What matters more for image matching and the comparison of descriptors: invariance and causality requirements or repeatability criteria? (a case study with SIFT, SURF, SIFER)

Ives Rey-Otero   Mauricio Delbracio   Jean-Michel Morel

Centre de Mathématiques et Leurs Applications (CMLA), Ecole Normale Supérieure de Cachan

Saarbrücken, September 2013

References

Paper on SIFT:

[1] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

Paper on SURF:

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

Paper on SIFER:

[3] P. Mainali, G. Lafruit, Q. Yang, B. Geelen, L. Gool, and R. Lauwereins, "Sifer: Scale-invariant feature detector with error resilience," *International Journal of Computer Vision*, pp. 1–26, 2013.

Discussion on scale invariance:

[4] J.-M. Morel and G. Yu, "Is sift scale invariant?," *Inverse Problems and Imaging*, vol. 5, no. 1, pp. 115–136, 2011.

Review of detectors:

[5] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

Comparison methodology:

[6] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

Modified keypoint detector comparison methodology (overlap)

[7] S. Ehsan, N. Kanwal, A. F. Clark, and K. D. McDonald-Maier, "Measuring the coverage of interest point detectors," in *Image Analysis and Recognition*, pp. 253–261, 2011.

Modified keypoint comparison methodology (precision)

[8] K. Cordes, B. Rosenhahn, and J. Ostermann, "Increasing the accuracy of feature evaluation benchmarks using differential evolution," in *Differential Evolution (SDE), 2011 IEEE Symposium on*, pp. 1–8, 2011.

Summary

- SIFT, SURF, SIFER, their invariances properties
- The repeatability criteria
- Possible bias in the performance measure.
- A suggested correction

SIFT, SURF, SIFER
share a general “scale space” framework:

Detection	Extract the 3D extrema from $v(\sigma, x, y)$, a multi-scale detector of the image $u(x, y)$.
Description	Extract an image patch around each keypoint (σ, x, y) to compute the feature vector.

SIFT / SURF / SIFER: extrema of the multiscale detector yield the <u>key points</u> or <u>points of interest</u> *(x,y)* with an associated scale σ

	Multi-scale detectors $v(\sigma, x, y) = K_\sigma * u(x, y)$
SIFT	$K_\sigma(x, y) = G_{k\sigma}(x, y) - G_\sigma(x, y)$
SURF	box filters used to approximate $\sigma^4 det(\mathcal{H}(G_\sigma))$
SIFER	$K_\sigma(x, y) = 2\pi\sigma^2 G_\sigma(x, y)\left(\cos\left(\frac{cx}{\sigma}\right) + \cos\left(\frac{cy}{\sigma}\right)\right)$

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

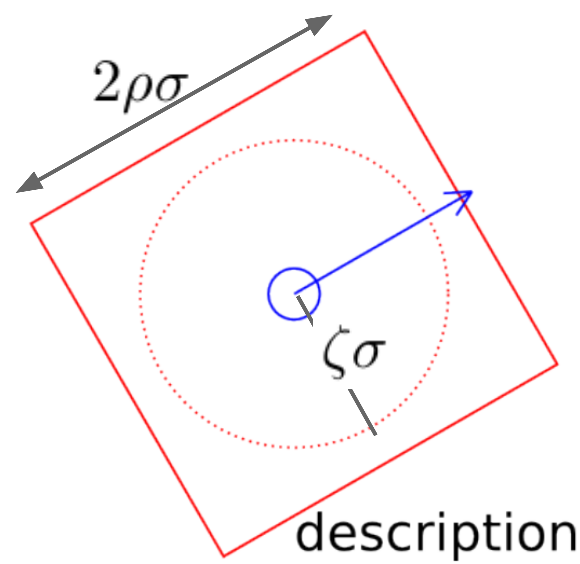
SIFT / SURF / SIFER

For each detection in scale-space (σ, x, y)

- Assign a principal orientation θ.
  (or several orientations in the case of SIFT)

- Extract a truncated Gaussian window centered on (x, y) and aligned with the orientation θ.
  Its standard deviation is $(\zeta\sigma)$.
  Its width is $(2\rho\sigma)$

	ρ	ζ
SIFT	6	6
SURF	10	3.3
SIFER	6	6



The classic repeatability criteria: a transform is simulated on a benchmark image, and the detector is applied before and after transform. Then keypoints are compared:

Depending on the adopted criteria, two detections $(\sigma_a, \mathbf{x}_a)$ and $(\sigma_b, \mathbf{x}_b)$ are one repeated detection if

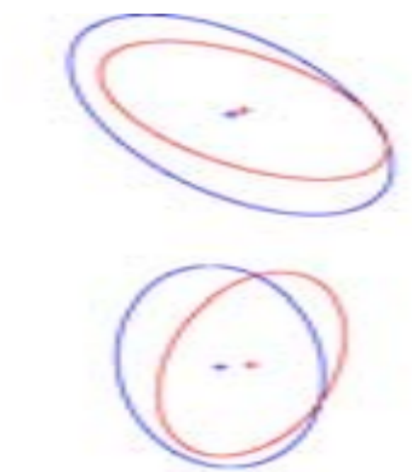
$$1 - \frac{R_{\mu_a} \cap R_{(H^T \mu_b H)}}{R_{\mu_a} \cup R_{(H^T \mu_b H)}} \leq \text{ overlap error}_{\max}$$

μ_a, μ_b the characteristic ellipses $(\mathbf{x}^T \mu \mathbf{x})$.

$$\left| 1 - s(H)^2 \frac{\min(\sigma_a^2, \sigma_b^2)}{\max(\sigma_a^2, \sigma_b^2)} \right| \leq \text{overlap error}_{\max}$$

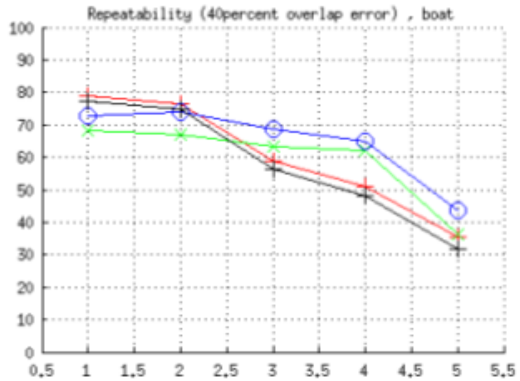
$s(H)$ the measured scale factor between $u_a(\mathbf{x})$ and $u_b(\mathbf{x})$.

or

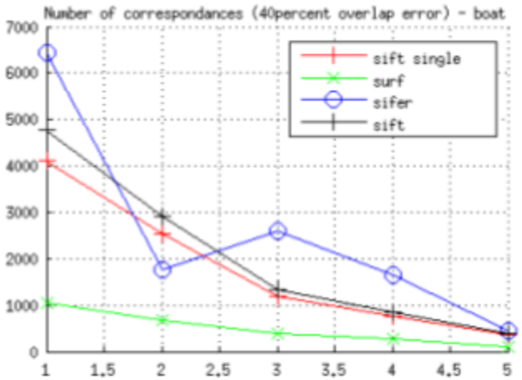


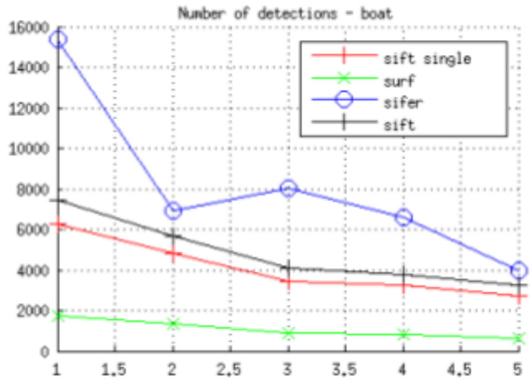
Classic detection/repeatability results
SIFT, SURF, SIFER

perturbation: rotation and scale

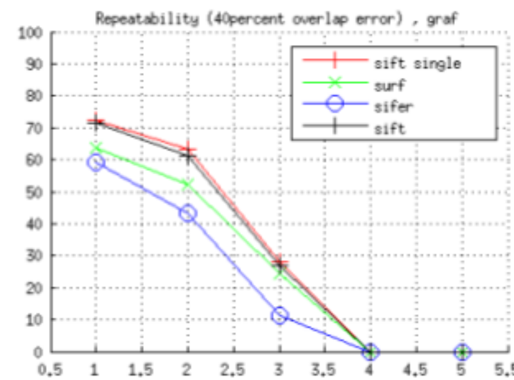


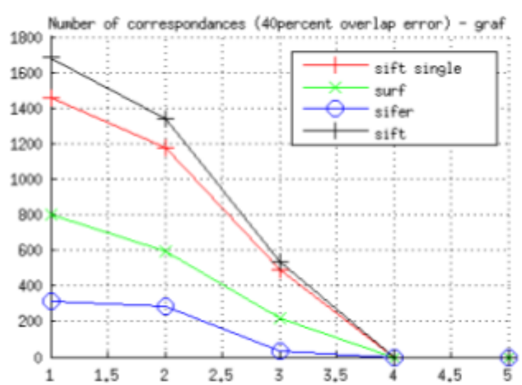
Problem: SIFER is NOT rotation or scale invariant and nevertheless beats two theoretically scale and rotation invariant detectors !





Classic detection/repeatability results
SIFT, SURF, SIFER perturbation: tilt

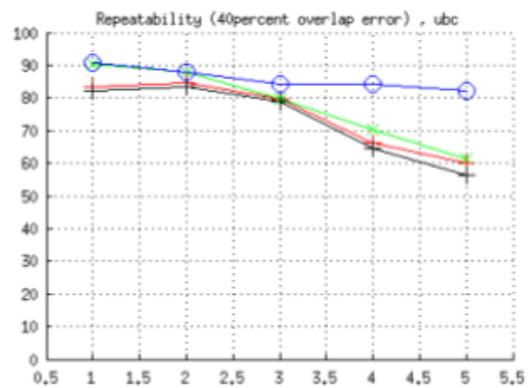




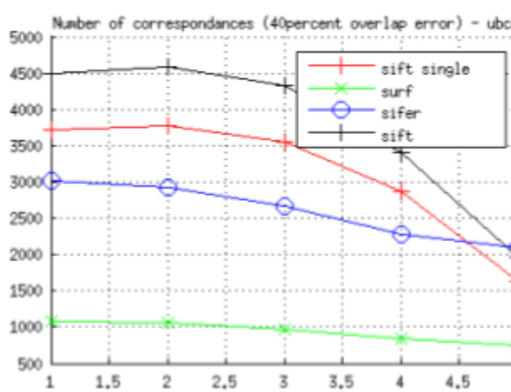
Classic detection/repeatability results

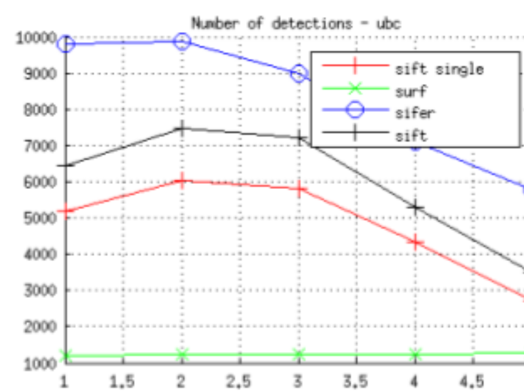
SIFT, SURF, SIFER

perturbation: JPG compression

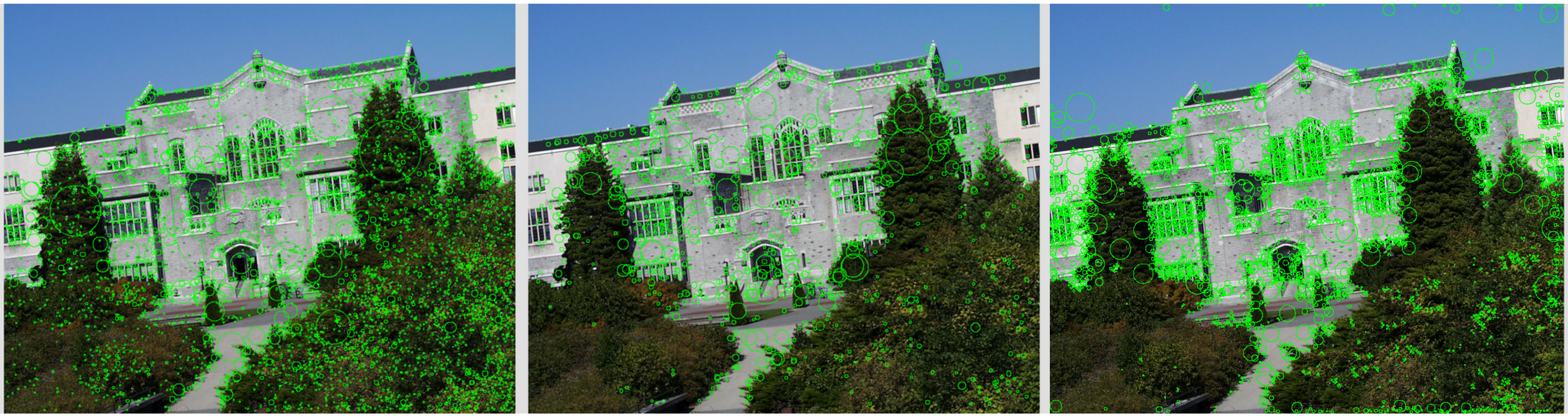


Apparent conclusion: SIFER is more repeatable and has more detections than SIFT or SURF. It is therefore better.





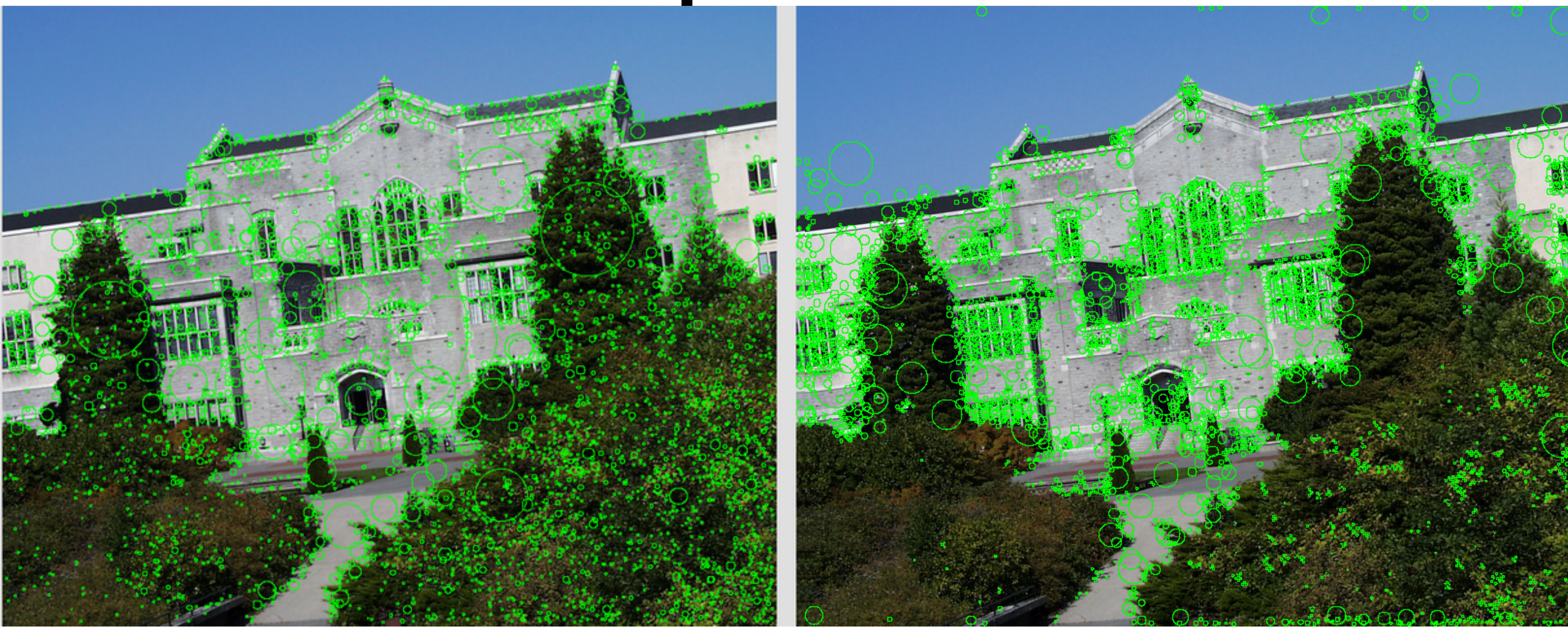
Detections maps



SIFT SURF SIFER

The SIFT descriptors are more spread out than the SURF descriptors

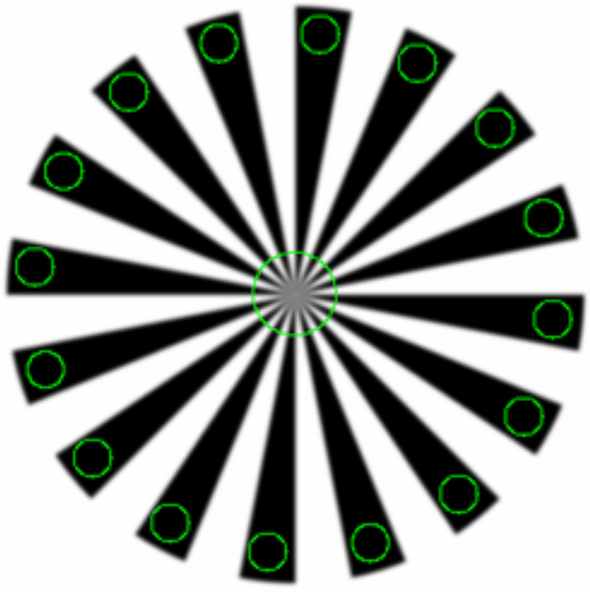
Detections maps



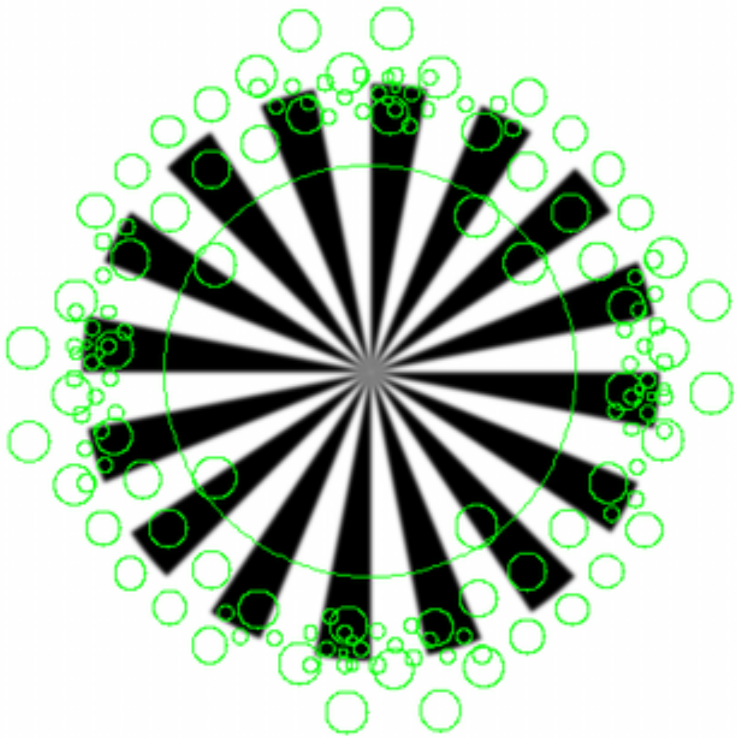
SIFT SIFER

The SIFER descriptors are more cluttered than the SIFT descriptors

Detections maps (siemens star)

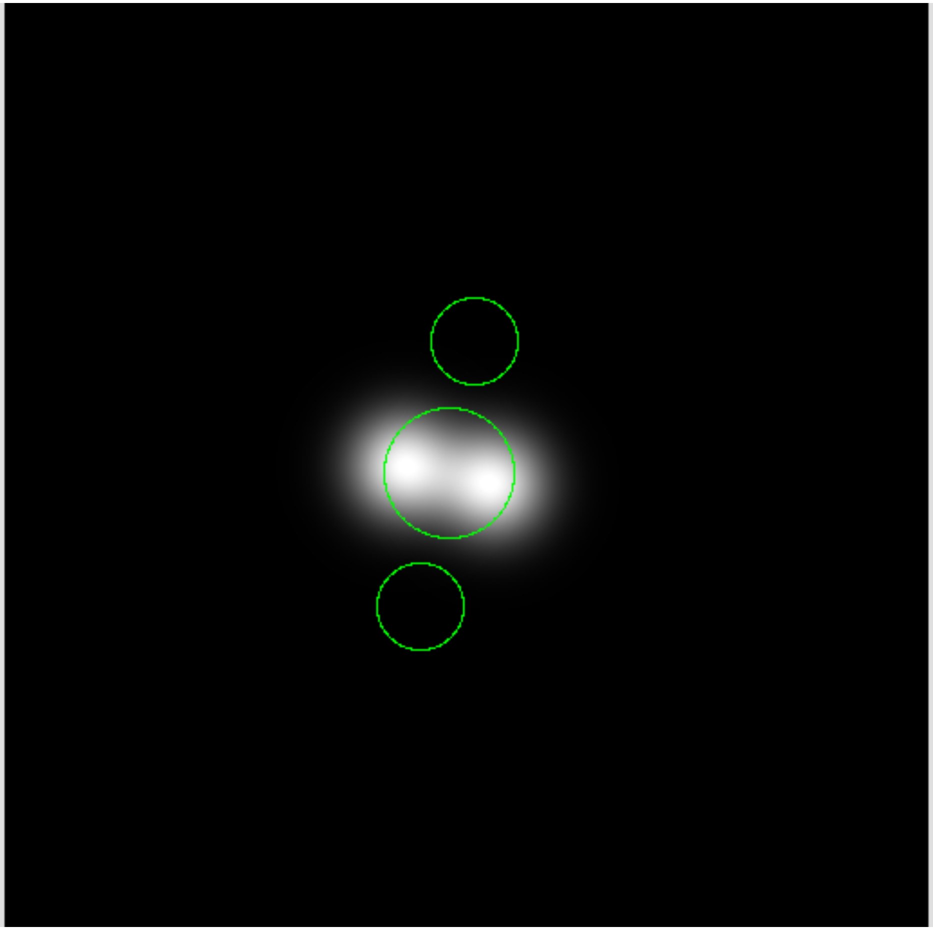


SIFT

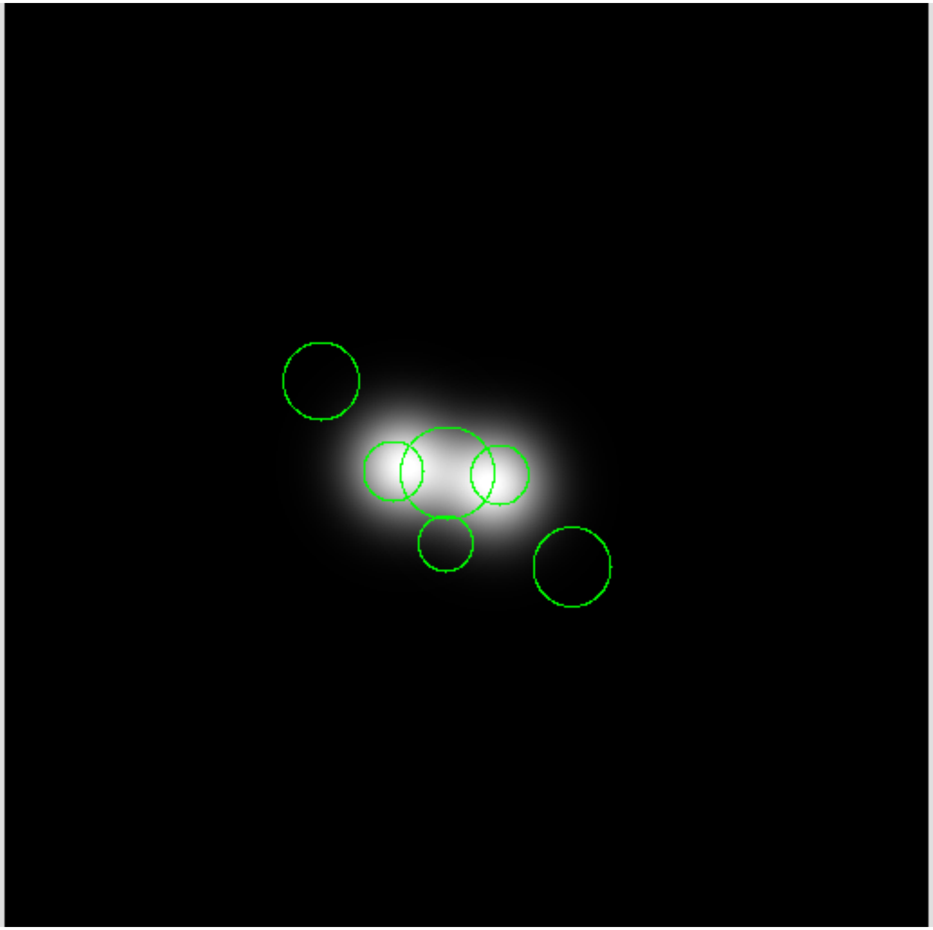


SIFER

Detections maps (two Gaussian blobs)

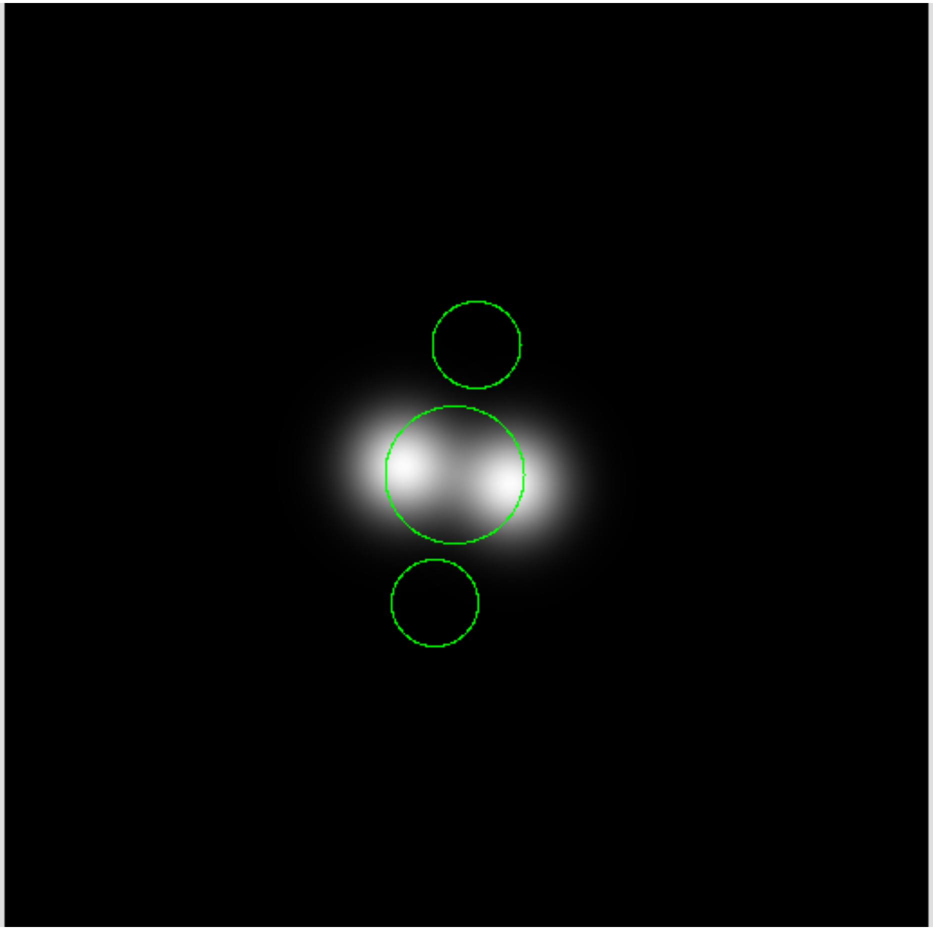


SIFT

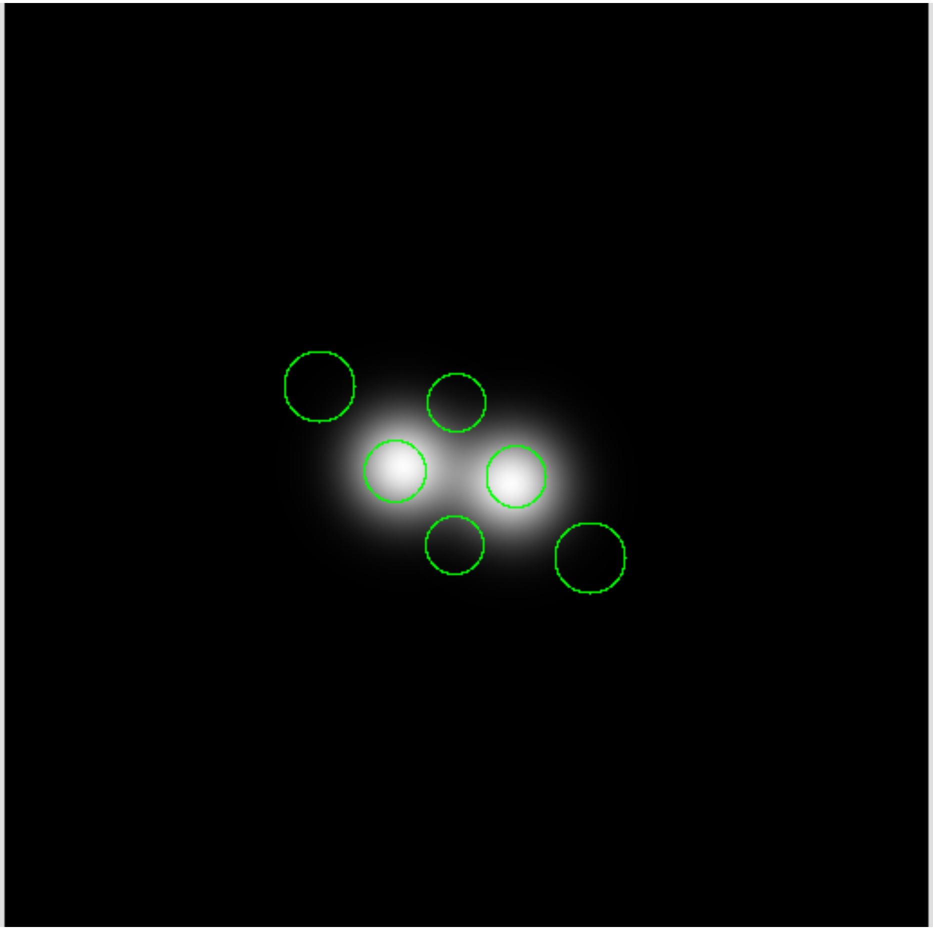


SIFER

Detections maps (two Gaussian blobs)

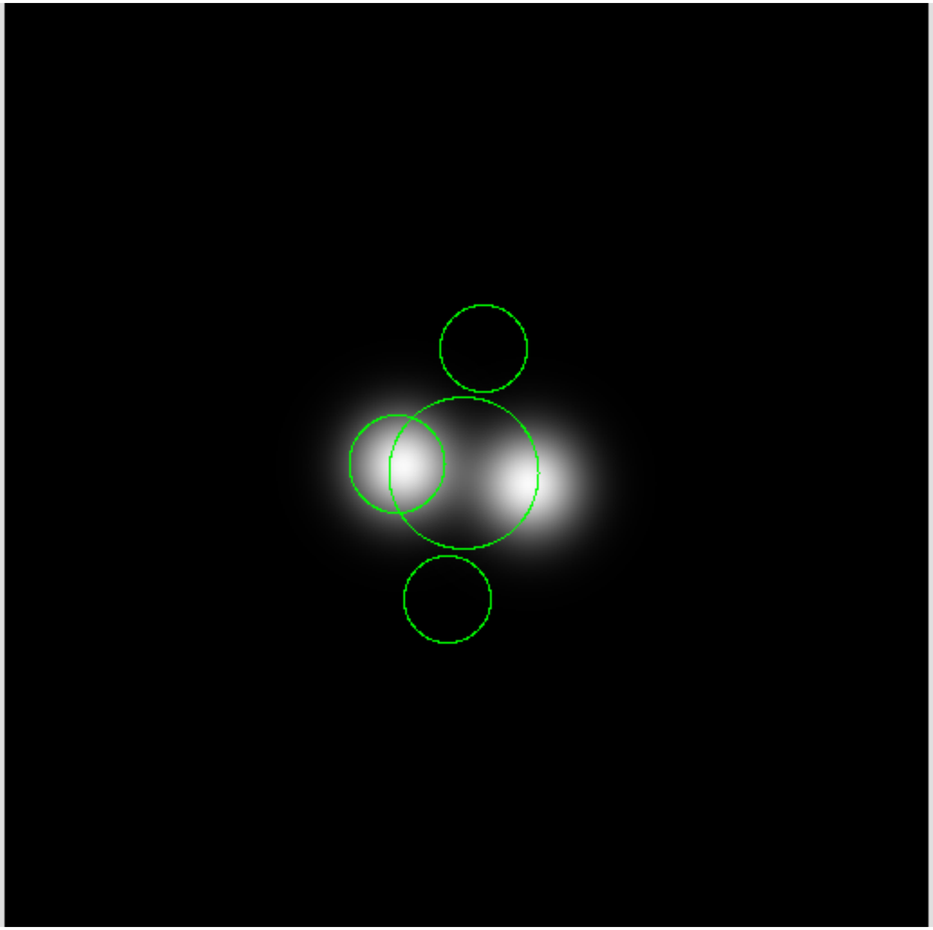


SIFT

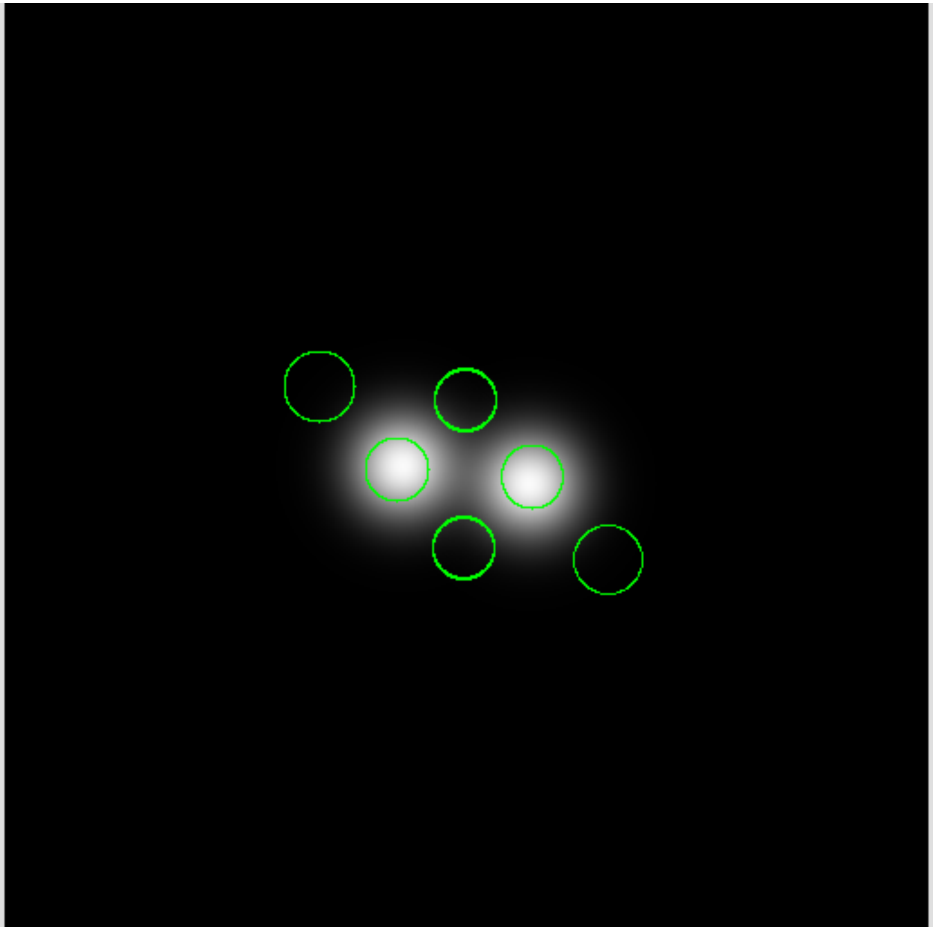


SIFER

Detections maps (two Gaussian blobs)

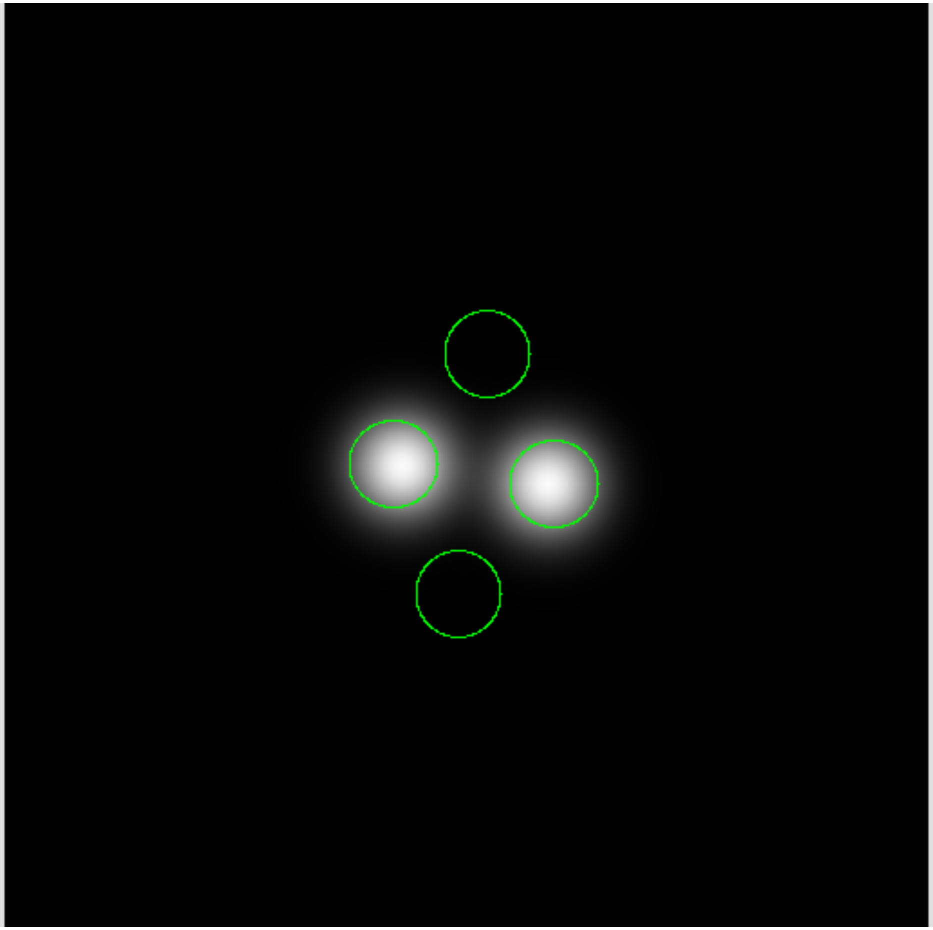


SIFT

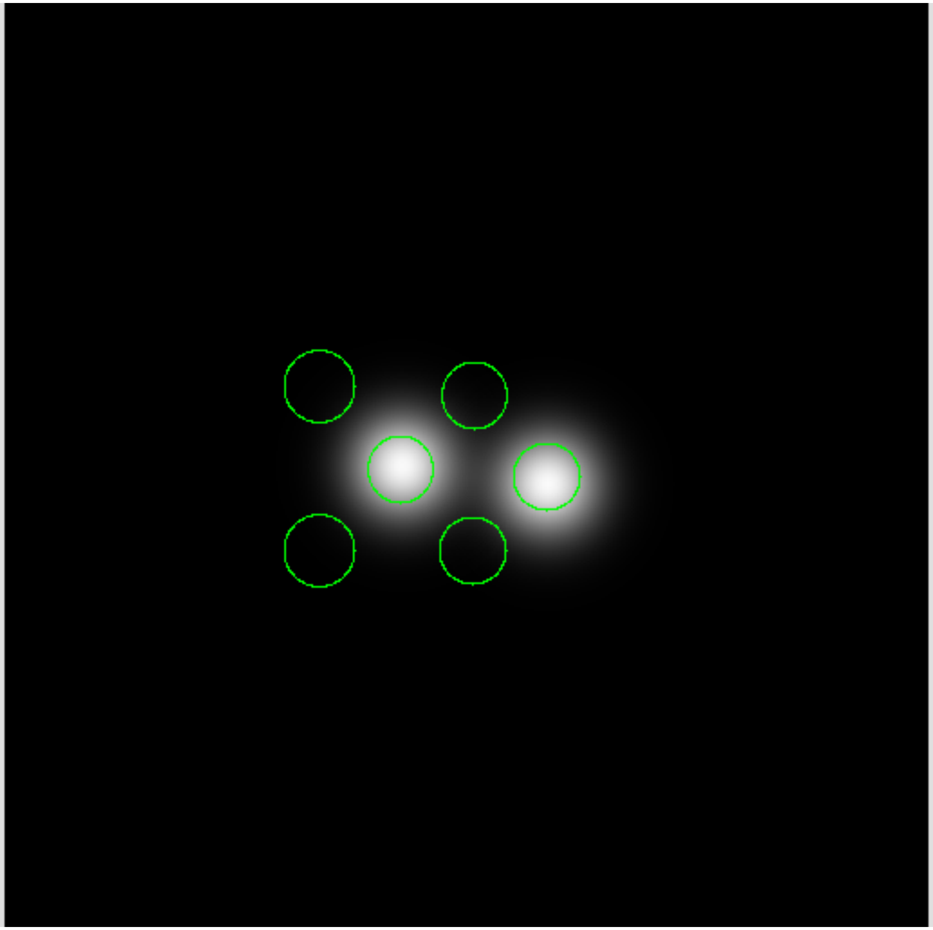


SIFER

Detections maps (two Gaussian blobs)

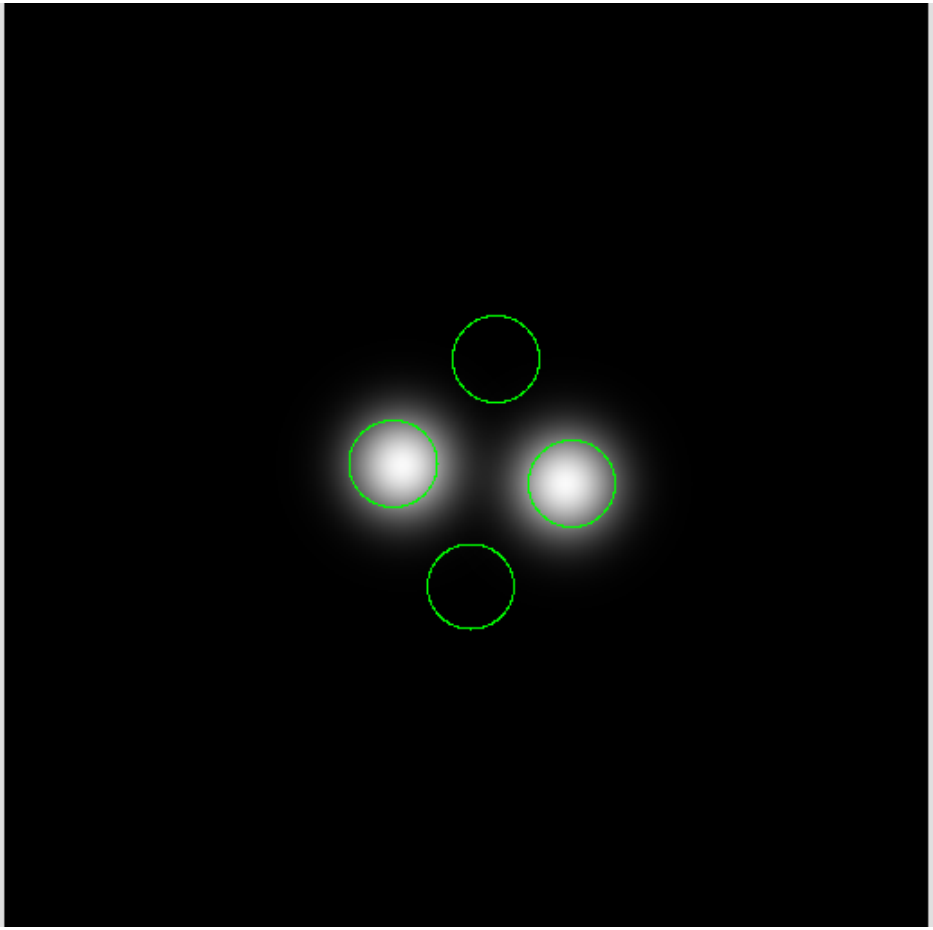


SIFT

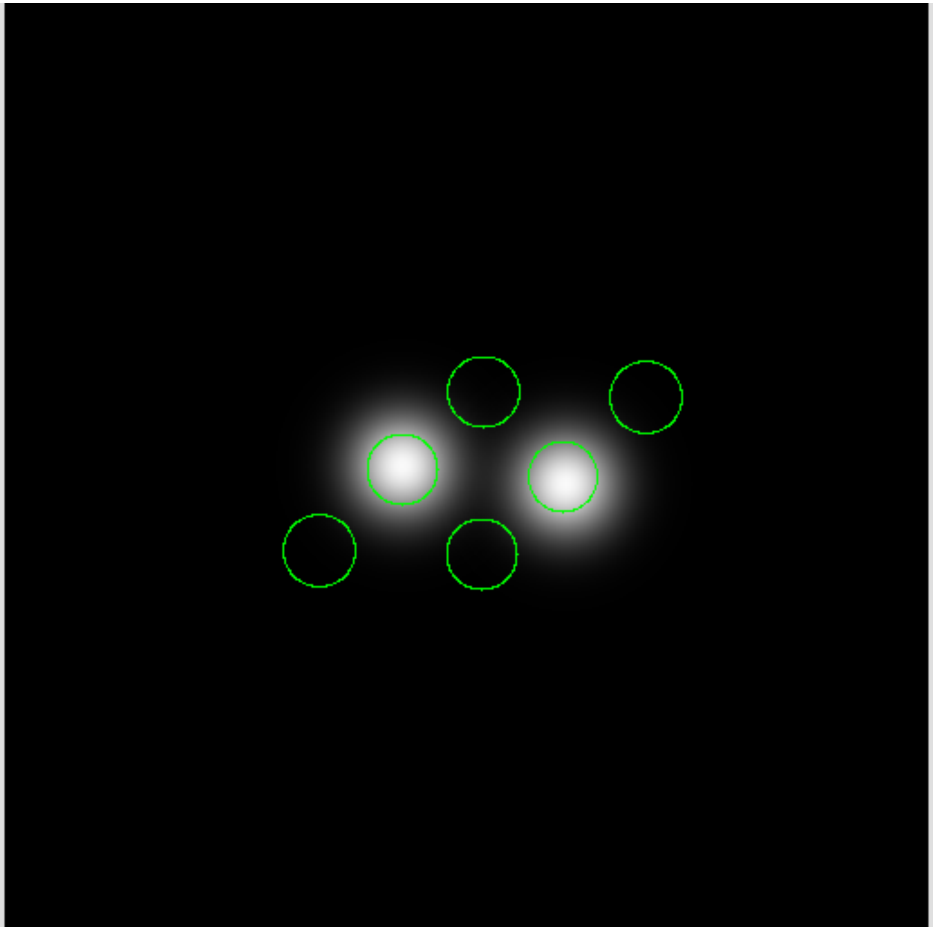


SIFER

Detections maps (two Gaussian blobs)

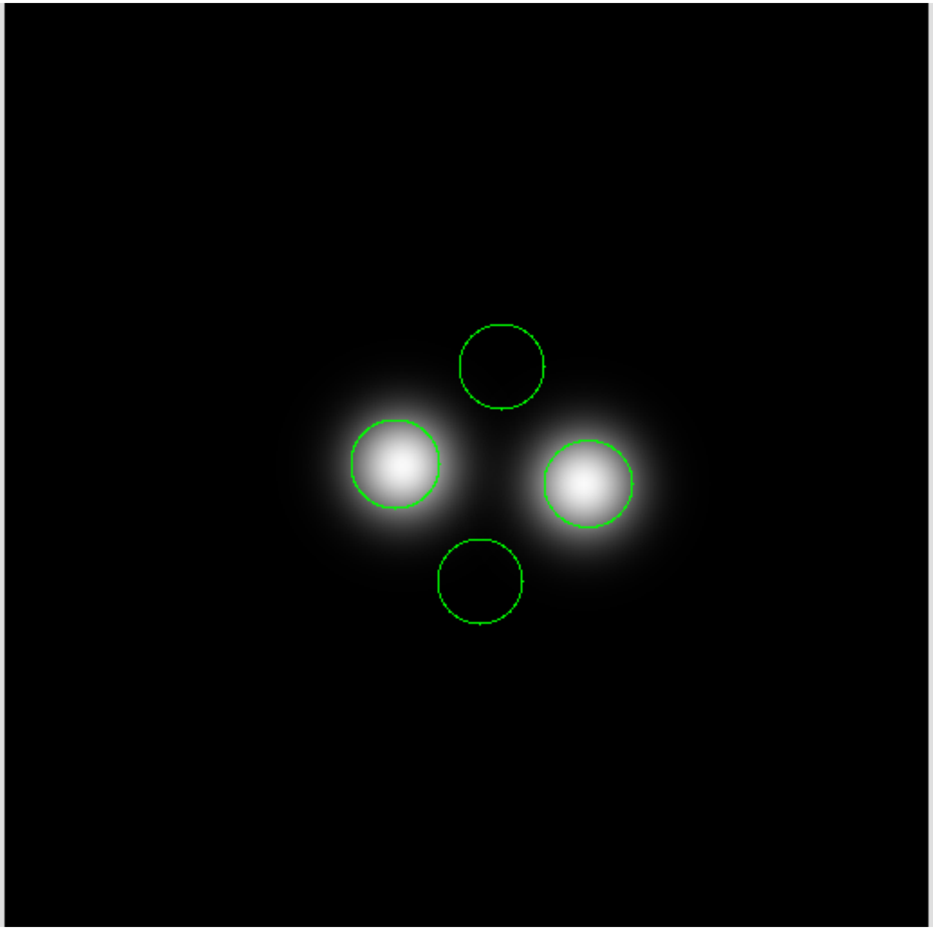


SIFT

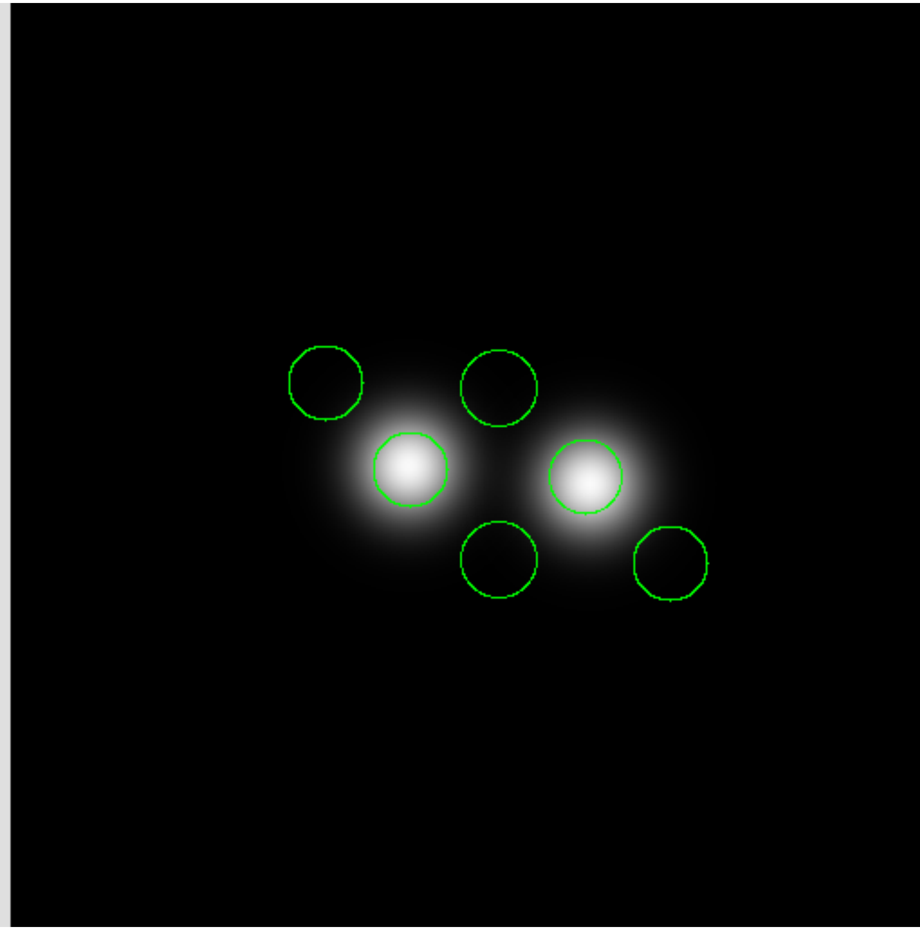


SIFER

Detections maps (two Gaussian blobs)

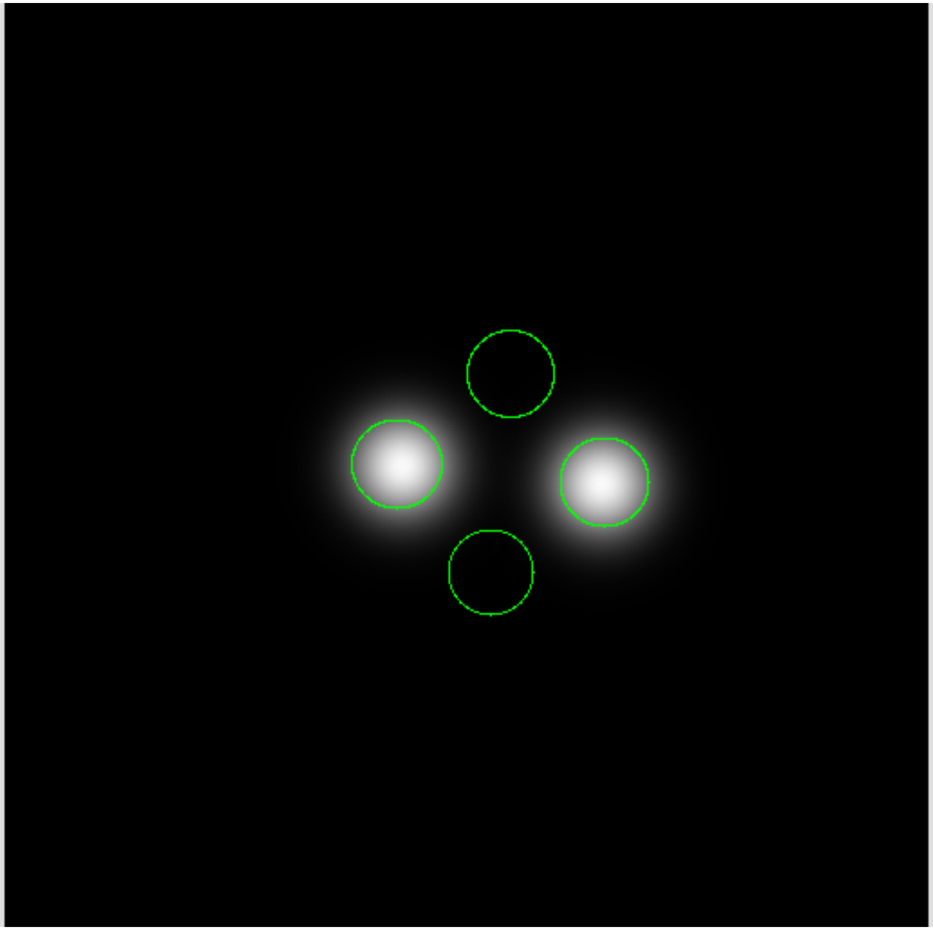


SIFT
SIFT

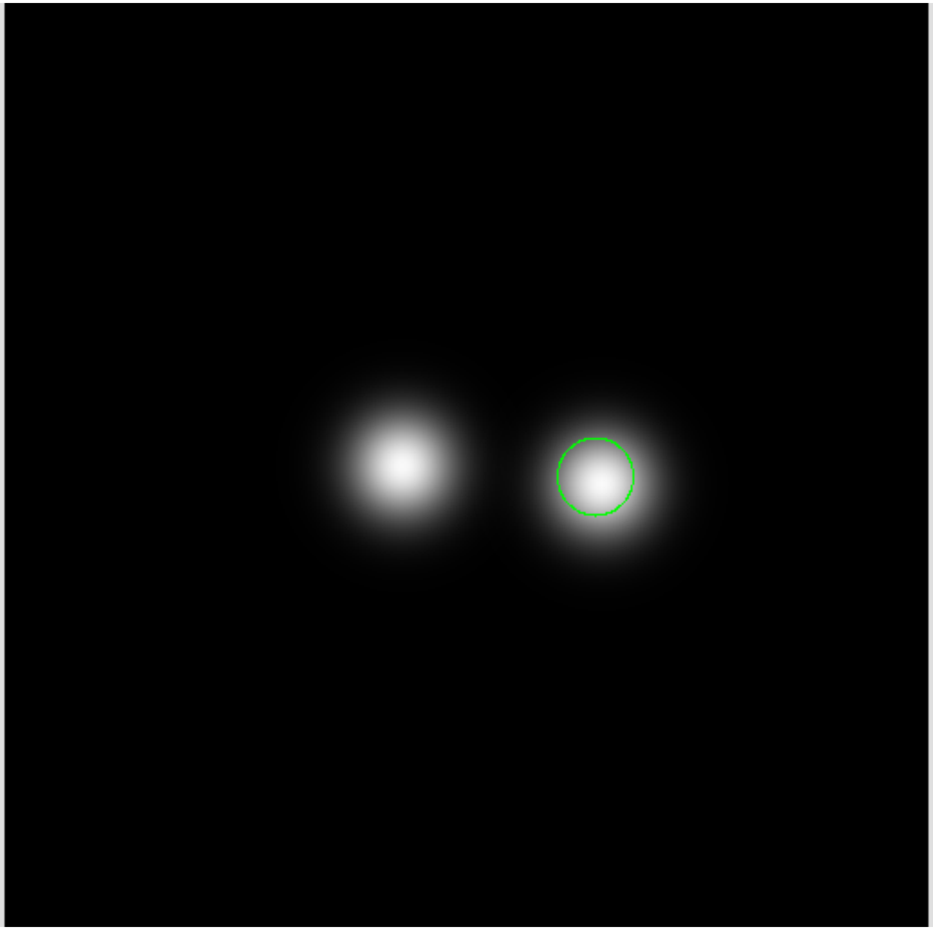


SIFER
SIFER

Detections maps (two Gaussian blobs)

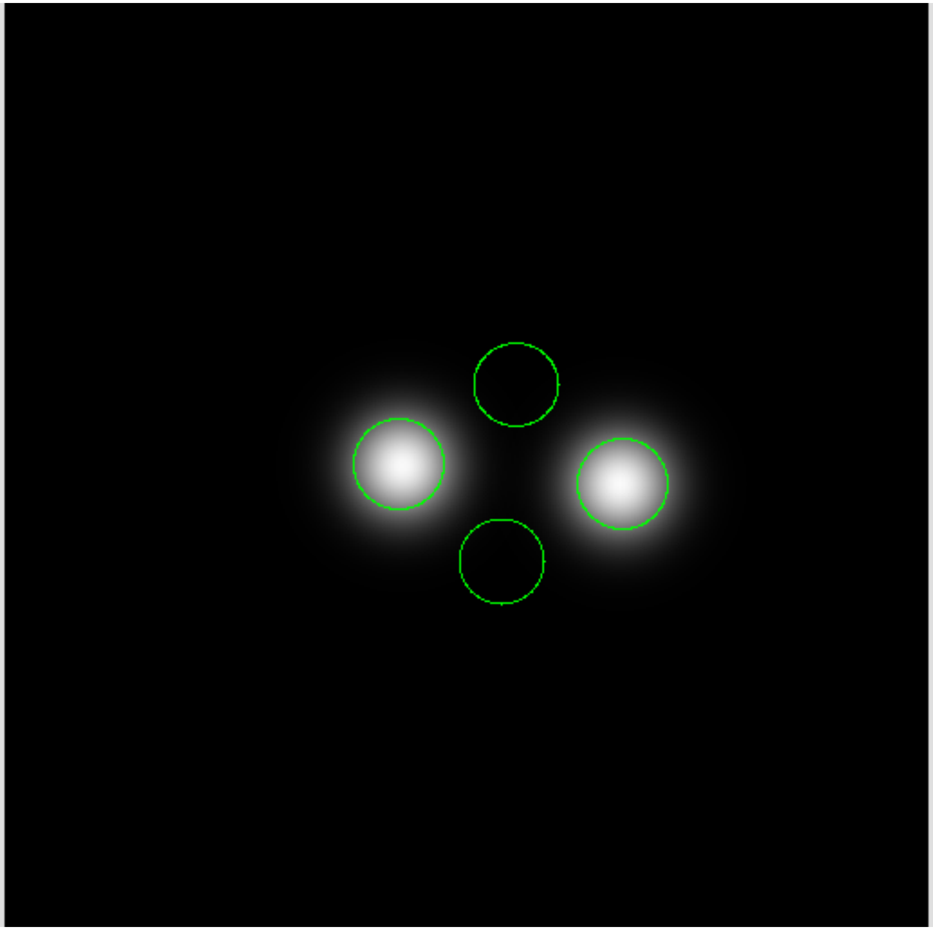


SIFT

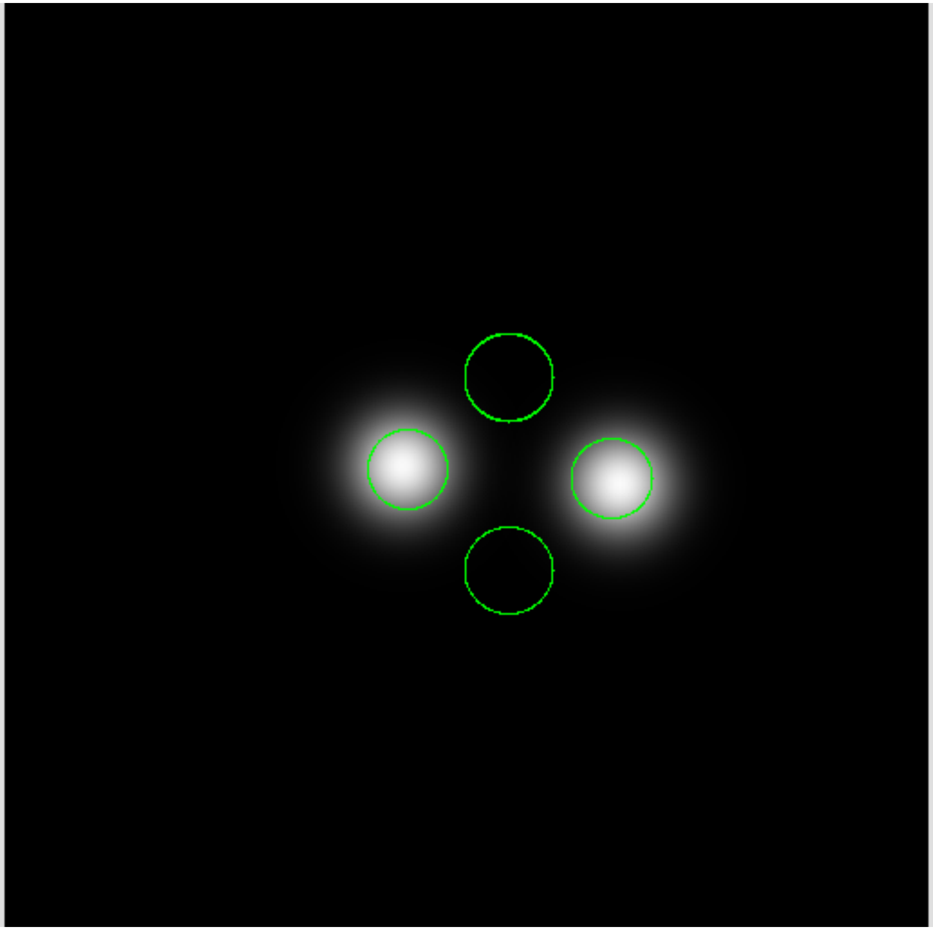


SIFER

Detections maps (two Gaussian blobs)

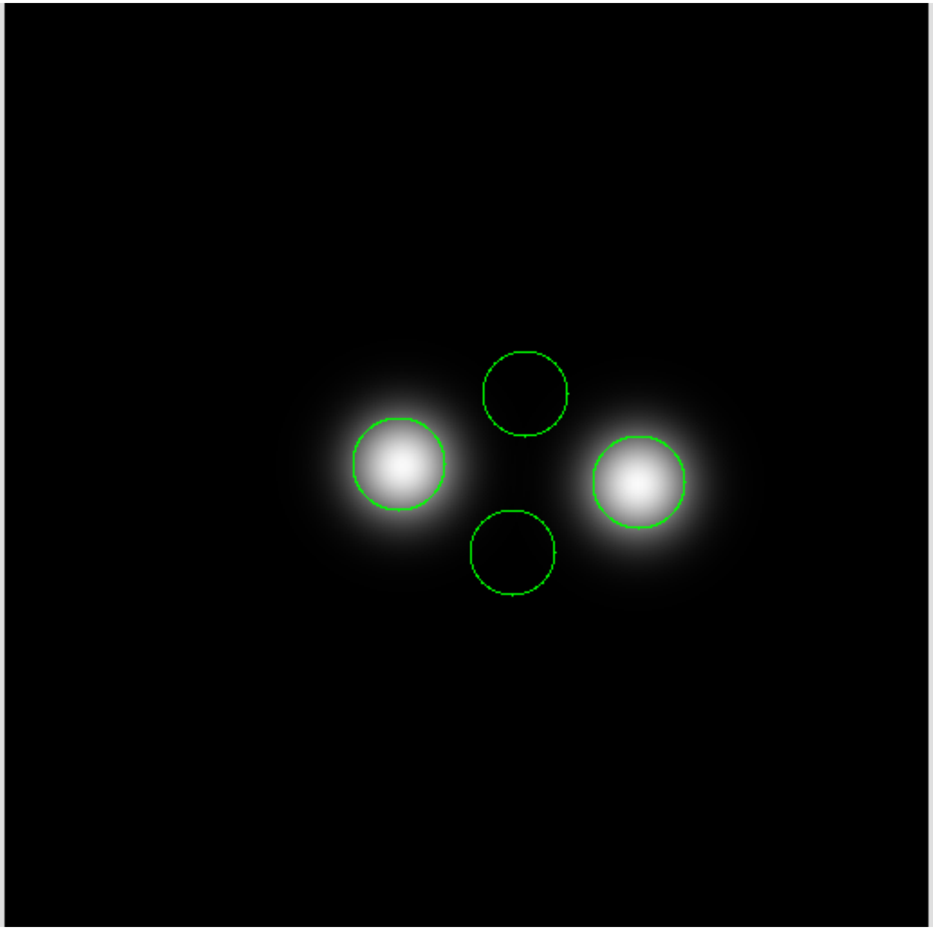


SIFT

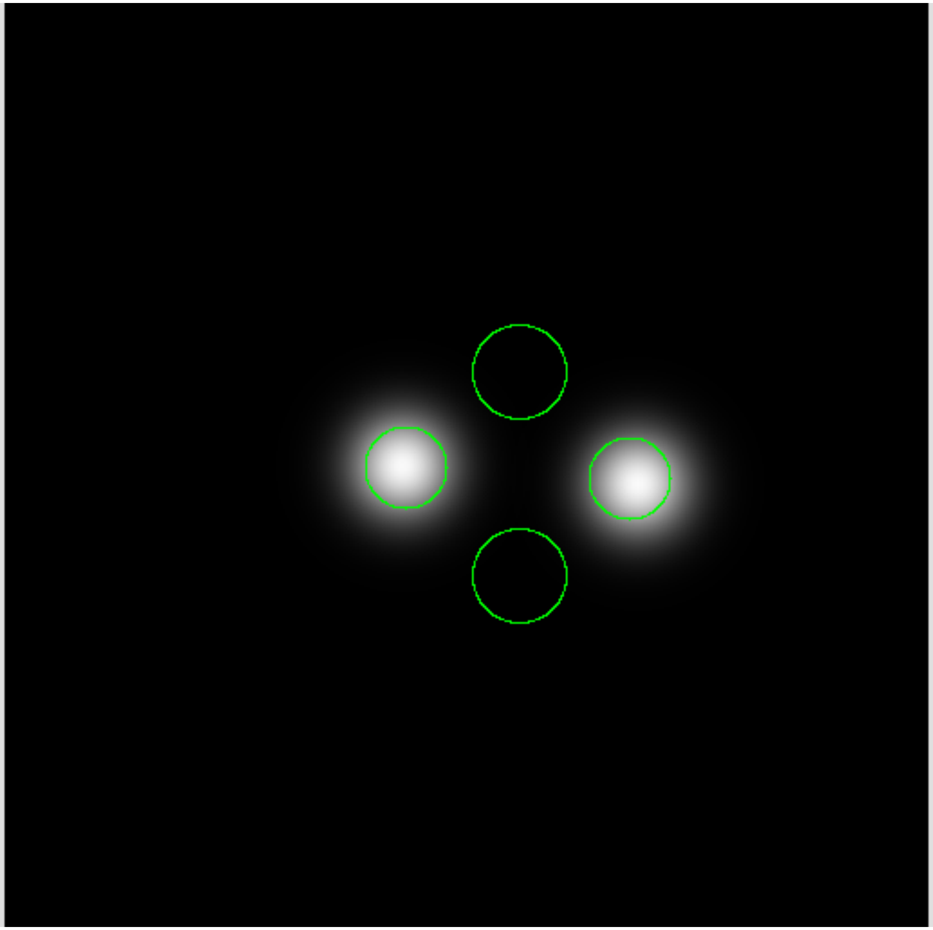


SIFER

Detections maps (two Gaussian blobs)

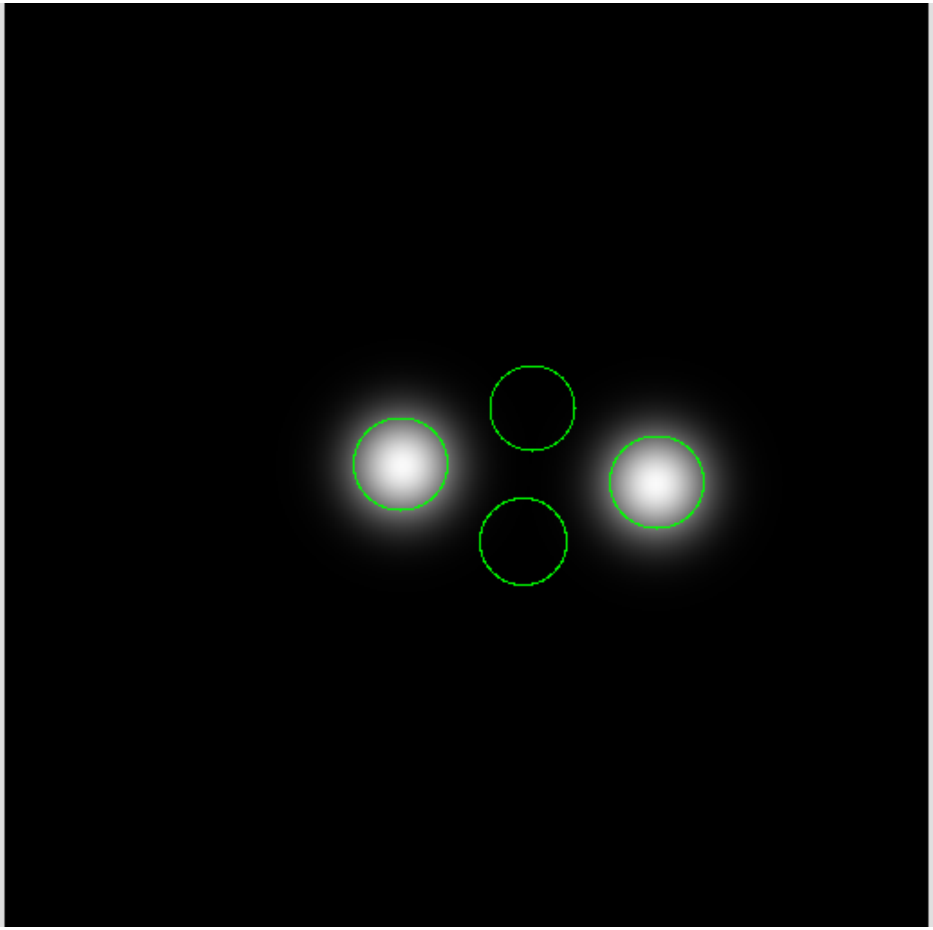


SIFT

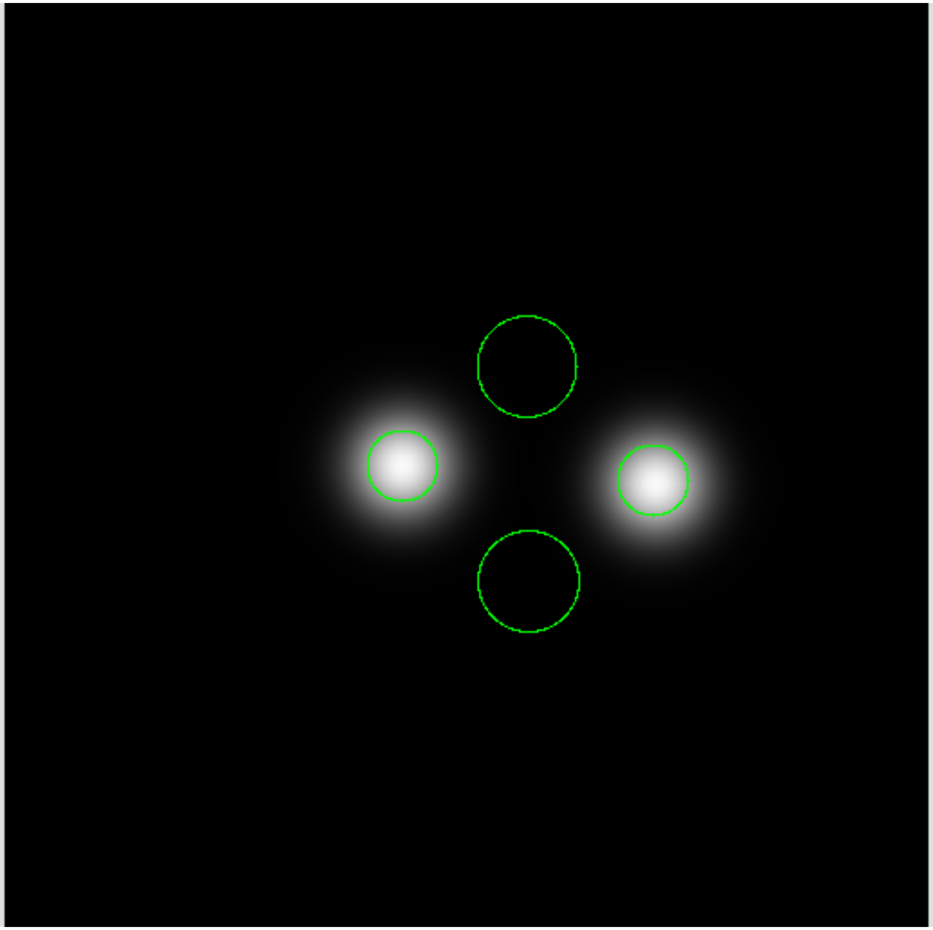


SIFER

Detections maps (two Gaussian blobs)

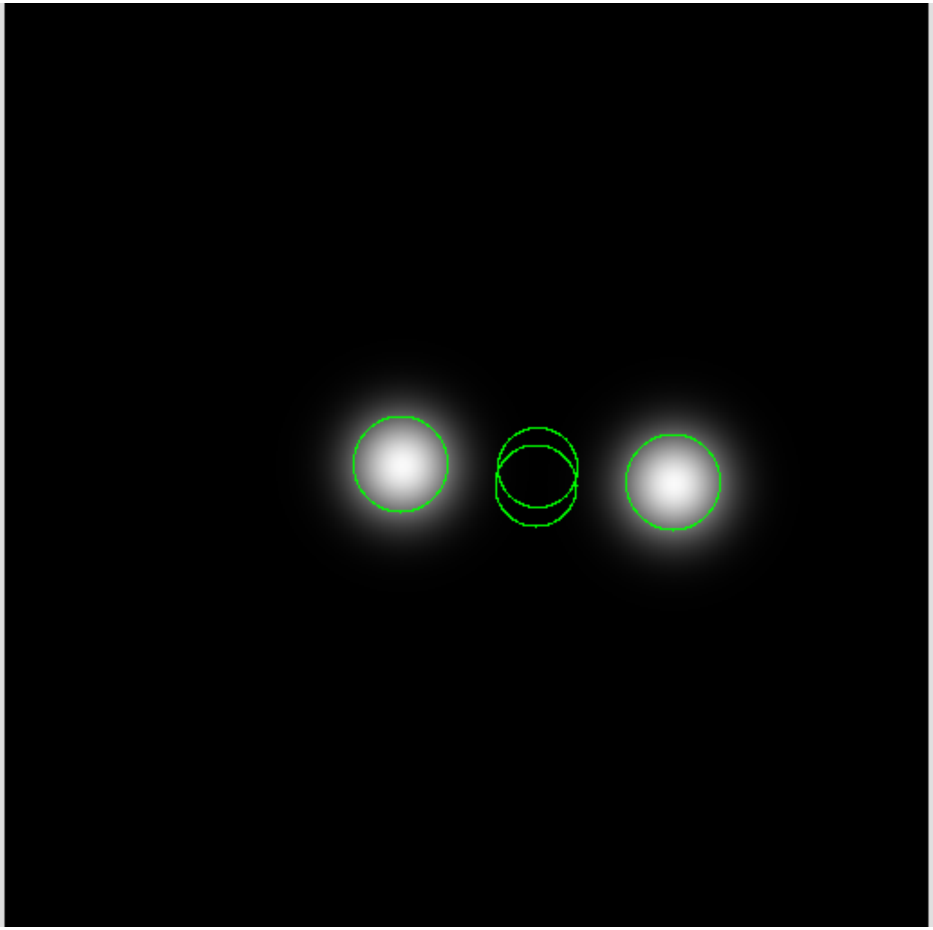


SIFT

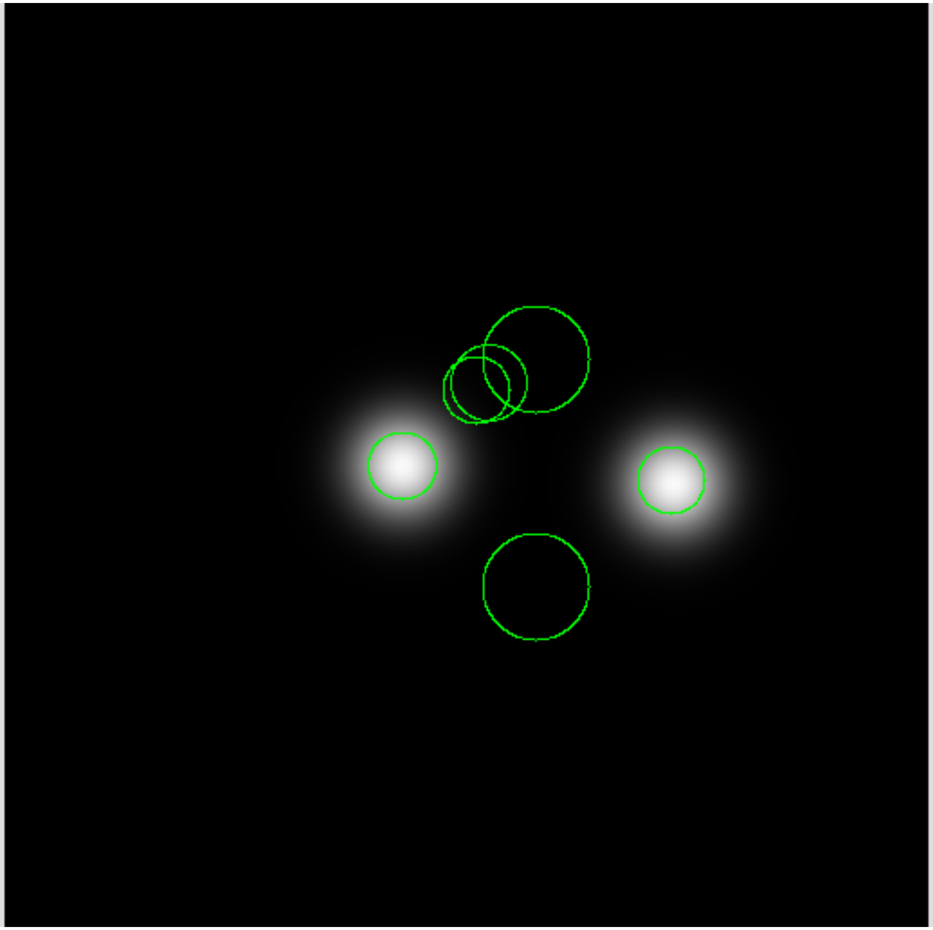


SIFER

Detections maps (two Gaussian blobs)

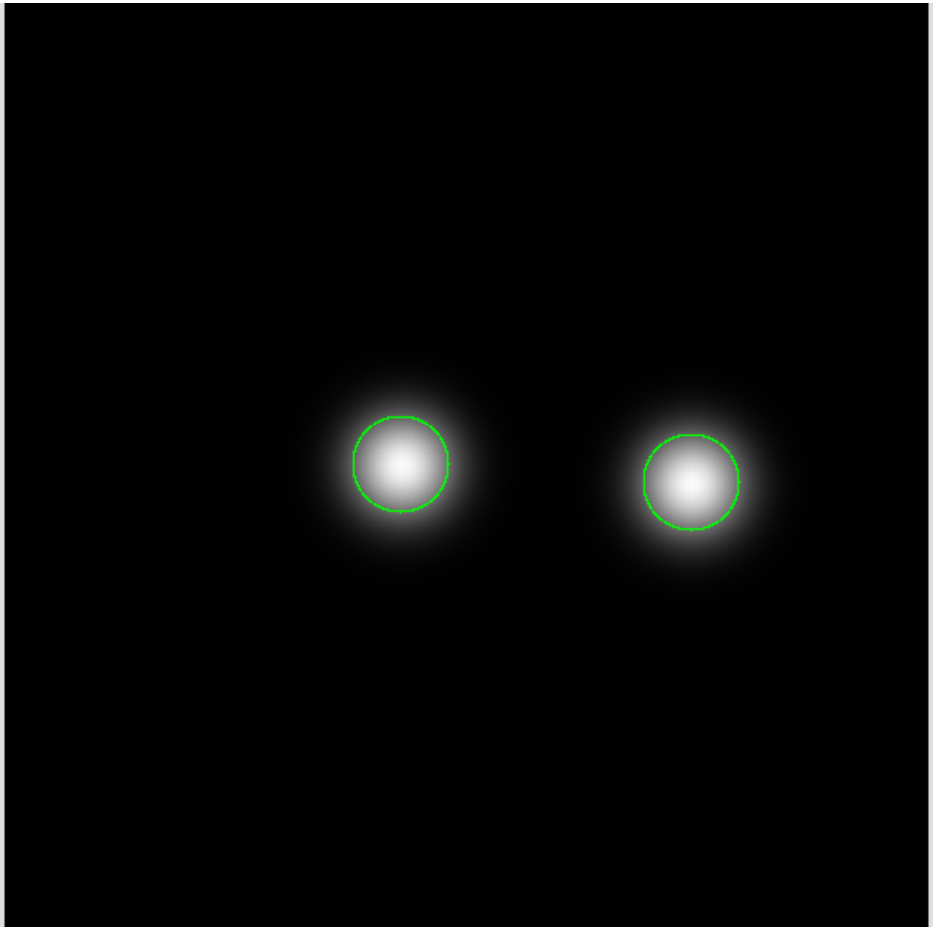


SIFT

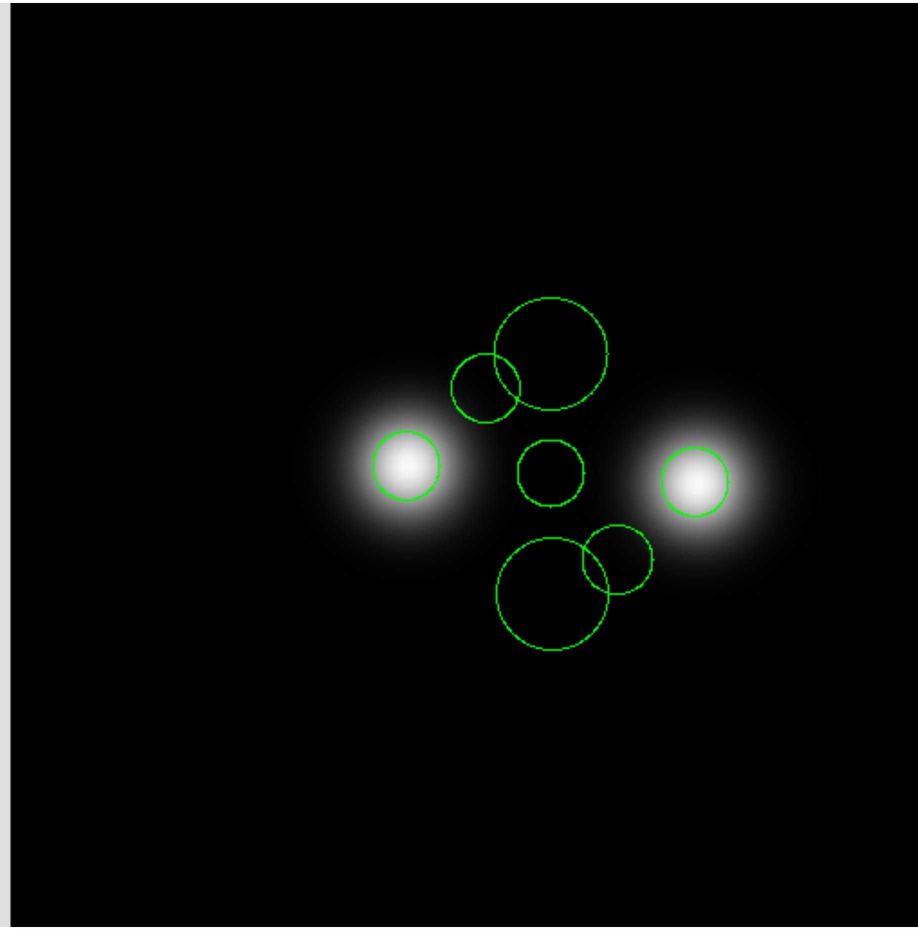


SIFER

Detections maps (two Gaussian blobs)

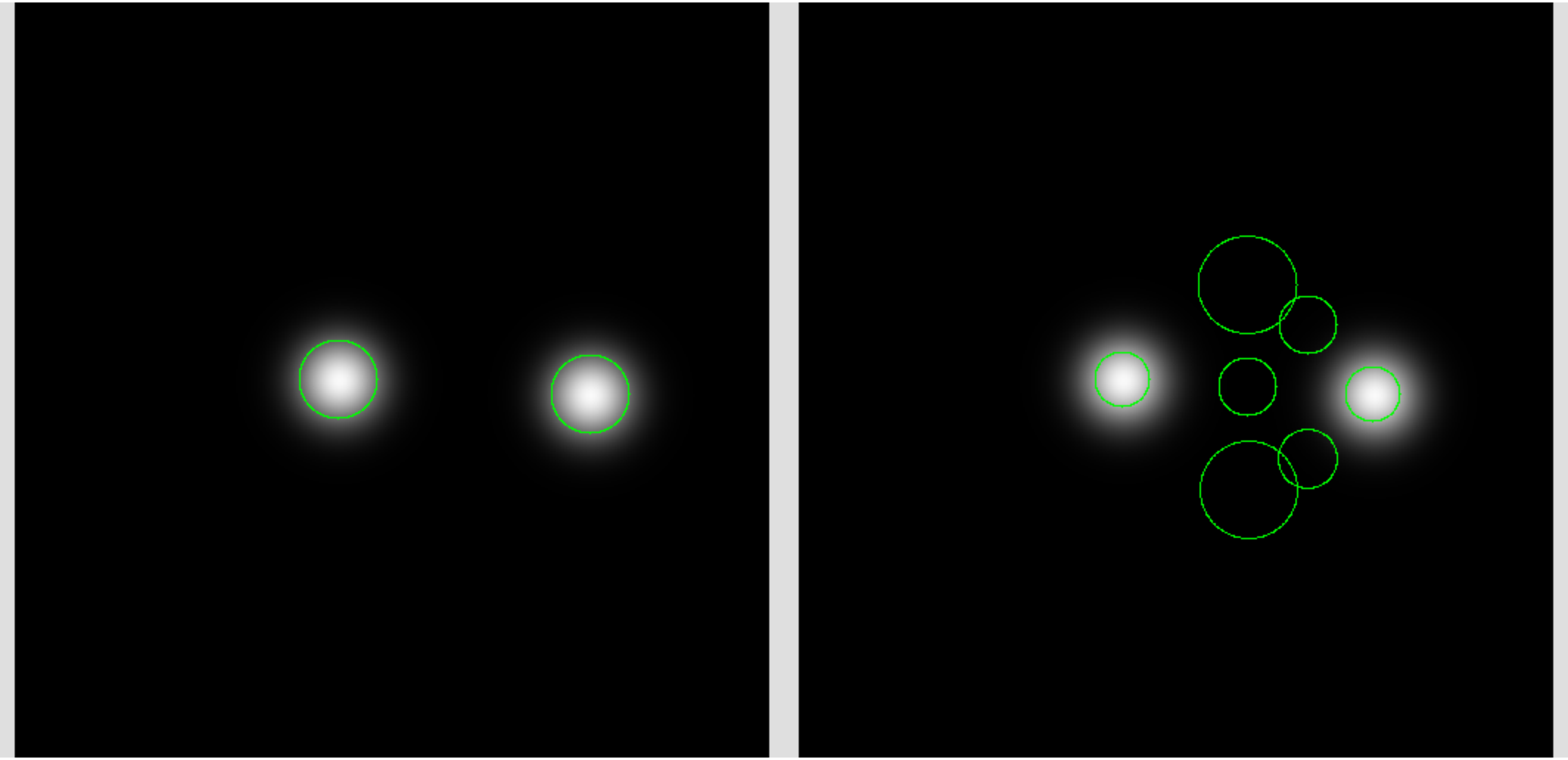


SIFT



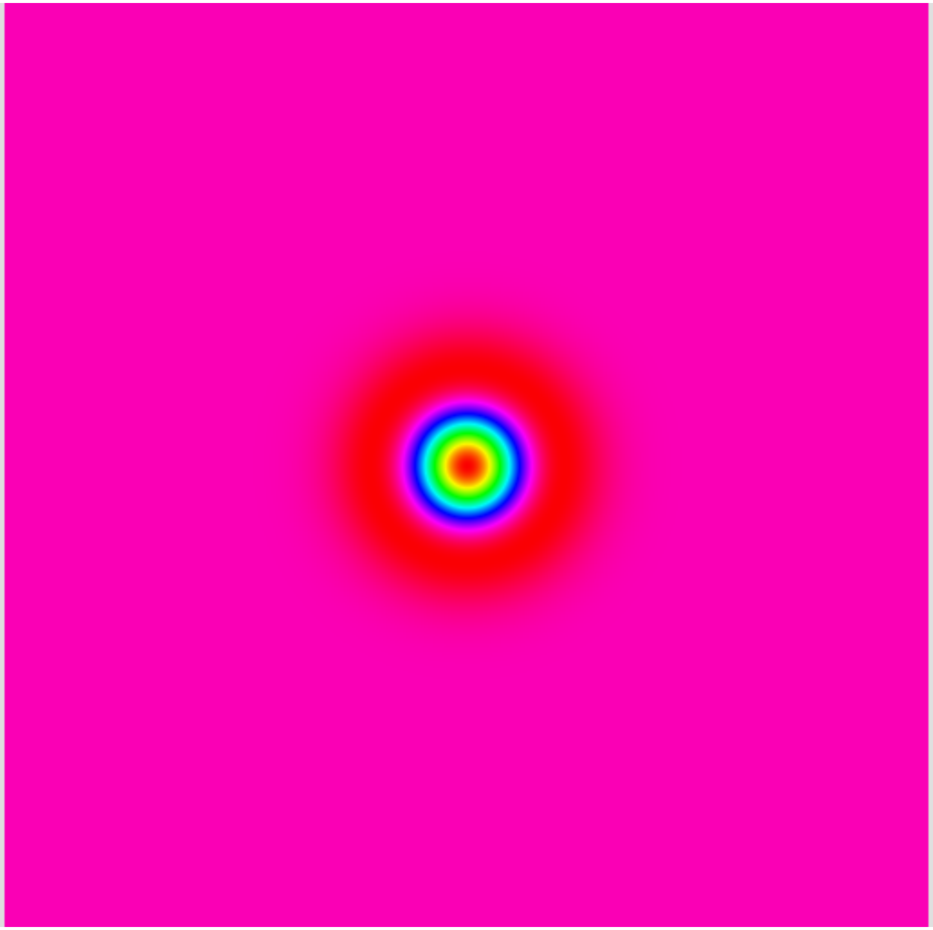
SIFER

Detections maps (two blobs)

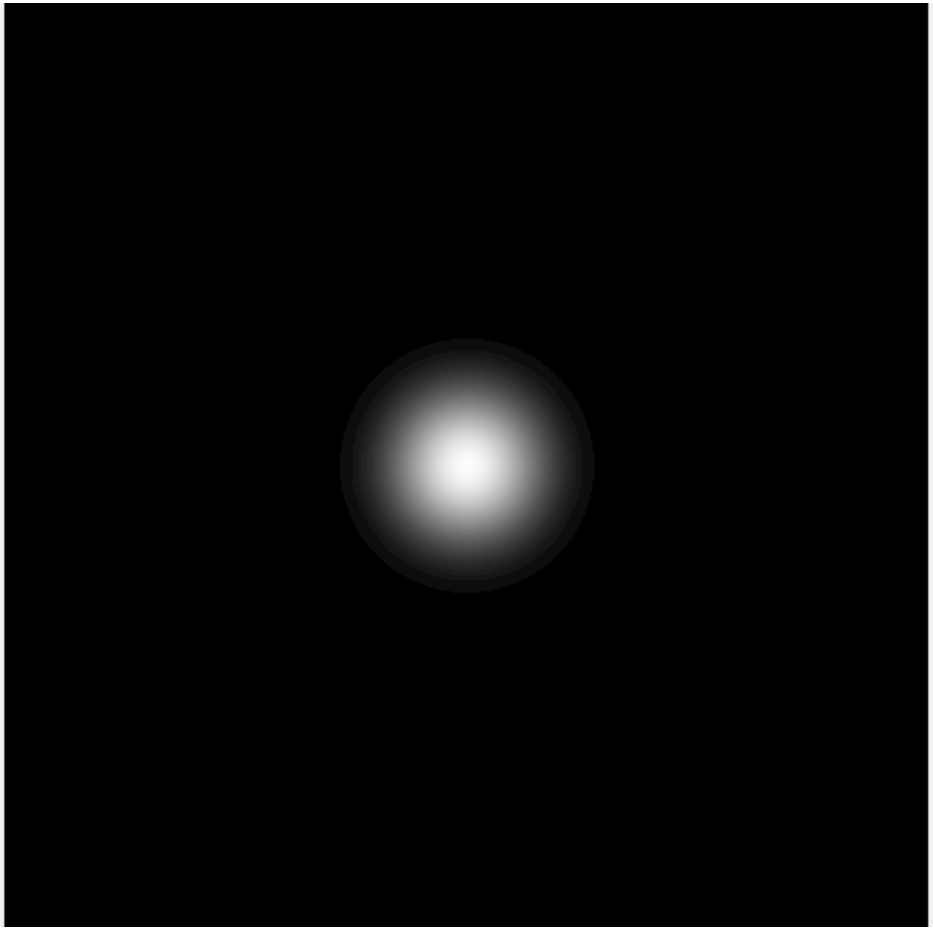


SIFT SIFER

Multi-scale response (Gaussian blob)



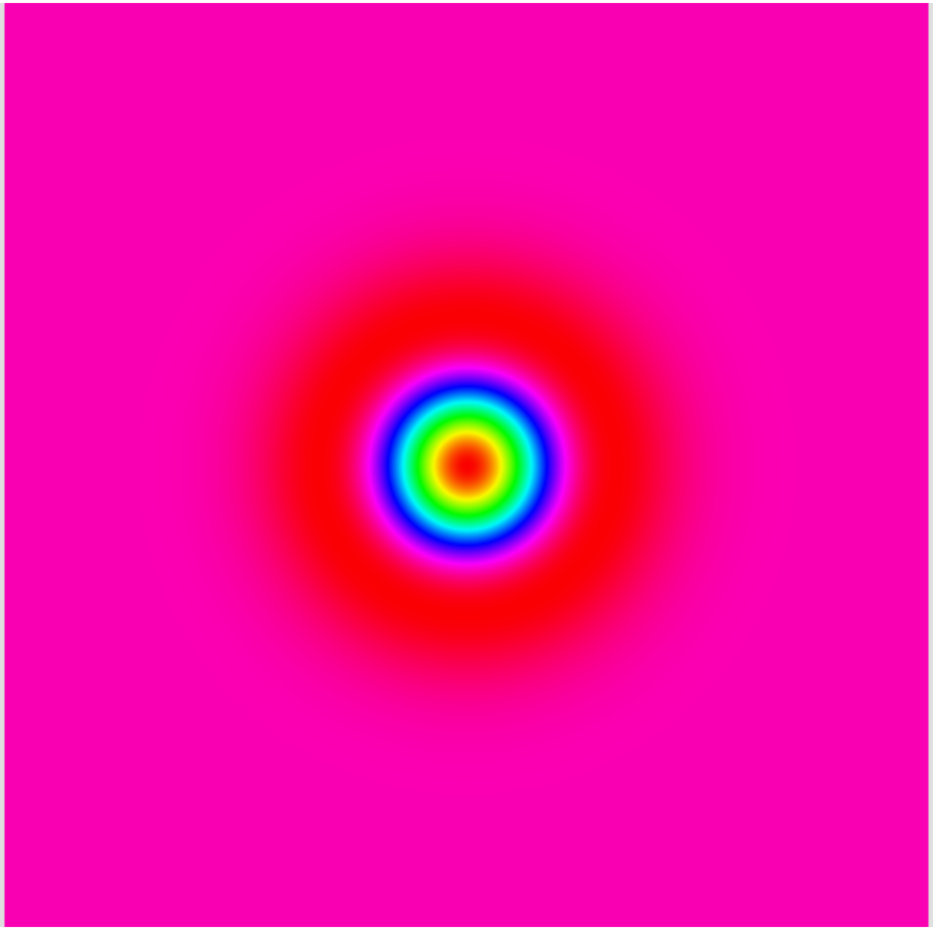
$$K_\sigma(x,y) = G_{k\sigma}(x,y) - G_\sigma(x,y)$$



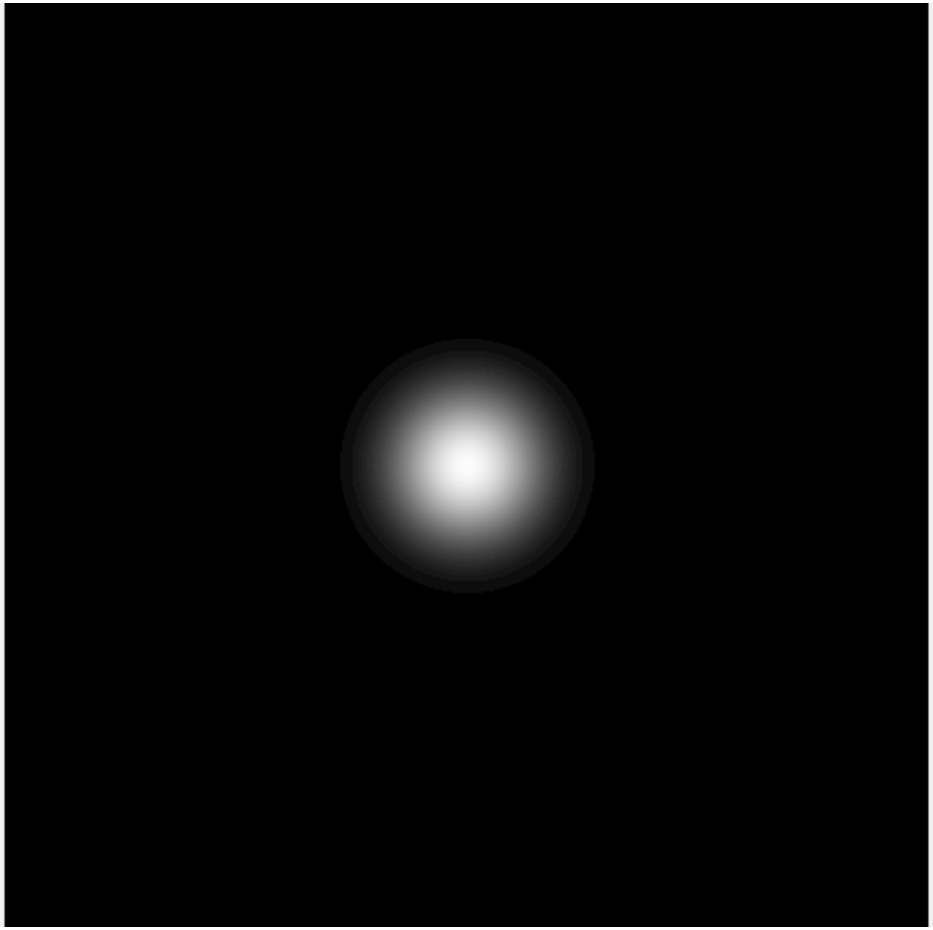
$$u(\mathbf{x})$$

SIFT

Multi-scale response (Gaussian blob)



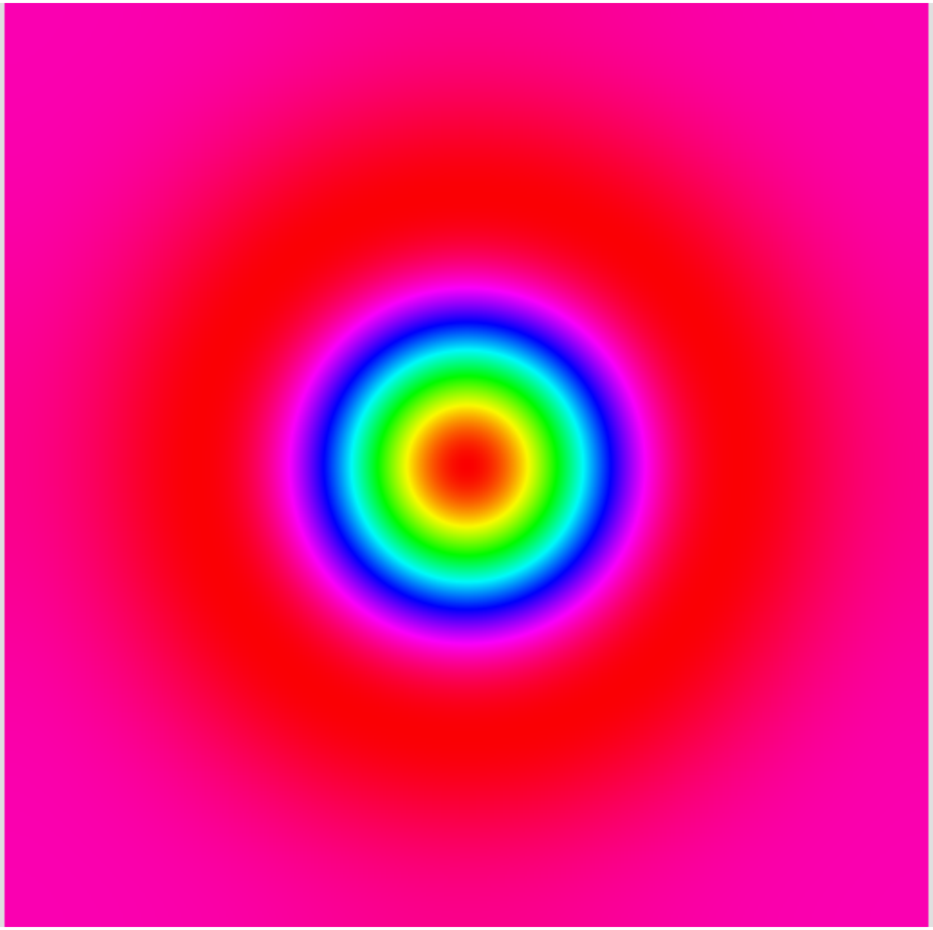
$$K_\sigma(x,y) = G_{k\sigma}(x,y) - G_\sigma(x,y)$$



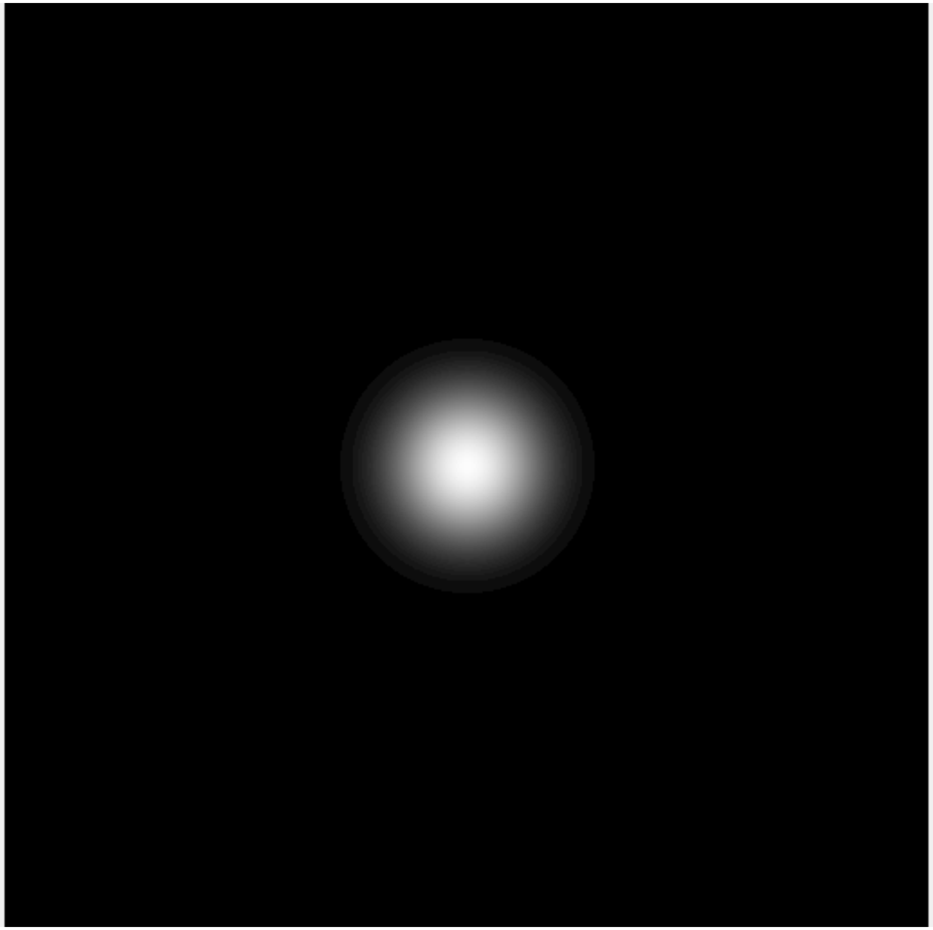
$$u(\mathbf{x})$$

SIFT

Multi-scale response (Gaussian blob)



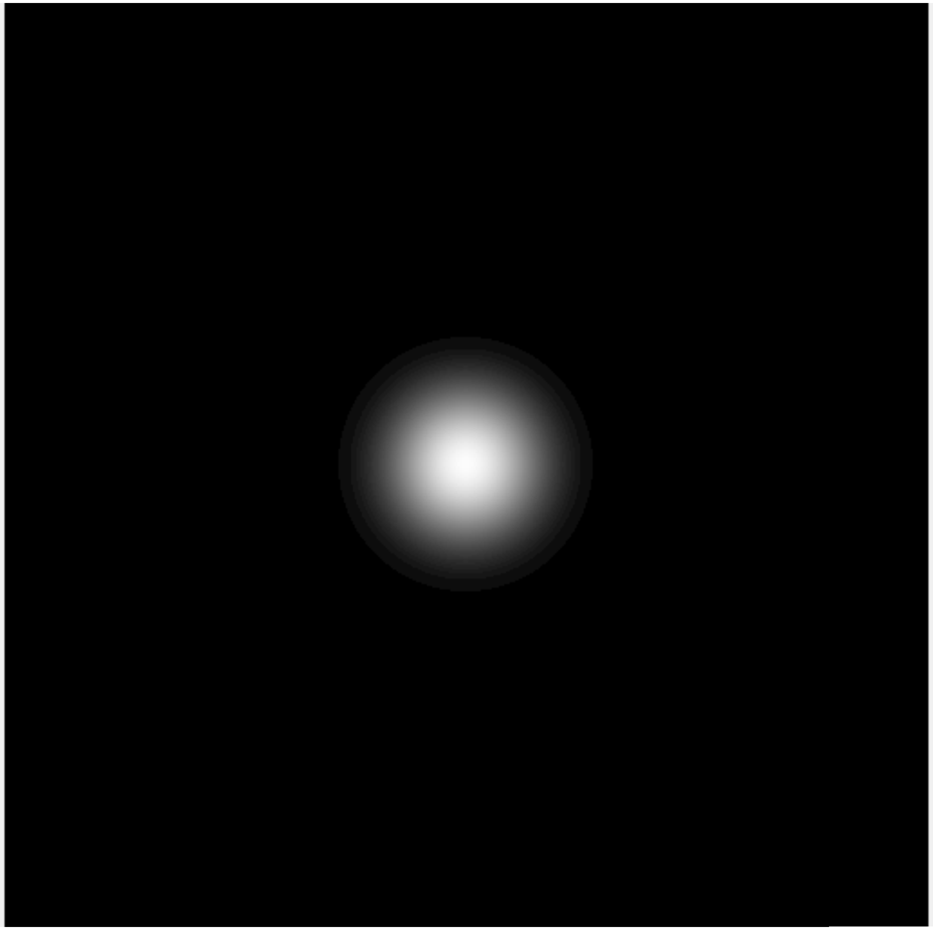
$$K_\sigma(x,y) = G_{k\sigma}(x,y) - G_\sigma(x,y)$$

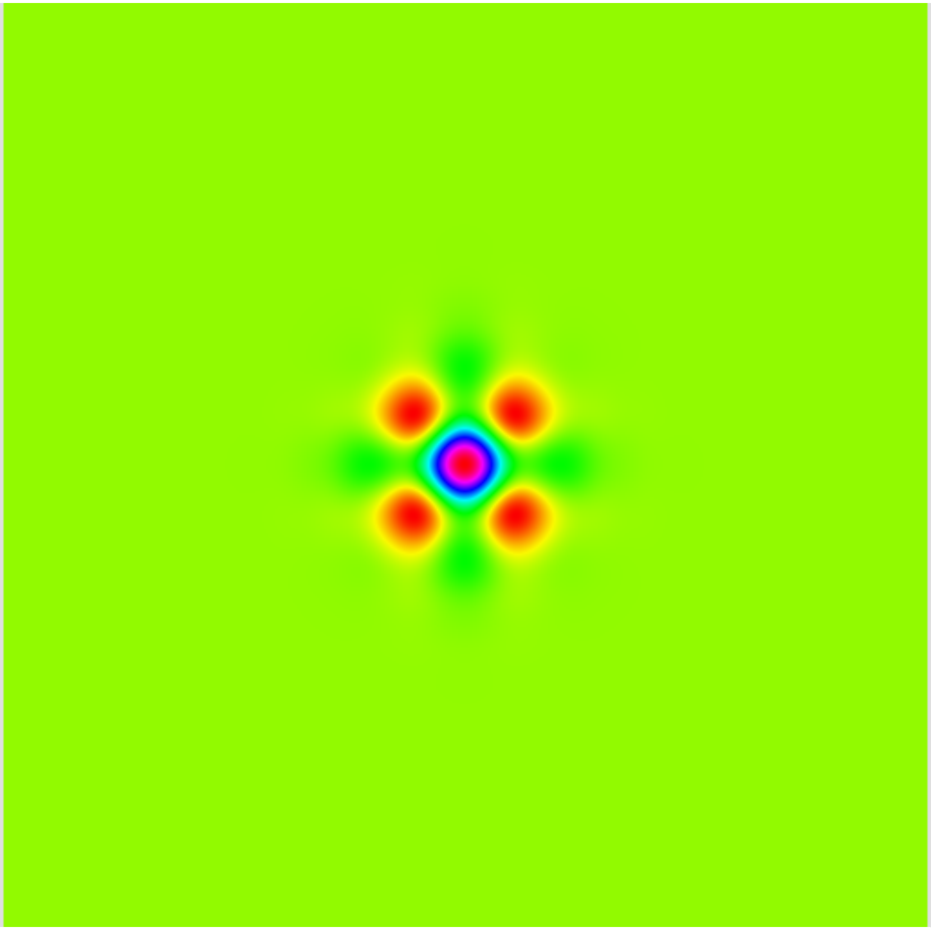


$$u(\mathbf{x})$$

SIFT

Multi-scale response

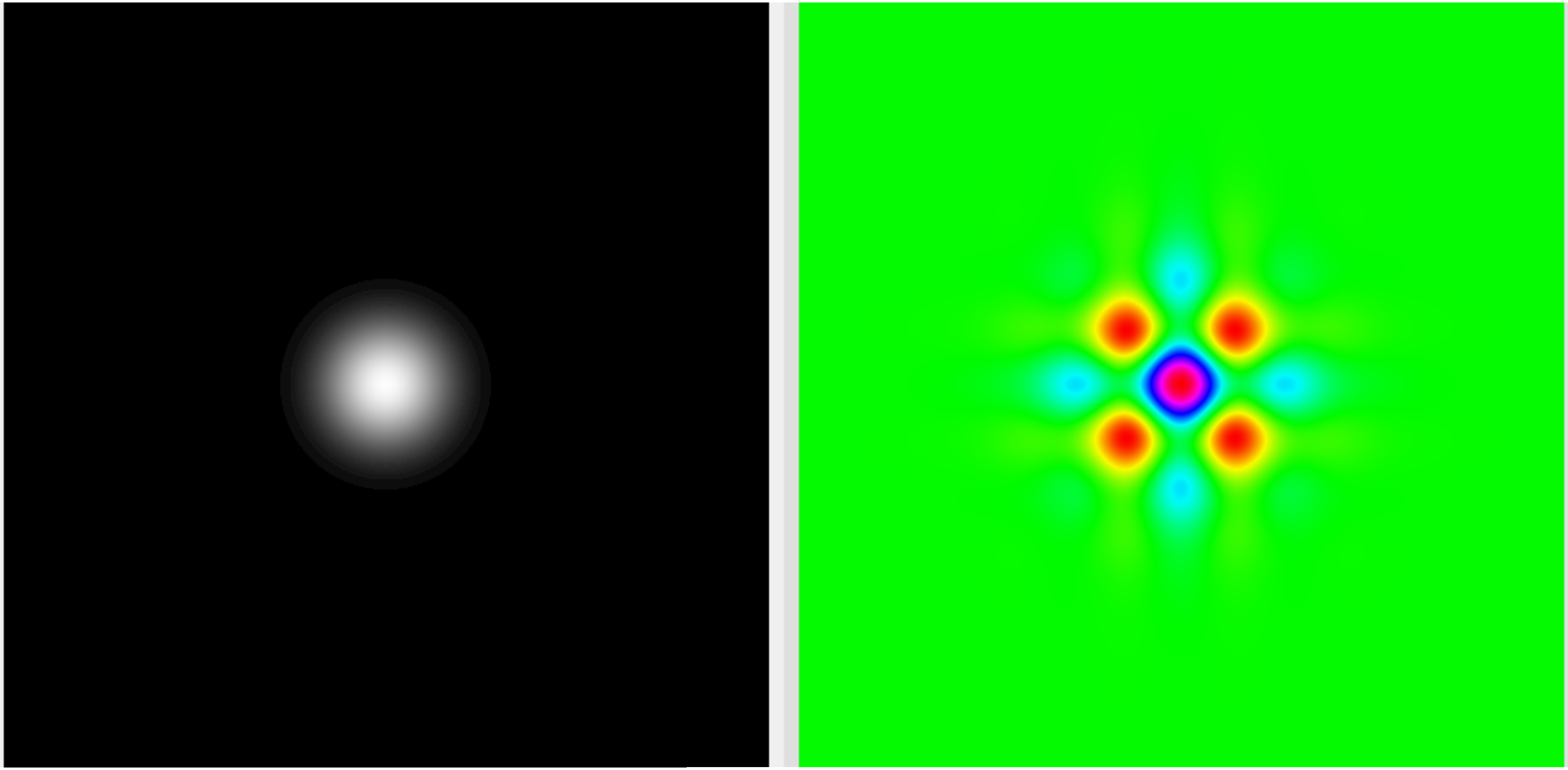




Anisotropic response

$$K_\sigma(x,y) = 2\pi\sigma^2 G_\sigma(x,y)\left(\cos\left(\frac{cx}{\sigma}\right) + \cos\left(\frac{cy}{\sigma}\right)\right)$$

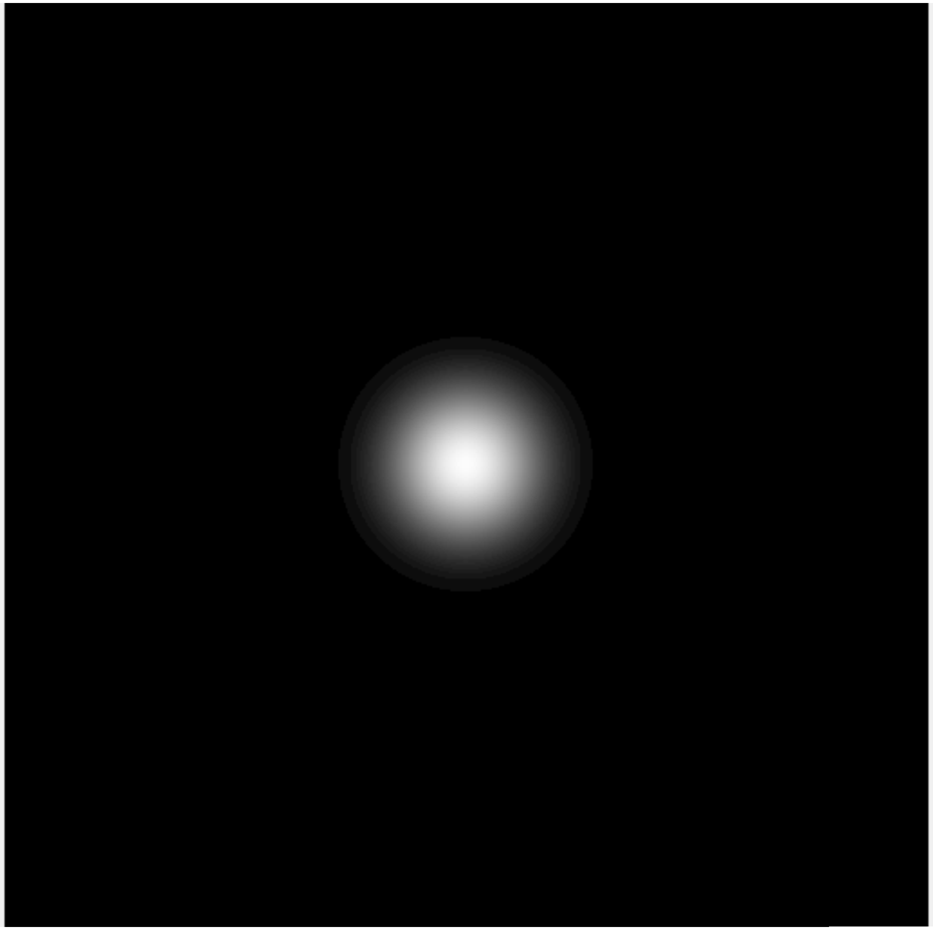
Multi-scale response (Gaussian blob)

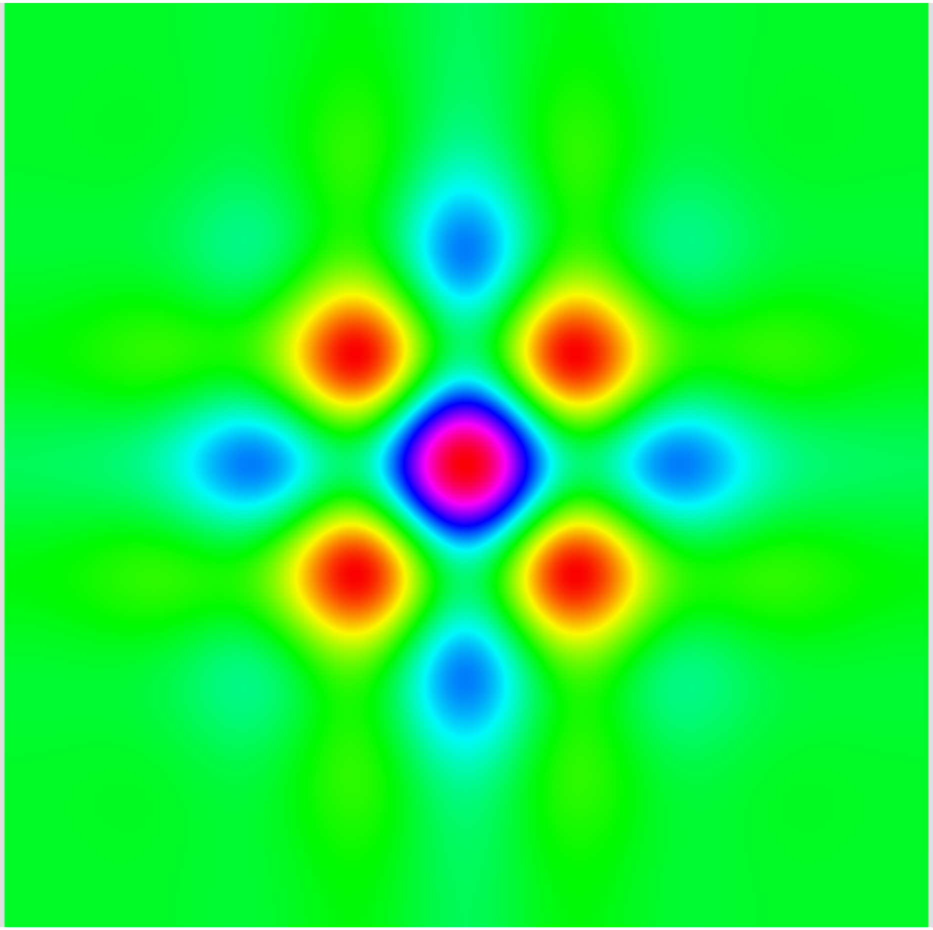


Anisotropic response

$$K_\sigma(x,y) = 2\pi\sigma^2 G_\sigma(x,y)\left(\cos\left(\frac{cx}{\sigma}\right) + \cos\left(\frac{cy}{\sigma}\right)\right)$$

Multi-scale response (Gaussian blob)

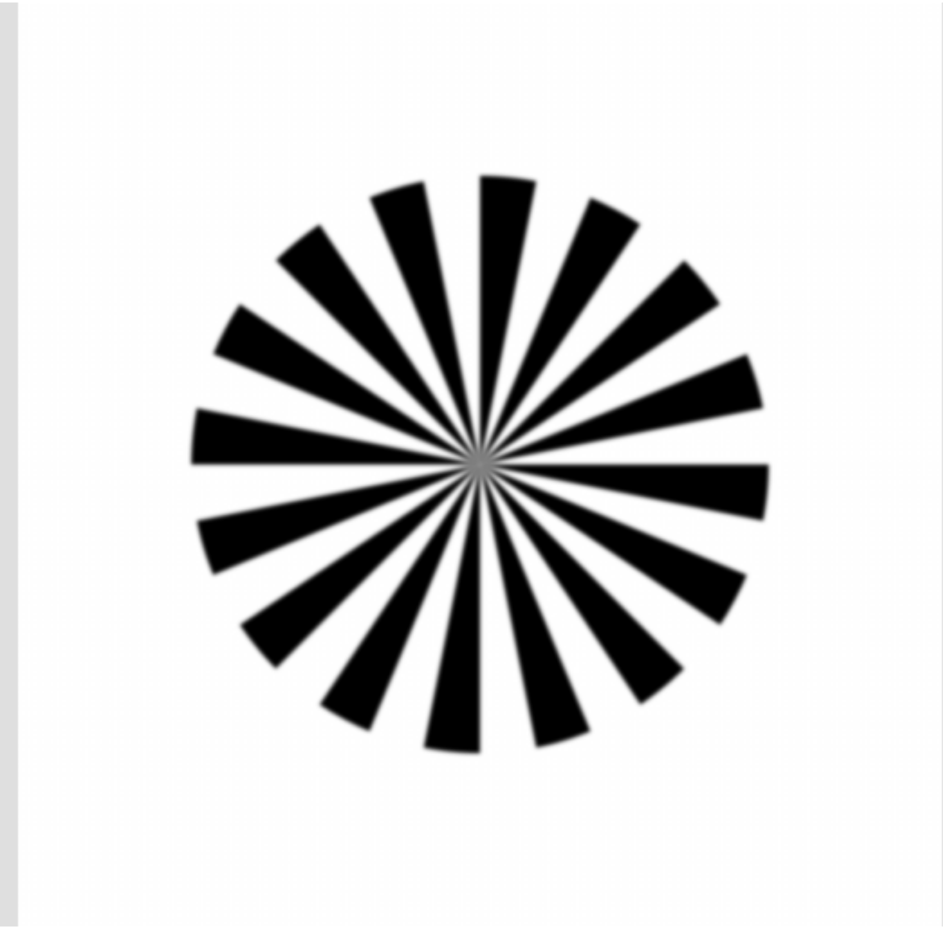




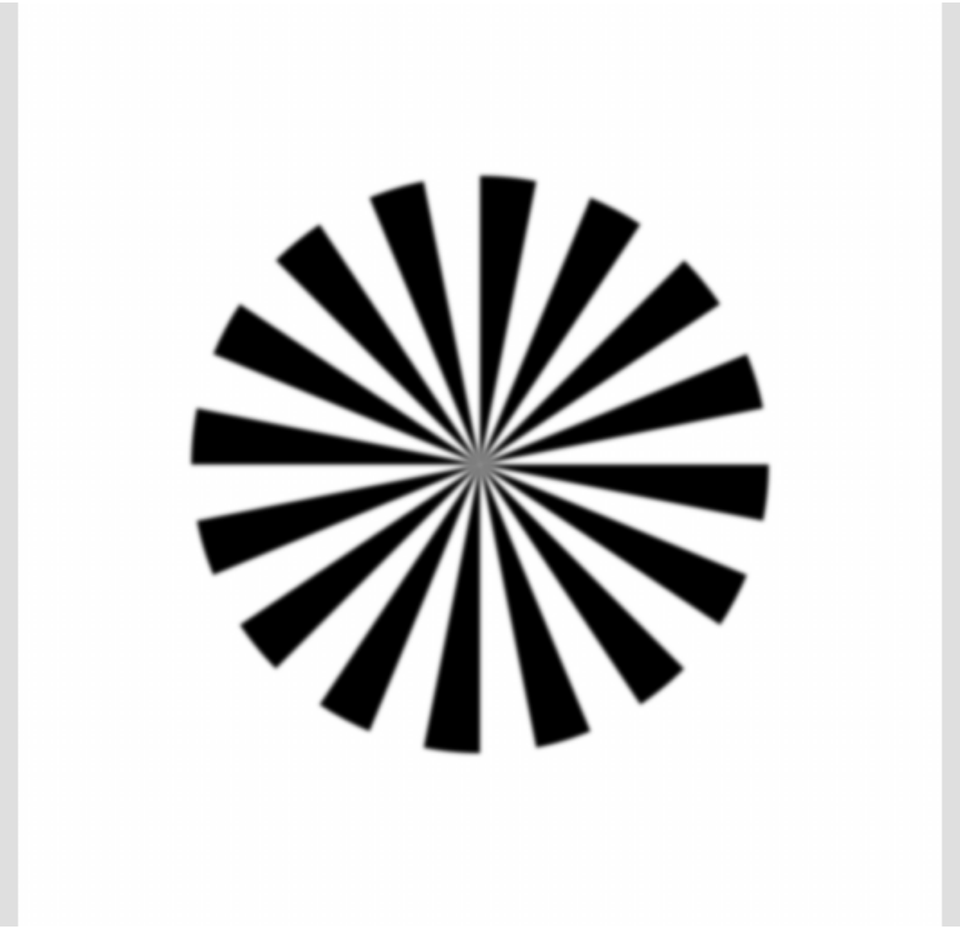
Anisotropic response

$$K_\sigma(x,y) = 2\pi\sigma^2 G_\sigma(x,y)\left(\cos\left(\frac{cx}{\sigma}\right) + \cos\left(\frac{cy}{\sigma}\right)\right)$$

Multi-scale response (siemens star)

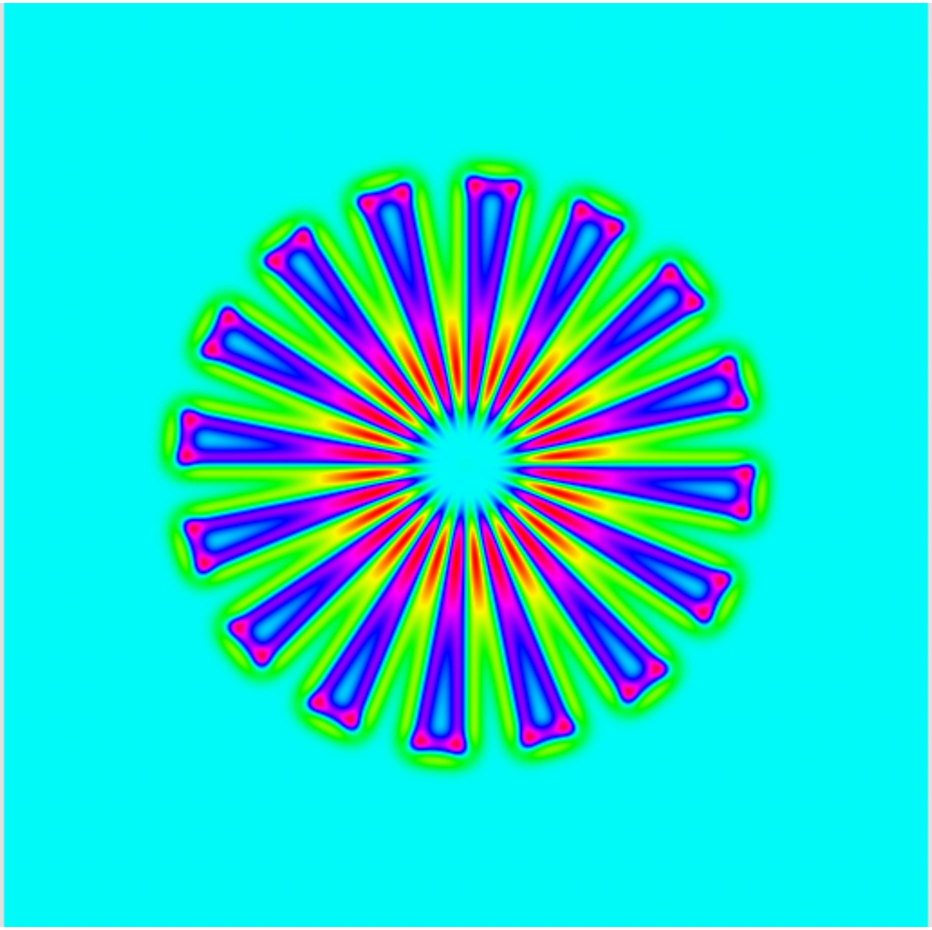


SIFT

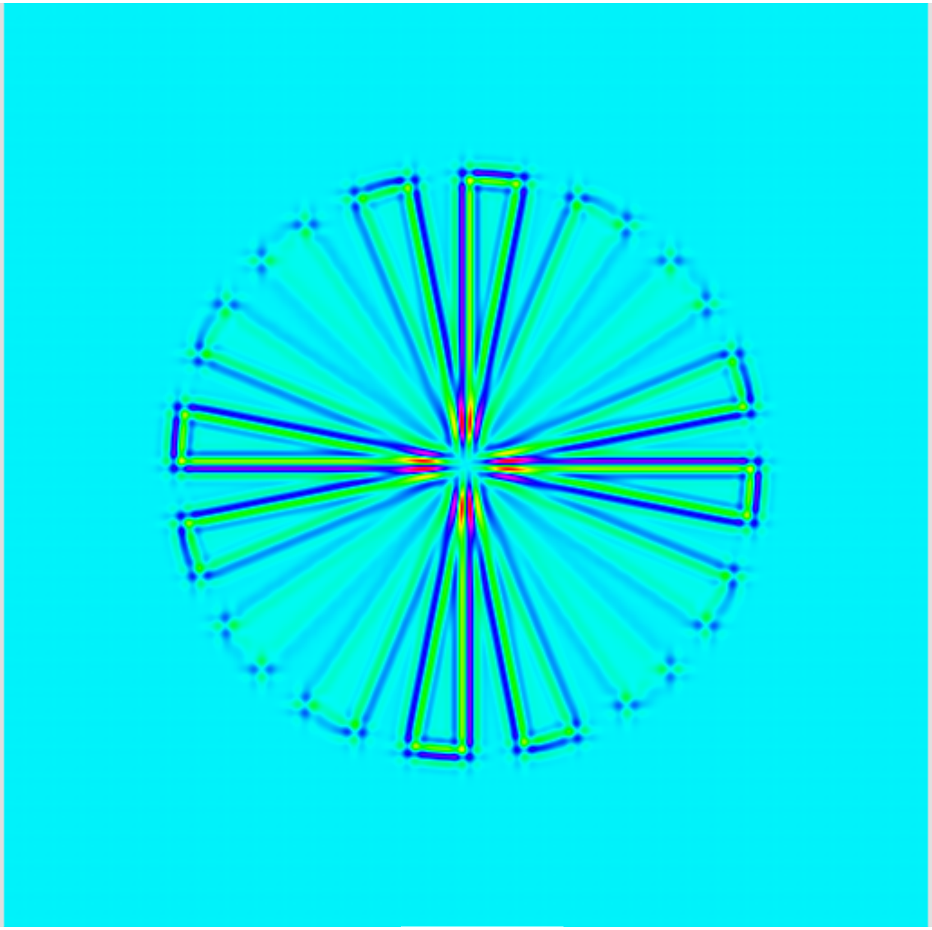


SIFER

Multi-scale response (siemens star)

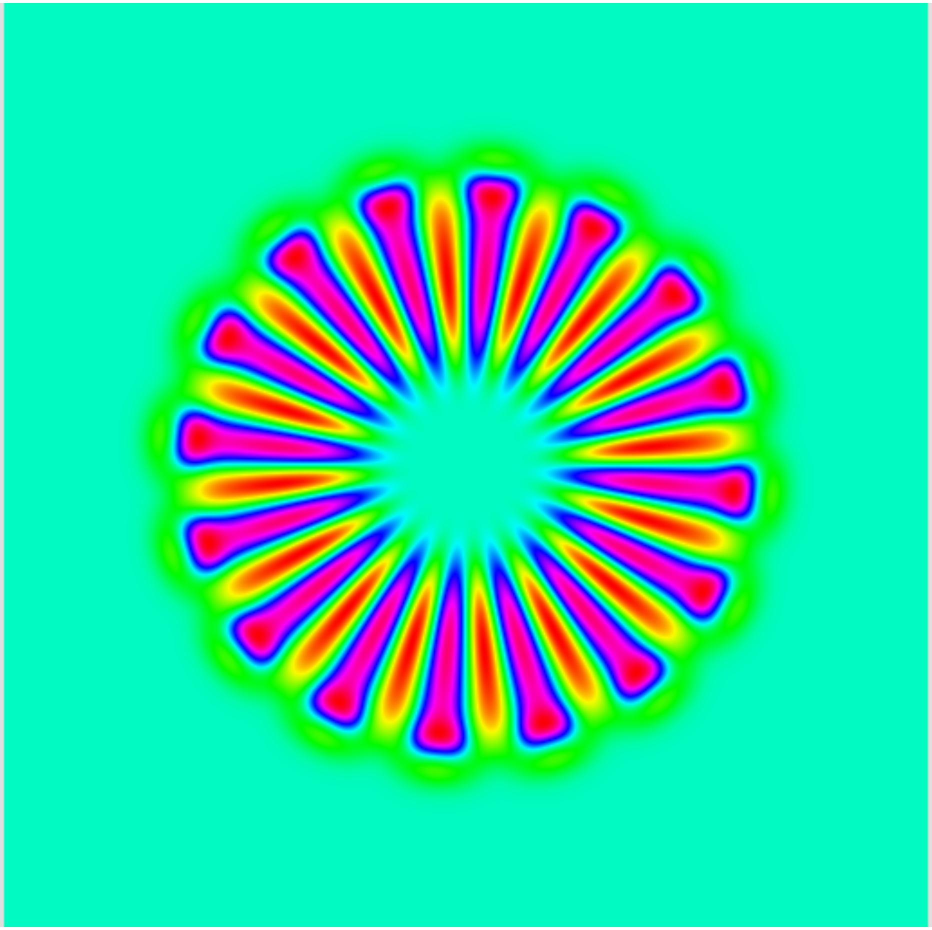


SIFT

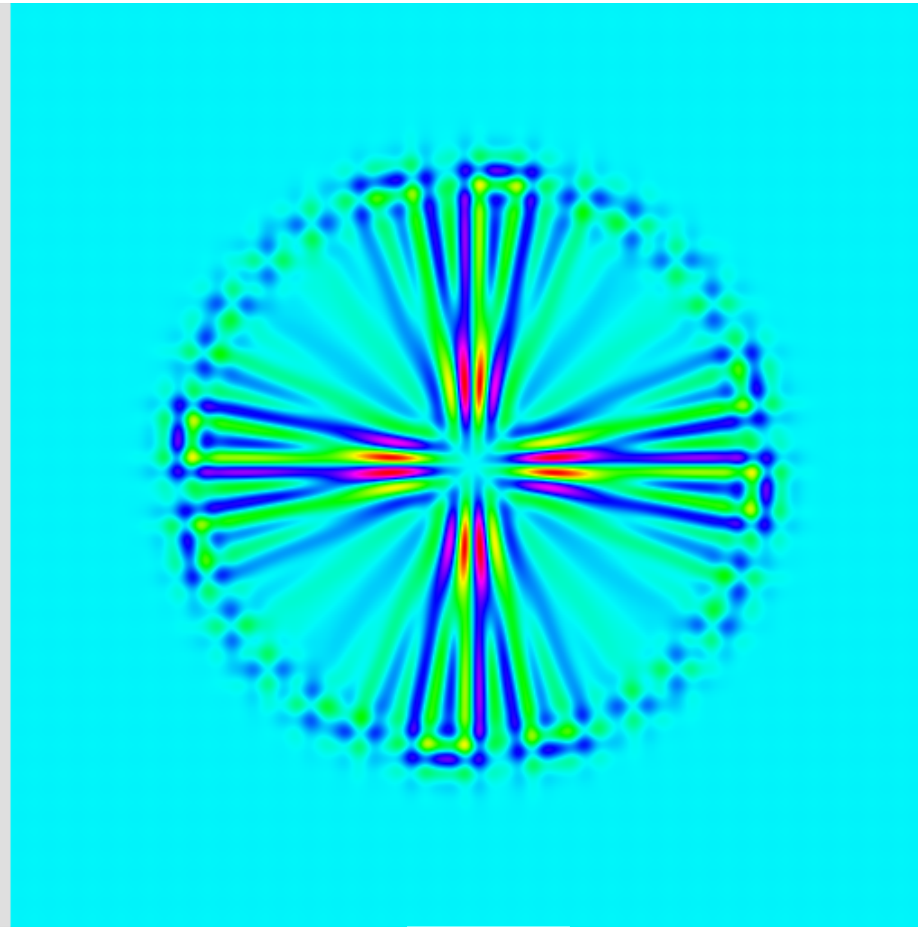


SIFER

Multi-scale response (siemens star)

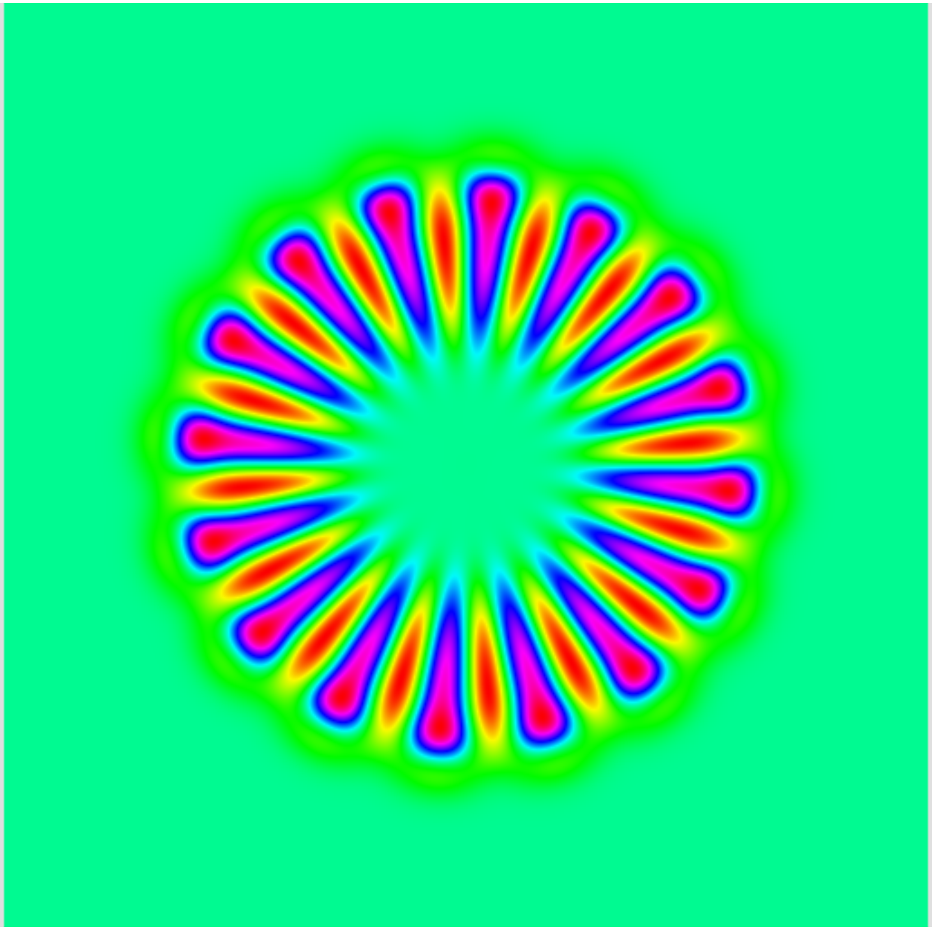


SIFT

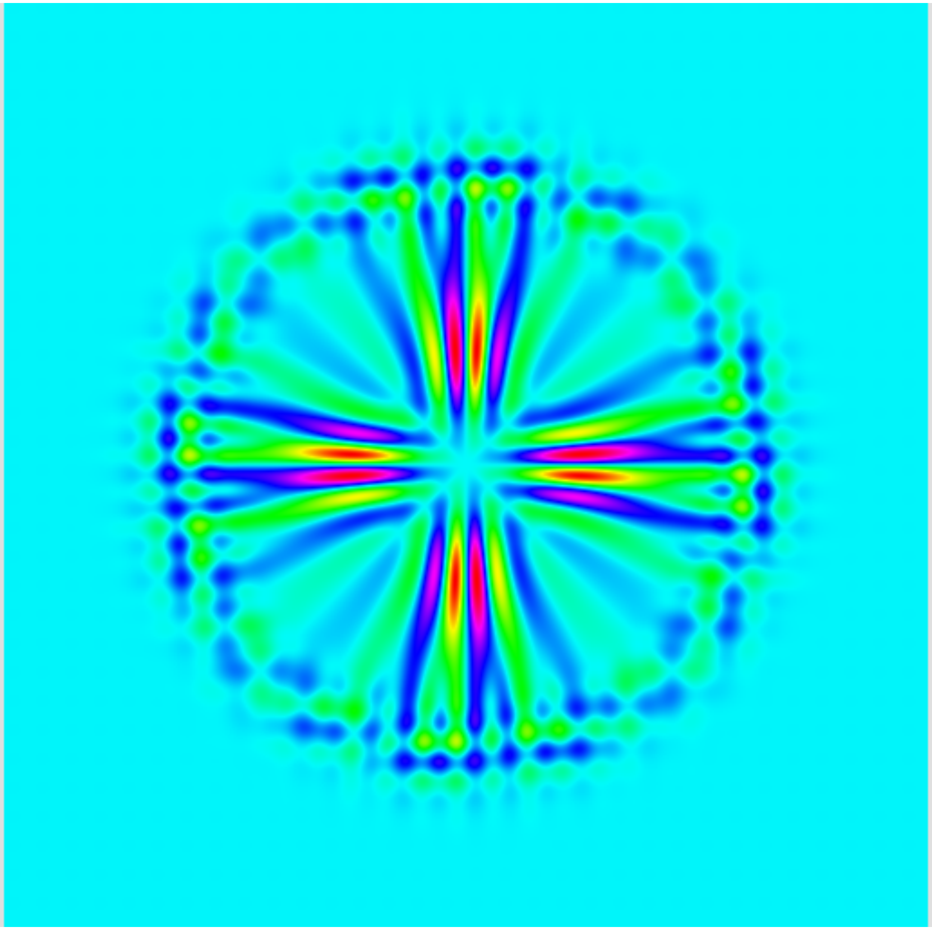


SIFER

Multi-scale response (siemens star)

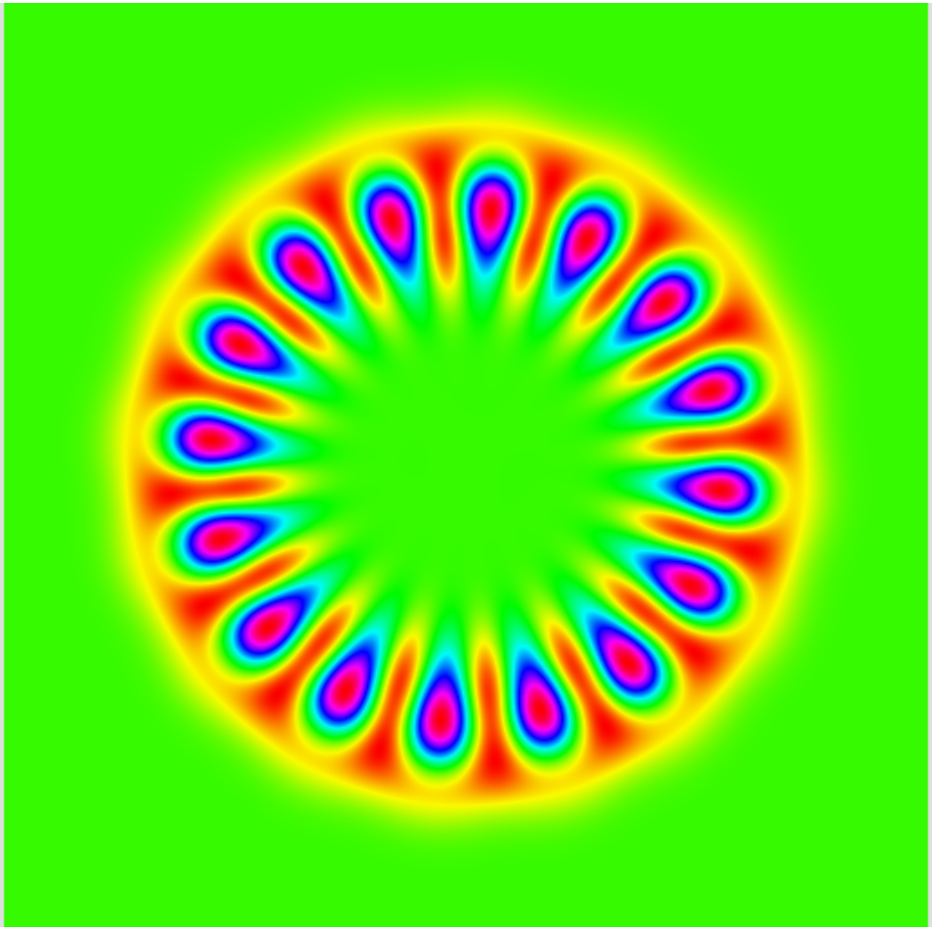


SIFT

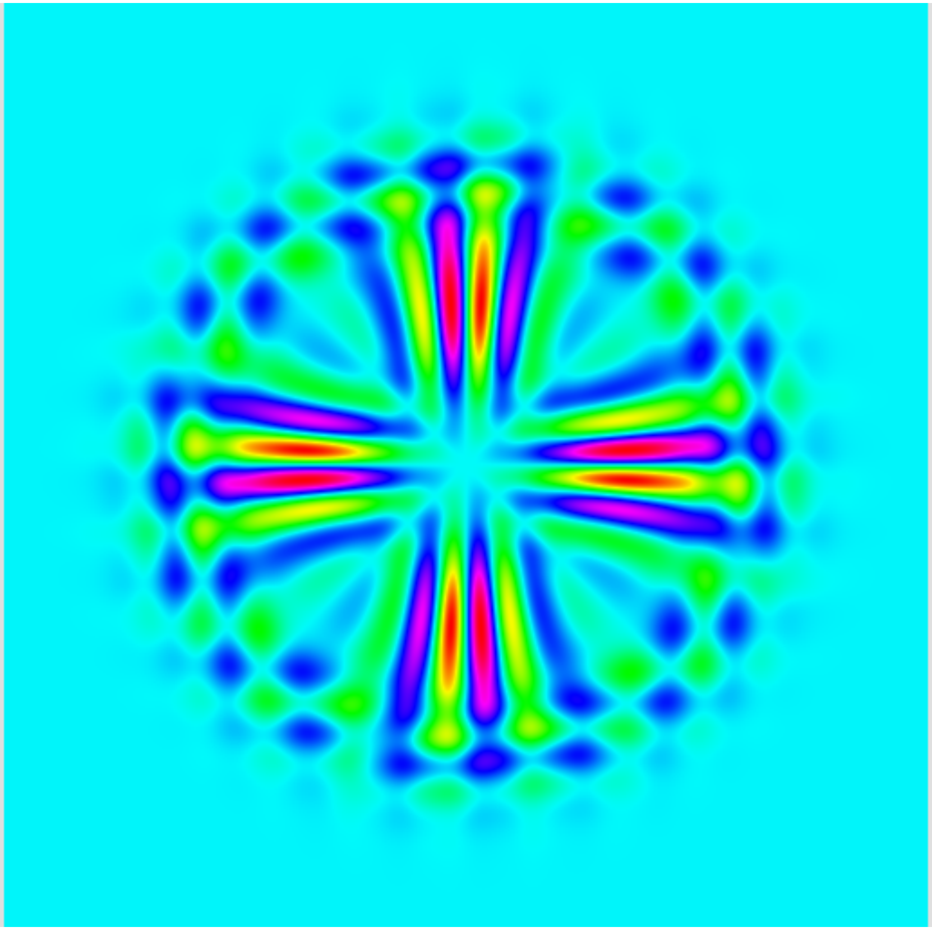


SIFER

Multi-scale response (siemens star)

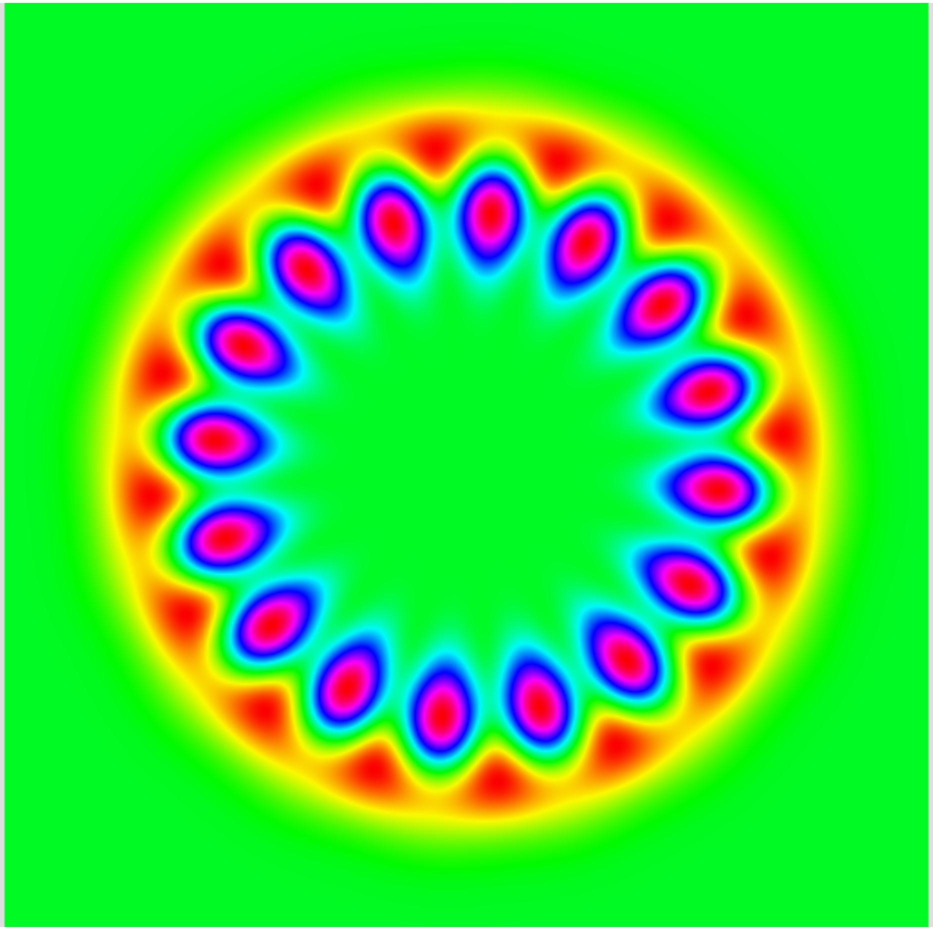


SIFT

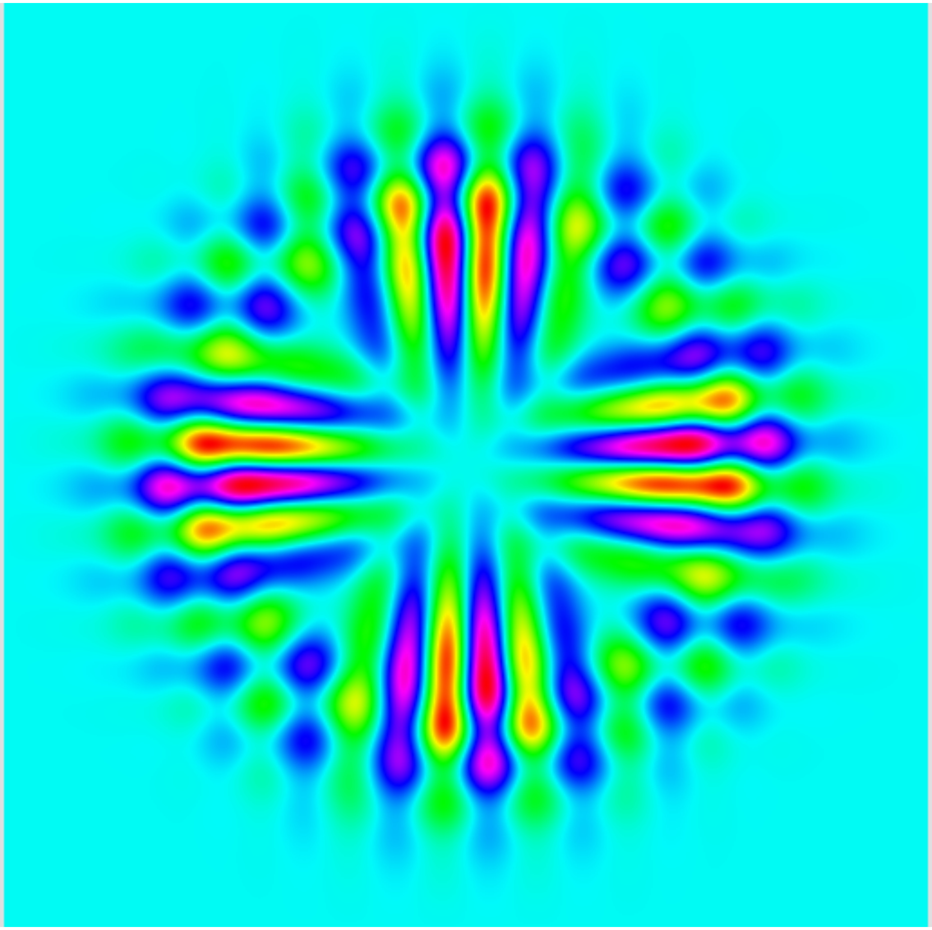


SIFER

Multi-scale response (siemens star)



SIFT



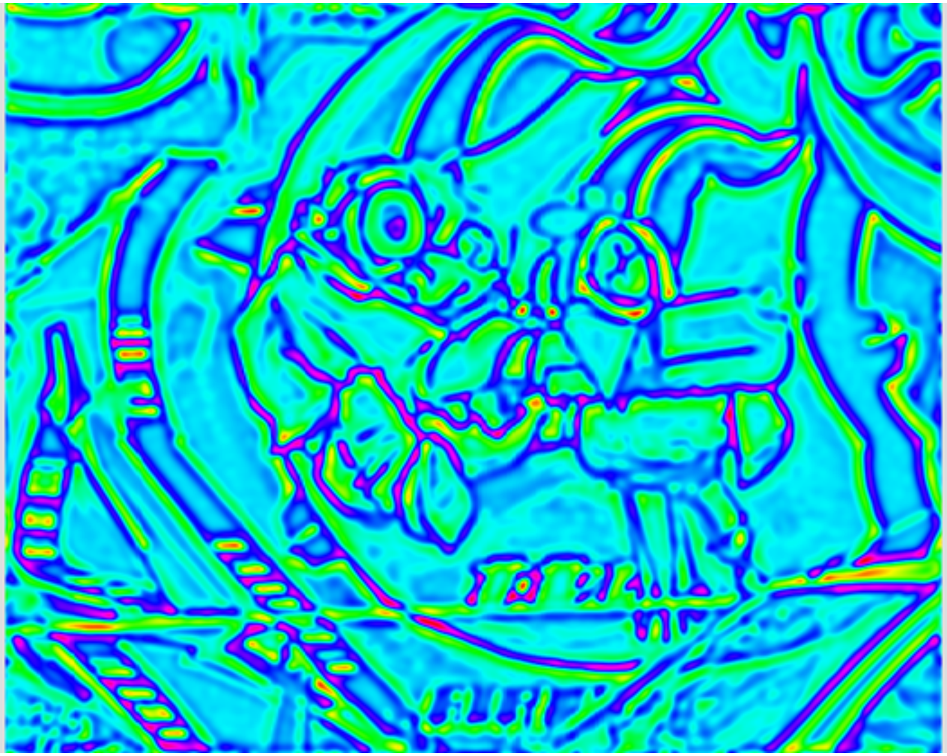
SIFER

Multi-scale response (graf)

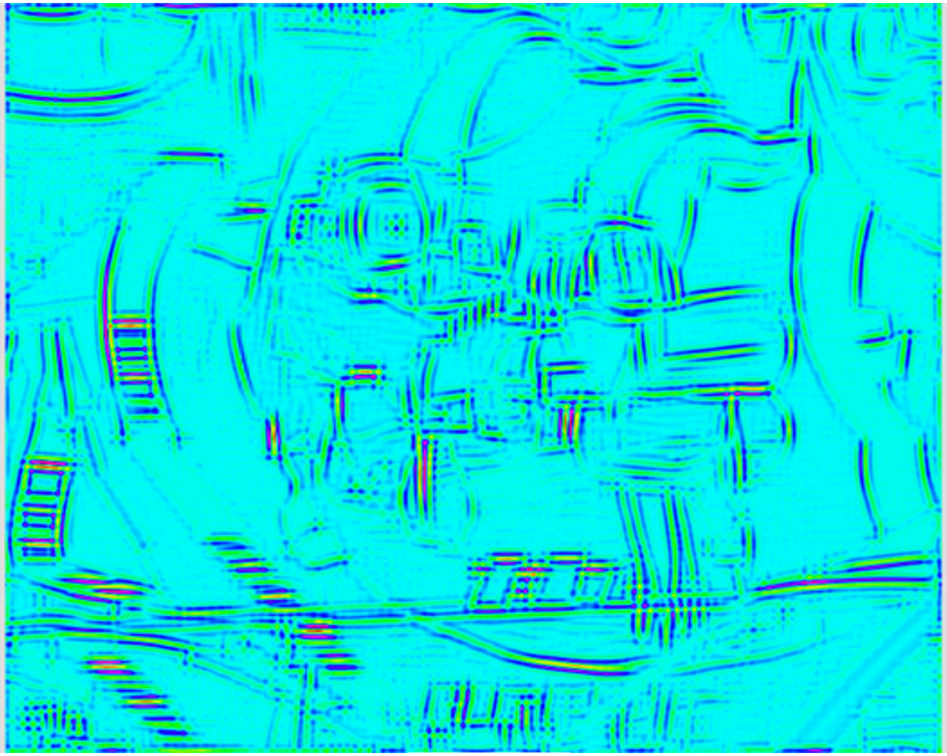
SIFT

SIFER

Multi-scale response (graf)

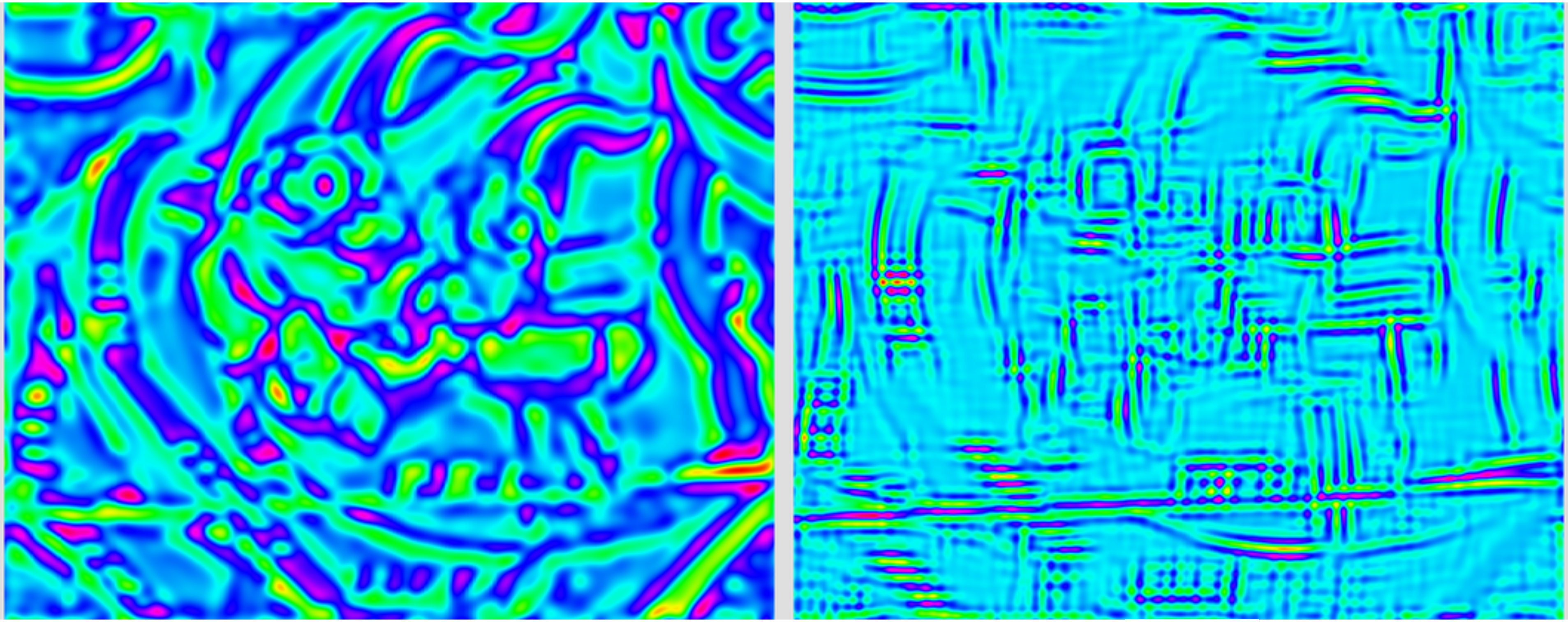


SIFT



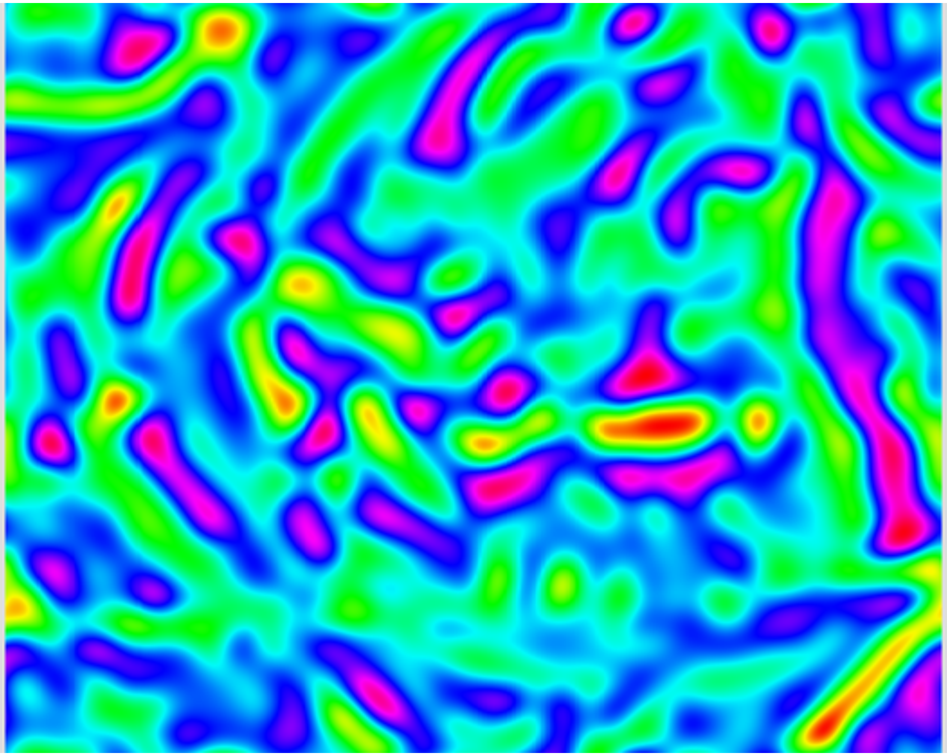
SIFER

Multi-scale response (graf)



SIFT SIFER

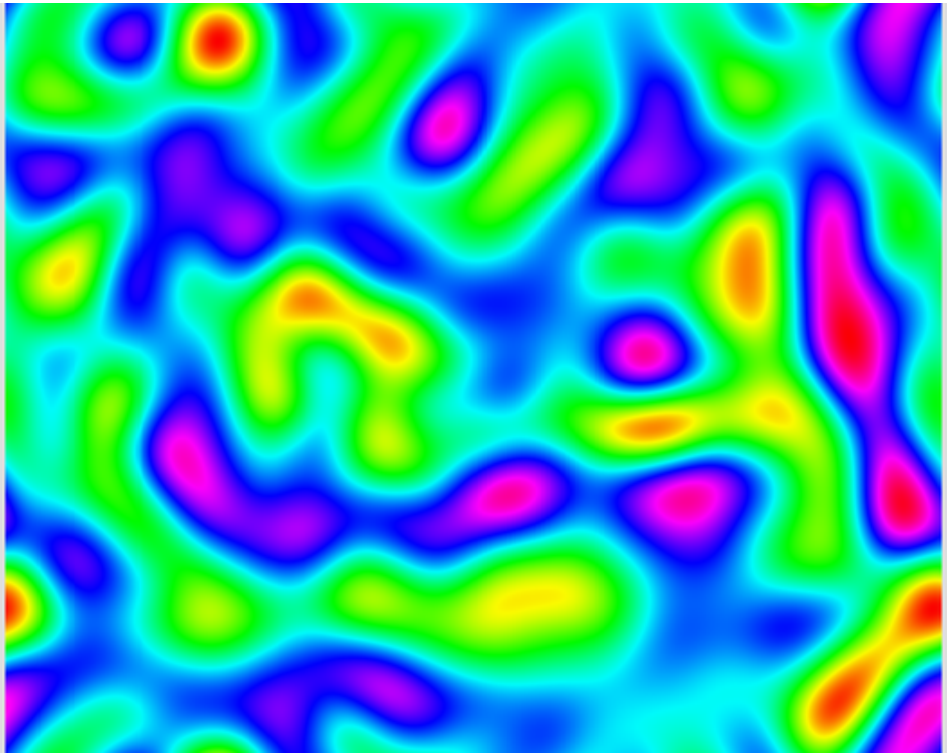
Multi-scale response (graf)



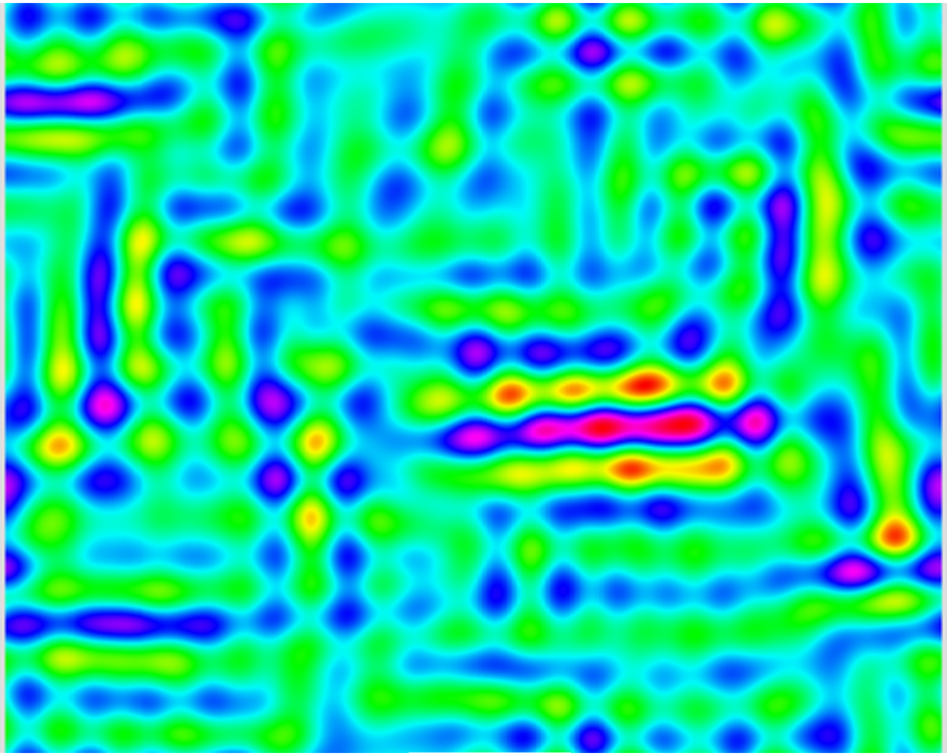
SIFT

SIFER

Multi-scale response (graf)



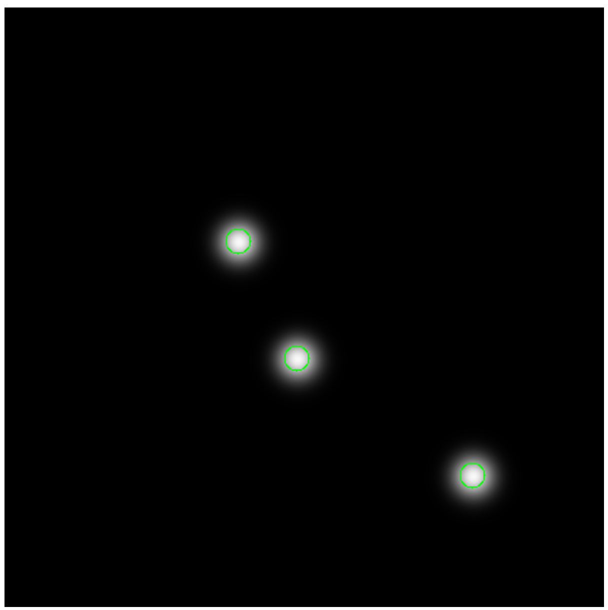
SIFT

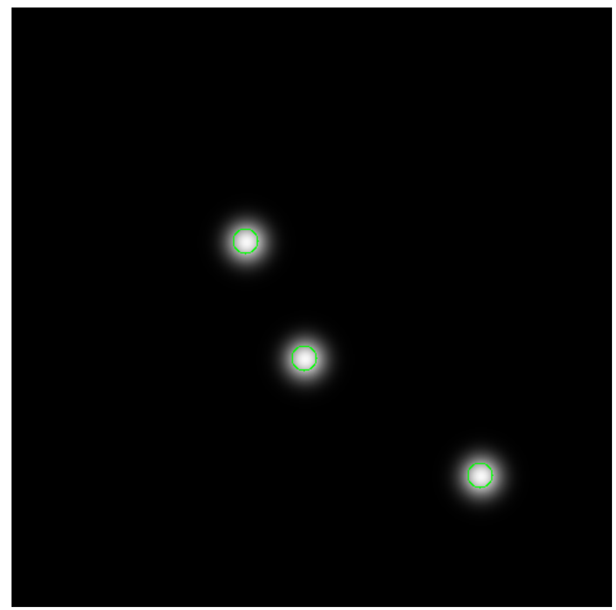


SIFER

Suggested simple correction

$f_k(x, y)$: the Gaussian window extracted for the description of the keypoint $(\sigma_k, x_k, y_k, \theta_k)$.



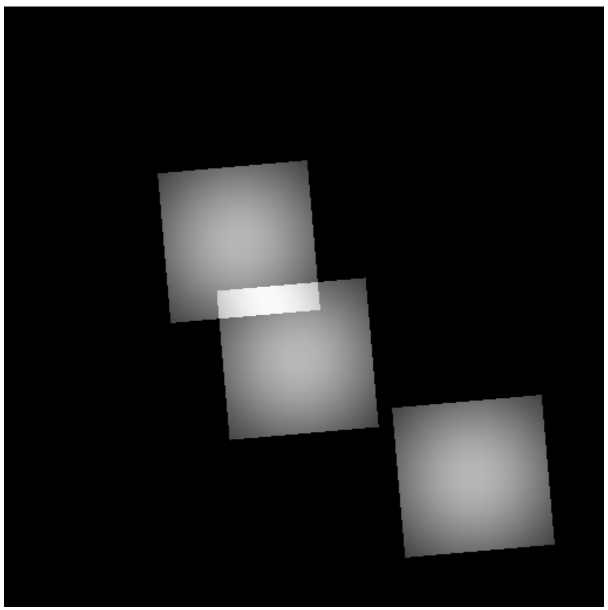


3 detected keypoints

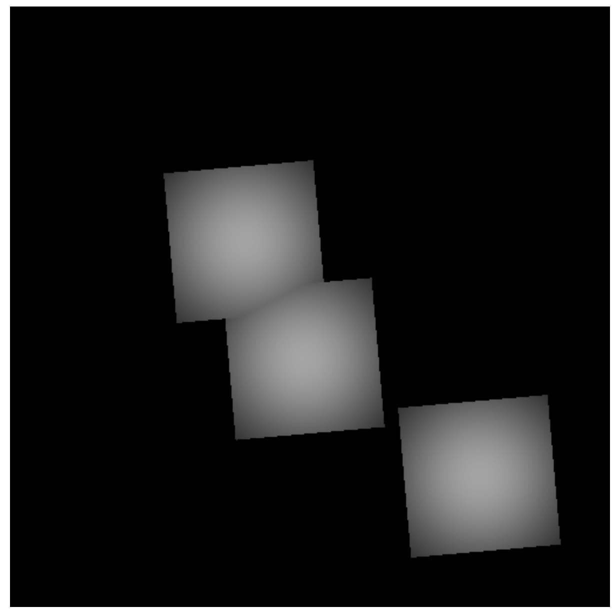
3 detected keypoints

Suggested simple correction

$f_k(x, y)$: the Gaussian window extracted for the description of the keypoint $(\sigma_k, x_k, y_k, \theta_k)$.



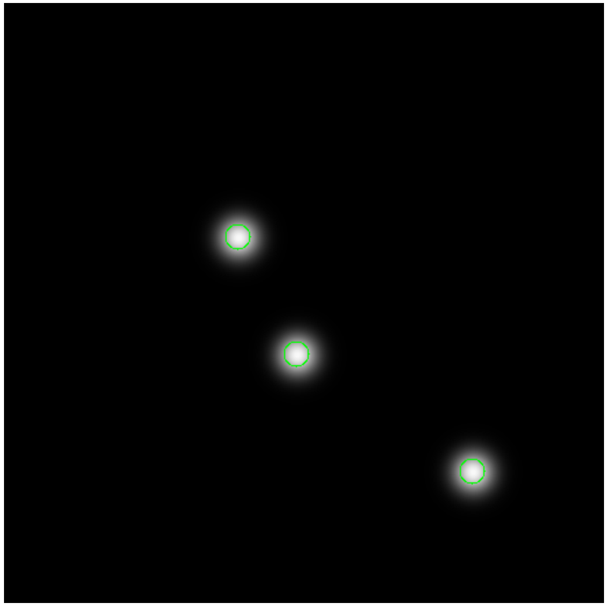
$$\sum_k f_k(x, y)$$

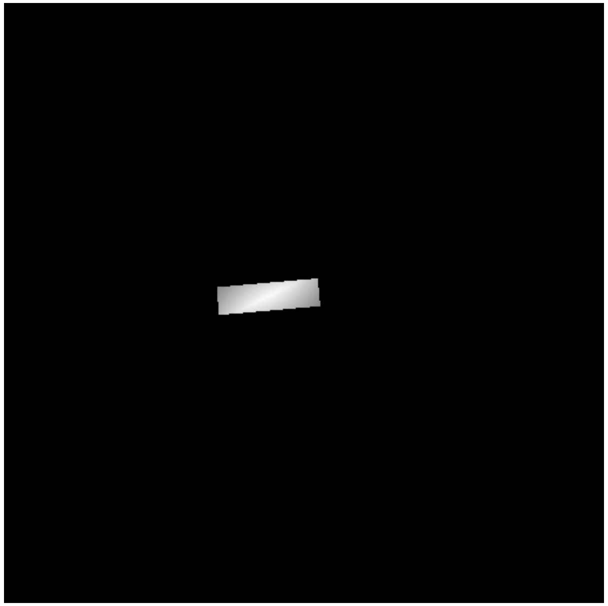


$$\max_k f_k(x, y)$$

Proposed correction of the repeatability criterion

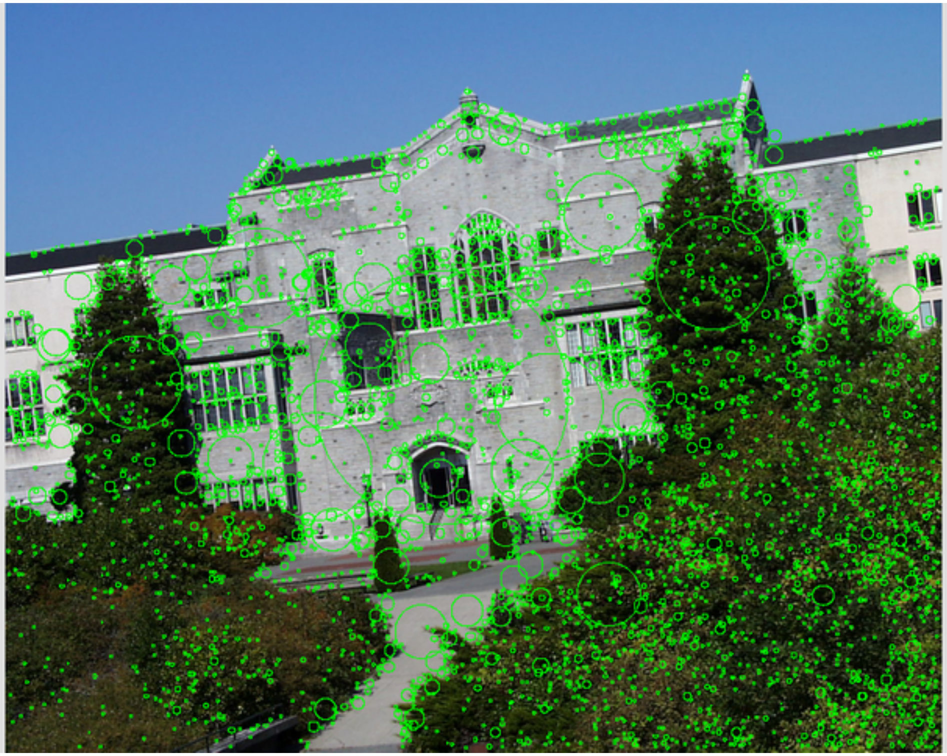
$$\left(\sum_k f_k(x,y) - \max_k f_k(x,y)\right)$$ maps the detections redundancy



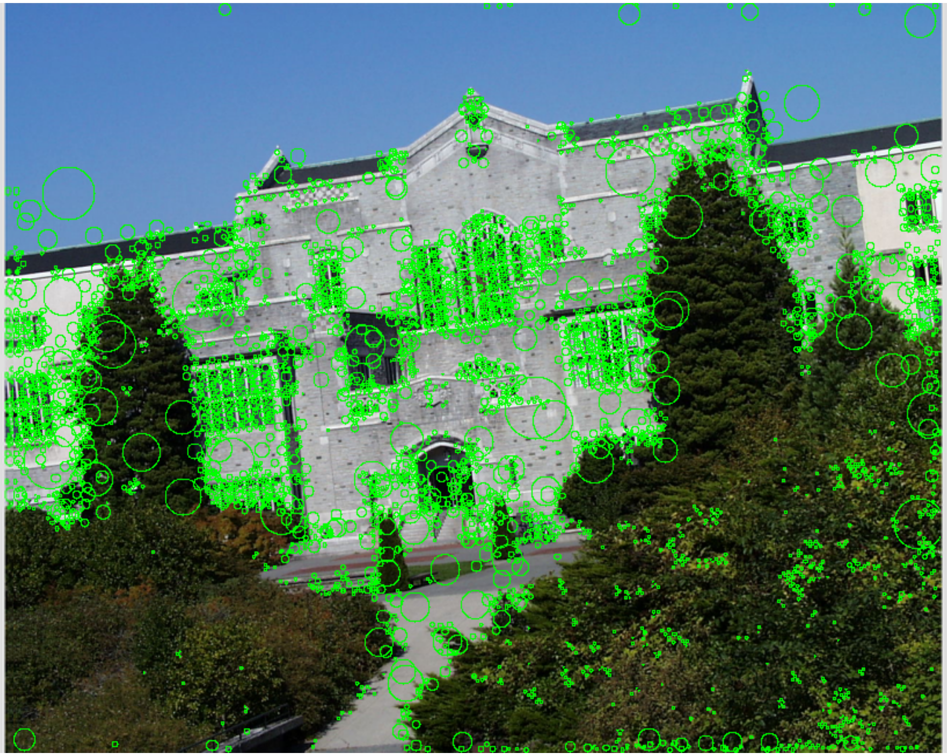


$$\left(\sum_k f_k(x,y) - \max_k f_k(x,y)\right)$$

Proposed correction of the repeatability criterion: Detections overlap



SIFT

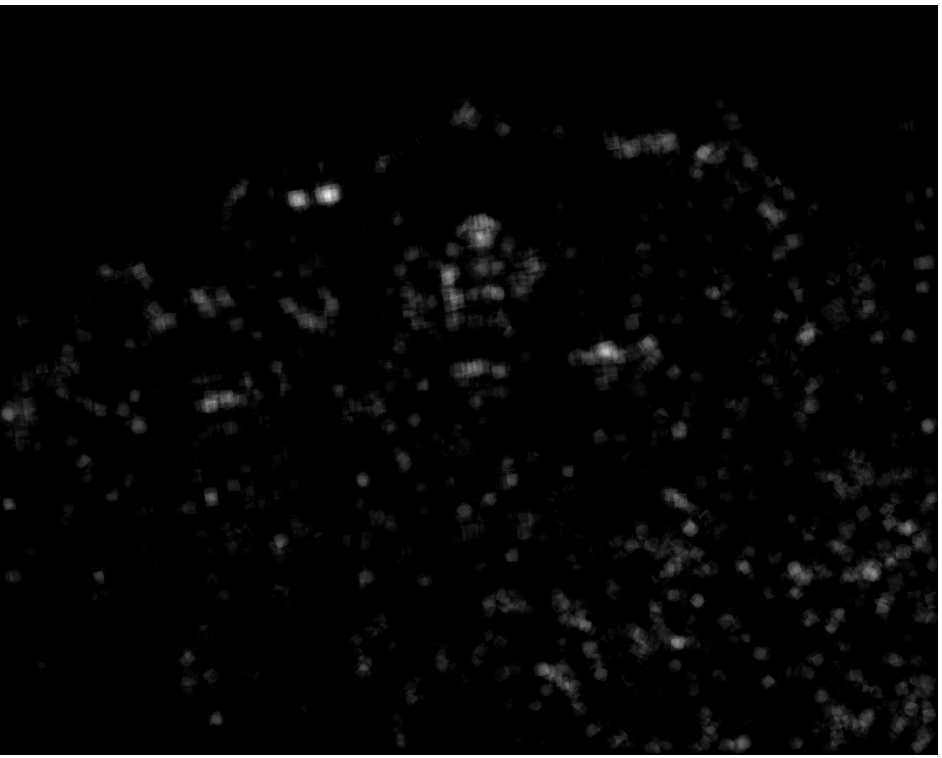


SIFER

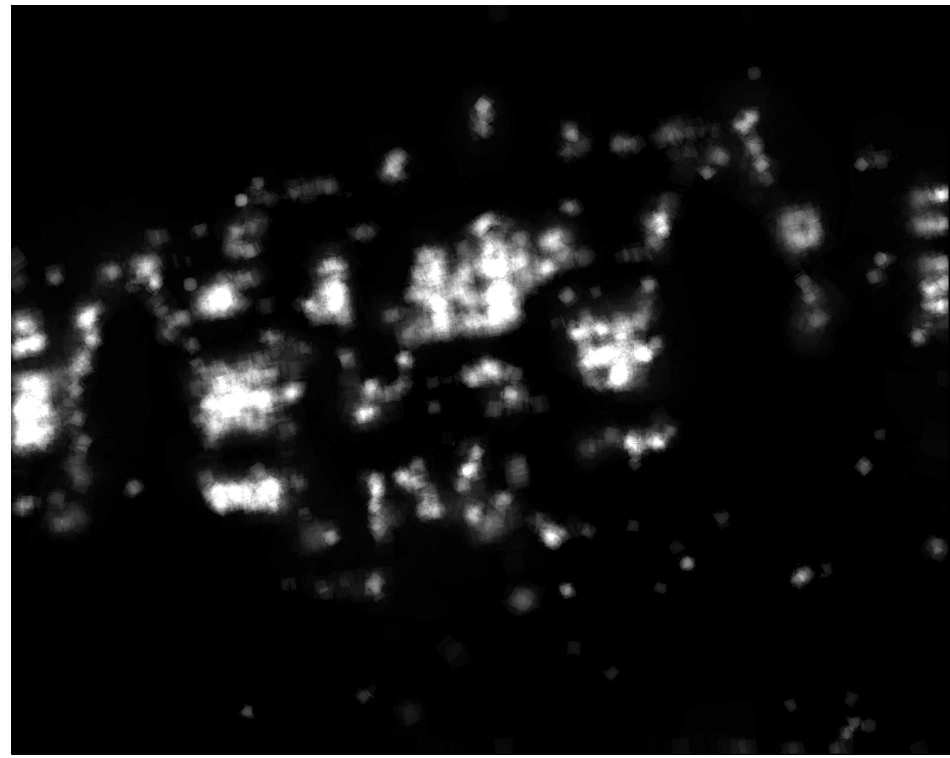
Detection maps

Suggested simple correction

Detections overlap



SIFT



SIFER

$$\left(\sum_k f_k(x,y) - \max_k f_k(x,y)\right)$$

Suggested simple correction

$$\int_\Omega \sum_k f_k(x,y)dxdy = \text{ number of detections}$$

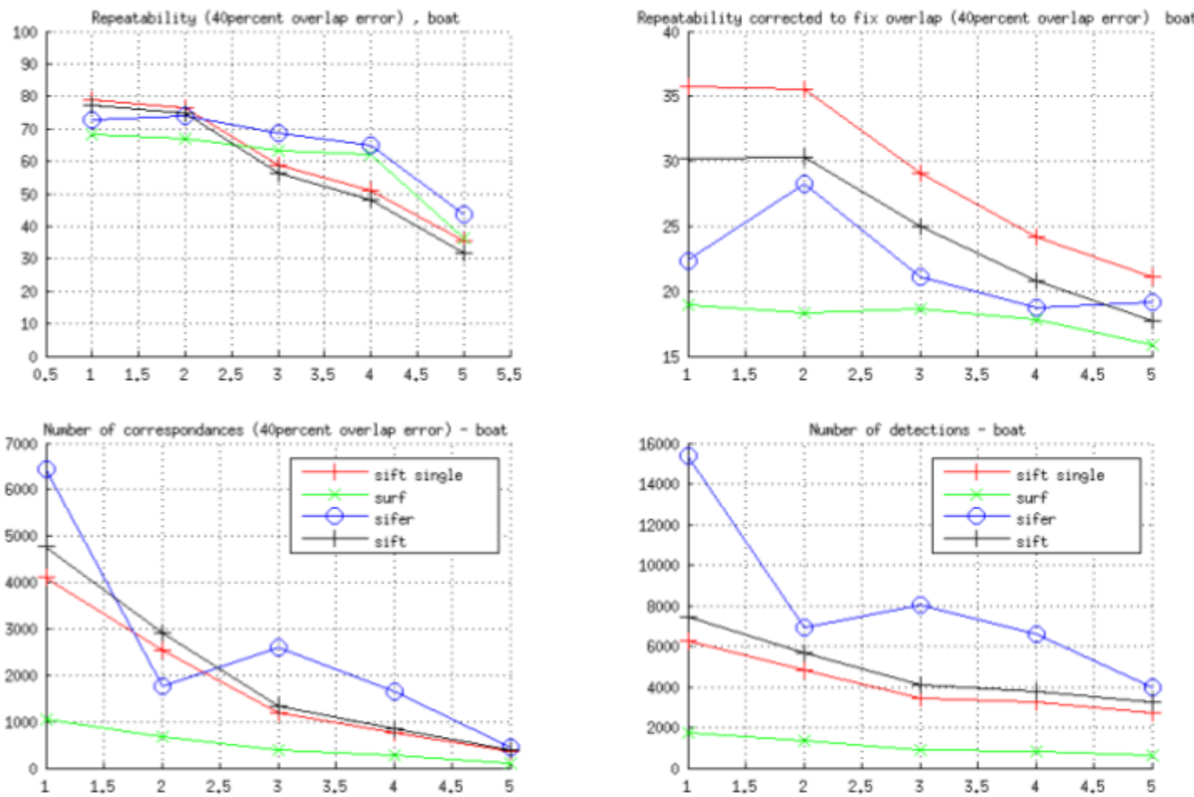
$$\int_\Omega \max_k f_k(x,y)dxdy \approx \text{ number of detetections without overlap}$$

$$\text{repeatability rate} = \frac{\int_\Omega \max_{k \in K_{\text{rep}}} f_k(x,y)dxdy}{\text{total number of detection in use}}$$

K_{rep}: set of repeated keypoints.

The new repeatability curves
SIFT, SURF, SIFER

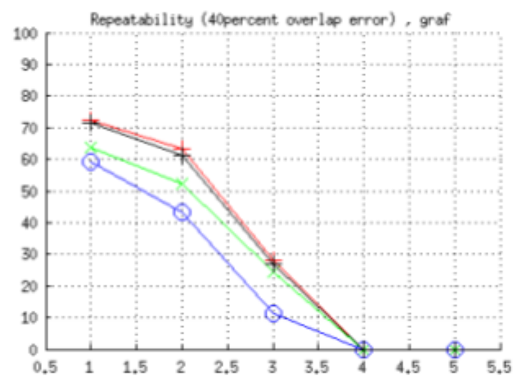
perturbation: rotation and scale

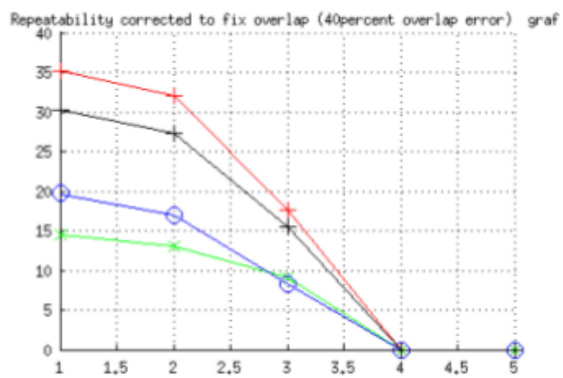


New repeatability curves

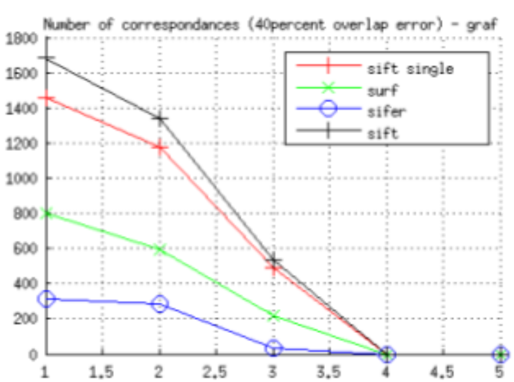
The new repeatability curves
SIFT, SURF, SIFER

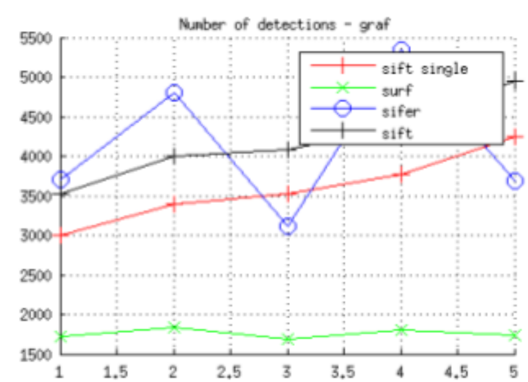
perturbation: tilt





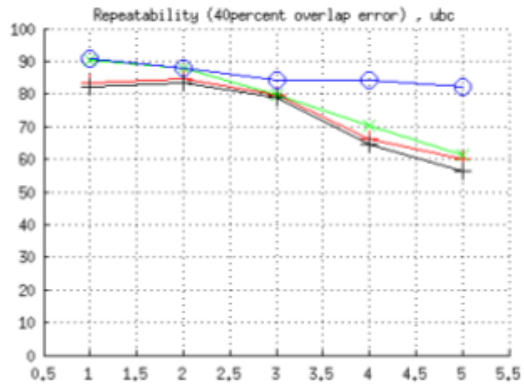
New repeatability curves

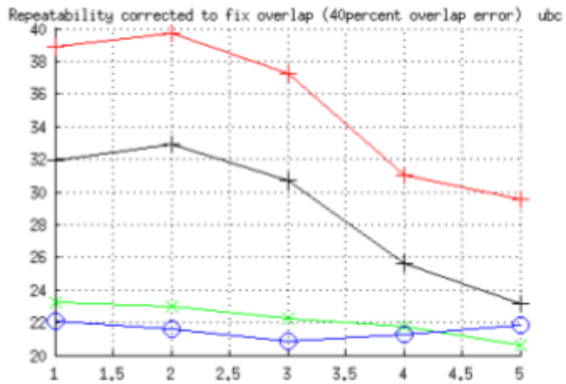


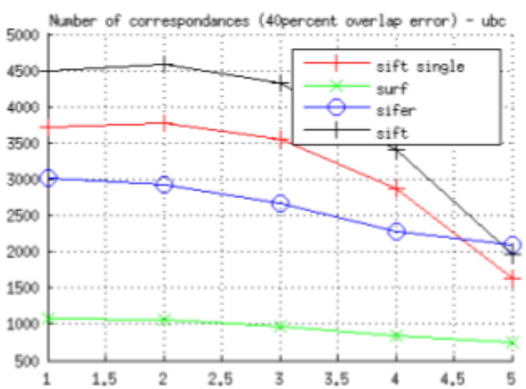


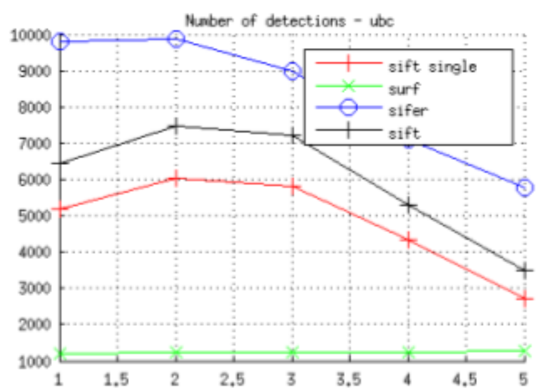
The new repeatability curves SIFT, SURF, SIFER

perturbation: JPG compression









New repeatability curves

Conclusion and open problems

Conclusion: One cannot be satisfied with the proliferation of unprincipled detectors/descriptors. For many of them, the benchmark data demonstrating than they “win” may well be misleading.

Open problem 1:

By simply modifying the parameters of the most invariant method (so far SIFT), one may reach improvements in the performance curves equivalent to those obtained by modifying the detector/descriptor pair. All things equal, we will prefer the really invariant methods.

Open problem 2:

Make the mathematical theory and check if the numerous interest-point/descriptor methods are really scale invariant or not in the Lowe sense. Classify them by their proven invariances.

Open problem 3:

SIFER seems to suggest that scale invariance is not necessary. Thus, all homothety-invariant families of filters are candidates to construct keypoints! If this is true, the chase is open for the best filter family. SIFER is just one of them.

Open problem/proposition 4:

Should we not check repeatability, invariance, and robustness on a short list of reliable patterns where we can also view and discuss the position of interest points?

Open question 5: make requested invariances and benchmarks match!!!

Why are we testing and comparing detectors/descriptors for invariances that they do not have in theory?? (E.g. blur invariance or affine invariance)

Detections maps

SIFT SURF SIFER

Detections maps

SIFT

SURF

SIFER