Region detectors

Requirements for region detection

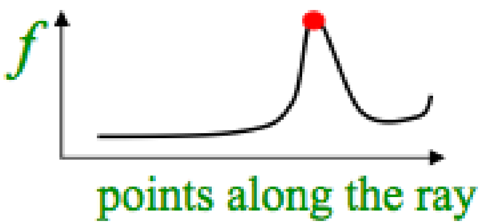
- For region detection invariance transformations that should be considered are *illumination changes, translation*, *rotation, scale* and *full affine transform* (i.e. a region should correspond to the same pre-image for different viewpoints. Viewpoint changes can be locally approximated by affine transform if assuming locally planar objects and orthographic camera, that is perspective effects ignored)

- Region detection should be repeatable and stable, and capable to discriminate between regions

Affine Invariant Intensity Extrema–based

- The algorithm
  - Start from a local intensity extremum point
  - Go in every direction until the point of extremum of some function f. The curve connecting the points is the region boundary
  - Compute geometric moments of orders up to 2 for this region
  - Replace the region with ellipse

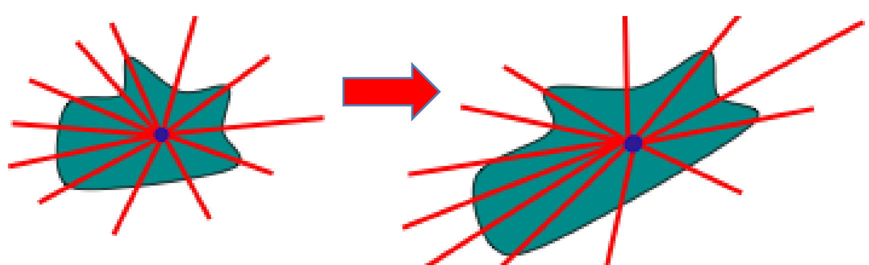
- Detecting extremal regions:
  - detect anchor points (f.e. using Harris detector for corners).
    Anchor points detected at multiple scales are local extremas of intensity
  - explore image around rays from each anchor point. Go along every ray starting from this point until an extremum of function f is reached.



$$f(t) = \frac{|I(t) - I_0|}{\frac{1}{t}\int_0^t |I(t) - I_0|\, dt}$$

f characteristic function of the region
(1 inside, 0 outside)

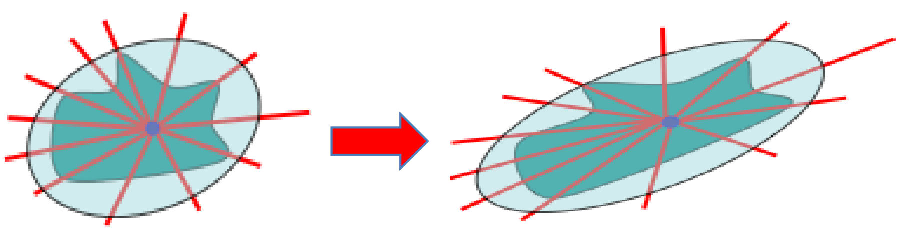
  - all points create some irregularly-shaped region. Approximately corresponding regions are obtained for affine-transformed regions



- Approximating the region
  - Being function f the characteristic function of the region, moments up to 2^{nd} order allow to approximate the region with an ellipse (the ellipse has the same 2^{nd} order moments as the region)

$$m_{pq} = \int x^p y^q f(x,y)\, dxdy$$

  - The ellipse of an affine-transformed region corresponds to the ellipse of the original region under the same transformation



$$q = Ap$$

$p = [x, y]^T$ relative to the center of mass

$$\Sigma_2 = A\Sigma_1 A^T$$

Σ covariance matrix of region points

  - As regions are illumination based warp ellipse to circle for affine invariance

MSER (Maximally Stable Extremal Regions)

- MSER is a method for blob detection in images. The MSER algorithm extracts from an image a number of co-variant regions, called MSERs: an MSER is a *stable connected component of some gray-level sets of the image* .

- MSER is based on the idea of taking regions which stay nearly the same through a wide range of thresholds.
  - All the pixels below a given threshold are *white* and all those above or equal are *black*.
  - If we are shown a sequence of thresholded images I_t with frame t corresponding to threshold t, we would see first a *black* image, then *white* spots corresponding to local intensity minima will appear then grow larger.
  - These *white* spots will eventually merge, until the whole image is *white*.
  - The set of all connected components in the sequence is the set of all extremal regions.
- Optionally, elliptical frames are attached to the MSERs by fitting ellipses to the regions. Those regions descriptors are kept as features

- The word *extremal* refers to the property that all pixels inside the MSER have either higher (bright extremal regions) or lower (dark extremal regions) intensity than all the pixels on its outer boundary

- This operation can be performed by first sorting all pixels by gray value and then incrementally adding pixels to each connected component as the threshold is changed. The area is monitored. Regions such that their variation wrt the threshold is minimal are defined maximally stable:
  - Let's make all the pixels below a threshold white. The others black
  - Considering a sequence of thresholded images with increasing thresholds sweeping from black to white we pass from a black image to images where white blobs appear and grow larger by merging, up to the final image.
  - Over a large range of thresholds the local binarization is stable and shows some invariance to affine transformation of image intensities and scaling

MSER processing

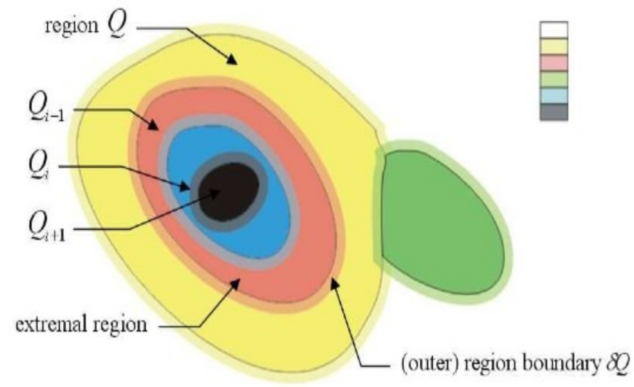
The MSER extraction implements the following steps:

- Sweep threshold of intensity from black to white, performing a simple luminance thresholding of the image
- Extract connected components ("*Extremal Regions*")
- Find a threshold when an extremal region is "Maximally Stable", i.e. local minimum of the relative growth of its square. Due to the discrete nature of the image, the region below / above may be coincident with the actual region, in which case the region is still deemed maximal.
- Approximate a region with an ellipse (*this step is optional*)
- Keep those regions descriptors as features

However, even if an extremal region is maximally stable, it might be rejected if:

- it is too big (there is a parameter MaxArea);
- it is too small (there is a parameter MinArea);
- it is too unstable (there is a parameter MaxVariation);
- it is too similar to its parent MSER

Margin = the number of thresholds for which the region is stable



$$Q_{i^*} : i^* = \arg\min_i |Q_{i+\Delta} \setminus Q_{i-\Delta}| / |Q_i|$$

Math. Details

Image I is a mapping $I : \mathcal{D} \subset \mathbb{Z}^2 \rightarrow \mathcal{S}$. Extremal regions are well defined on images if:

1. $\mathcal{S}$ is totally ordered, i.e. reflexive, antisymmetric and transitive binary relation $\leq$ exists. In this paper only $S = \{0, 1, \ldots, 255\}$ is considered, but extremal regions can be defined on e.g. real-valued images ($\mathcal{S} = R$).
2. An adjacency (neighbourhood) relation $A \subset \mathcal{D} \times \mathcal{D}$ is defined. In this paper 4-neighbourhoods are used, i.e. $p, q \in \mathcal{D}$ are adjacent (pAq) iff $\sum_{i=1}^{d} |p_i - q_i| \leq 1$.

Region $\mathcal{Q}$ is a contiguous subset of $\mathcal{D}$, i.e. for each $p, q \in \mathcal{Q}$ there is a sequence $p, a_1, a_2, \ldots, a_n, q$ and $pAa_1, a_iAa_{i+1}, a_nAq$.

(Outer) Region Boundary $\partial\mathcal{Q} = \{q \in \mathcal{D} \setminus \mathcal{Q} : \exists p \in \mathcal{Q} : qAp\}$, i.e. the boundary $\partial\mathcal{Q}$ of $\mathcal{Q}$ is the set of pixels being adjacent to at least one pixel of $\mathcal{Q}$ but not belonging to $\mathcal{Q}$.

Extremal Region $\mathcal{Q} \subset D$ is a region such that for all $p \in \mathcal{Q}, q \in \partial\mathcal{Q} : I(p) > I(q)$ (maximum intensity region) or $I(p) < I(q)$ (minimum intensity region).

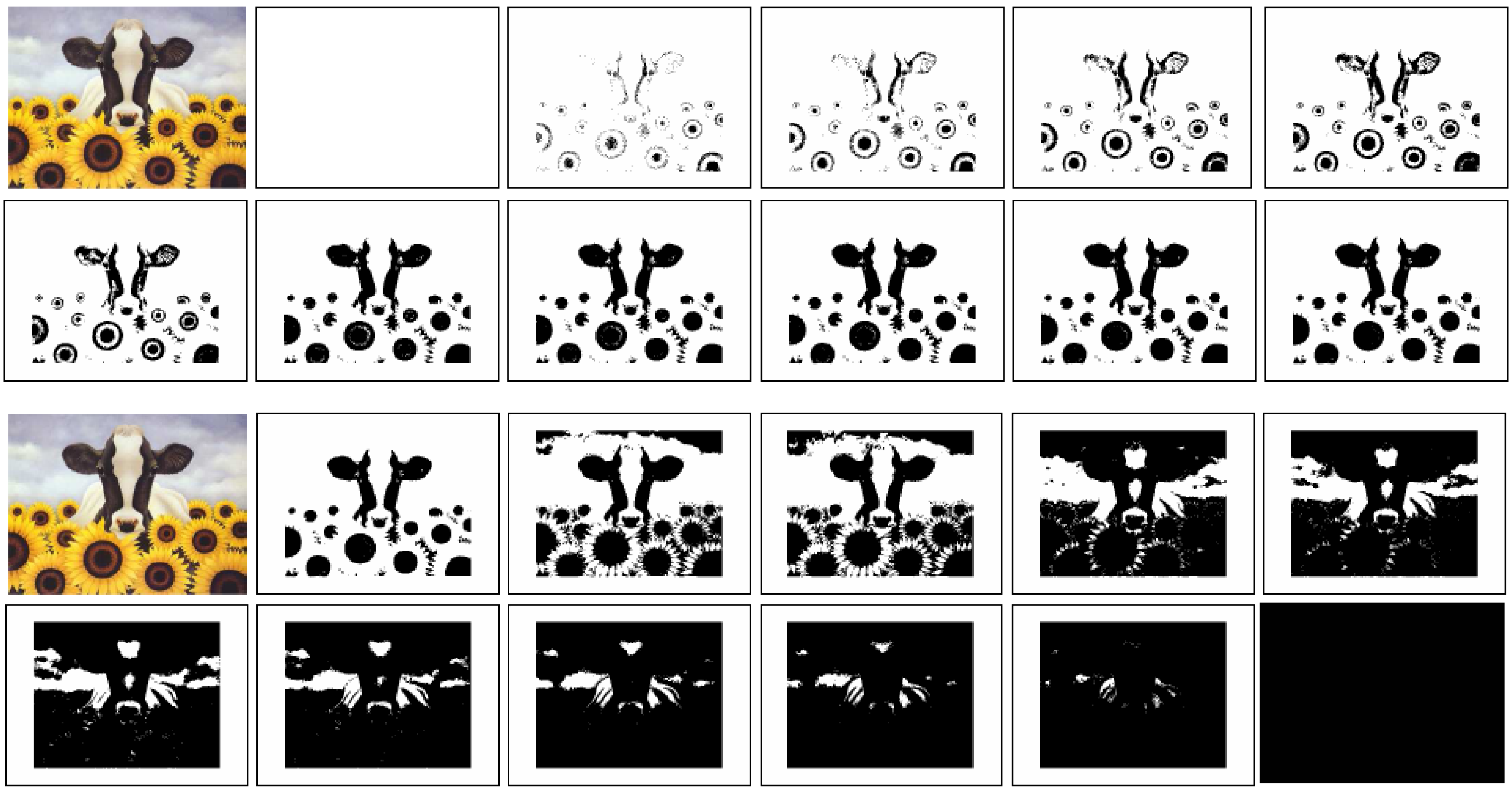
Maximally Stable Extremal Region (MSER). Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_{i-1}, \mathcal{Q}_i, \ldots$ be a sequence of nested extremal regions, i.e. $\mathcal{Q}_i \subset \mathcal{Q}_{i+1}$. Extremal region $\mathcal{Q}_{i*}$ is maximally stable iff $q(i) = |\mathcal{Q}_{i+\Delta} \setminus \mathcal{Q}_{i-\Delta}|/|\mathcal{Q}_i|$ has a local minimum at i^* ($|.|$ denotes cardinality). $\Delta \in \mathcal{S}$ is a parameter of the method.

Table 1: **Definitions** used in Section 2

Sweeping image threhsolds

- Apply a series of thresholds – one for each grayscale level.
- Threshold the image at each level to create a series of black and white images.
- One extreme will be all white, the other all black. In between, blobs grow and merge.

From white to black



From black to white

- *Extremal regions* have the important properties, that the set is closed under continuous (and thus projective) transformation of image coordinates, i.e. it is affine invariant and it doesn't matter if the image is warped or skewed.

- Because the regions are defined exclusively by the intensity function in the region and the outer border, this leads to many key characteristics of the regions which make them useful. Over a large range of thresholds, the local binarization is stable in certain regions, and have the properties listed below:
  - *Invariance to affine transformation of image intensities*
  - *Stability*: only regions whose support is nearly the same over a range of thresholds is selected.
  - *Multi-scale detection* both fine and large structure is detected.

  - The approach is instead sensitive to natural lighting effects as change of day light or moving shadows.

MSER examples

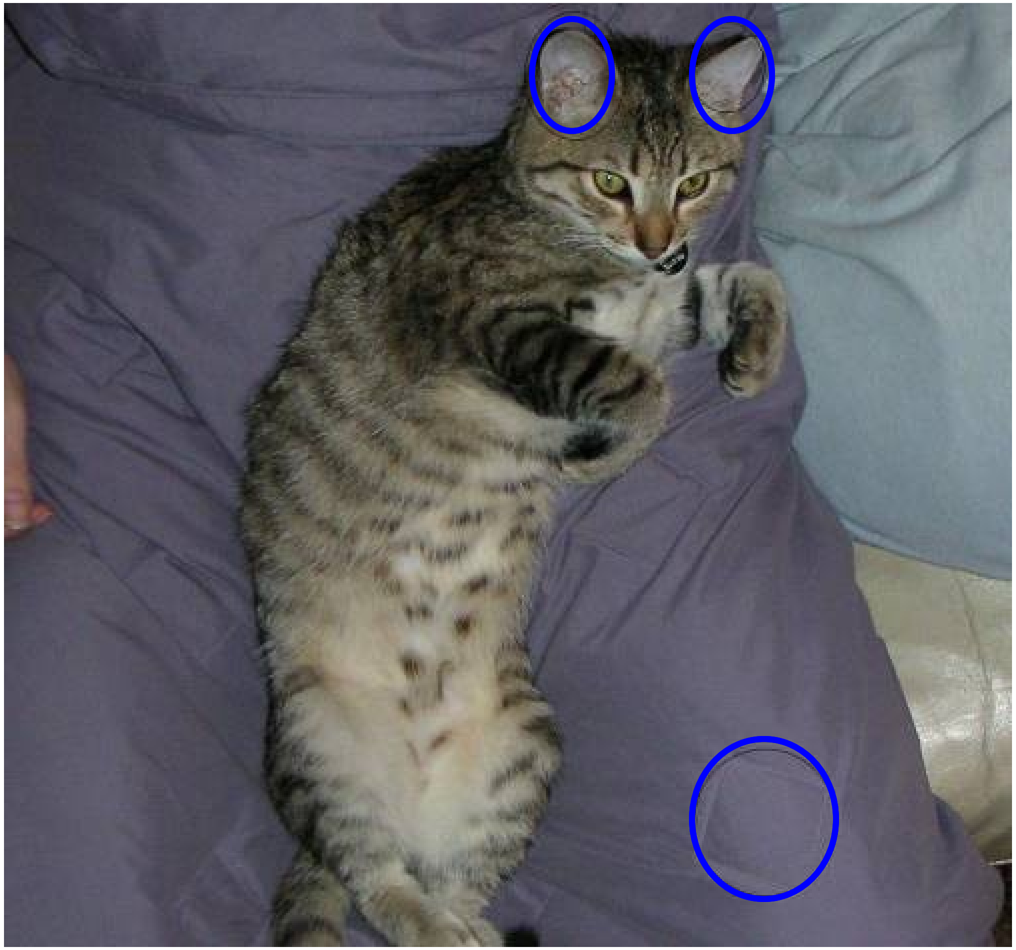
Example1

Threshold 1

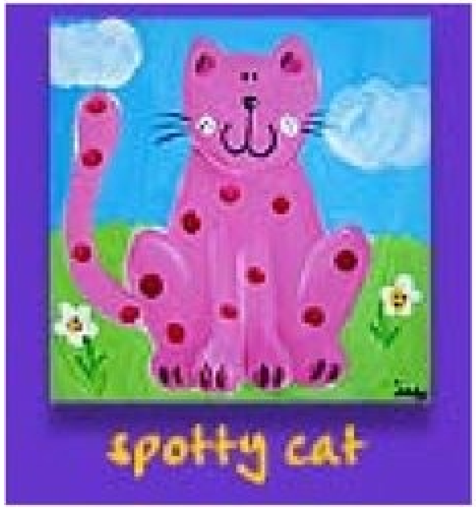
Threshold 2

Threshold 3

Ears and Square bounded by ellipses

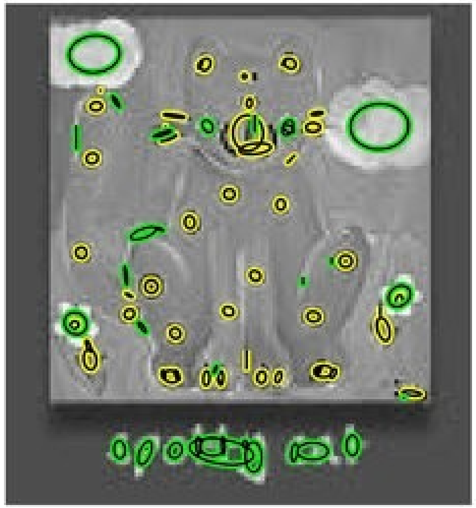


Example 2



Original image

MSER regions

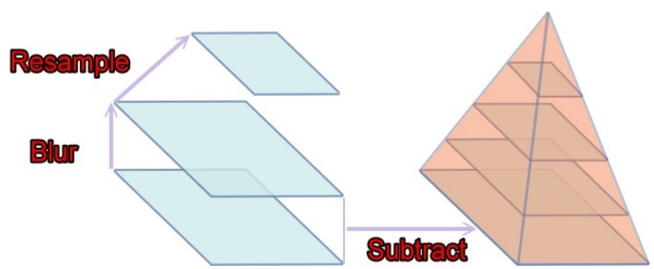


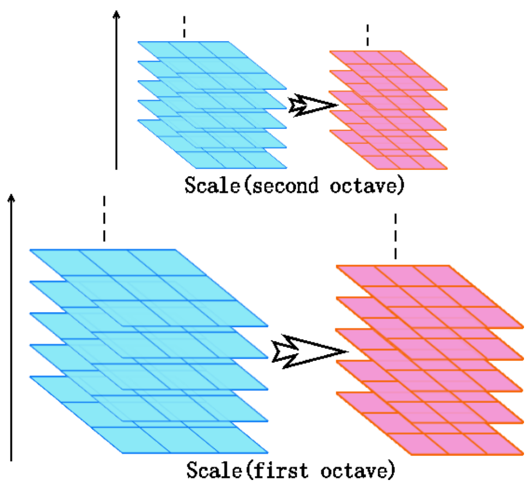
MSER ellipses

Extracted MSERs for both bright-on-dark (green) and dark-on-bright (yellow).

Multi-Resolution MSER

- Method：
  - Step1：Instead of detecting features only in the input image, construct a scale pyramid with one octave between scales
  - Step2：Detect MSERs separately at each resolution
  - Step3：Duplicate MSERs are removed by eliminating fine scale MSERs with similar locations and sizes as MSERs detected at the next coarser scale





- The scale pyramid is constructed by blurring and resample with a Gaussian kernel , σ = 1.0 pixels.

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

P. Forss´en, D. Lowe, S-H Wang

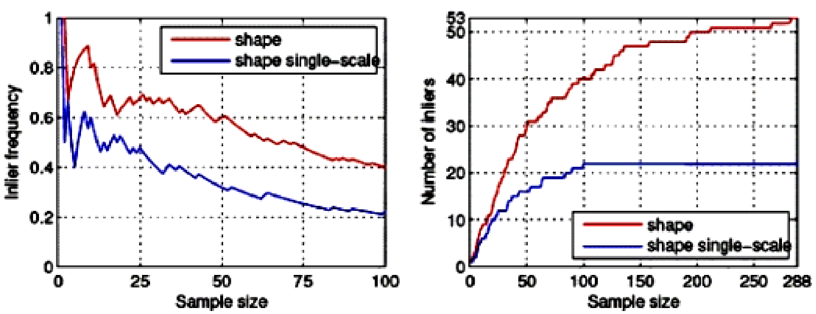
MSER properties

- MSER performs well on images containing homogeneous regions with distinctive boundaries.
- MSER works well for small regions
- MSER doesn’t work well with images with any motion blur
- Multi-resolution MSER provides better robustness to large scale changes and blurred images and improves matching performance over large scale changes and for blurred images

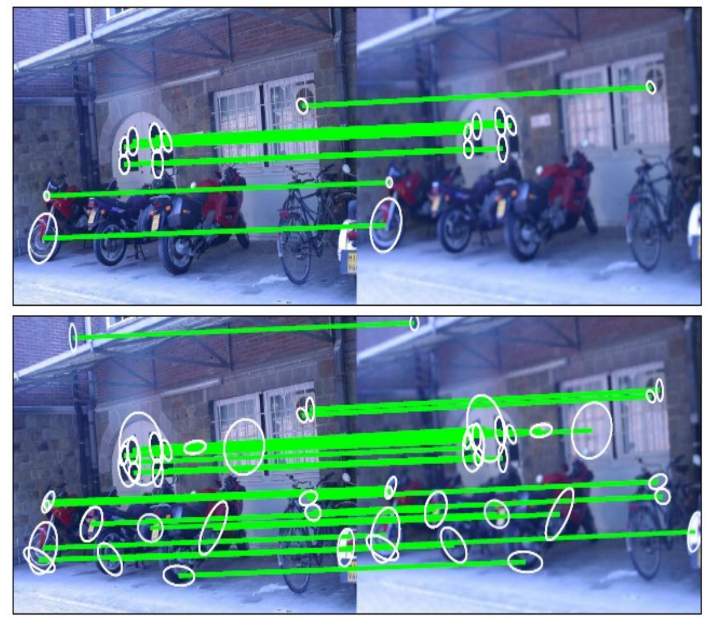
- Good repeatability
- Affine invariant
- A smart implementation makes it one of the fastest region detectors

Multi-resolution MSER experiment

- Planar and parallax-free scenes



single-resolution MSERs

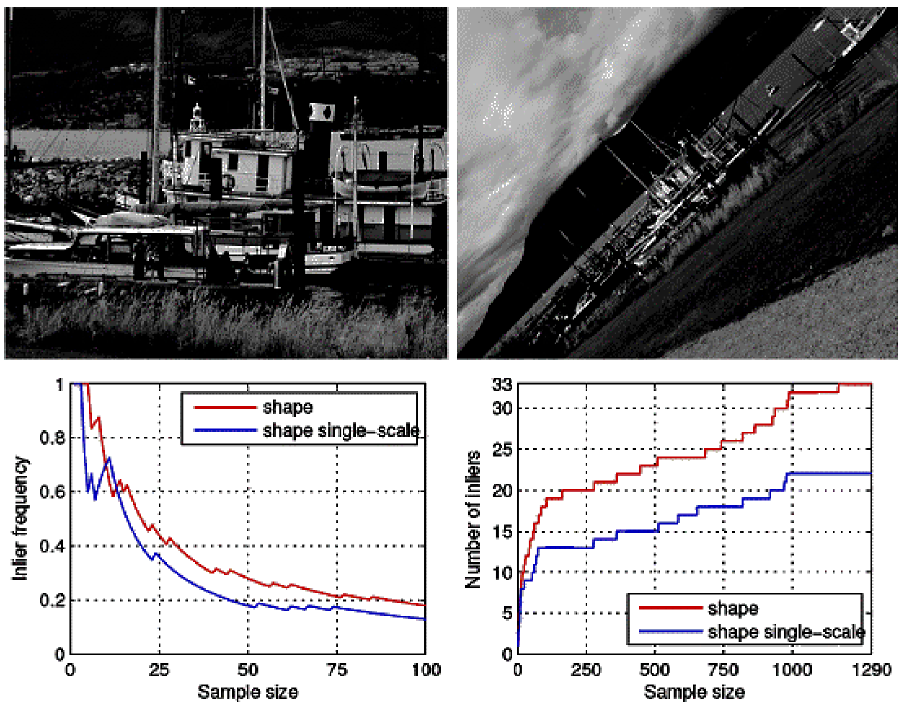


multi-resolution MSERs

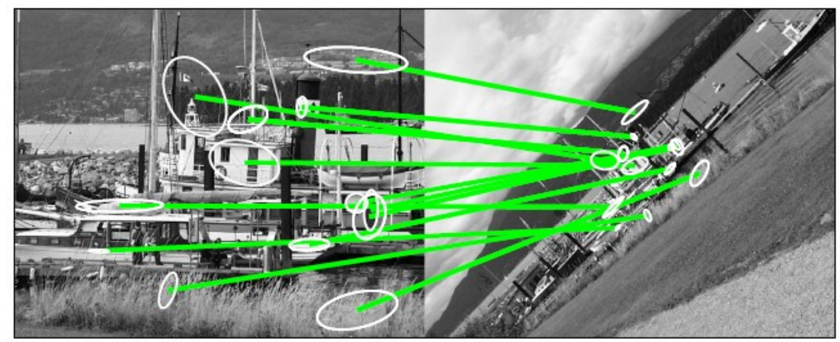
- A scene with de-focus blur. The multi-resolution MSER provides better performance than using only the original resolution

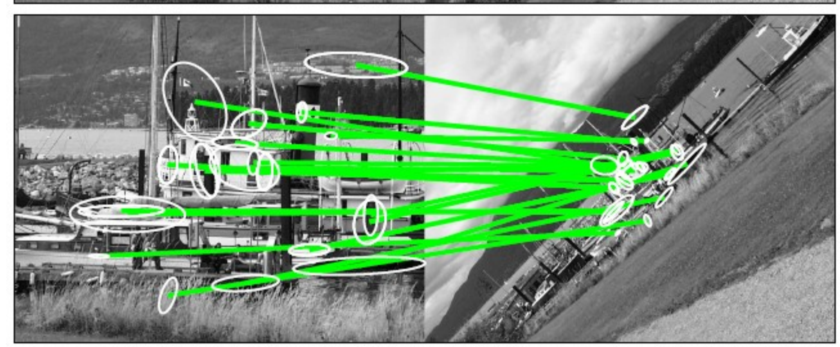
P. Forss´en, D. Lowe, S-H Wang

- Planar and parallax-free scenes



single-resolution MSERs





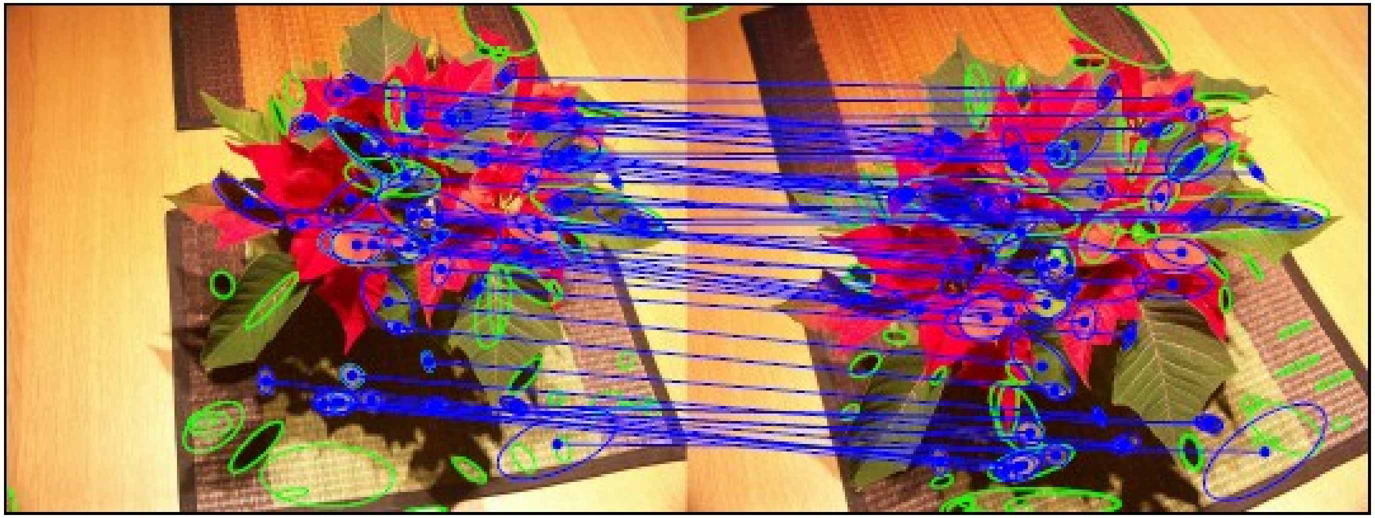
multi-resolution MSERs

- scale change: multi-resolution MSER gives better performance

P. Forss´en, D. Lowe, S-H Wang

Multi-resolution MSER experiment

- 3D scene correspondence evaluation



- The 56 accepted correspondences in blue. Rejected correspondences in green.

P. Forss´en, D. Lowe, S-H Wang

MSER against other region detectors

- *View point change*: MSER performs the best.
- *Scale change*: Hessian-Affine, MSER, Harris-Affine are the best
- *Blur*: all detectors are invariant to image blur. MSER not invariant with any motion blur
- *JPEG change*: MSER is the best