ASIFT: Fully Affine Invariant !

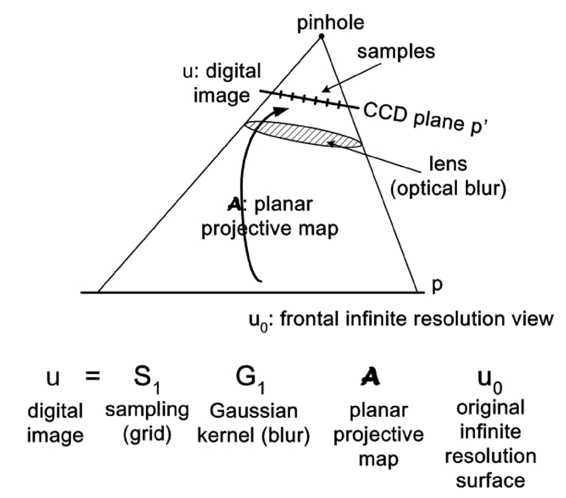
Guoshen Yu
CMAP, Ecole Polytechnique
91128 Palaiseau Cedex, France

Jean-Michel Morel
CMLA, ENS Cachan, 61 avenue du President Wilson,
94235 Cachan Cedex, France

A fully affine invariant image comparison method, Affine-SIFT (ASIFT) is introduced. While SIFT is fully invariant with respect to only four parameters namely zoom, rotation and translation, the new method treats the two left over parameters : the angles defining the camera axis orientation. Against any prognosis, simulating all views depending on these two parameters is feasible. The method permits to reliably identify features that have undergone very large affine distortions measured by a new parameter, the *transition tilt*. State-of-the-art methods hardly exceed transition tilts of 2 (SIFT), 2.5 (Harris-Affine and Hessian-Affine) and 10 (MSER). ASIFT can handle transition tilts up 36 and higher.

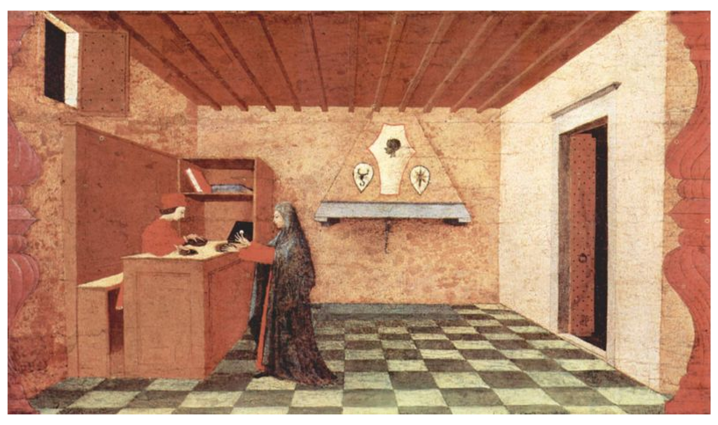
Why affine invariance?

Image formation model



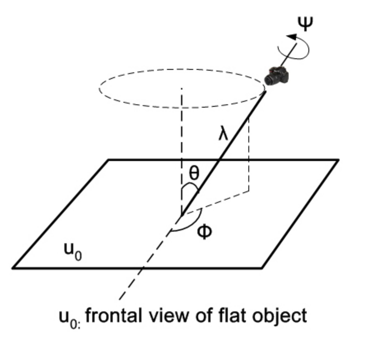
Perspective is local affine!

Local perspective effects can be modeled by local affine transforms $u(x,y) \to u(ax+by+e, cx+dy+f)$ in each image region.



The global deformation of the ground is strongly projective (a rectangle becomes a trapezoid), but the local deformation is affine: each tile on the pavement is almost a parallelogram.

Main decomposition formula



$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = H_\lambda R_1(\psi) T_t R_2(\phi) = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

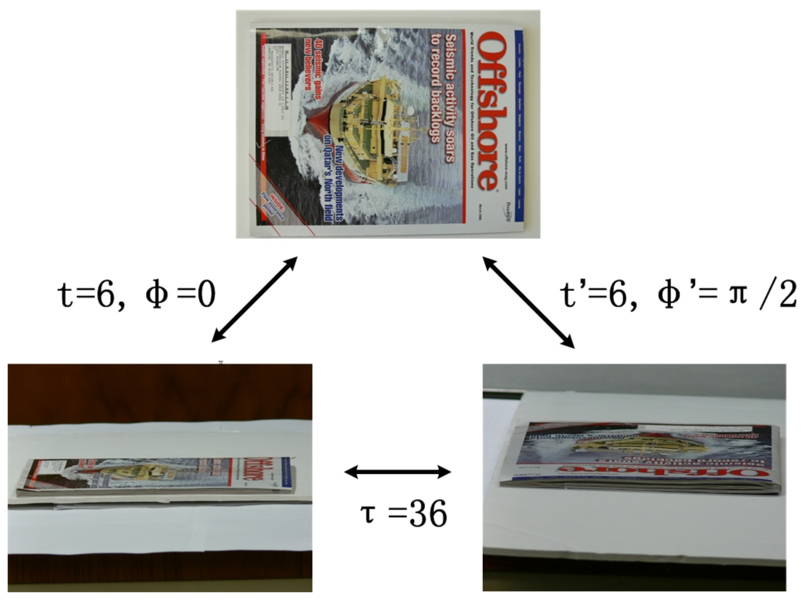
- A: affine map with strictly positive determinant.
- ϕ: *longitude* angle between optical axis and a fixed vertical plane.
- $\theta = \arccos(1/t)$: *latitude* angle between optical axis and the normal to the image plane. **Tilt** $t > 1 \leftrightarrow \theta \in [0°, 90°]$.
- ψ: rotation angle of camera around optical axis.
- λ: *zoom* parameter.

Transition tilts can be very high!

Both compared images $u_1(x,y) = u(A(x,y))$ and $u_2(x,y) = u(B(x,y))$ are usually slanted views. The *transition tilt* quantifies the tilt between two such images.

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

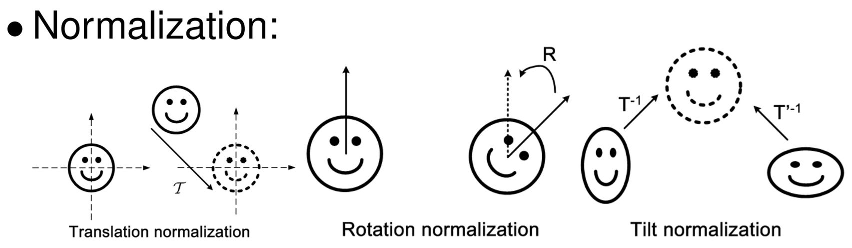
The **transition tilt** satisfies $t_1/t_2 \leq \tau \leq t_1 t_2$.



Competitors prefer normalization.

Simulation or normalization?

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



– Translation $\mathcal{T}$ and rotation $\mathbf{R}$: OK. $\mathbf{G}_1\mathbf{R}u_0 = \mathbf{R}\mathbf{G}_1 u_0$.
– Zoom $\mathbf{H}_\lambda$ and tilt $\mathbf{T}$: not exact. $\mathbf{H}_\lambda \mathbf{G}_1 u_0 \neq \mathbf{G}_1 \mathbf{H}_\lambda u_0$.

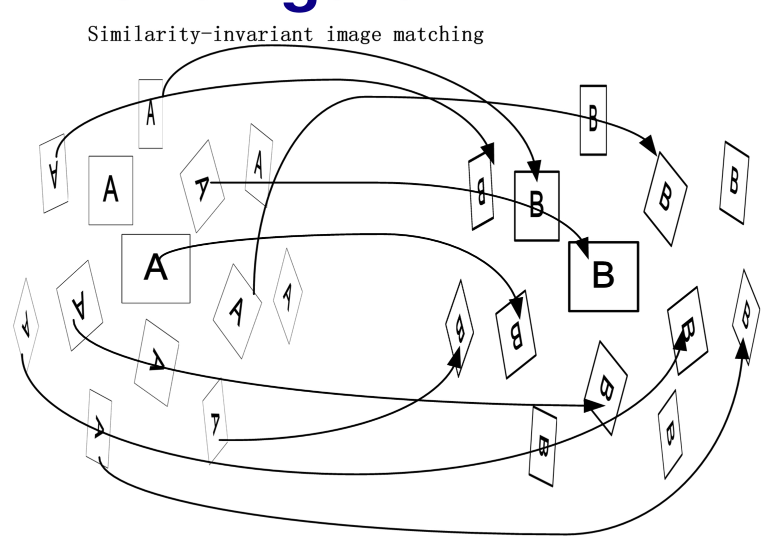
SIFT does both.

- Rotation and translation are *normalized*.
- Zoom is *simulated* in the scale space.
- No treatment on latitude and longitude: $\tau_{\max} < 2$.

MSER, Harris-Affine, Hessian-Affine

- *Normalize* all affine parameters.
- Limited performance on scale- and tilt-invariance.
- MSER: $\tau_{\max} < 10$ in optimal conditions.

ASIFT in one figure



ASIFT—simulate latitude and longitude, then apply SIFT.

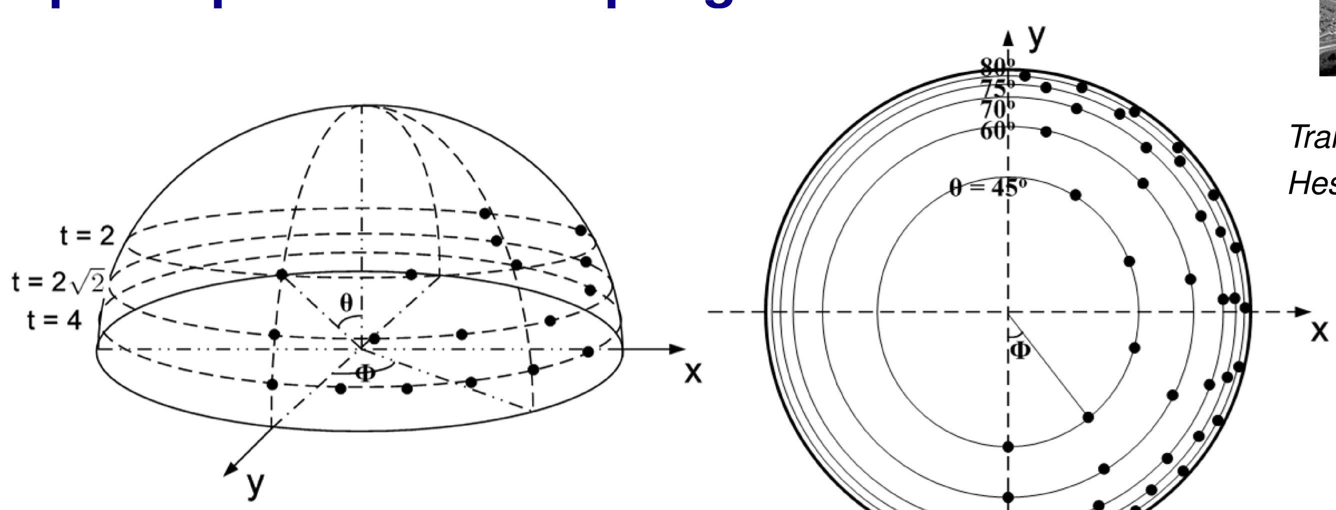
ASIFT is fully affine invariant.

Theorem 1 *Let* $u = G_1 A \mathcal{T}_1 u_0$ *and* $v = G_1 B \mathcal{T}_2 u_0$ *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.*

Why it works: Tilt reverts tilt.

A tilt in one direction is reversed by simulating a tilt of same amount in the orthogonal direction, up to a zoom-out scale change.

Sparse parameter sampling



When the latitude angle θ (or tilt t) increases:
- Sampling density increases.
- Simulated image size decreases.

Two-resolution acceleration — a good deal.

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

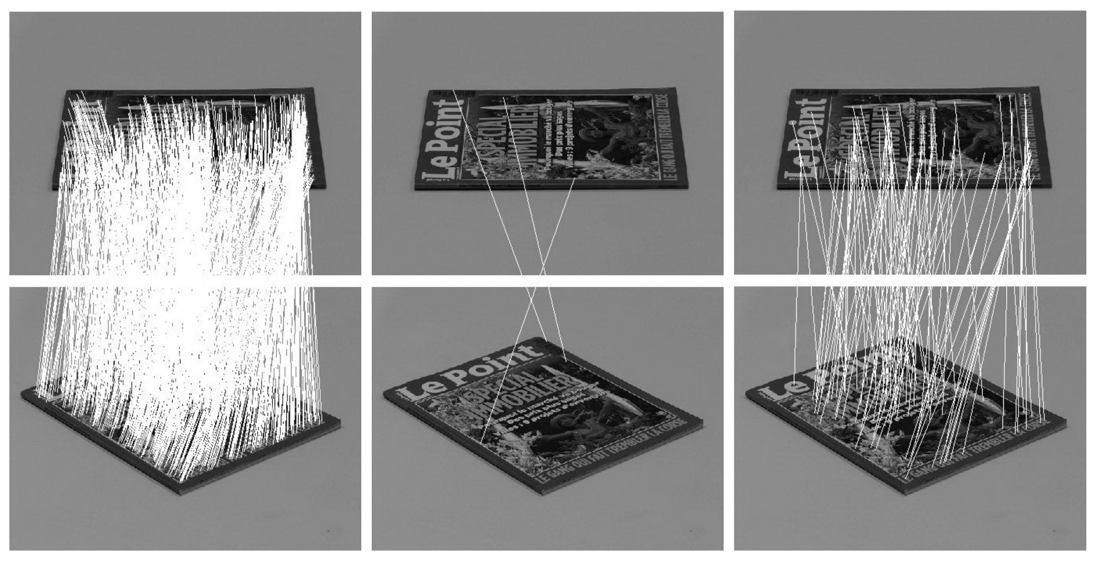
ASIFT has just twice SIFT complexity.

- 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 longitude samplings for tilt t is about $2.5t$.
  – At tilt t, simulated 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, $\tau_{max} = 32$.

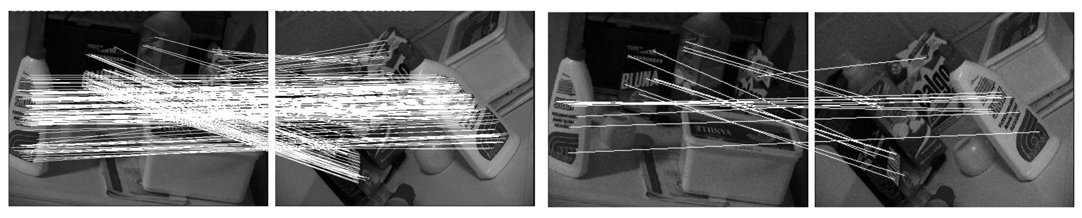
ASIFT attains tilts 10 times larger!

Transition tilt $t \approx 36$. *ASIFT (shown) – 116 correct matches. SIFT, Harris-Affine, Hessian-Affine and MSER fail completely.*

Transition $\tau \approx 5.8$. *ASIFT (shown) – 116, SIFT – 1, Harris-Affine (shown) – 1, Hessian-Affine – 0, and MSER (shown) – 2 correct matches.*

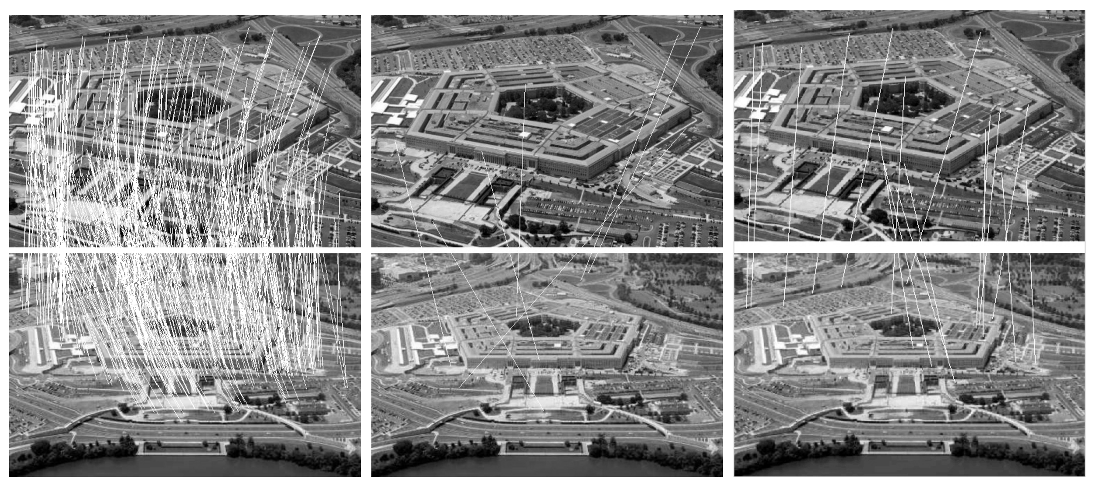


Transition $\tau \approx 3$. *ASIFT (shown) – 881, SIFT (shown) – 3, Harris-Affine – 1, Hessian-Affine – 3, and MSER (shown) – 87 correct matches.*



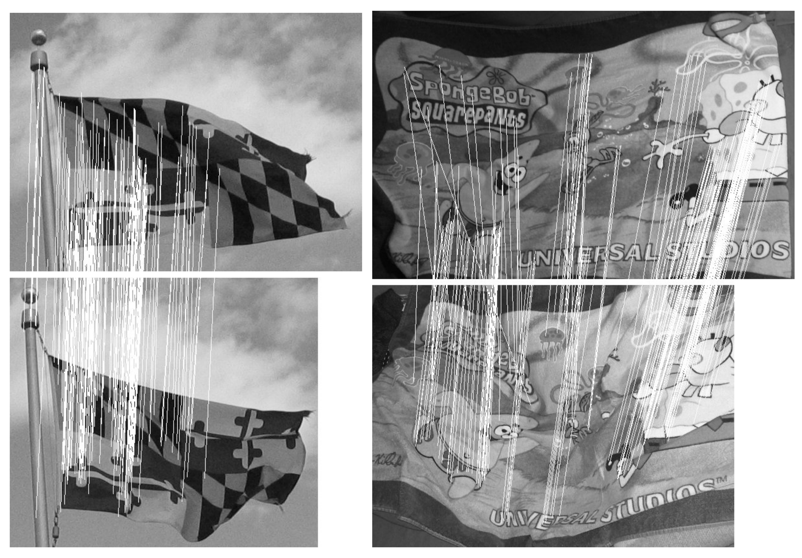
Transition tilt: $\tau \in [1.6, 3.0]$ *(images used by the authors of MSER). ASIFT (shown) – 254, SIFT–10, Harris-Affine–23, Hessian-Affine–11 and MSER (shown) – 22 correct matches.*

Transition tilt: $\tau \approx 2.6$. *ASIFT (shown) – 50, SIFT – 0, Harris-Affine – 0, Hessian-Affine – 0 and MSER (shown) – 1 correct matches.*



Transition $\tau \in [1.3, \infty)$. *ASIFT (shown) – 378, SIFT (shown)– 6, Harris-Affine – 2, Hessian-Affine – 8, and MSER (shown) – 17 correct matches.*

Transition $\tau \approx 5.8$. *ASIFT (shown) – 22, SIFT (shown)– 1, Harris-Affine – 0, Hessian-Affine – 0, and MSER – 0 correct matches.*



Deformable objects (images proposed by Ling and Jacobs). Left: flag. ASIFT (shown) – 141, SIFT – 31, Harris-Affine – 15, Hessian-Affine – 10 and MSER – 2 correct matches. Right: SpongeBob. ASIFT (shown) – 370, SIFT – 75, Harris-Affine – 8, Hessian-Affine – 6 and MSER – 4correct matches.

References:

- G. Yu and J.M. Morel, A Fully Affine Invariant Image Comparison Method, *IEEE ICASSP, Taipei*, 2009.
- J.M. Morel and G.Yu, ASIFT: A New Framework for Fully Affine Invariant Image Comparison, to appear in *SIAM Journal on Imaging Sciences*, 2009.

For more information,

Google ASIFT

- Try if your images match with an ASIFT *online demo*!
- Free software and source code.
- More fascinating examples and movies!
- An image dataset available.