

# Tutorial Outline

---

- **Part 1:** Background and challenges (20 min)
- **Part 2:** Preliminaries of invariance (20 min)
- *Q&A / Break (10 min)*
- **Part 3: Invariance in the era before deep learning (30 min)**
- **Part 4:** Invariance in the early era of deep learning (10 min)
- *Q&A / Coffee Break (30 min)*
- **Part 5:** Invariance in the era of rethinking deep learning (50 min)
- **Part 6:** Conclusions and discussions (20 min)
- *Q&A (10 min)*



A Historical Perspective of Data Representation  
Rethinking Deep Learning with Invariance: The Good, The Bad, and The Ugly

# Invariance in The Era Before Deep Learning

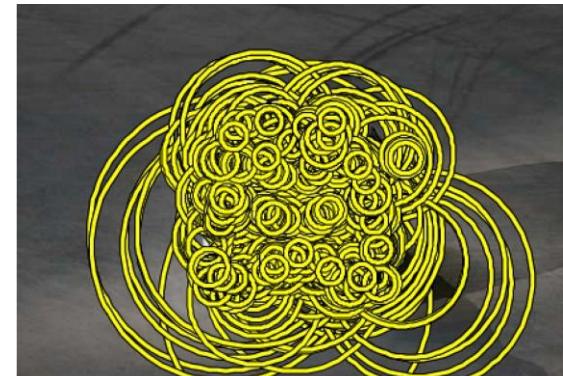
- In the era before deep learning, data representations were almost always designed by experts manually, driven by **knowledge** in math, physics, signal processing, and computer vision.
- Depending on the **spatial scope** of the action, these representations can be classified as **global**, **locally sparse** and **locally dense**. Such assumptions are different and lead to different realizations of invariance.



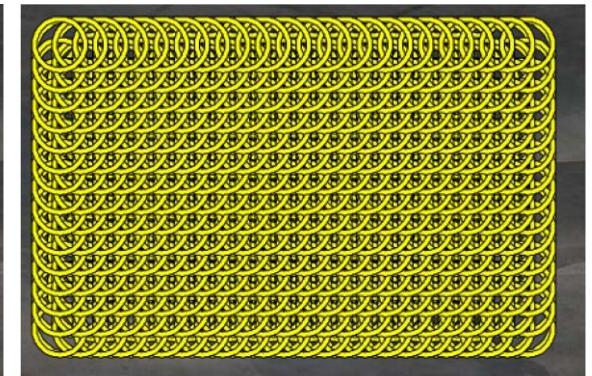
*Original Image*



*Global Representation*



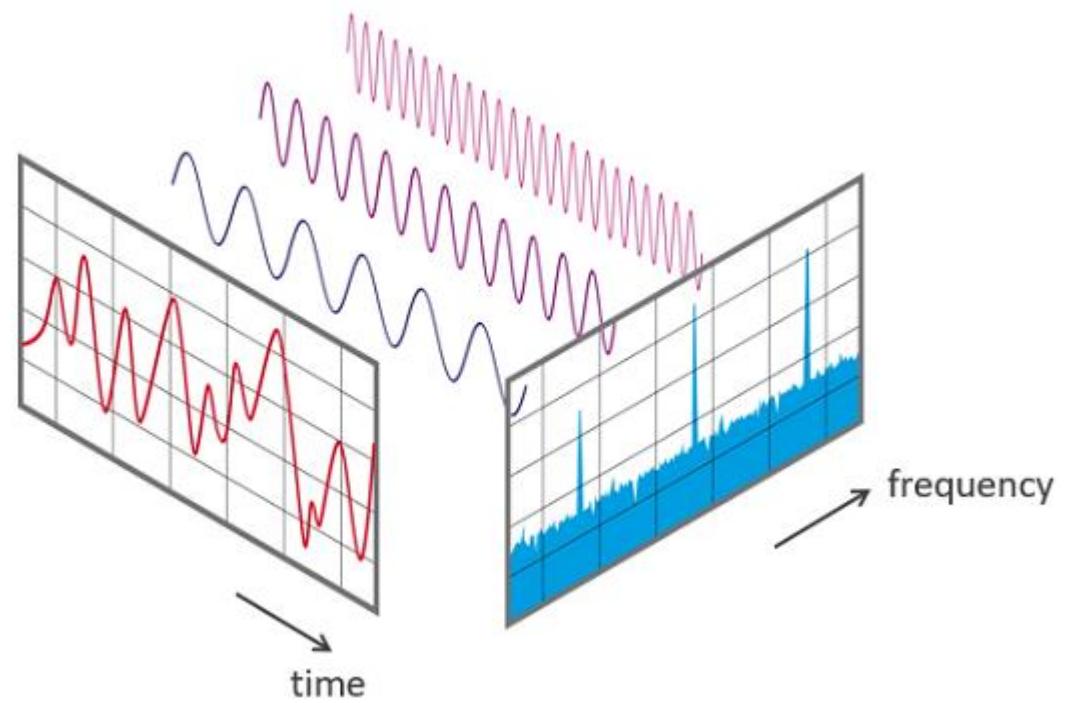
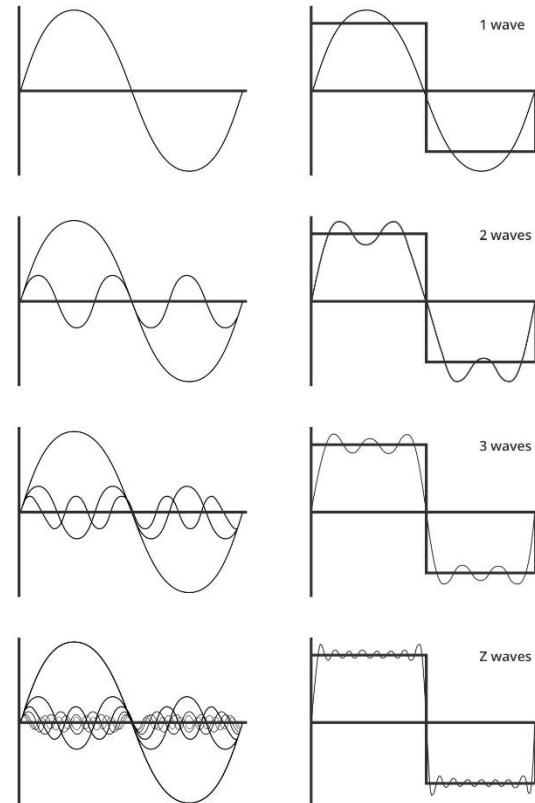
*Locally Sparse  
Representation*



*Locally Dense  
Representation*

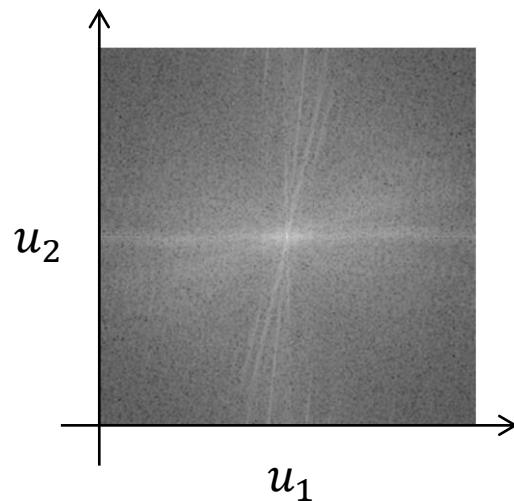
# Global Representations: Fourier Transform

- **Fourier Transform** is a tool that rewrite a (continuous and smooth) function as a (coefficient-weighted) sum of sine/cosine functions.



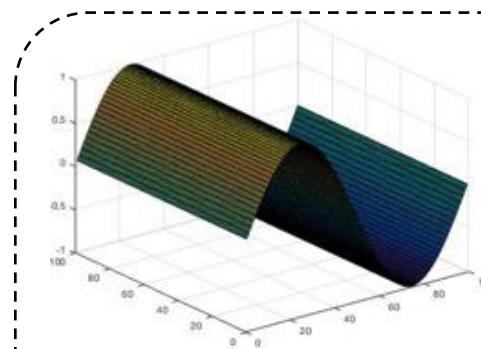
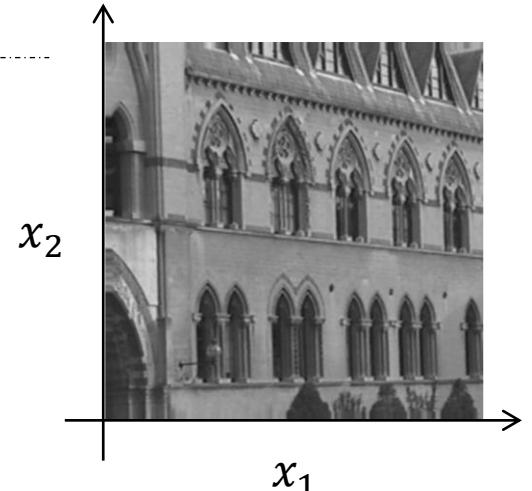
# Global Representations: Fourier Transform

- Image, as a 2D function, can also be rewritten as a sum of 2D sine/cosine functions:

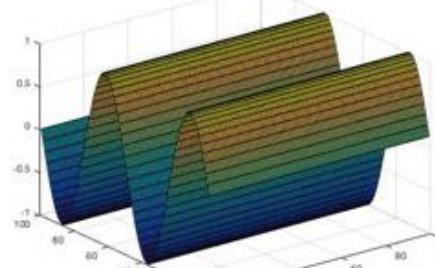


$$\mathcal{F}(f)(\mathbf{u}) \equiv F(\mathbf{u}) = \int_{\mathbb{R}^d} [e^{-2\pi i \mathbf{u} \cdot \mathbf{x}}] f(\mathbf{x}) d\mathbf{x}$$

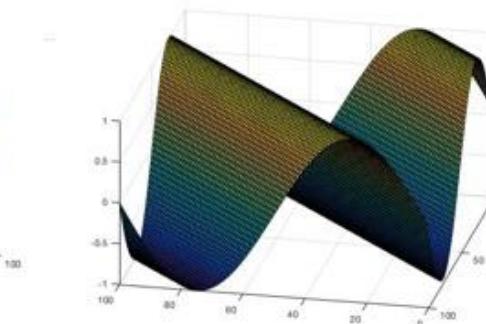
$$\mathcal{F}^{-1}(F)(\mathbf{x}) = \int_{\mathbb{R}^d} e^{2\pi i \mathbf{u} \cdot \mathbf{x}} F(\mathbf{u}) d\mathbf{u} = f(\mathbf{x})$$



$$u_1 = 1, u_2 = 0$$



$$u_1 = 0, u_2 = 2$$



$$u_1 = 1, u_2 = 1$$

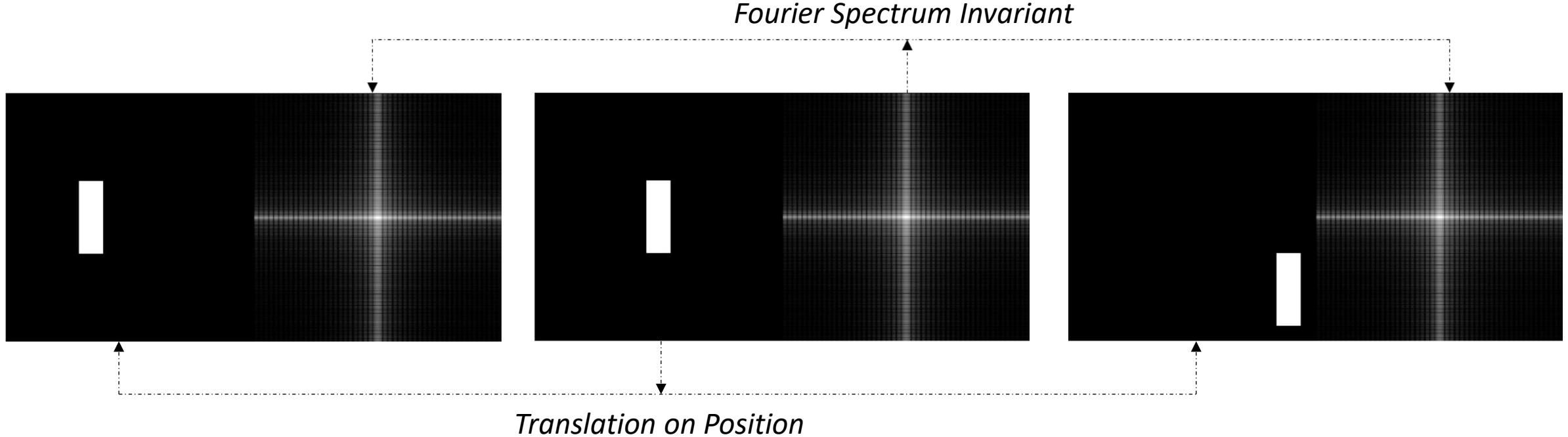
So, How About Invariance?

# Translation Invariance of Fourier Transform

- Translating the function leads to multiplying the Fourier transform by a phase factor:

$$\mathcal{F}(f(\mathbf{x} - \mathbf{t}))(\mathbf{u}) = [e^{-2\pi i \mathbf{u} \cdot \mathbf{t}}] \mathcal{F}(f(\mathbf{x}))(\mathbf{u})$$

- As a consequence, the absolute values of Fourier transform are invariant to translation.



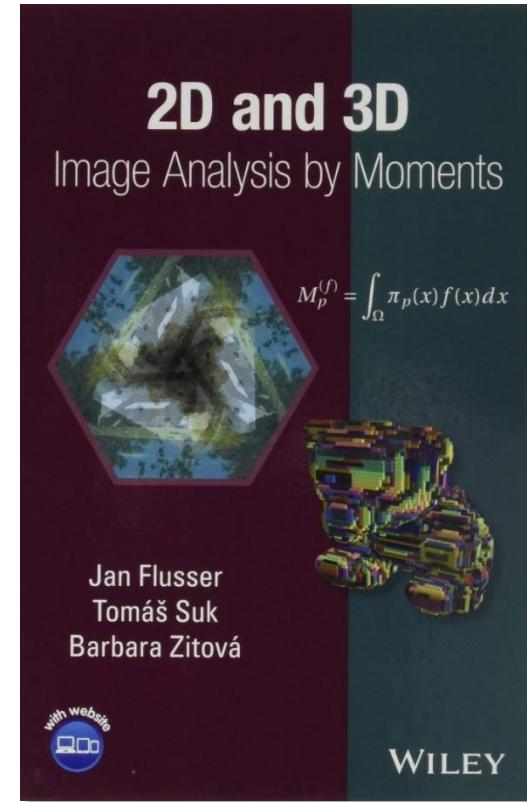
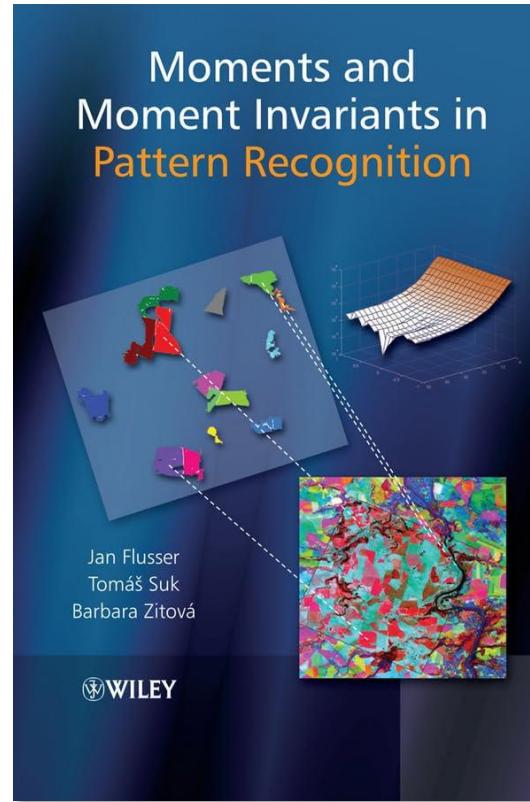
Can Global Invariance Be Generalized To Other  
Geometric Transformations?

# Global Representations: Moment Invariants

- **Moment Invariants** are similar to Fourier transforms in that they also rewrite the function as a (coefficient-weighted) sum of basis functions, but with a different purpose — more generalized invariants.



J. Flusser, B. Zitova, & T. Suk, 2009  
Moment Invariants



- J. Flusser, B. Zitova, T. Suk. *Moments and Moment Invariants in Pattern Recognition*. John Wiley & Sons, 2009.

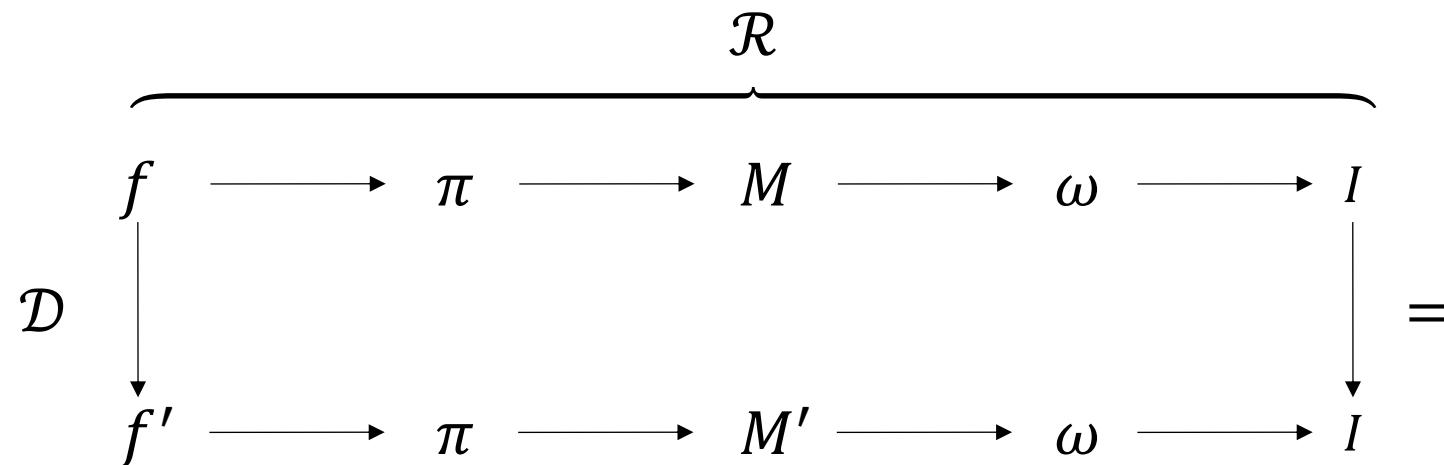
# Moments as a Generic Form of Global Representation

---

- **Moments** have a very simple definition, and is in fact a **generic form of the global representation**:

$$M_p^{(f)} = \int_{\Omega} \pi_p(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

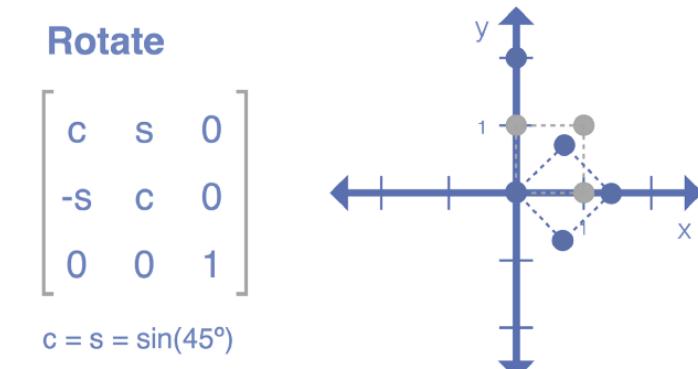
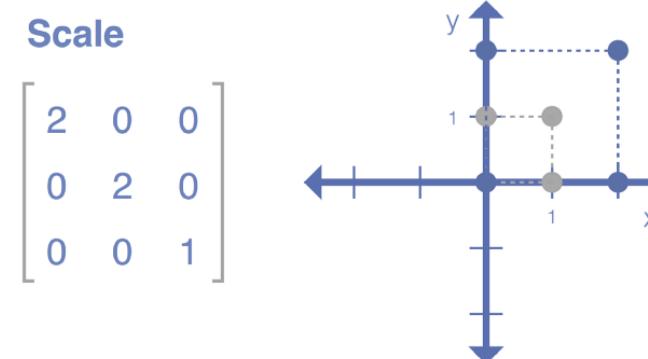
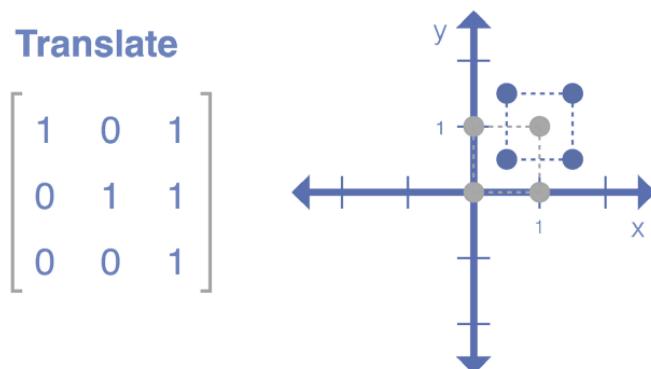
- Here, the core is how such basis functions  $\pi$  are designed so that more generalized invariants  $I$  can be derived from the corresponding moments  $M$  by a certain operation  $\omega$ .



# Geometric Transformations and Geometric Moments

- Let us consider the basic geometric transformations, including **translation, rotation and scaling**, which can be modeled as:

$$\mathbf{x}' = s\mathbf{R}_\alpha \mathbf{x} + \mathbf{t} \quad \mathbf{R}_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$$



- We can also define the so-called **geometric moments** with very simple basis functions:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

# Translation and Scaling Invariants

---

- With the above definitions, **translation invariants  $\mu$**  can be derived from the geometric moments:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x, y) dx dy \quad x_c = m_{10}/m_{00}, \quad y_c = m_{01}/m_{00}$$

- where  $(x_c, y_c)$  should be considered as the **centroid** of the image. The invariance is achieved by aligning the coordinate origin of the basis functions with the centroid.
- Let us further consider **scaling invariants  $v$** , which again can be derived from geometric moments, by normalizing the scaling factor on moments:

$$\mu'_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x/s, y/s) dx dy \quad v_{pq} = \frac{\mu_{pq}}{\mu_{00}^w} \quad w = \frac{p+q}{2} + 1$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s^p (x - x_c)^p s^q (y - y_c)^q f(x, y) s^2 dx dy = \boxed{s^{p+q+2} \mu_{pq}} \quad v'_{pq} = \frac{\mu'_{pq}}{(\mu'_{00})^w} = \frac{s^{p+q+2} \mu_{pq}}{(s^2 \mu_{00})^w} = v_{pq}$$

# Rotation Invariants by Hu and Hilbert

- Are **rotation invariants  $\phi$**  equally derivable from geometric moments? Yes, **Hu** gives 7 invariants based on **Hilbert's algebraic invariants**, which seems very complex. But it makes sense, due to the nonlinear action of the rotations on  $x$  and  $y$ .

$$\phi_1 = m_{20} + m_{02},$$

$$\phi_2 = (m_{20} - m_{02})^2 + 4m_{11}^2,$$

$$\phi_3 = (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2,$$

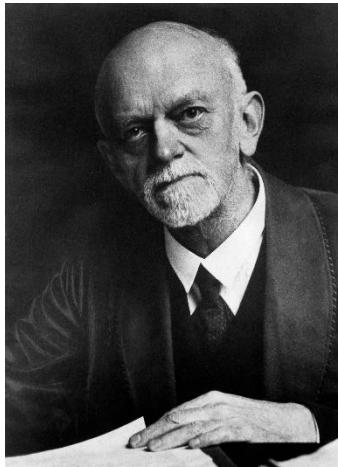
$$\phi_4 = (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2,$$

$$\begin{aligned}\phi_5 = & (m_{30} - 3m_{12})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) \\ & + (3m_{21} - m_{03})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2),\end{aligned}$$

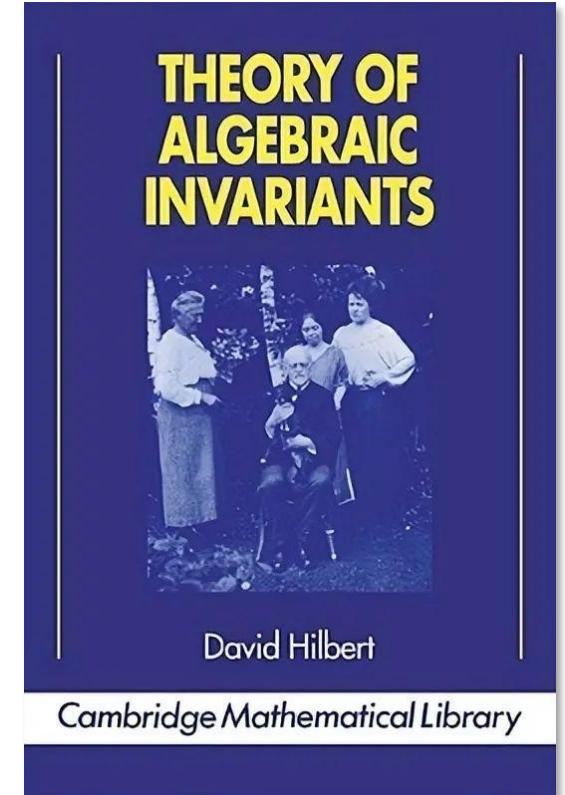
$$\begin{aligned}\phi_6 = & (m_{20} - m_{02})((m_{30} + m_{12})^2 - (m_{21} + m_{03})^2) \\ & + 4m_{11}(m_{30} + m_{12})(m_{21} + m_{03}),\end{aligned}$$

$$\begin{aligned}\phi_7 = & (3m_{21} - m_{03})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) \\ & - (m_{30} - 3m_{12})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2).\end{aligned}$$

- MK Hu. Visual pattern recognition by moment invariants. *TIT*, 1962.



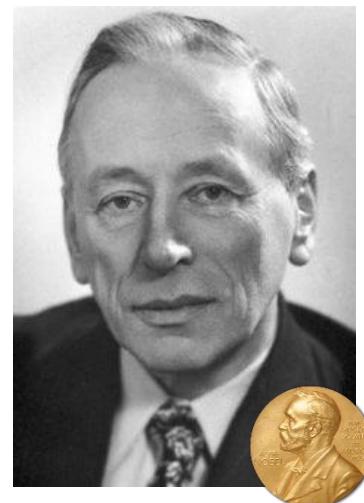
D. Hilbert, 1897  
Algebraic Invariants



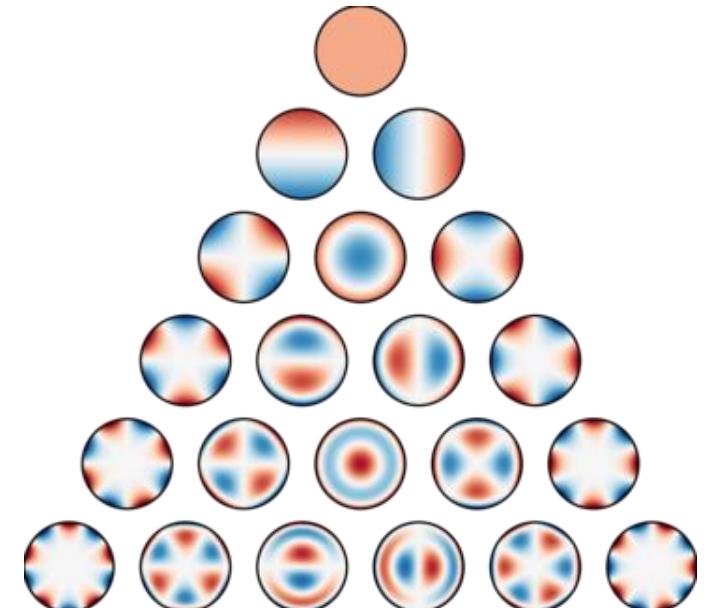
# Rotation Invariants by Zernike

- Can rotation invariants be simpler? Let us define the basis functions in **polar** coordinates, where rotations are more easily managed, by the Fourier translation invariance in the angular form.
- In this respect, **Zernike polynomials** are typical — they are complete orthogonal bases on the unit circle and easily realize rotation invariance, from **Zernike's optical research**.

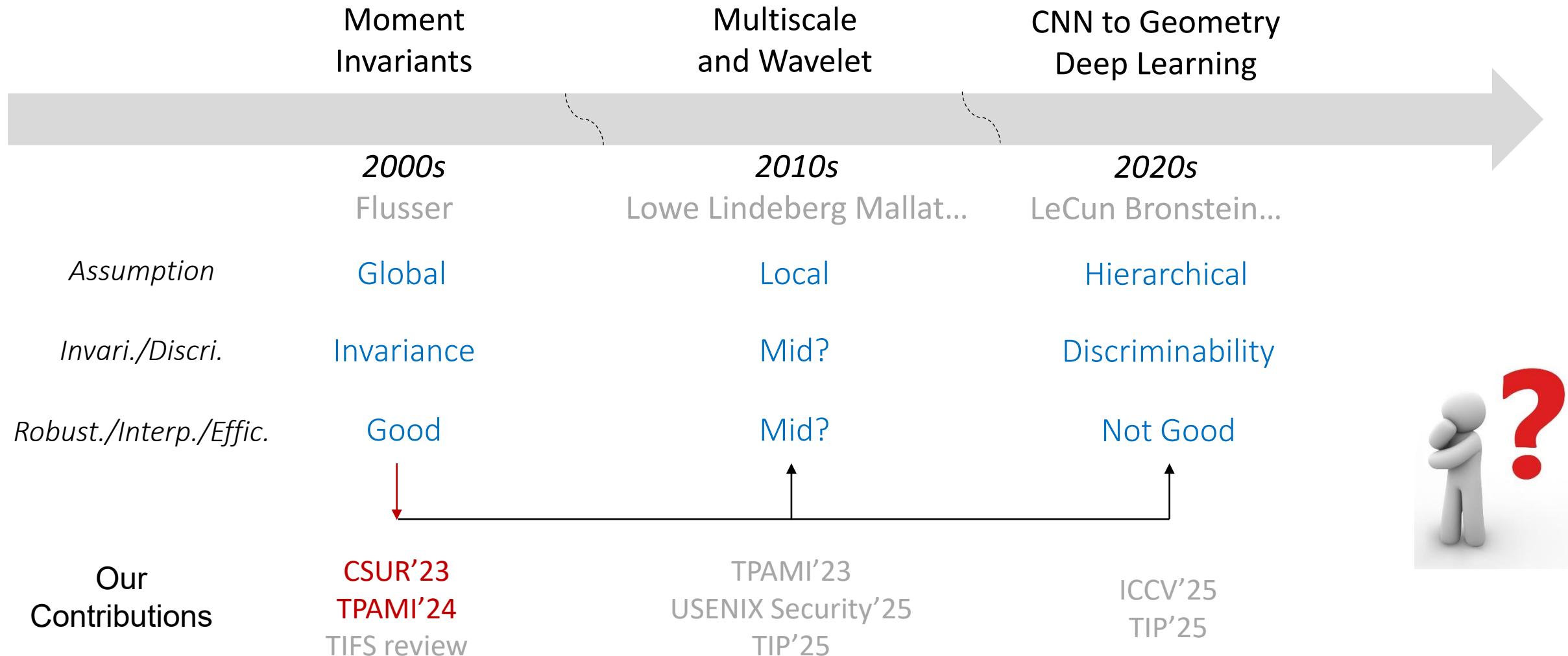
$$\begin{aligned} C_{pq} &= \int_0^\infty \int_0^{2\pi} R_{pq}(r) e^{i\xi(p,q)\theta} f(r, \theta) r d\theta dr \\ C'_{pq} &= \int_0^\infty \int_0^{2\pi} R_{pq}(r) e^{i\xi(p,q)\theta} f(r, \theta + \alpha) r d\theta dr \\ &= \int_0^\infty \int_0^{2\pi} R_{pq}(r) e^{i\xi(p,q)(\theta-\alpha)} f(r, \theta) r d\theta dr \\ &= \underbrace{\left[ e^{-i\xi(p,q)\alpha} \right]}_{\text{---}} C_{pq}. \end{aligned}$$



F. Zernike, 1934  
Zernike Polynomials

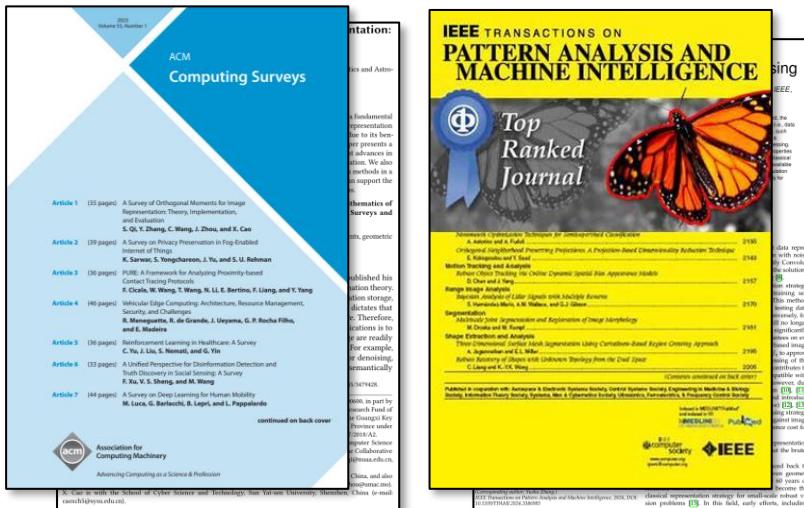


# Our Contributions

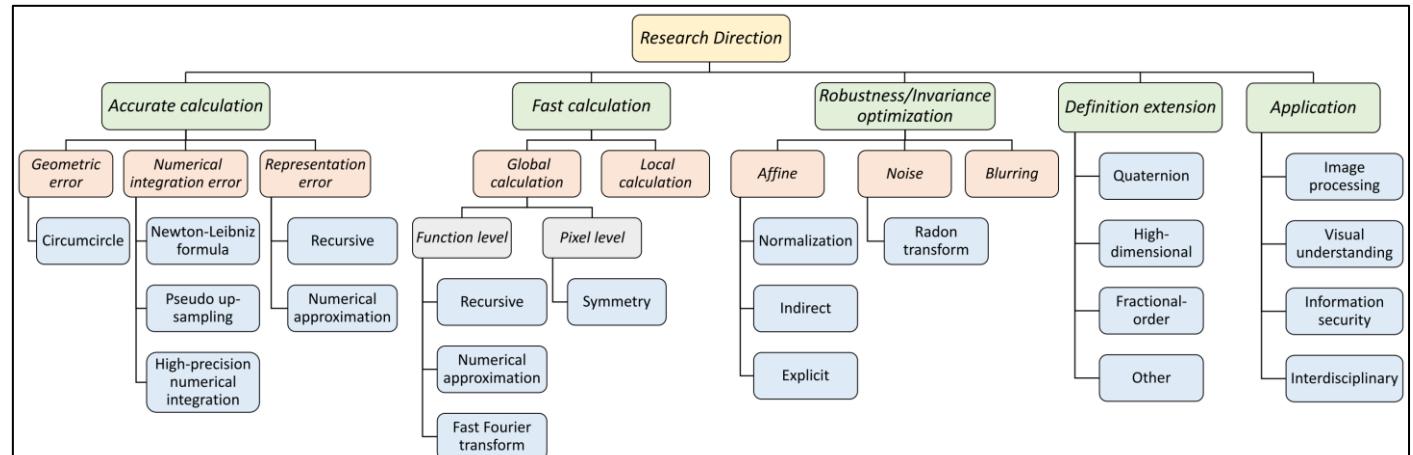


# Refining Global Invariants

- We give papers on the practical aspects of moments for refining global invariants, covering **numerical analyses**, **software implementations**, **benchmark evaluations**, and **recent advances**.



- S. Qi, Y. Zhang, C. Wang, et al. A Survey of Orthogonal Moments for Image Representation: Theory, Implementation, and Evaluation. *ACM Computing Surveys (CSUR)*, 2023, 55(1): 1-35.
- S. Qi, Y. Zhang, C. Wang, et al. Representing Noisy Image Without Denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2024, 46(10): 6713 - 6730

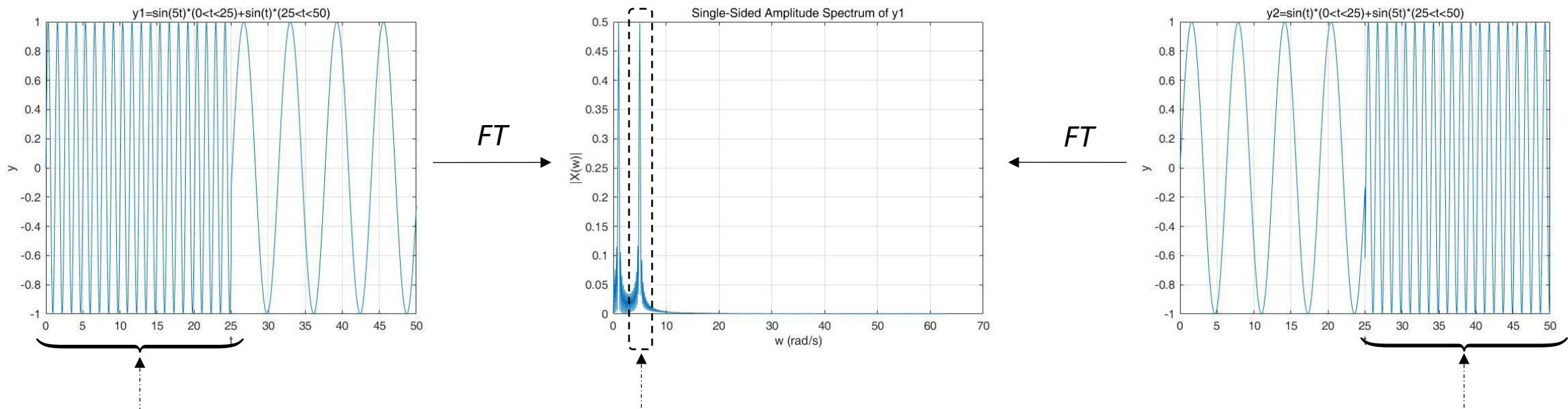


A screenshot of a GitHub repository page for 'MomentToolbox' (Public). The repository has 1 branch and 0 tags. The README.md file was updated 4 years ago. The repository URL is <https://github.com/ShurenQi/MomentToolbox>. The repository description is: 'Matlab code for the paper "A survey of orthogonal moments for image moment theory, implementation, and evaluation"'. The repository has 2 forks and 28 stars. It includes tags for image-processing, pattern-recognition, image-analysis, and orthogonal-moments.

From Global To Local

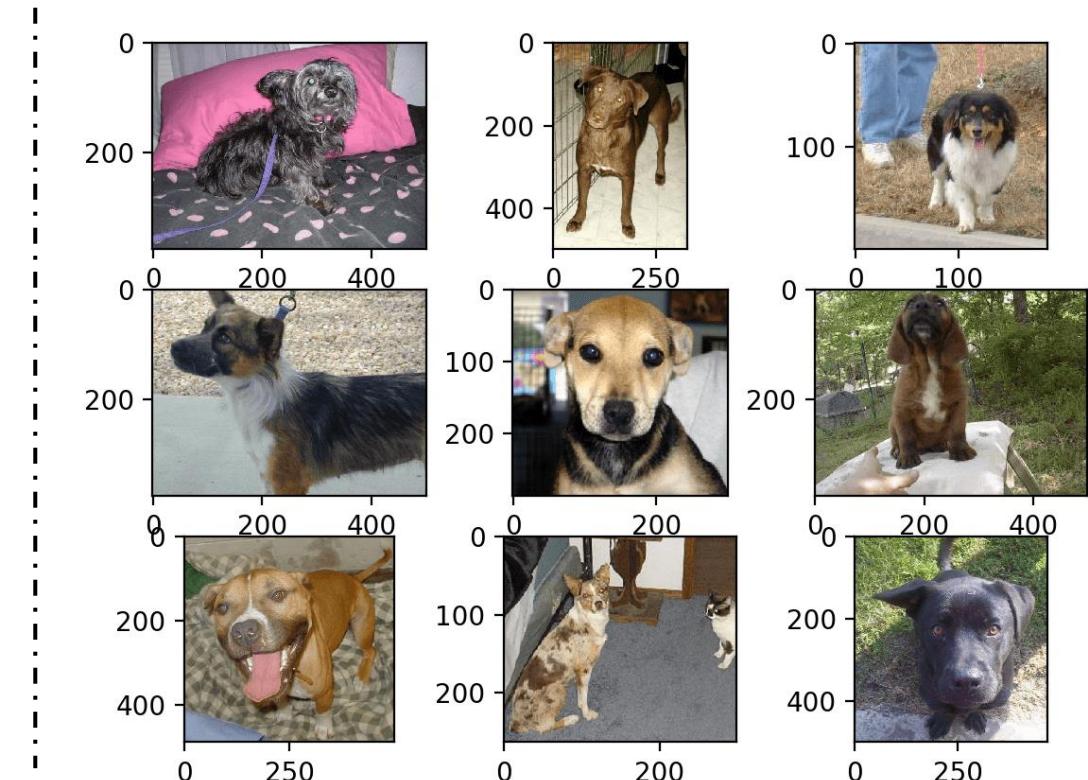
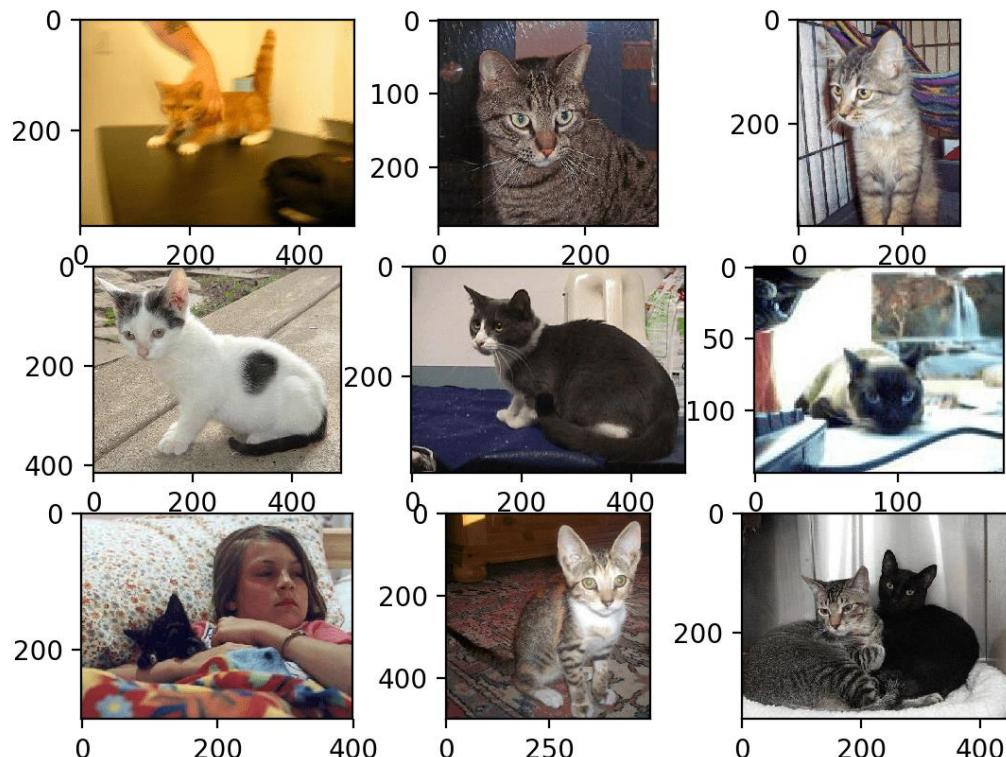
# Why We Need Local Representations

- Fourier transform-like global representations are typically **(under)-complete** and are just designed for **low-level processing**, struggling to express high-level semantics with over-completeness.
- As a toy example, the Fourier transform cannot even distinguish the order in which the two signals appear.



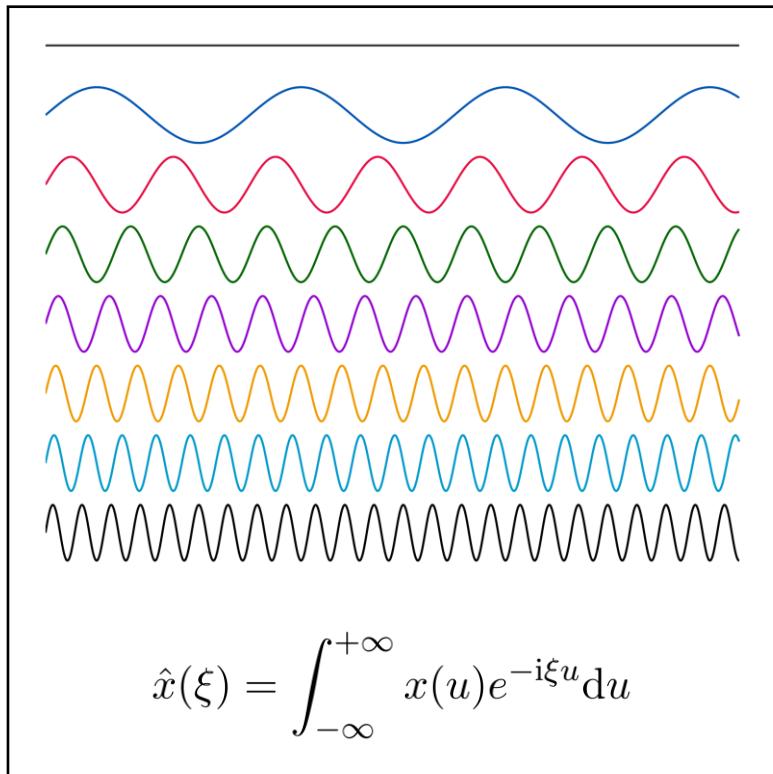
# Why We Need Local Representations

- Under realistic considerations, there are too many tasks concerned with local semantic properties — **recognition and classification** (distinguish images of cats and dogs), where global representations are likely unable to provide enough information to support discriminability.

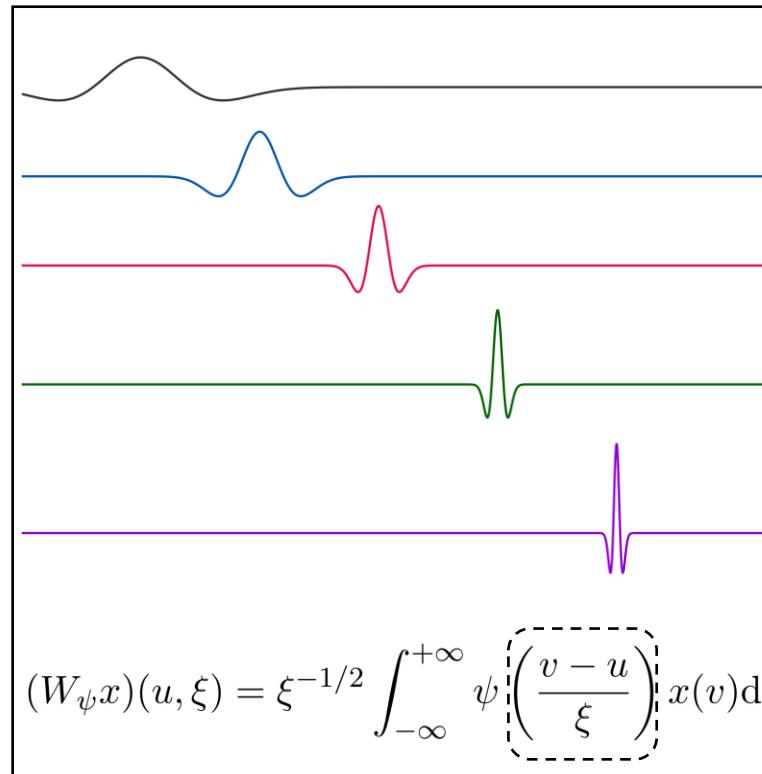


# Local Representations: Wavelet Transform

- Different from Fourier, basis functions of **Wavelet Transform** are **local** and **multi-scale**.



*Fourier*

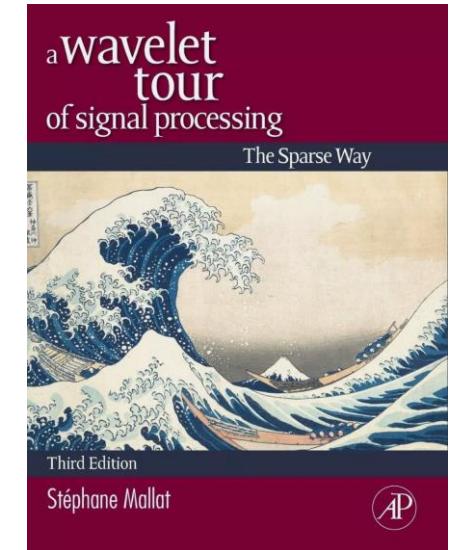


*Wavelets*



S. Mallat, 1999

Wavelets



# Local Representations: Wavelet Transform

---

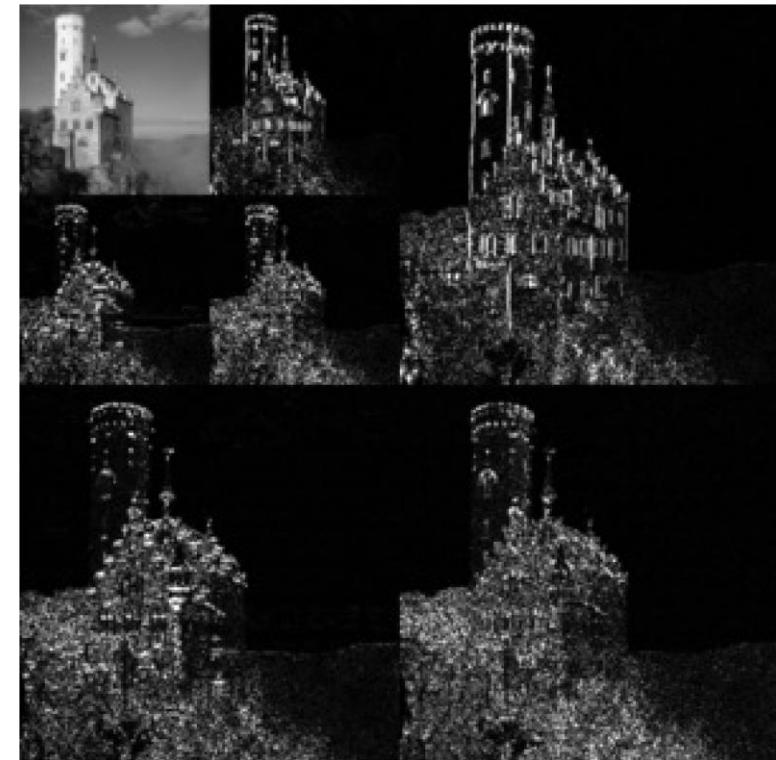
- Wavelet transform can capture local information, with **better discriminative properties** — time-frequency discriminability and over-completeness.



*Original Image*



*Fourier Representations*

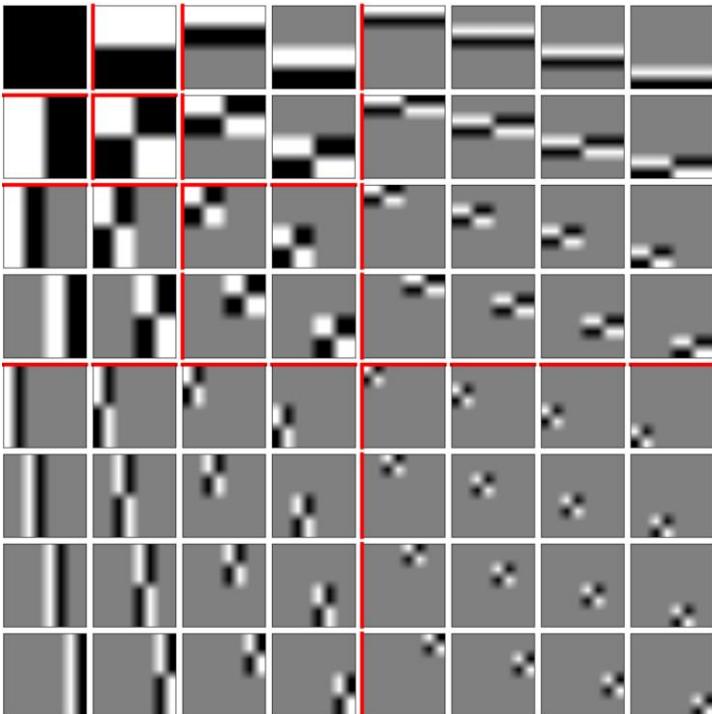


*Wavelet Representations*

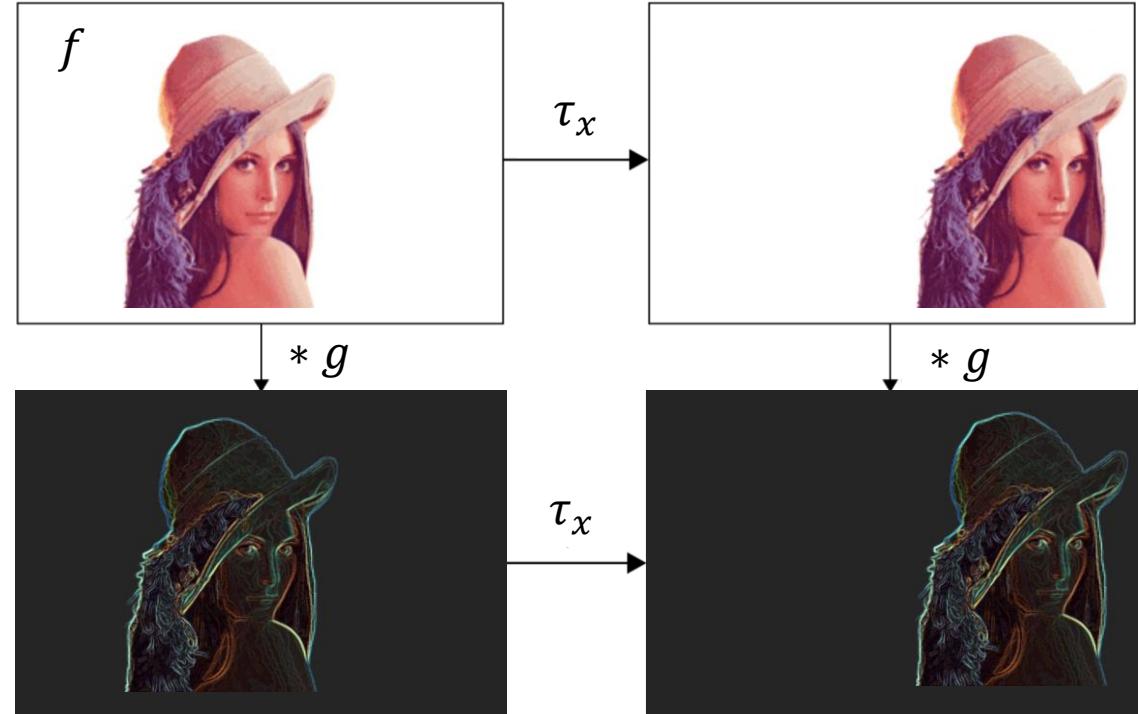
So, How About Invariance?

# Translation Equivariance of Wavelet Transform

- The wavelet basis functions define **convolution operators  $g$**  — the wavelet transform of an image  $f$  means the convolution of  $f$  and  $g$ . Therefore, the wavelet transform has a translation equivariance with the convolution.



Wavelet Convolutional Operators  $g$



$$(\tau_x f) * g = \tau_x(f * g)$$

Can Local Invariance Be Generalized To Other  
Geometric Transformations?

# Local Representations: SIFT

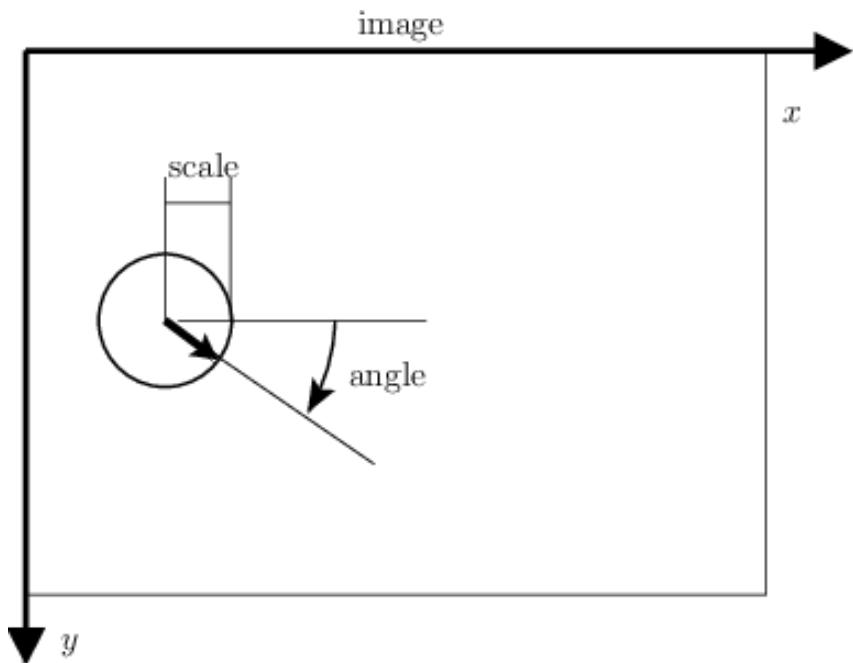
- The local and multiscale concepts of the wavelet transform were **followed** by later local representations.
- For example, the well-known **Scale-Invariant Feature Transform (SIFT)** aims at the **local invariance of rotation and scaling in multiscale spaces**.



- DG Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.

# Local Representations: SIFT

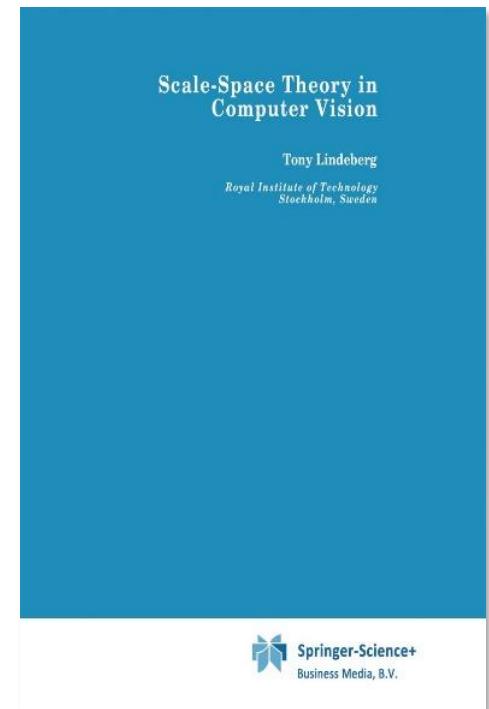
- SIFT describes **local regions that have their own scale and orientation**, with the **scale space theory** as a foundation.
- Here, once the scale and orientation of the regions can be evaluated stably, then invariant features can be constructed by **normalizing** the scale and orientation.



T. Lindeberg, 1993  
Scale Space Theory



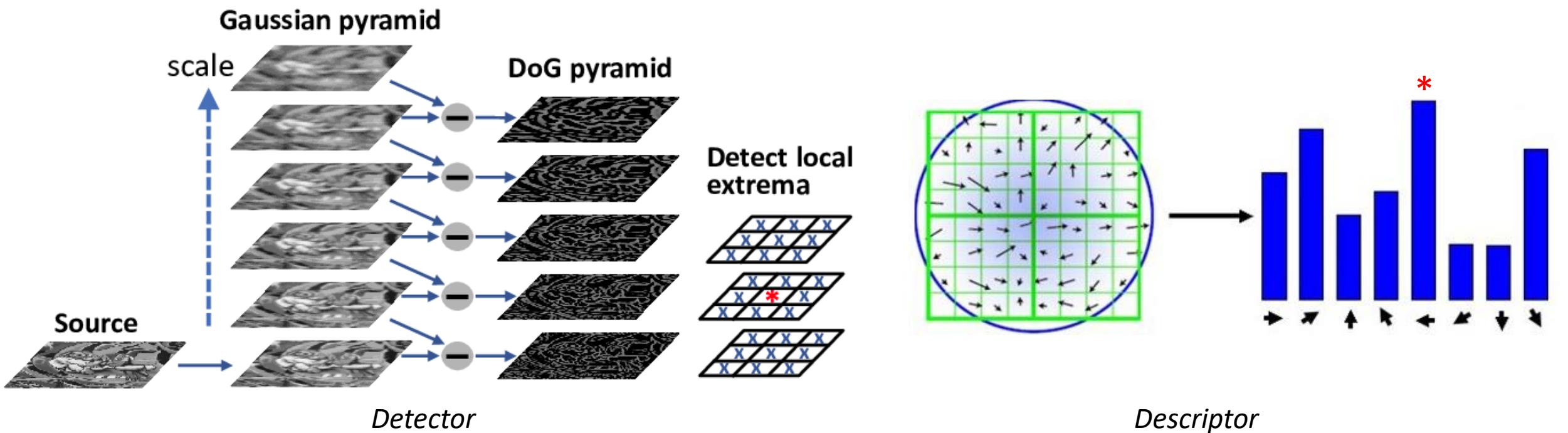
D. Lowe, 1999  
SIFT



- T Lindeberg. *Scale-space Theory in Computer Vision*. Springer Science & Business Media, 1993.

# Local Representations: SIFT

- SIFT has two main components: **detector** and **descriptor**.
- The detector is responsible for finding the interest point with evaluated scale to achieve **scaling invariance**. The descriptor is responsible for describing the interest point with evaluated orientation to further achieve **rotation invariance**.



Scale is evaluated by finding the extreme in the scale space Orientation is evaluated by computing the histogram of gradients

# From Sparse To Dense

# Why We Need Dense Representations

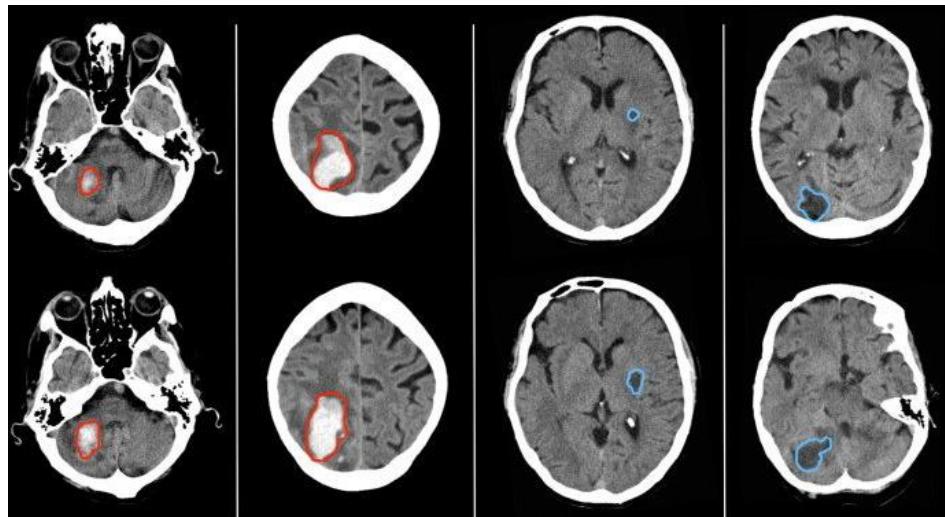
- SIFT-like interest points are **sparse** in the image and are designed to **focus only on the main subject (ignoring all other regions)**.



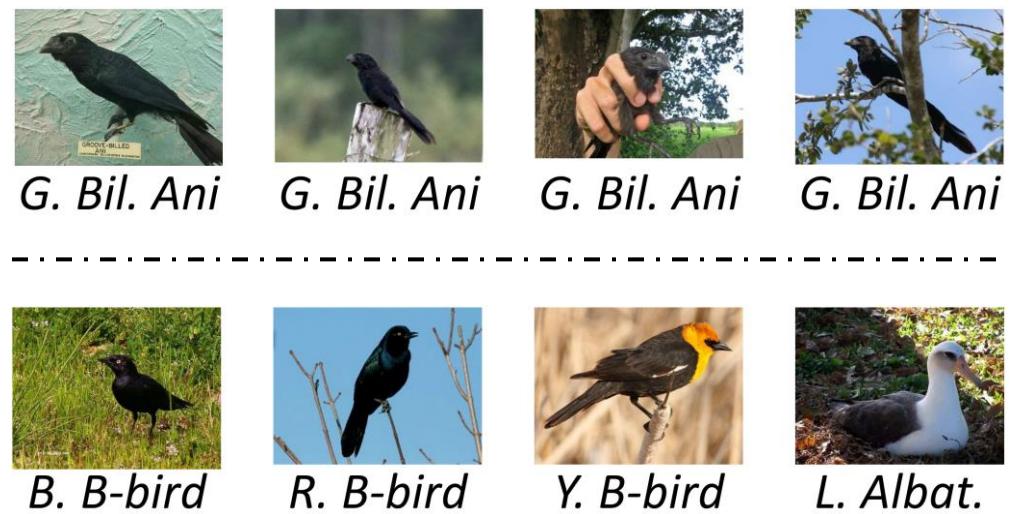
- A Iscen, G Tolias, PH Gosselin, et al. A comparison of dense region detectors for image search and fine-grained classification. *TIP*, 2015.

# Why We Need Dense Representations

- Under realistic considerations, there are too many tasks concerned with dense semantic properties — **detection/localization** (detect lesions in CT images), **fine-grained classification** (distinguish large-scale bird images), where sparse points are likely to miss important local information.



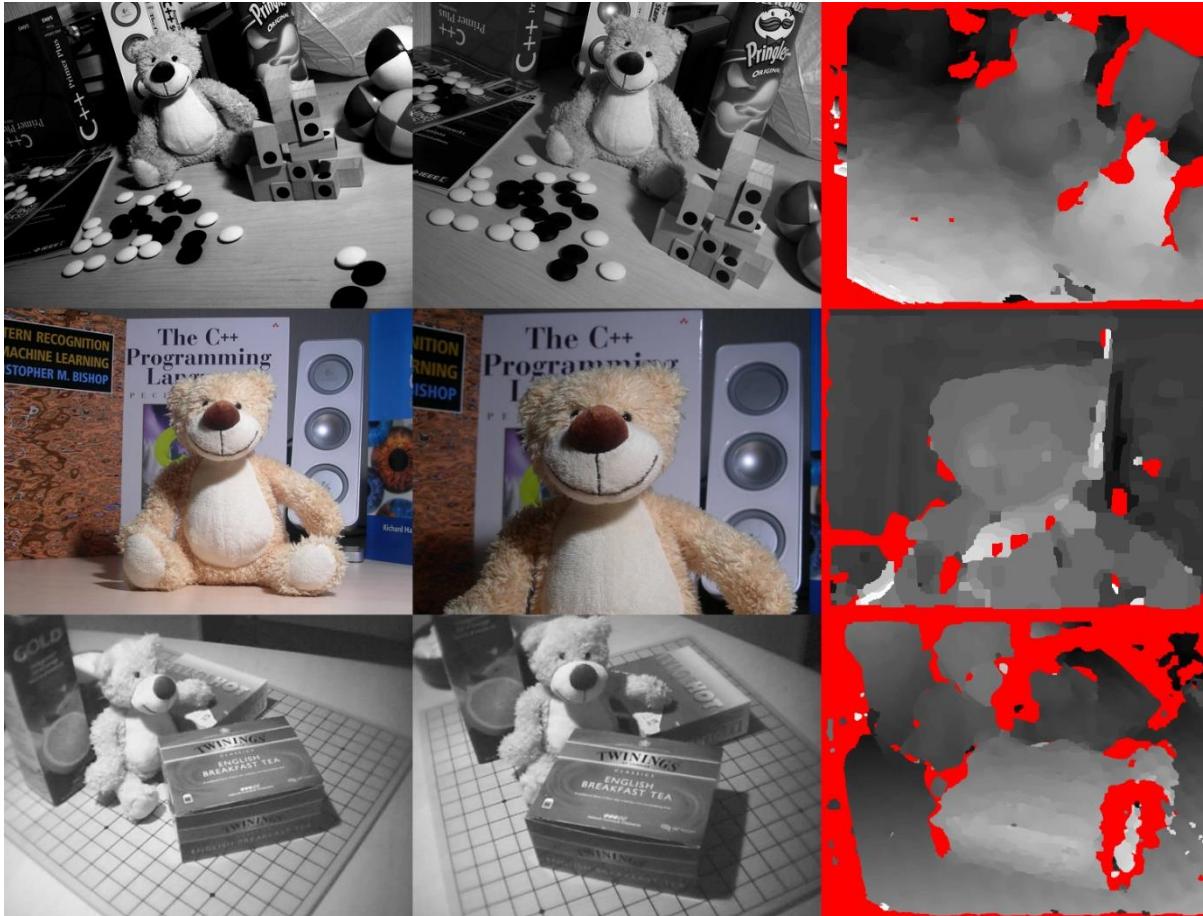
*Detection/Localization*



*Fine-grained Classification*

# Local Representations: DAISY

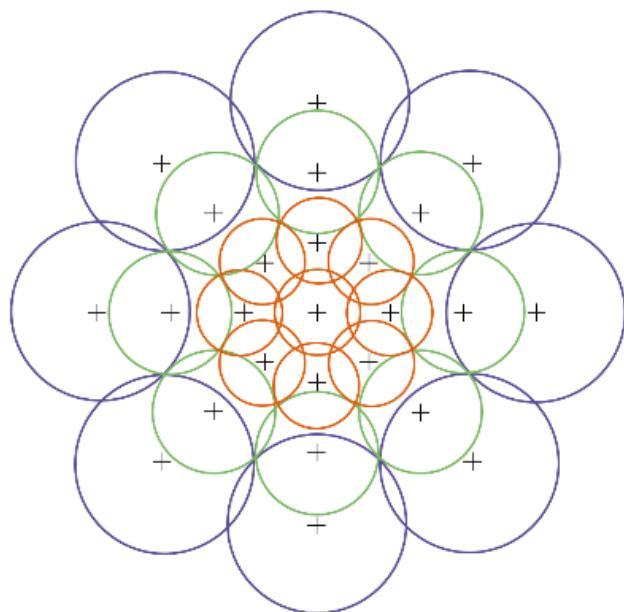
- DAISY aims to extend SIFT from sparse to dense, achieving local invariance of rotation and scaling for each pixel position.



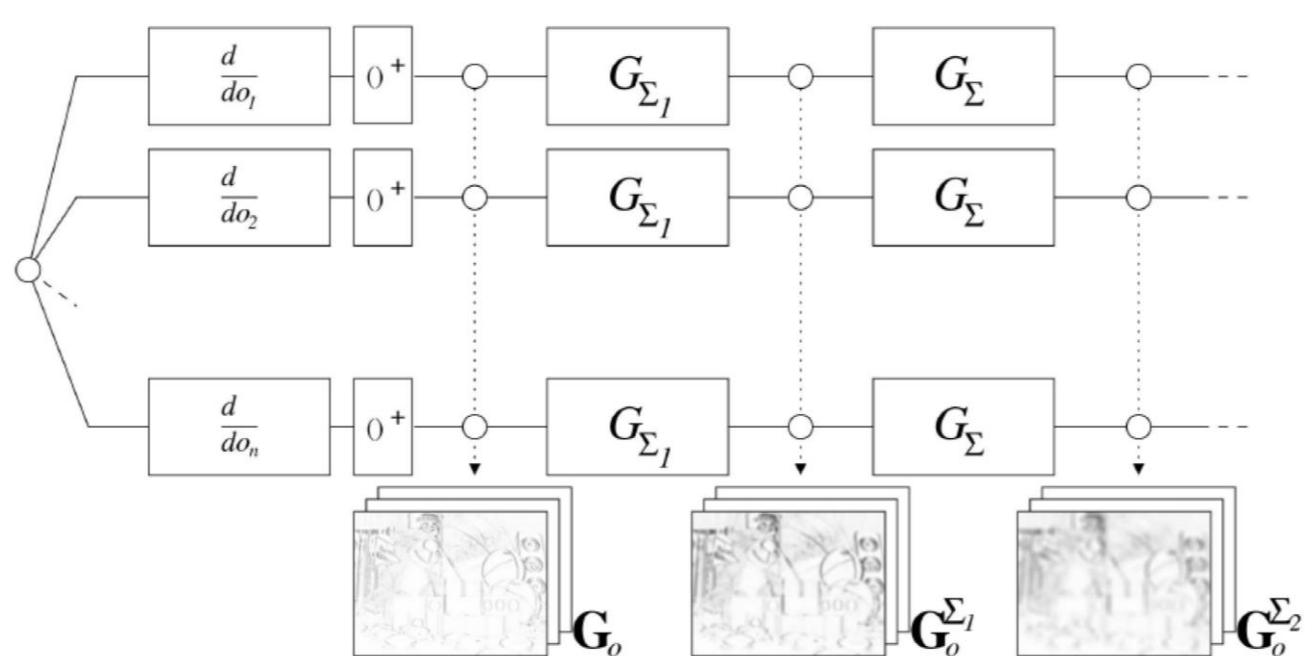
- E Tola, V Lepetit, P Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *TPAMI*, 2009.

# Local Representations: DAISY

- The main difficulty is that the complex operations of SIFT in scale and orientation evaluation **cannot be performed directly for dense positions**, due to high complexity.
- Therefore, DAISY introduces a series of simplified designs for scale and orientation, but at the same time invariance is reduced.

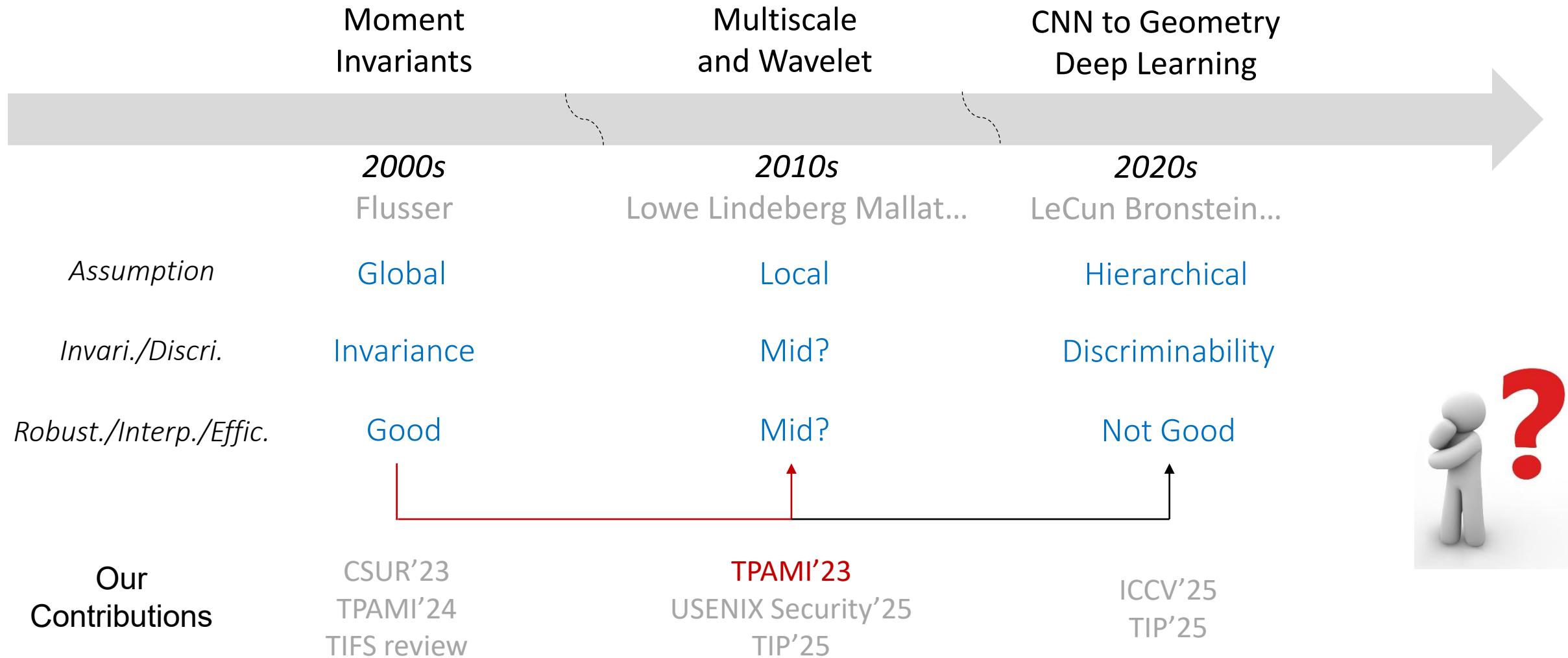


DAISY Descriptor



Simplified Designs for Scale and Orientation

# Our Contributions

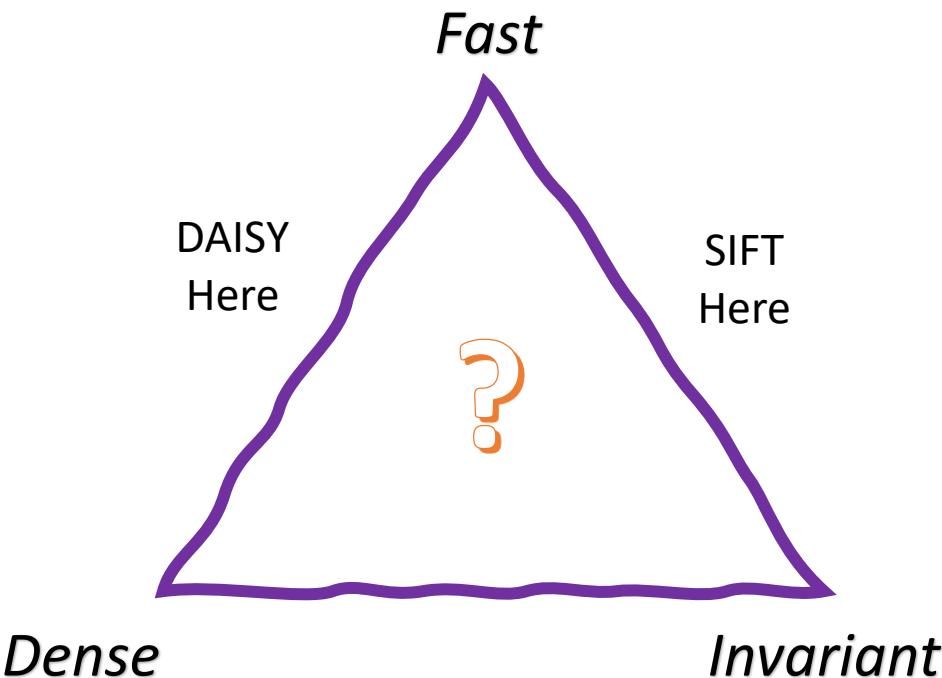


# Designing Local Invariants

- Reviewing the above local invariants, one can note **a gap**: SIFT is fast and invariant, but not suitable for dense tasks; DAISY is fast and dense, but largely compresses invariance.
- We tried to define **truly dense invariants while being fast enough**. We achieved this goal by exploring the potential of **classical moment invariants**.

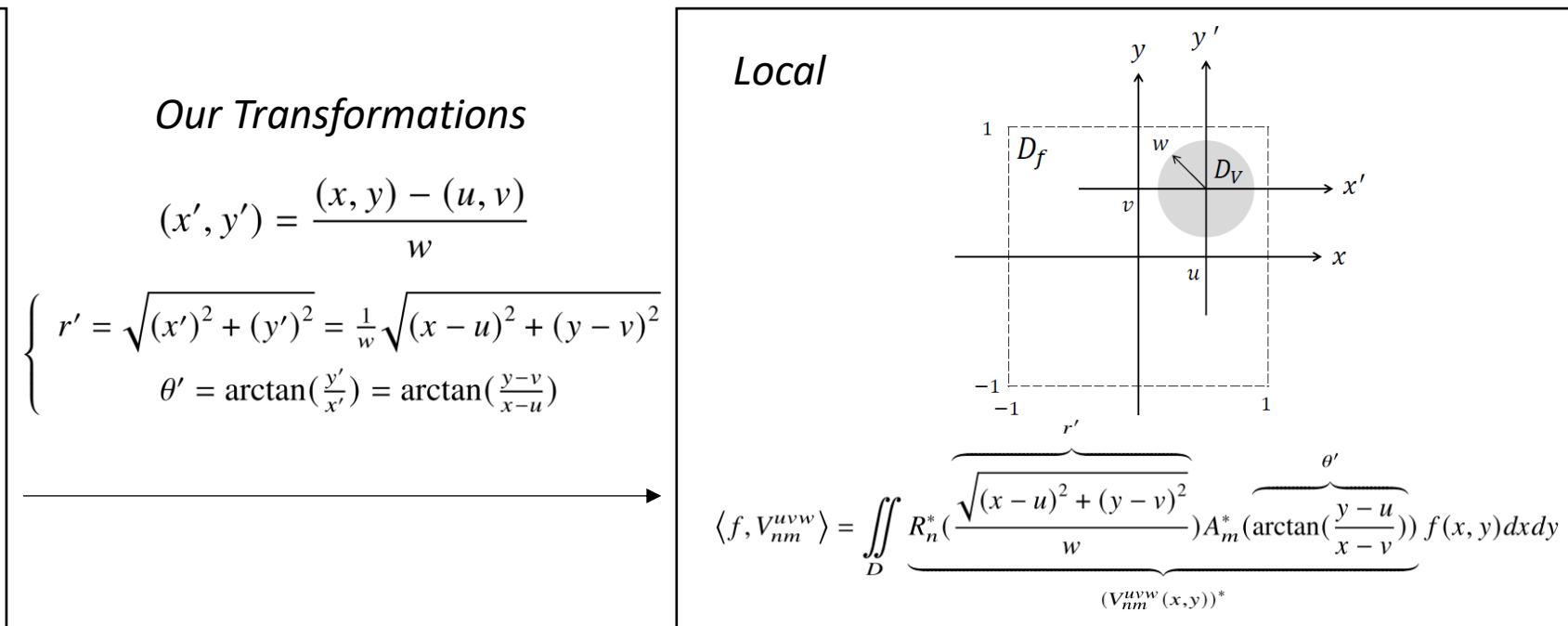
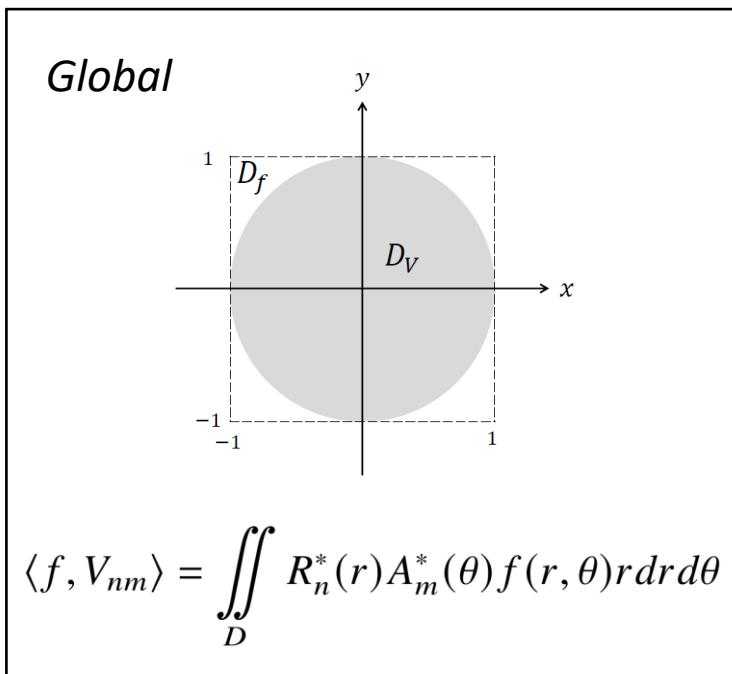


- S. Qi, Y. Zhang, C. Wang, et al. A Principled Design of Image Representation: Towards Forensic Tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023, 45(5): 5337 - 5354



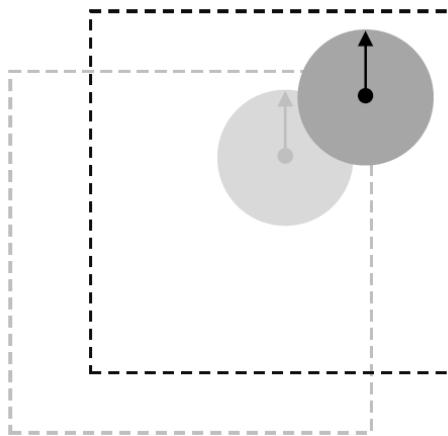
# Moments: From Global to Local

- First, we extend the definition of classical moments from the global to the local with scale space. Here, local coordinate system  $(x', y')$  is a translated and scaled version of the global coordinate system  $(x, y)$ , with translation offset  $(u, v)$  and scale factor  $w$ .
- Two interesting properties: generic nature and local representation capability.



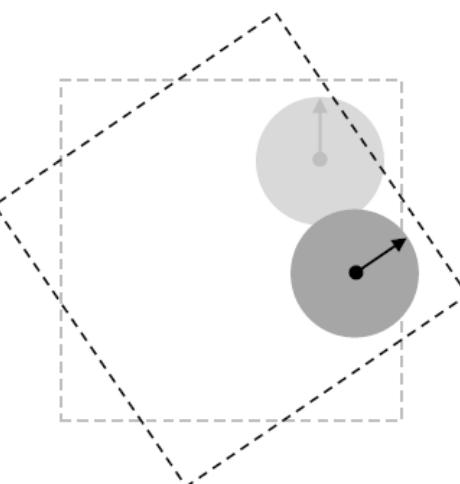
# Moment Invariants: From Global to Local

- Then, we found the symmetry properties of the local definition for several geometric transformations.
- Therefore, rotation and flipping invariants can be obtained by taking the absolute values; translation and scaling invariants can be obtained by pooling over the  $(u, v)/w$ .



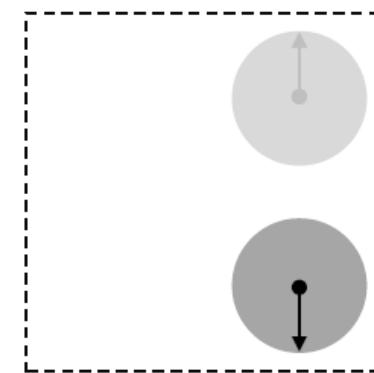
$$\begin{aligned} &\langle f(x + \Delta x, y + \Delta y), V_{nm}^{uvw}(x, y) \rangle \\ &= \langle f(x, y), V_{nm}^{(u+\Delta x)(v+\Delta y)w}(x, y) \rangle \end{aligned}$$

*Translation Equivariance*  
w.r.t.  $(u, v)$



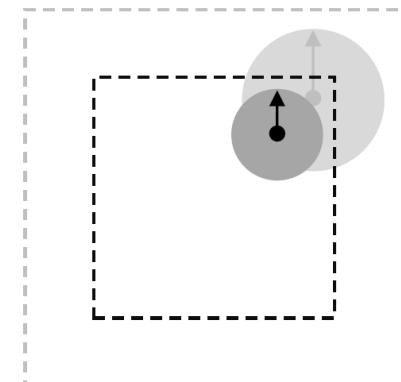
$$\begin{aligned} &\langle f(r, \theta + \phi), V_{nm}^{uvw}(r', \theta') \rangle \\ &= \langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle A_m^*(-\phi) \end{aligned}$$

*Rotation Invariance*  
w.r.t. absolute values



$$\begin{aligned} &\langle f(r, -\theta), V_{nm}^{uvw}(r', \theta') \rangle \\ &= (\langle f(r, \theta), V_{nm}^{uvw}(r', \theta') \rangle)^* \end{aligned}$$

*Flipping Invariance*  
w.r.t. absolute values

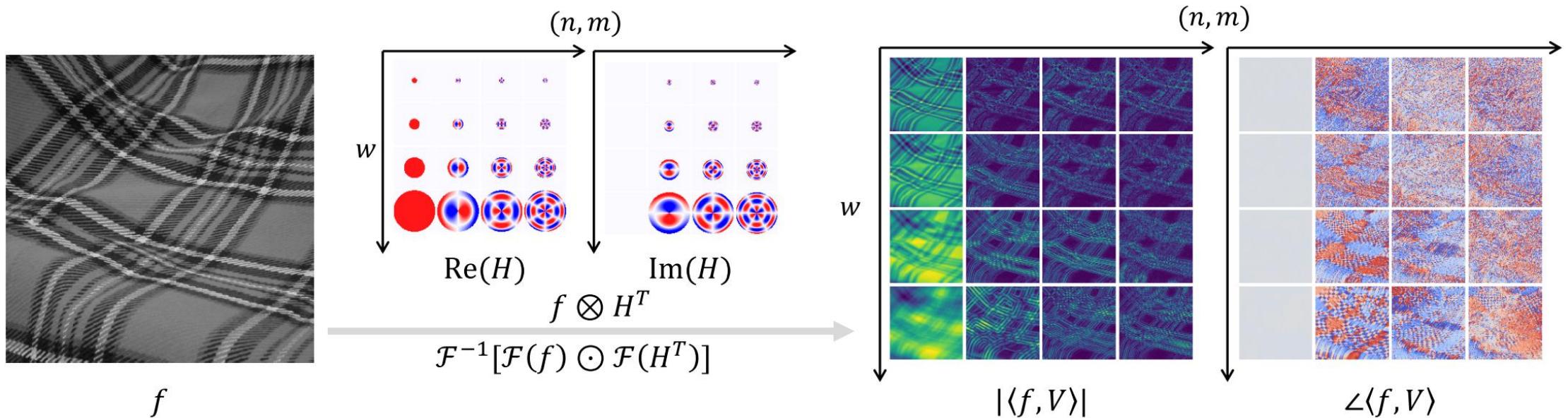


$$\begin{aligned} &\langle f(sx, sy), V_{nm}^{uvw}(x, y) \rangle \\ &= \langle f(x, y), V_{nm}^{uv(ws)}(x, y) \rangle \end{aligned}$$

*Scaling Covariance*  
w.r.t.  $w$

# Fast Implementation

- Finally, we give a fast implementation by the **convolution theorem**.



$$\mathcal{O}(w_{\max}^2 \#_{uv} \#_w) \quad \text{VS} \quad \mathcal{O}(\#_w \#_{uv} \log \#_{uv})$$

$w_{\max}^2$     **VS**     $\log(\#_{uv})$

