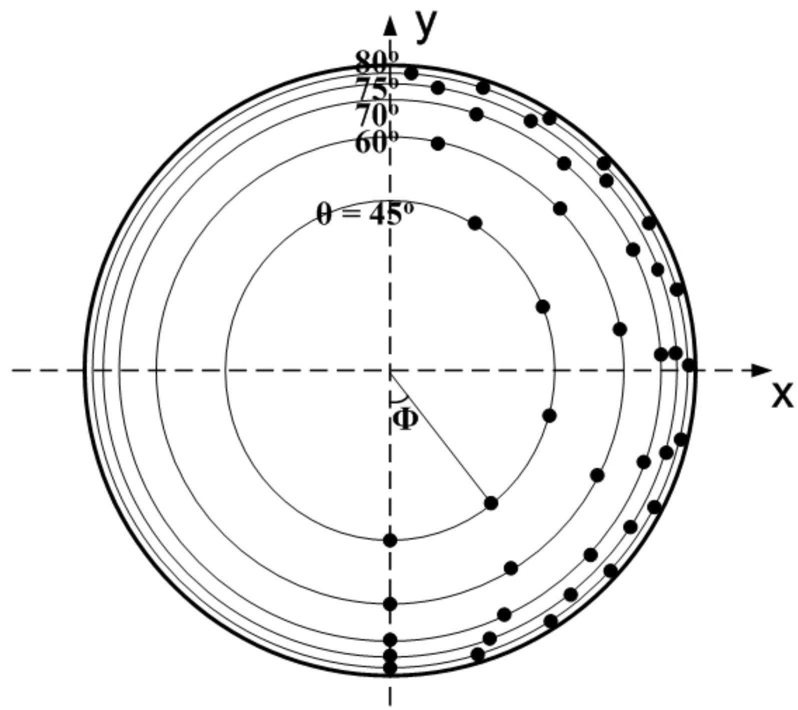
- ϕ: longitude angle.
- ϕ sampled with higher precision when $\theta \to 90°$: $t \uparrow \Rightarrow \triangle\phi \downarrow$.
- Arithmetical sampling of ϕ: $\triangle\phi = \phi_{k+1} - \phi_k$.
- $\triangle\phi = 2 \times \frac{36°}{t} = \frac{72°}{t}$ is obtained experimentally: compare $\mathbf{u} = \mathbf{T}_t\mathbf{R}_1(\phi)\mathbf{u}_0$ and $\mathbf{v} = \mathbf{T}_t\mathbf{u}_0$ with SIFT.



Parameter Sampling



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.).

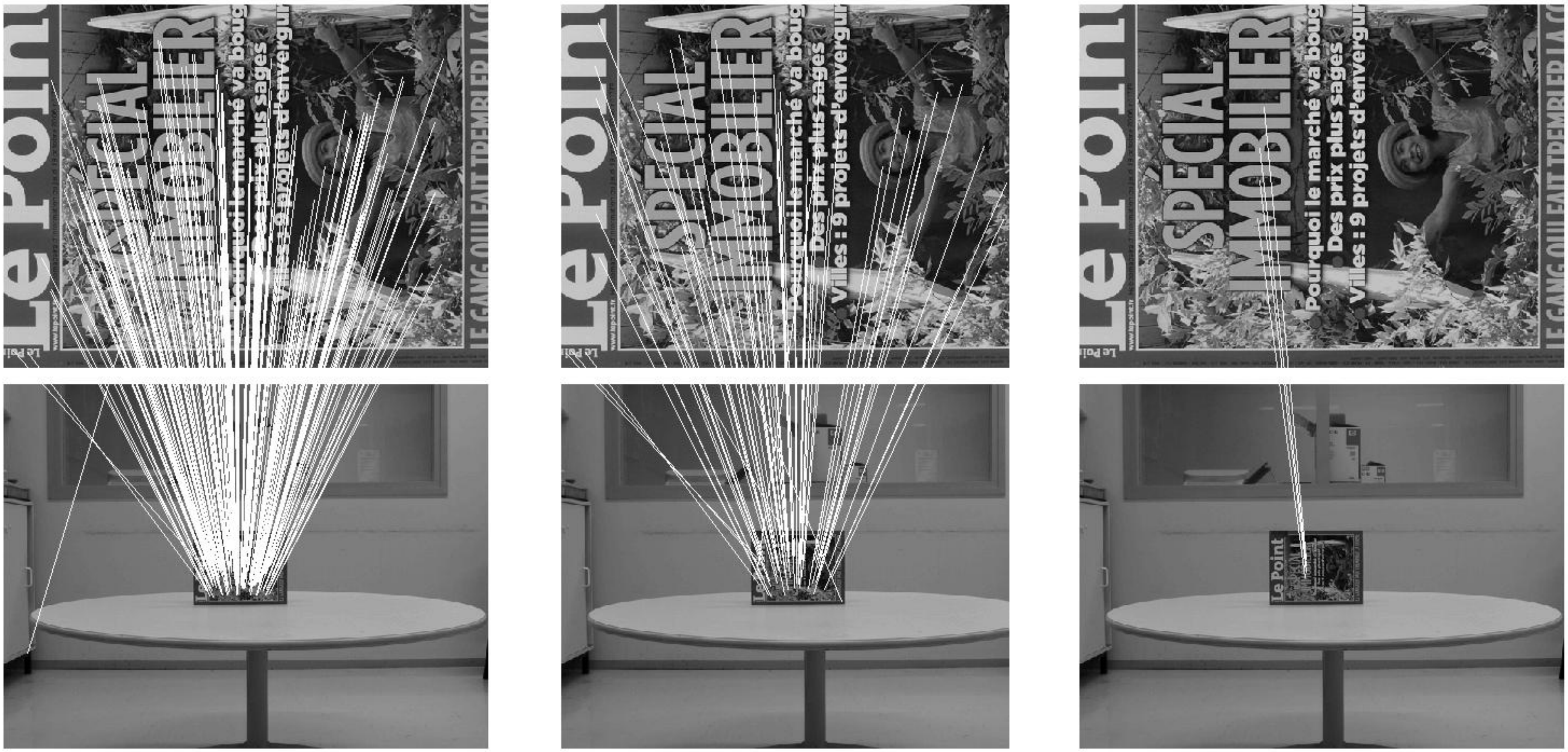
ASIFT Complexity

- Complexity proportional to (area of query) $\times$ (searched area).
- Image area proportional to number of simulated tilts.
  - $t = 2^{k/2}, \;\; k = 0, \ldots, K$.
  - $\phi \in [0°, 180°), \;\; \triangle\phi = \frac{72°}{t}$: $|\{\phi(t)\}| \sim t$.
  - At tilt t, image area $\sim 1/t$.
- Example: $t_{\max} = 4\sqrt{2}$ (i.e. $K = 5$), $r \times r = 3 \times 3$ subsampling.
  - Image area on one side:

$$\frac{1 + K\frac{180°}{72°}}{r^2} = 1.5 \times \text{original image}$$

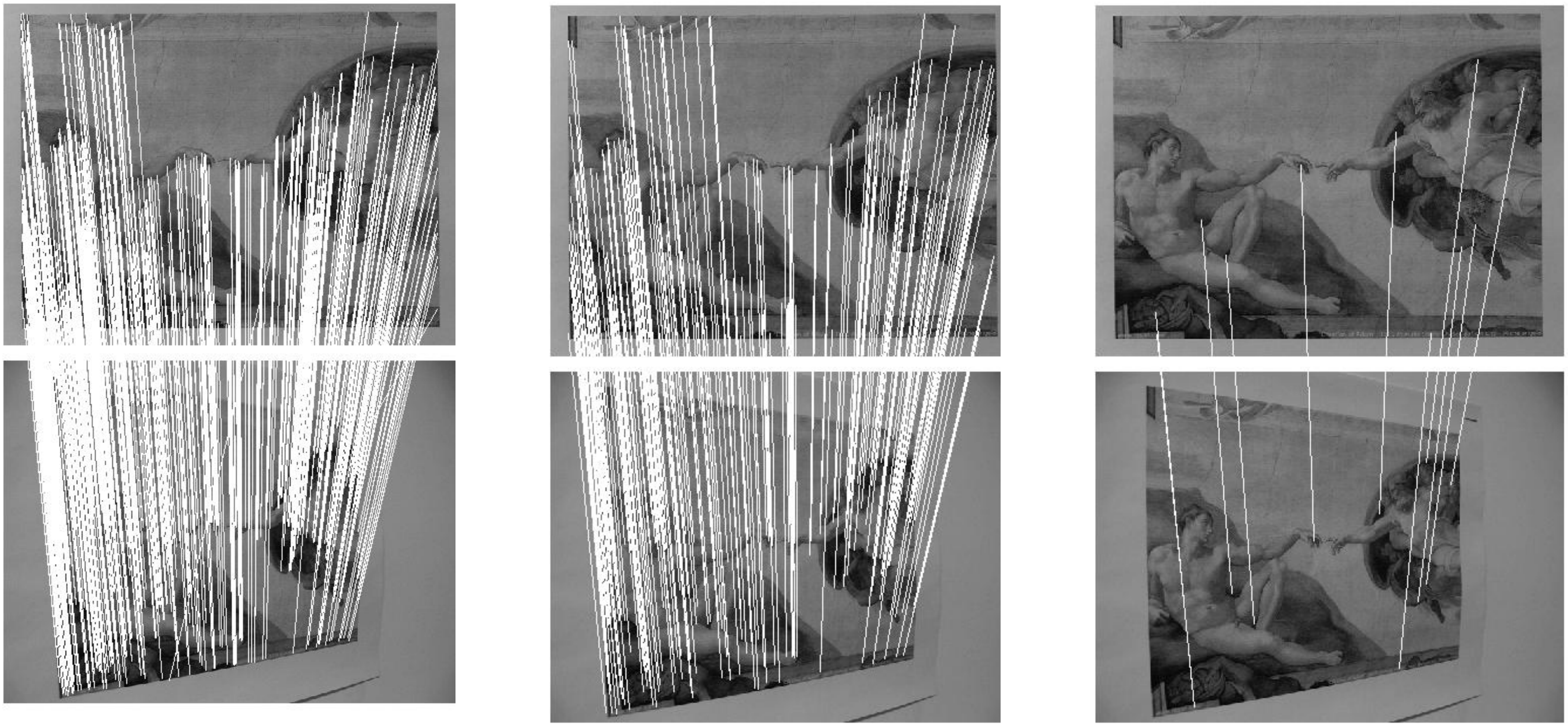
  - One sided ASIFT (tilts simulated on query only): total complexity $= 1.5\times$SIFT, $\tau_{max} = 4\sqrt{2} \simeq 5.6$.
  - Two sided ASIFT (tilts simulated on query and searched images): total complexity $= (1.5)^2\times$ SIFT $= 2.25$ SIFT, $\tau_{max} = 32$.

Experiments: Image Matching



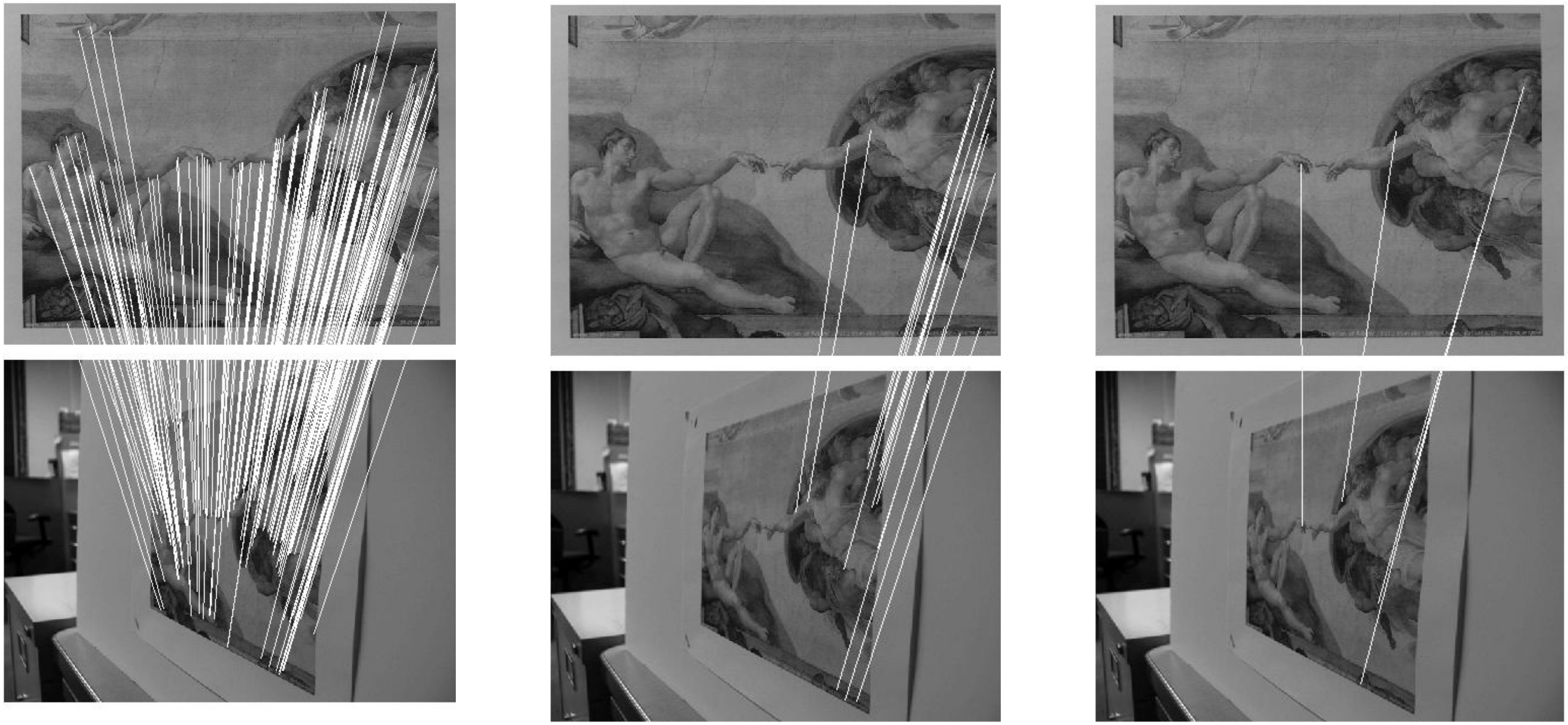
Zoom change. Number of correct matches: ASIFT (left)—222; SIFT (middle)—87; MSER (right)—4.

Experiments: Image Matching



Frontal v.s. -45° angle, zoom $\times 1$: absolute tilt $t = 2$ (middle), $t < 2$ (left part), $t > 2$ (right part). Number of correct matches: ASIFT (left)—624; SIFT (middle)—236; MSER (right)—11.

Experiments: Image Matching

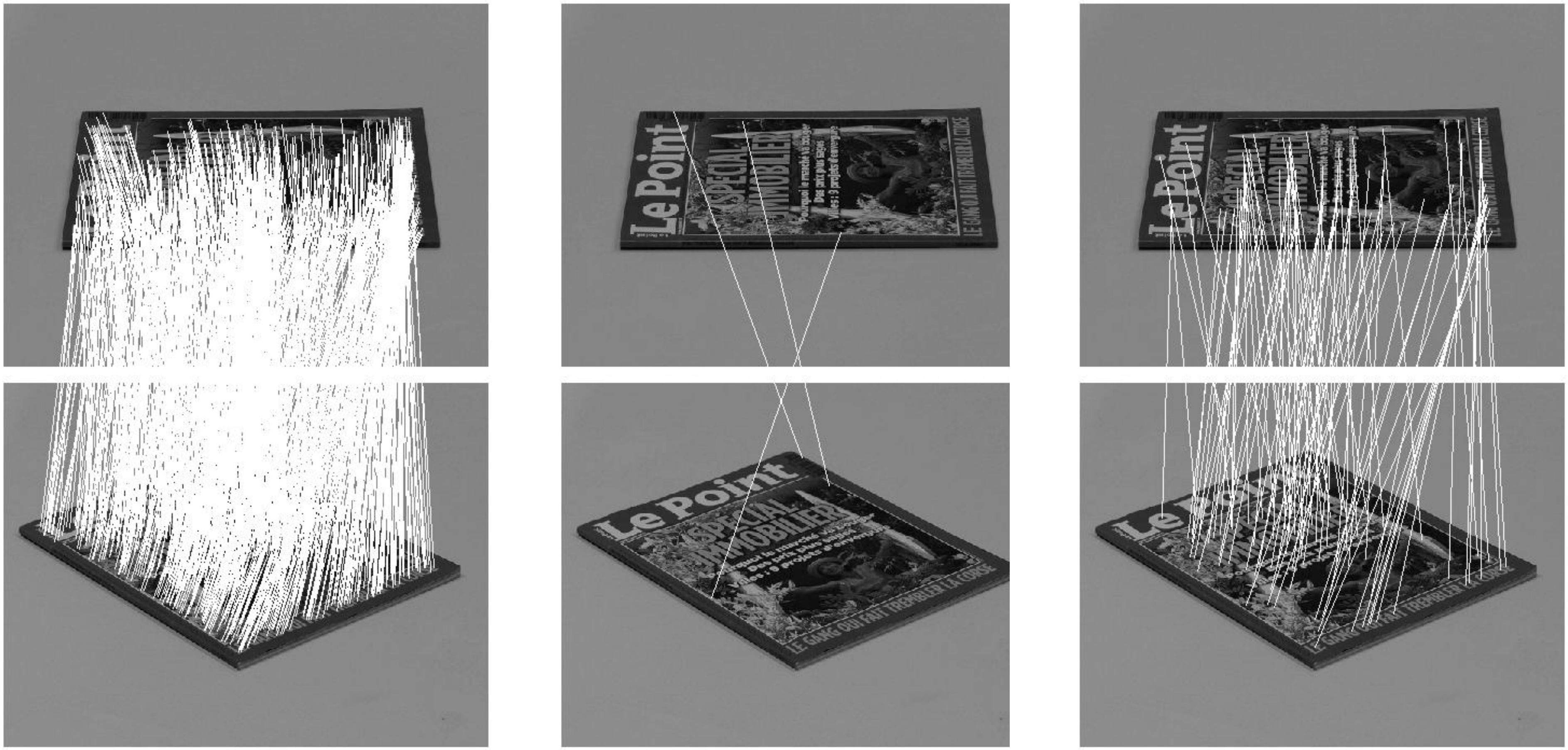


Frontal v.s. 75° angle, zoom ×1: absolute tilt $t = 4$ (middle), $t < 4$ (left part), $t > 4$ (right part). Number of correct matches: ASIFT (left)—202; SIFT (middle)—15; MSER (right)—5.

Experiments: Image Matching

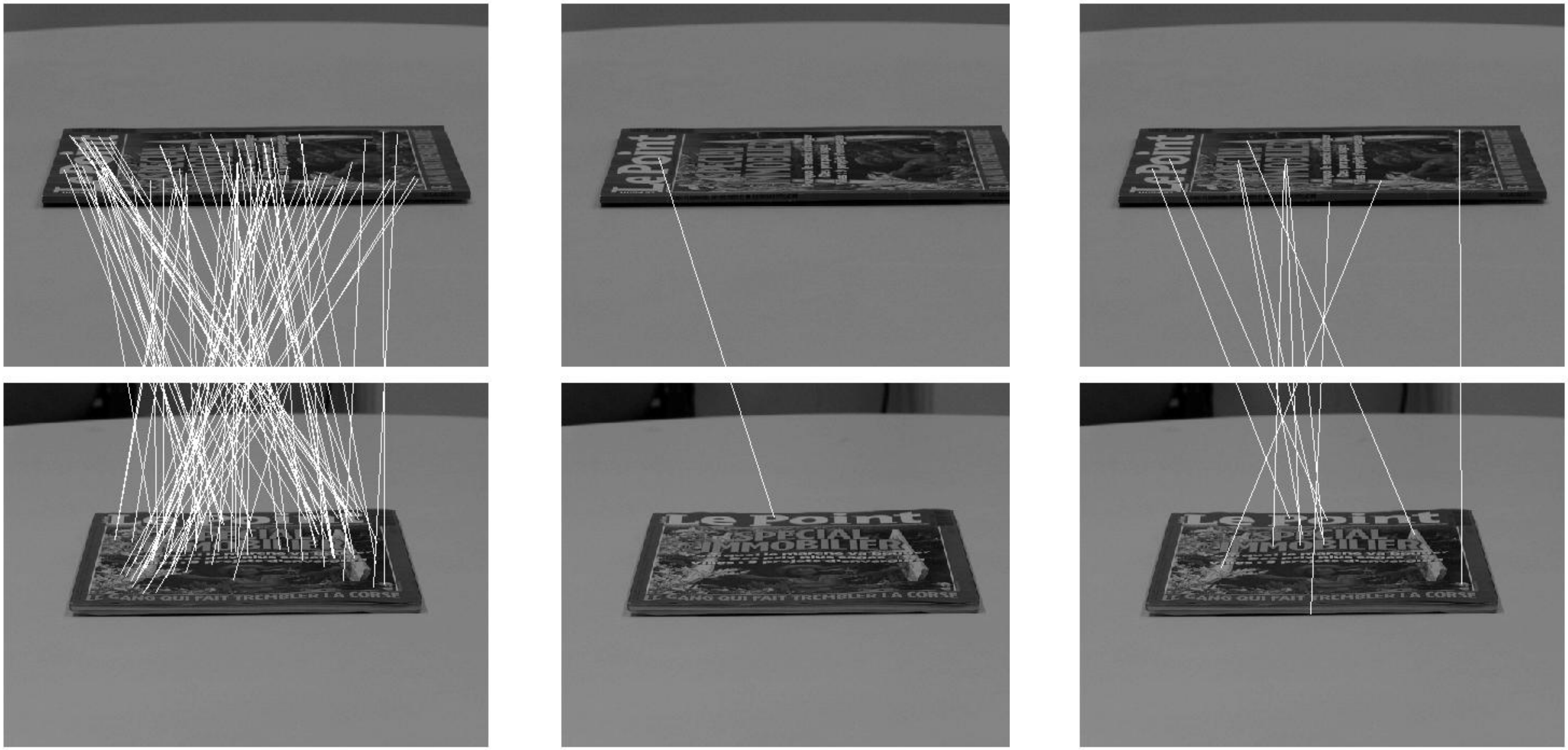
Frontal v.s. -80° angle, zoom $\times 10$: absolute tilt $t = 5.8$. Number of correct matches: ASIFT (left)—75; SIFT (middle)—1; MSER (right)—2.

Experiments: Image Matching



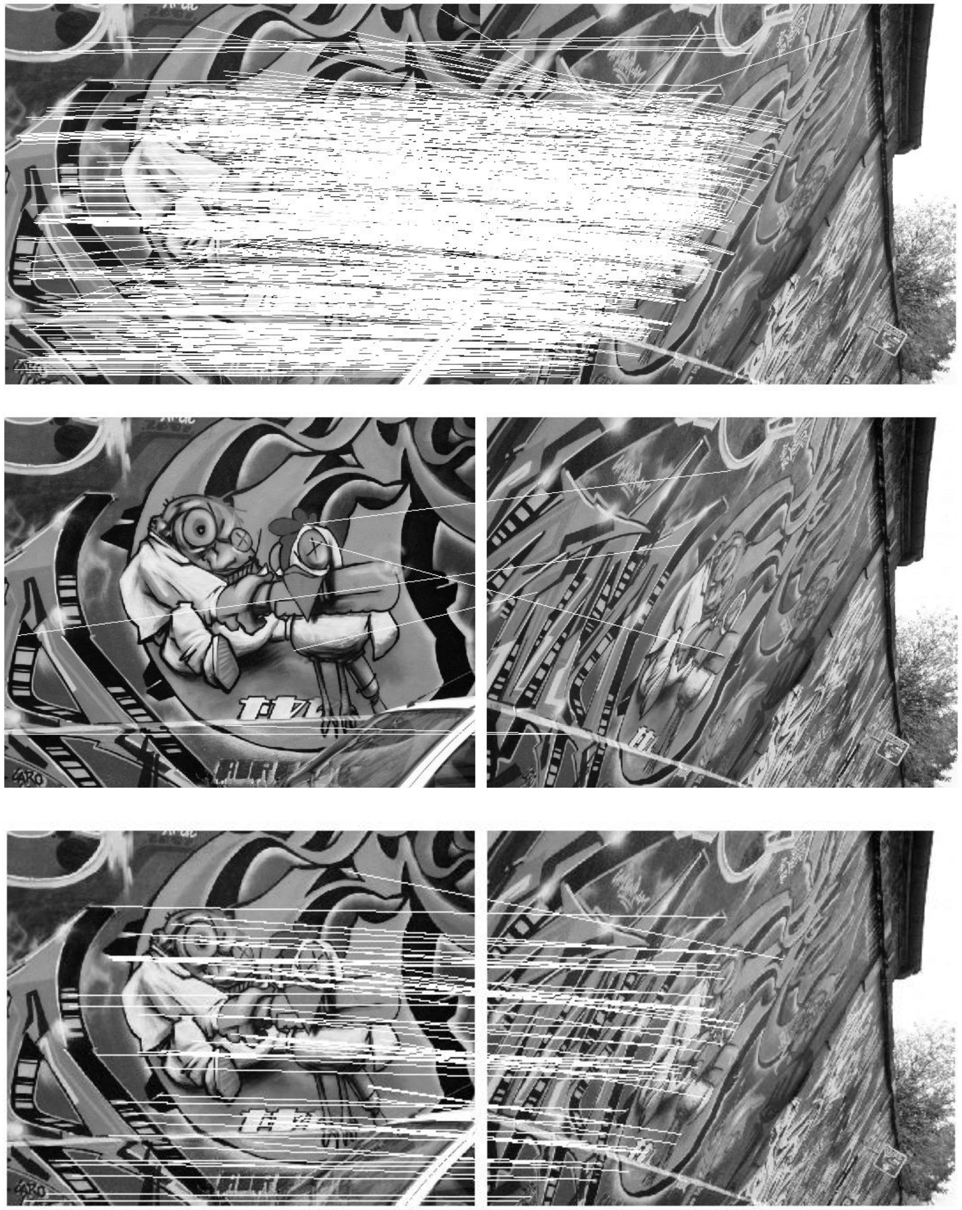
Correspondences between the magazine images taken with absolute tilts $t_1 = t_2 = 2$ with longitude angles $\phi_1 = 0^\circ$ and $\phi_2 = 50^\circ$, transition tilt $\tau = 3$. Number of correct matches: ASIFT (left)—881; SIFT (middle)—2; MSER (right)—87.

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



Graffiti 1 vs 6.

Transition tilt: $\tau \approx 3.2$.

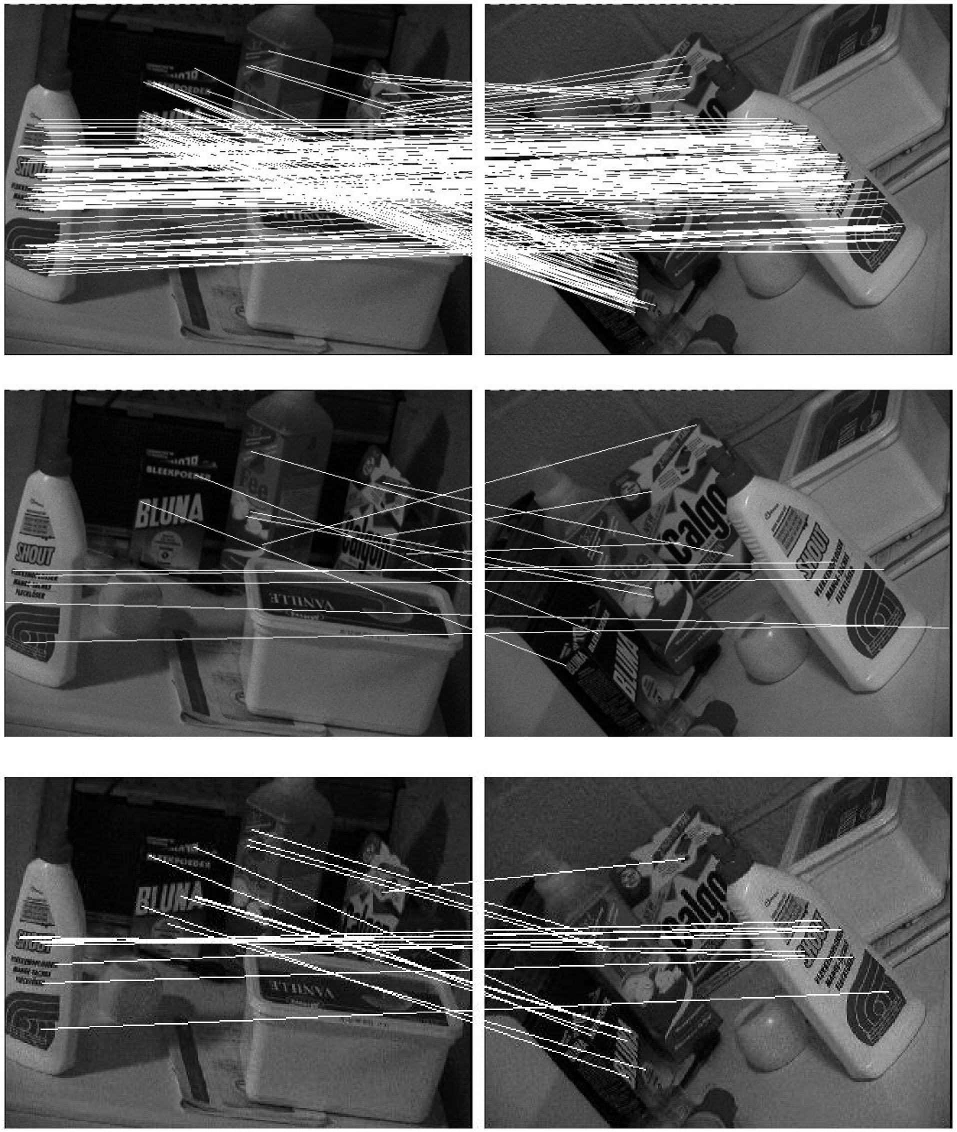
Number of correct matches:

ASIFT (top)—721;

SIFT (middle)—0;

MSER (bottom)—70.

Experiments: Image Matching



Images proposed by Matas et al.

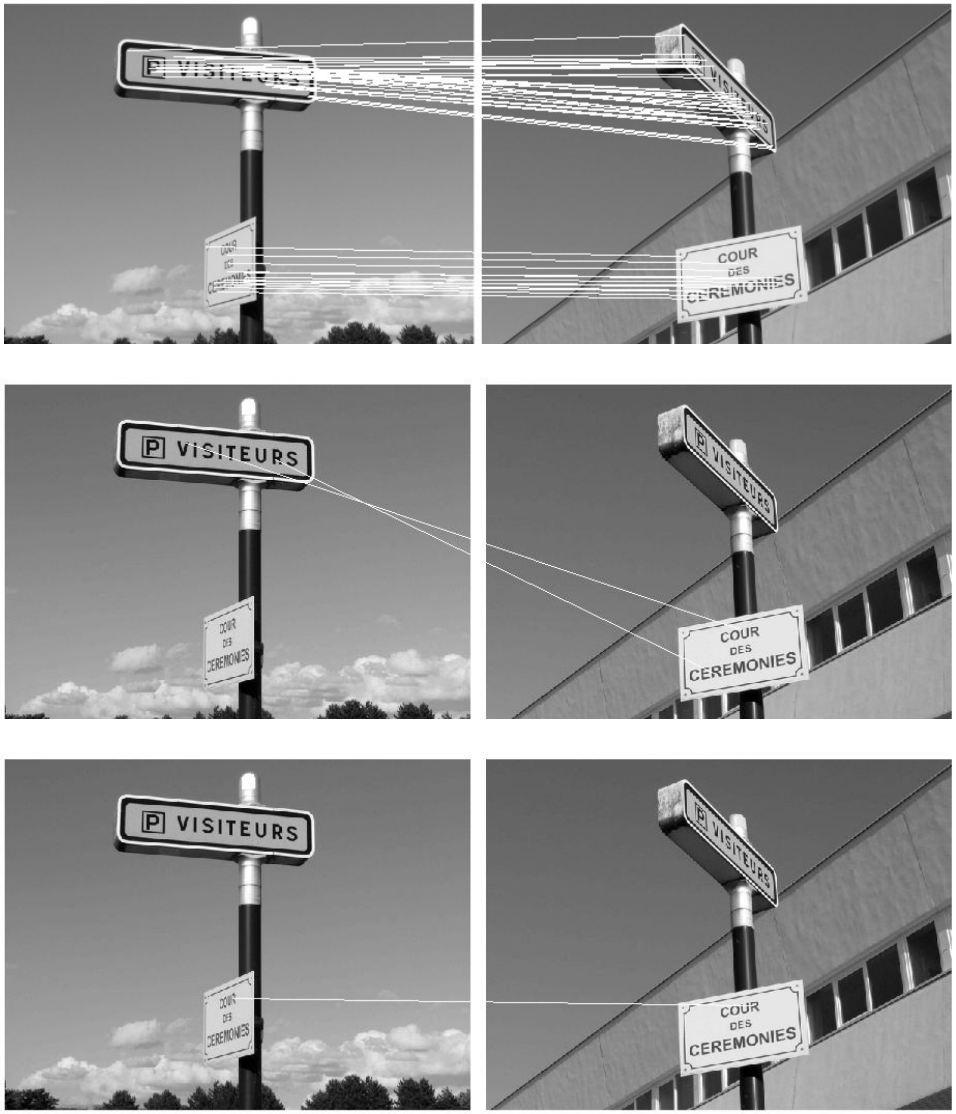
Number of correct matches:

ASIFT (top)—254;

SIFT (middle)—10;

MSER (bottom)—22.

Experiments: Image Matching



Road signs.

Transition tilt: $\tau \approx 2.6$.

Number of correct matches:

ASIFT (top)—50;

SIFT (middle)—0;

MSER (bottom)—1.

Experiments: Image Matching

Parkings.

Transition tilt: $\tau \approx 15$.

Number of correct matches:

ASIFT (top)—78;

SIFT (middle)—0;

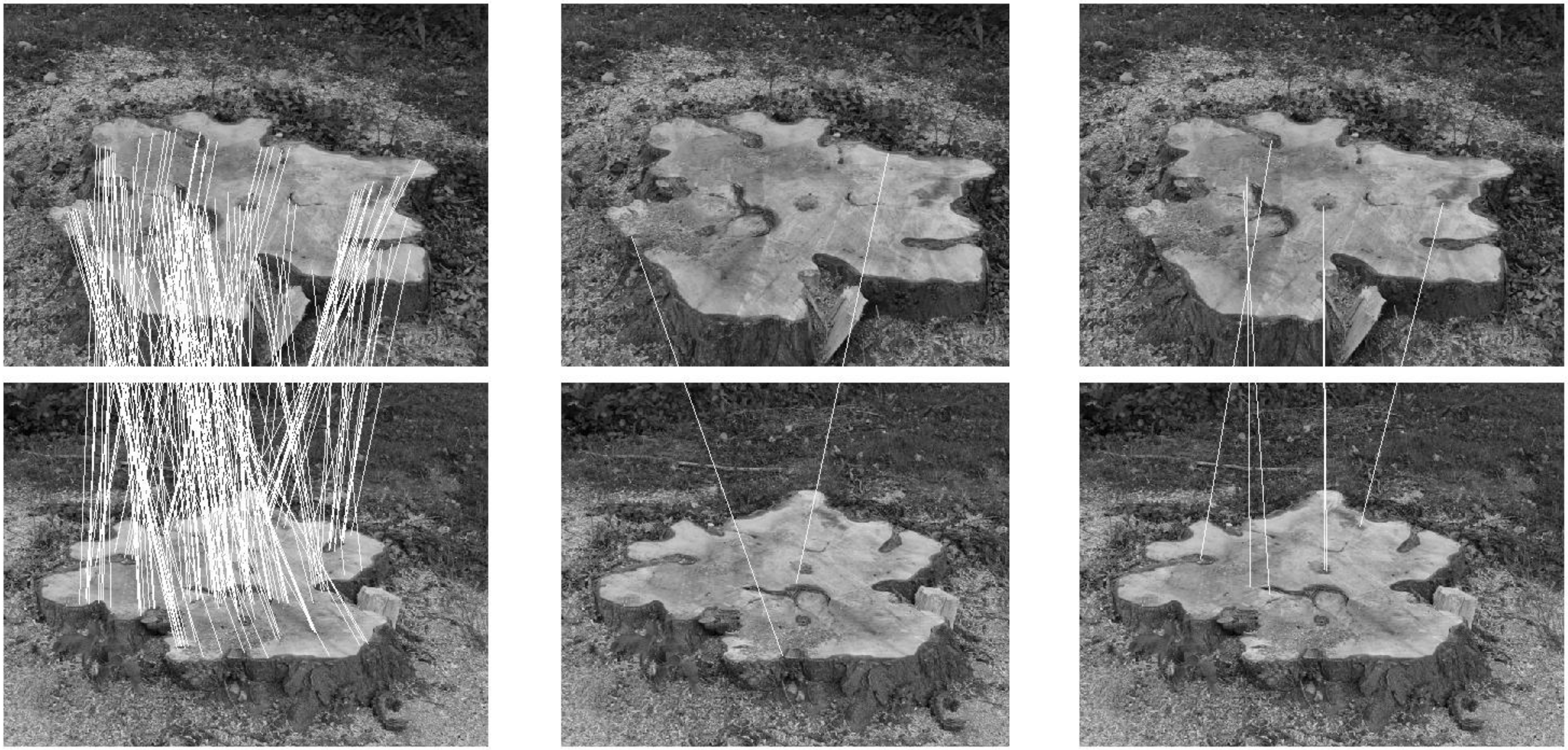
MSER (bottom)—0.

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



Stump. Transition tilt $\tau = 2.6$. Number of correct matches: ASIFT (left)—168; SIFT (middle)—1; MSER (right)—6.

Experiments: Image Matching

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

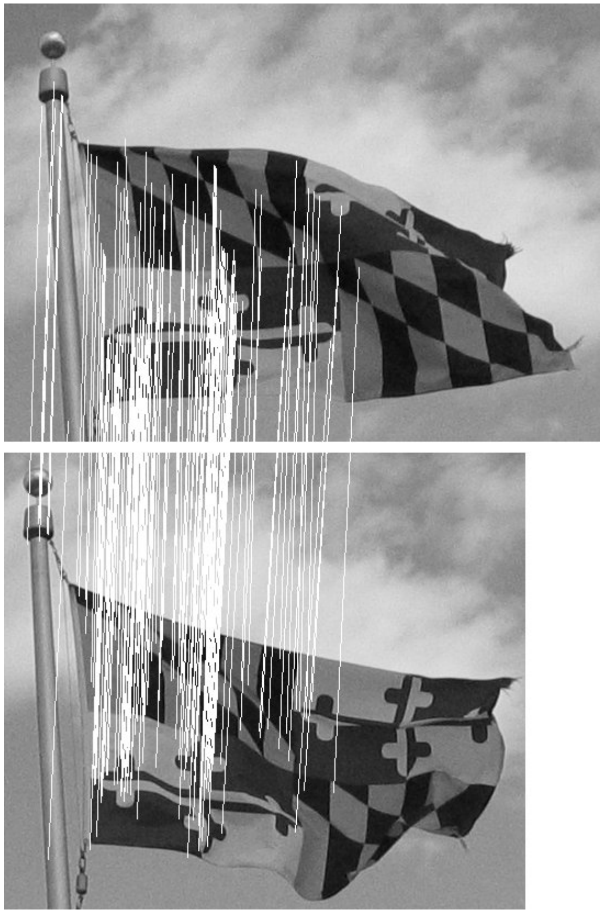
Number of correct matches: ASIFT (left)—378, SIFT (middle)—6, MSER(right)—17.

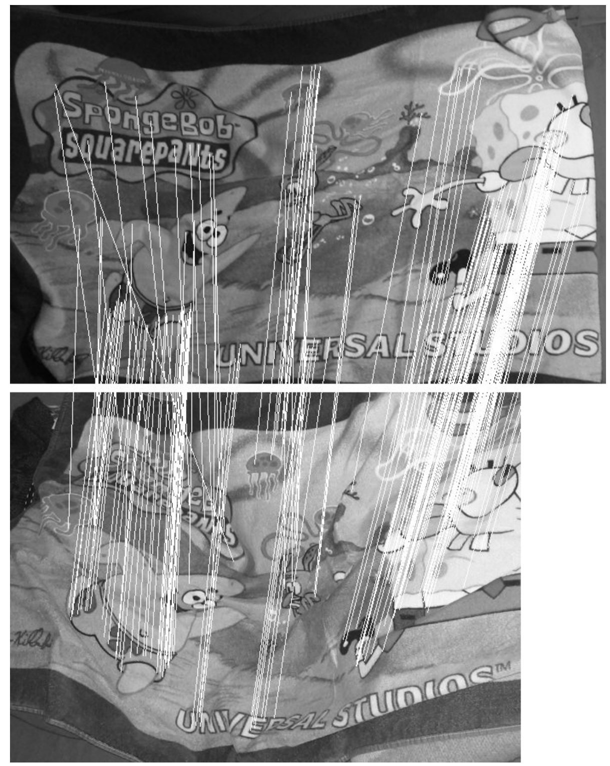
Experiments: Image Matching

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

Number of correct matches: ASIFT (left)—22, SIFT (right)—1.

Experiments: Image Matching





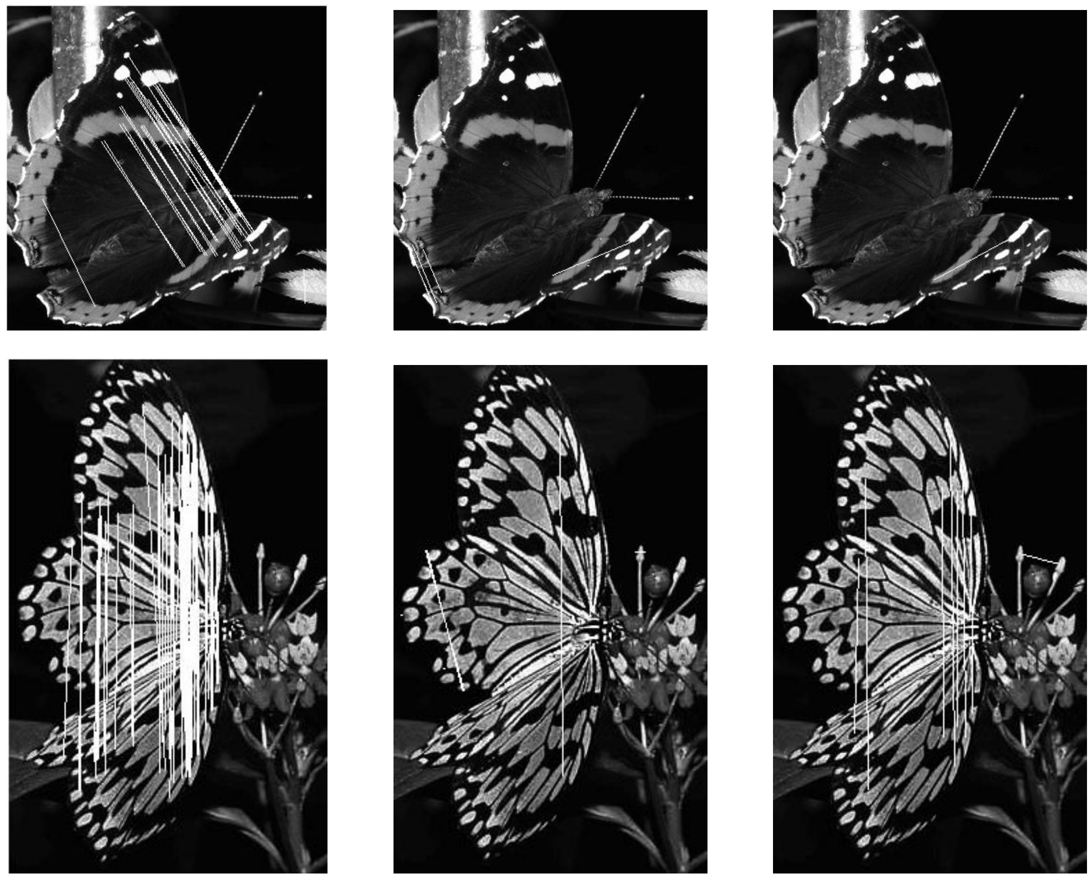
Left: flag. ASIFT (shown)—141, SIFT—31, MSER—2.

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

Experiments: Object Tracking

Symmetry Detection in Perspective

Symmetry detection = image comparison with its flipped version.



ASIFT SIFT MSER

Summary: fully affine-invariant image comparison

- Camera interpretation of affine space: 6 parameters.
- High transition tilts.
- Simulation v.s. normalization.
- Simulate scale, longitude and latitude.
- Normalize translation and rotation.
- Mathematical proof: fully affine-invariant.
- Sample the camera hemisphere (longitude and latitude).
- Multi-resolution acceleration.
- Reasonably small complexity.
- State-of-the-art results.

Reference:

- 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, proc. IEEE ICASSP, Taipei, 2009.
- J.M. Morel and G.Yu, On the consistency of the SIFT Method, Preprint, CMLA 2008-26, Sept 2008.

Patent:

- G. Yu and J.M. Morel, Viewpoint invariant object and shape recognition in digital images, pending, 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