Visual Object Recognition

state-of-the-art techniques and performance evaluation

Lecture 3: Detection of Canonical Frames

- The case against interest point groups
- Scale Selection and DoG
- Affine adaptation
- Maximally Stable Extremal Regions(MSER)
- Maximally Stable Colour Regions (MSCR)
- Edge Based Regions (EBR)

Canonical Frames

- In the previous lecture we saw how c-frames can be found from groups of feature points.
- This lecture is about detecting c-frames from single feature points/feature regions.
- Advantages:
  - smaller c-frames in image (better scale inv.)
  - higher frame repeatability

Canonical Frames

- ***Repeatability*** of a feature detector
  - $p(\text{feature detected in image}) = \epsilon$ (more on this in lecture 7)
- C-Frame repeatability:

$$p(F_1 \cap F_2 \ldots F_N) = \epsilon^N$$

N - Number of feature points in canonical frame.

Scale Space

- Scale space $\quad f(\mathbf{x}) \quad \Rightarrow \quad f_s(\mathbf{x}, \sigma)$
  - The image is extended with an extra dimension, for scale/image blur.

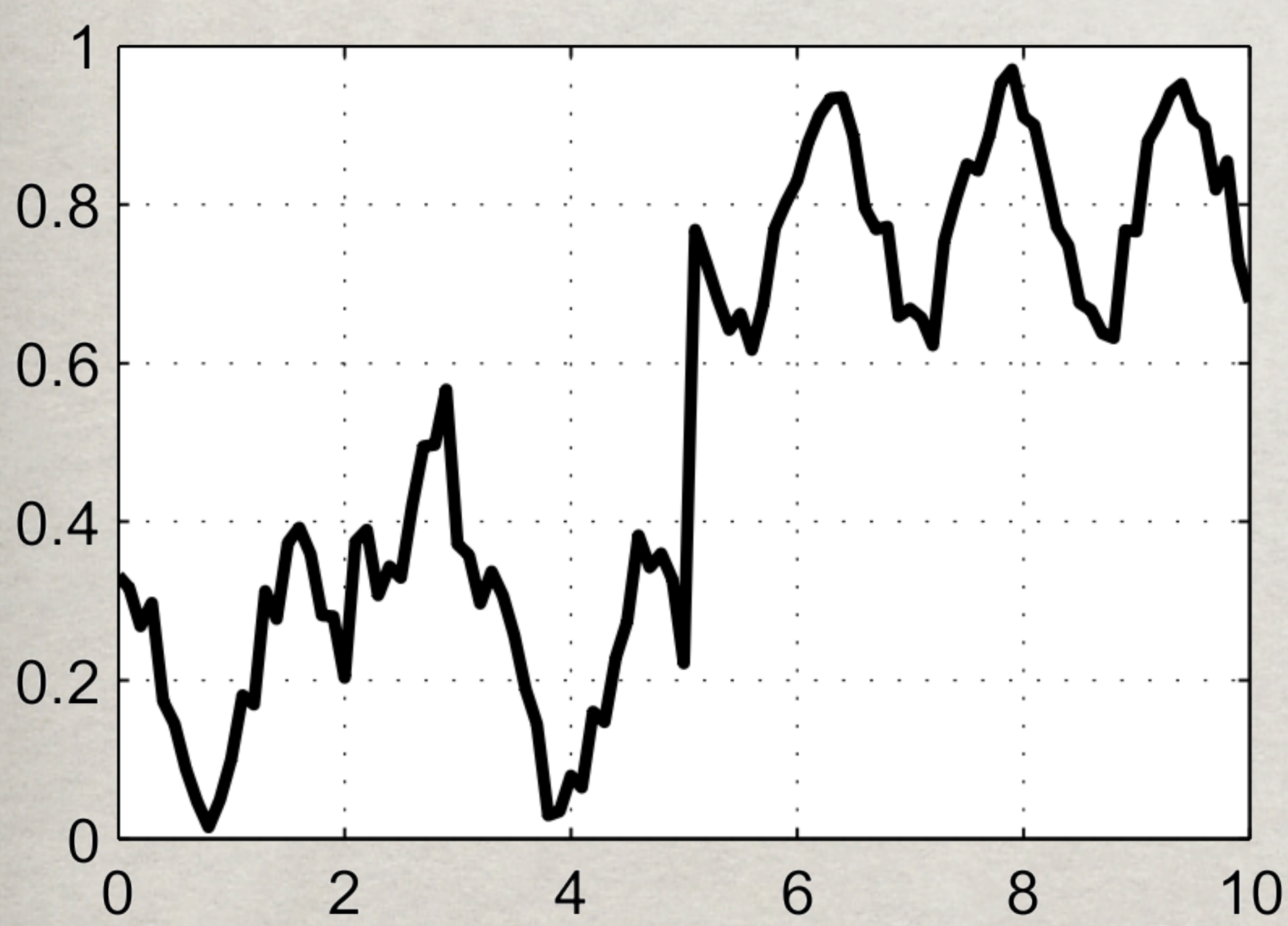
$$f_s(\mathbf{x}, \sigma) = (f * g(\sigma))(\mathbf{x})$$

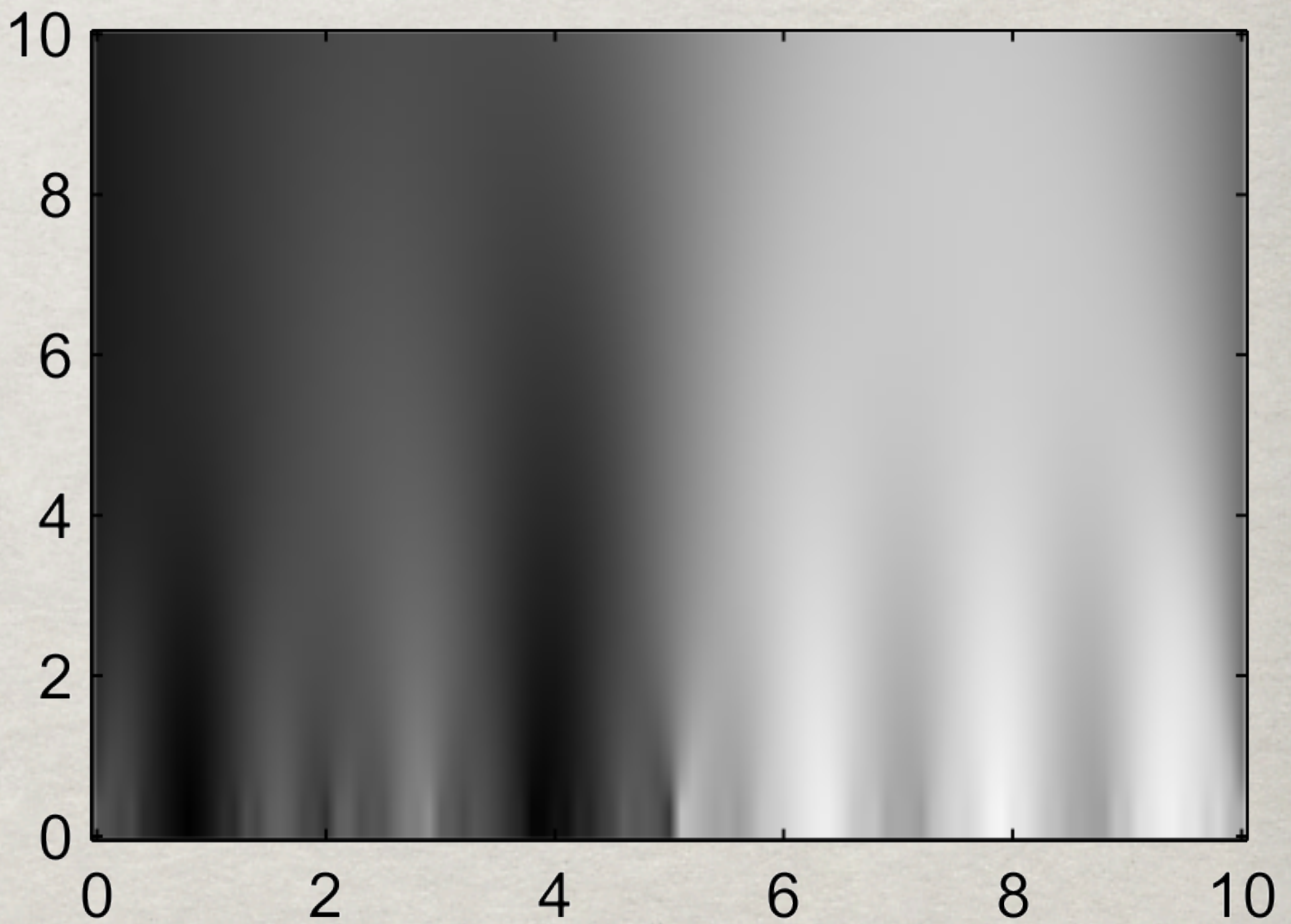
  - The blurring kernel $g(\sigma)$ is typically a Gaussian.

$$g(\mathbf{x}, \sigma) = \frac{1}{2\pi\sigma} e^{-\mathbf{x}^T\mathbf{x}/2\sigma^2}$$

Scale Space

- Illustration in 1D

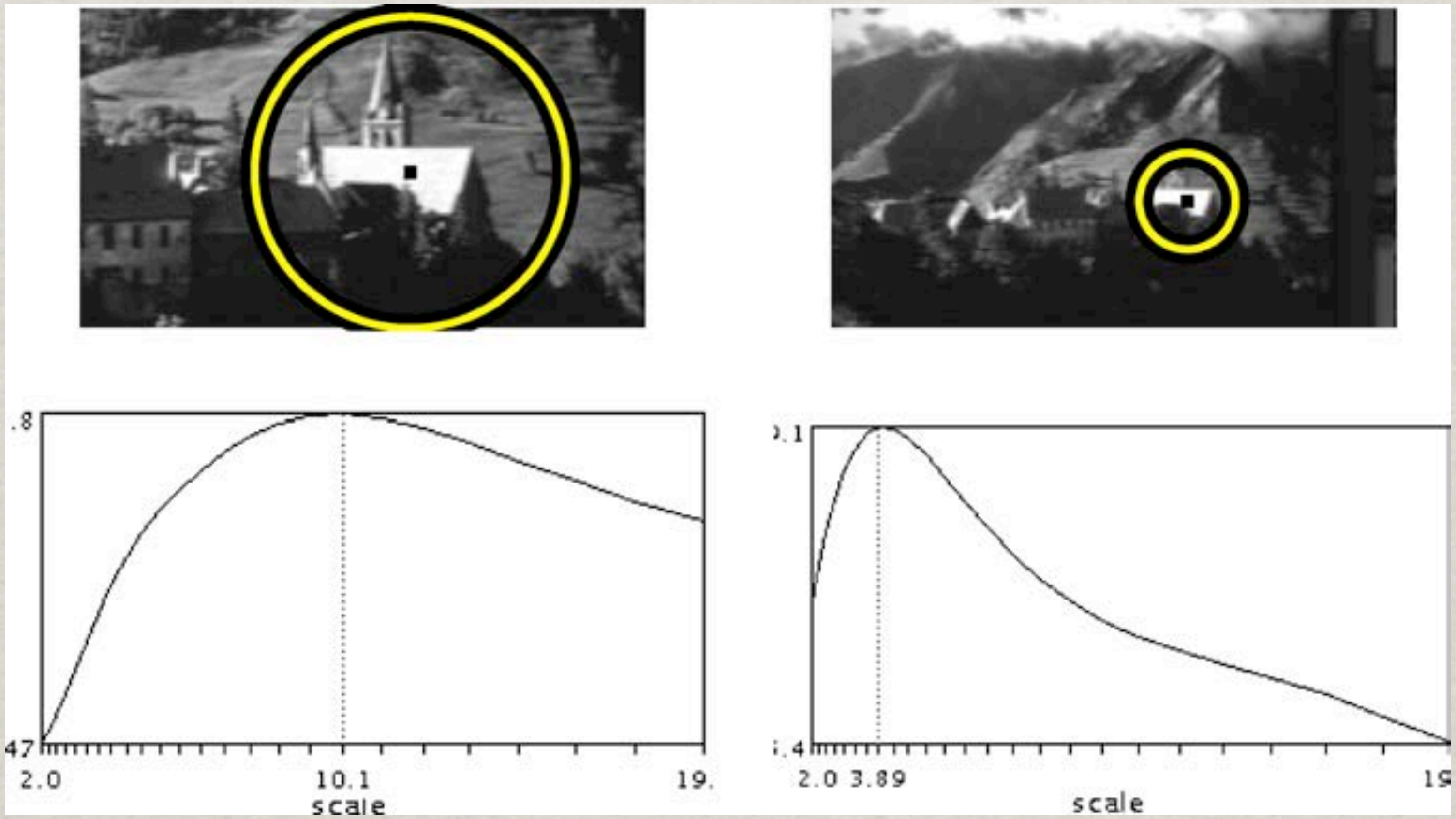




Scale Selection

- Find a characteristic point (e.g. max) on a function of position and scale

$$(\hat{\mathbf{x}}, \hat{\sigma}) = \arg\max h(f(\mathbf{x}, \sigma))$$



Idea from (Lindeberg 1993), illustration by (Mikolajczyk et al. 2005)

Scale Selection

- Example: maximum of normalised Laplacian:

$$h(f(\mathbf{x},\sigma)) = \sigma^2(f * \nabla^2 g(\sigma))(\mathbf{x})$$

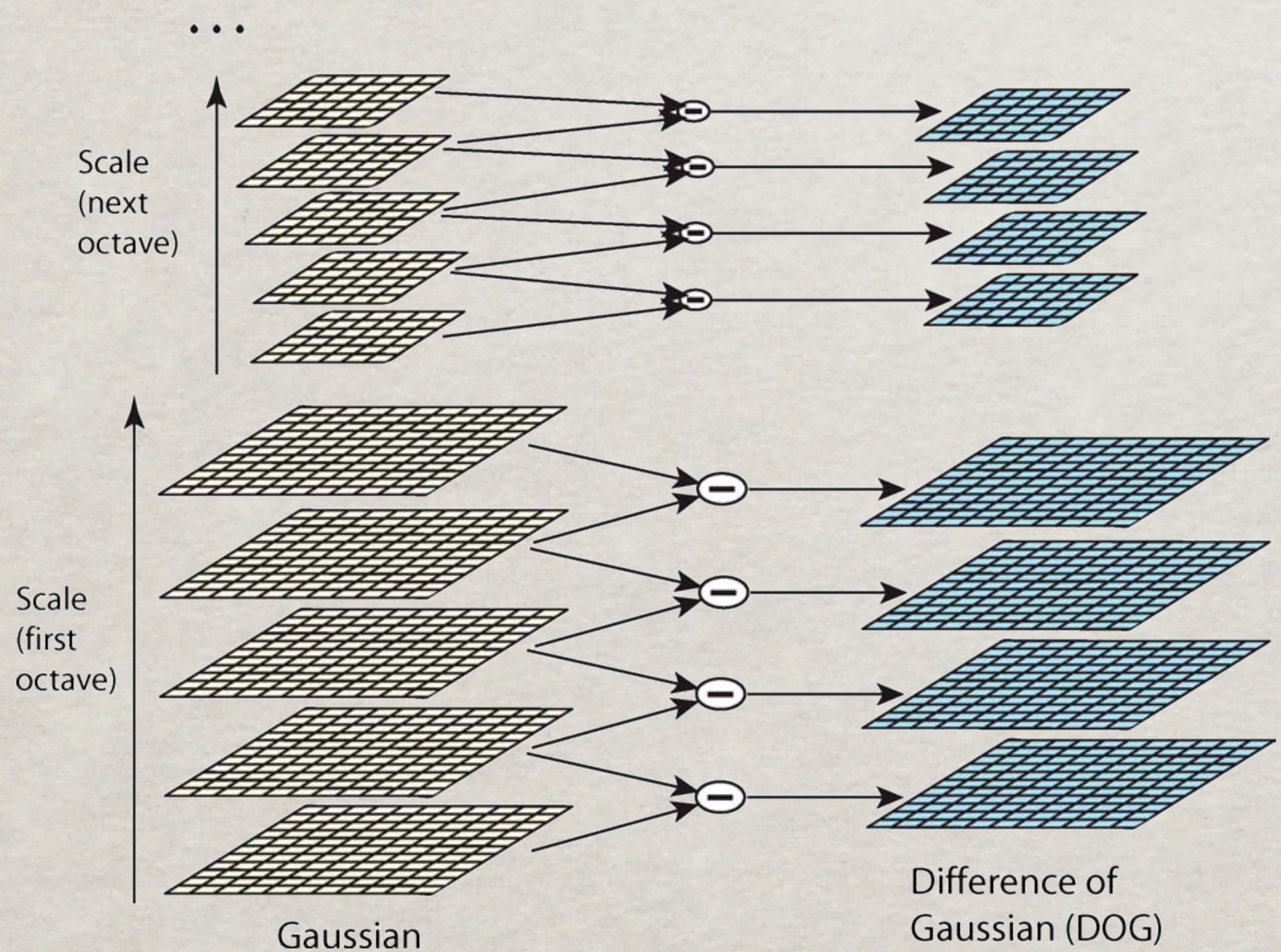
- Note the normalisation by σ^2, which is needed to compensate for decaying amplitude with scale.

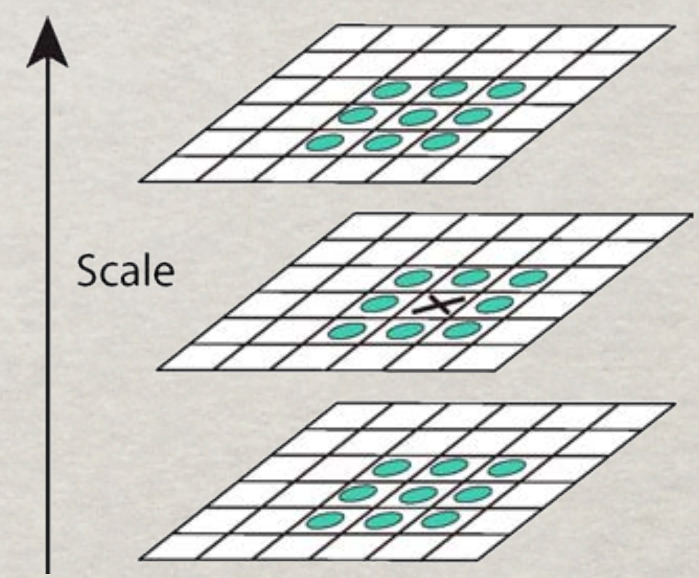
- Another option (used by SIFT) is difference-of-Gaussians:

$$h(f(\mathbf{x},\sigma)) = (f * (g(\sigma) - g(k\sigma)))(\mathbf{x})$$

Scale Selection

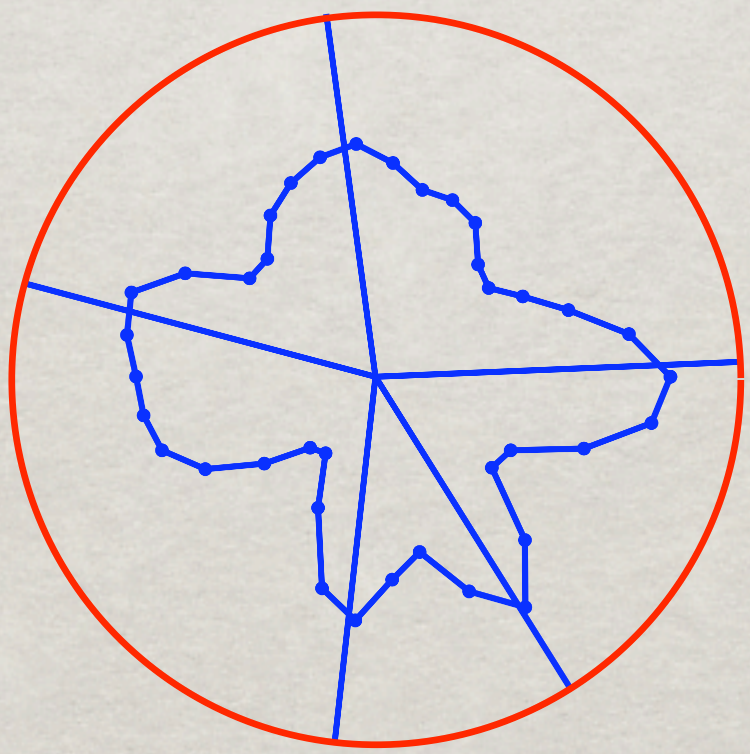
- Efficient implementation using pyramids (Lowe 99)





Scale Selection

- We now have position and scale determined. One or more reference directions can now be found using a gradient orientation histogram at the found location in scale space.



$$h_k = \sum_{\text{patch}} |\nabla \mathbf{f}(\mathbf{x})| B_k(\tan^{-1} \nabla \mathbf{f}(\mathbf{x}))$$

Affine Adaptation

- Scale Selection+Reference direction gives a similarity frame
- By iteratively adjusting the circle defined by position and scale to an ellipse, we can get a full affine frame instead.
- In practice done by finding a resampling $\mathbf{x} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{m}$ that gives a *structure tensor* with equal eigenvalues in the c-frame.

Affine Adaptation

- Structure tensor from gradient: $\nabla\mathbf{f} = [f_x \; f_y]^T$

$$\mathbf{T}(\mathbf{x}) = \left(\begin{bmatrix} f_x^2 & f_x f_y \\ f_x f_y & f_y^2 \end{bmatrix} * g\right)(\mathbf{x}) = \left(\nabla\mathbf{f}\nabla\mathbf{f}^T * g\right)(\mathbf{x})$$

- $\mathbf{T}$ is a measure of anisotropy. Set:

$$\mathbf{A} = \mathbf{T}^{1/2} \quad \text{here defined as} \quad \mathbf{A}^T\mathbf{A} = \mathbf{T}$$

- Inverse *whitening transform*. Needs to be iterated a few times, as g($\mathbf{x}$) should be anisotropic.

$$\tilde{\mathbf{A}} = \mathbf{A}\mathbf{R}\,,\; \mathbf{R} \in \mathrm{O}(2) \Rightarrow \tilde{\mathbf{A}}^T\tilde{\mathbf{A}} = \mathbf{T}$$

- Choose reference direction from gradient orientation histogram.

MSER

- Maximally Stable Extremal Regions
- Consider the set of all possible thresholdings of an image...
- Connected regions form segments.
- Cf. Watershed algorithm
- Look at stability of a function of segment across image evolution. e.g. $\texttt{area}(\texttt{component}(t))$

MSER

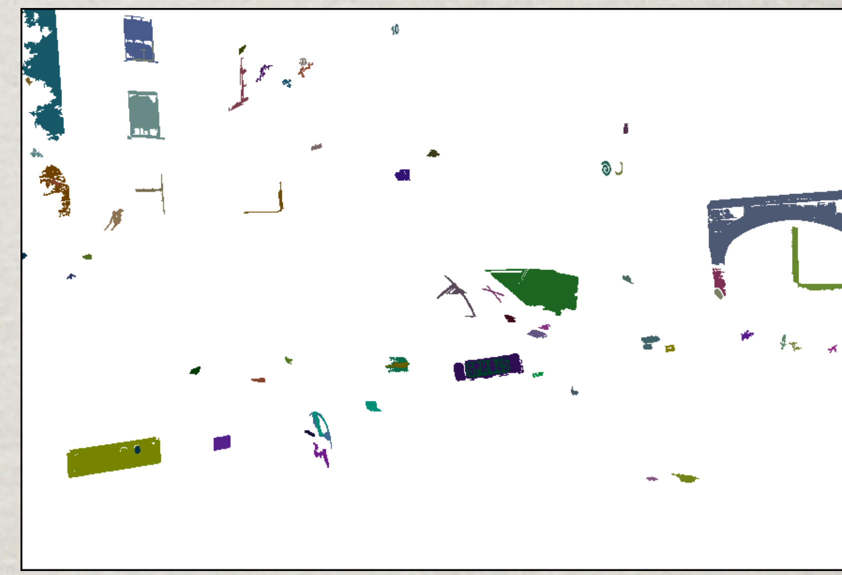
- MSERs are components that are *maximally stable*, i.e., have a local minimum of the rate of change: $$\frac{\partial \mathtt{area}(\mathtt{component}(\mathtt{t}))}{\partial t}$$
- c.f. Scale Selection
- Stability measure: Range of stable thresholds t_2-t_1 around min is called the *margin* of the region.

MSER

- Two possible thresholdings: $I(\mathbf{x}) < t\,,\;\; I(\mathbf{x}) > t$

Input image

64 MSER- (total 272)



64 MSER+ (total 294)

- Detection is fast, 0.1sec on 800x600 image (using the union/find algorithm).

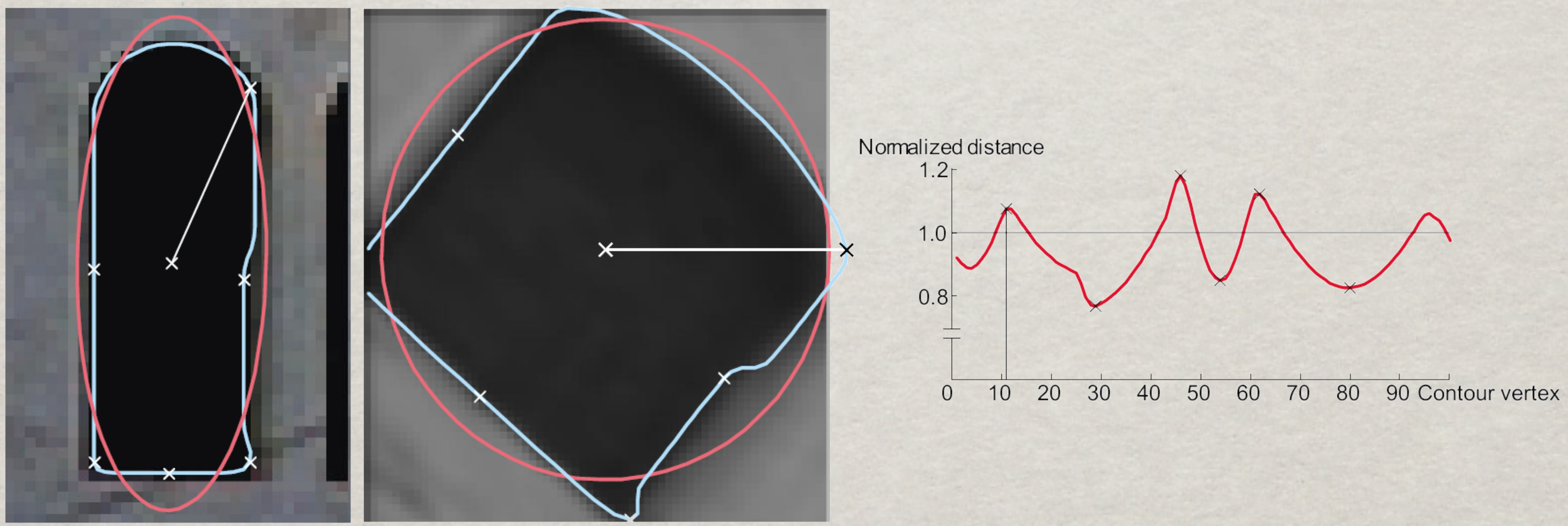
- MSER type (+/-) is useful for matching

MSER

- MSER is invariant to monotonic changes of intensity. i.e. I(x) and f(I(x)) have the same output if $f(x+k) > f(x) \; \forall \; k > 0$
- Wide range of sizes obtained without a scale pyramid. Better still with a pyramid (Forssén&Lowe CVPR'07)
- Can be used to track colour objects by computing MSERs on the Mahalanobis distance to a colour distribution. (Donoser&Bischof CVPR'06)

Local Affine Frames

- Find approximating ellipse of region.
- Contour extrema in normalised frame give reference directions.



Matas et al. ICPR'02

Local Affine Frames

- **Approximating ellipse**
  from moments of binary mask $v : \Omega \mapsto \{0, 1\}$

$$\mu_{k,l}(v) = \sum_x \sum_y x^k y^l v(x, y)$$

$$\mathbf{m} = \frac{1}{\mu_{0,0}} \begin{bmatrix} \mu_{1,0} \\ \mu_{0,1} \end{bmatrix} \quad \mathbf{C} = \frac{1}{\mu_{0,0}} \begin{bmatrix} \mu_{2,0} & \mu_{1,1} \\ \mu_{1,1} & \mu_{0,2} \end{bmatrix} - \mathbf{m}\mathbf{m}^T$$

$$\mathcal{R}(\mathbf{m}, \mathbf{C}) = \{\mathbf{x} : (\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}) \leq 4\}$$

- See appendix C in thesis by Forssén 2004

Local Affine Frames

- **Normalisation to a circle** (axis aligned)
  Compute the eigenfactorisation:

$$\mathbf{C} = \mathbf{R}\mathbf{D}\mathbf{R}^T\,, \quad \det \mathbf{R} > 0$$

The circle normalisation can now be performed as:

$$\mathbf{x} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{m}\,, \quad \text{for} \;\; \mathbf{A} = 2\mathbf{R}\mathbf{D}^{1/2}$$

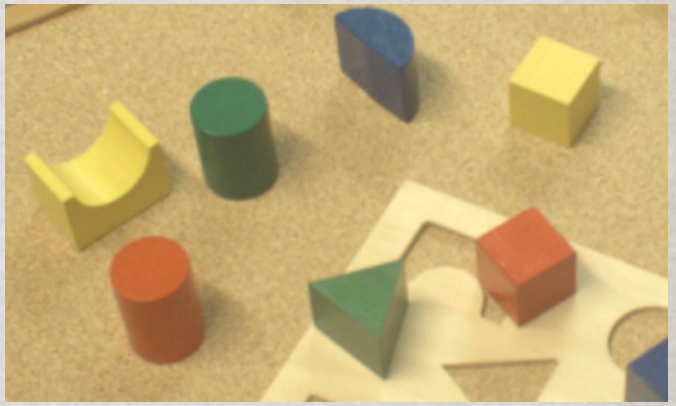
$\hat{\mathbf{x}}$ - canonical coordinates
$\mathbf{x}$ - image coordinates

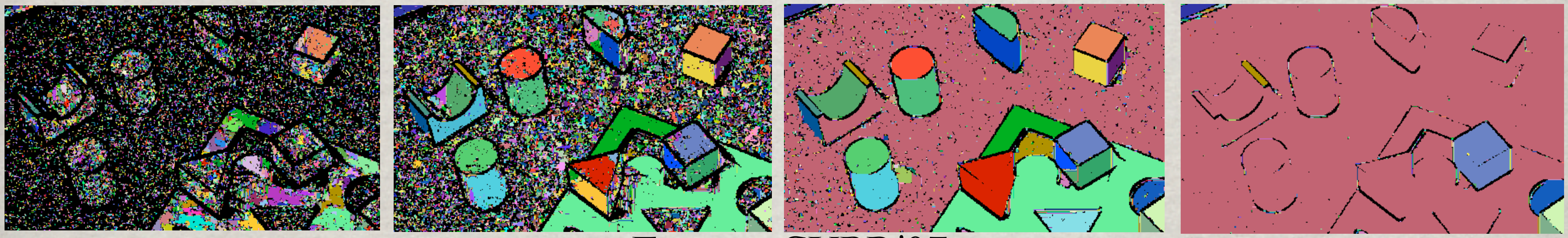
Local Affine Frames

- Ellipse+extrema of distance to centre is just one frame construction option.
- Other (affine covariant) choices:
  - Points of maximum curvature.
  - Bi-tangens.
- See Obdrzalek&Matas BMVC'02

MSCR

- Maximally Stable Colour Regions.
  - Define evolution function on an agglomerative clustering of image.





Forssén CVPR'07

MSCR

- Improved robustness to illumination changes, and changes of background

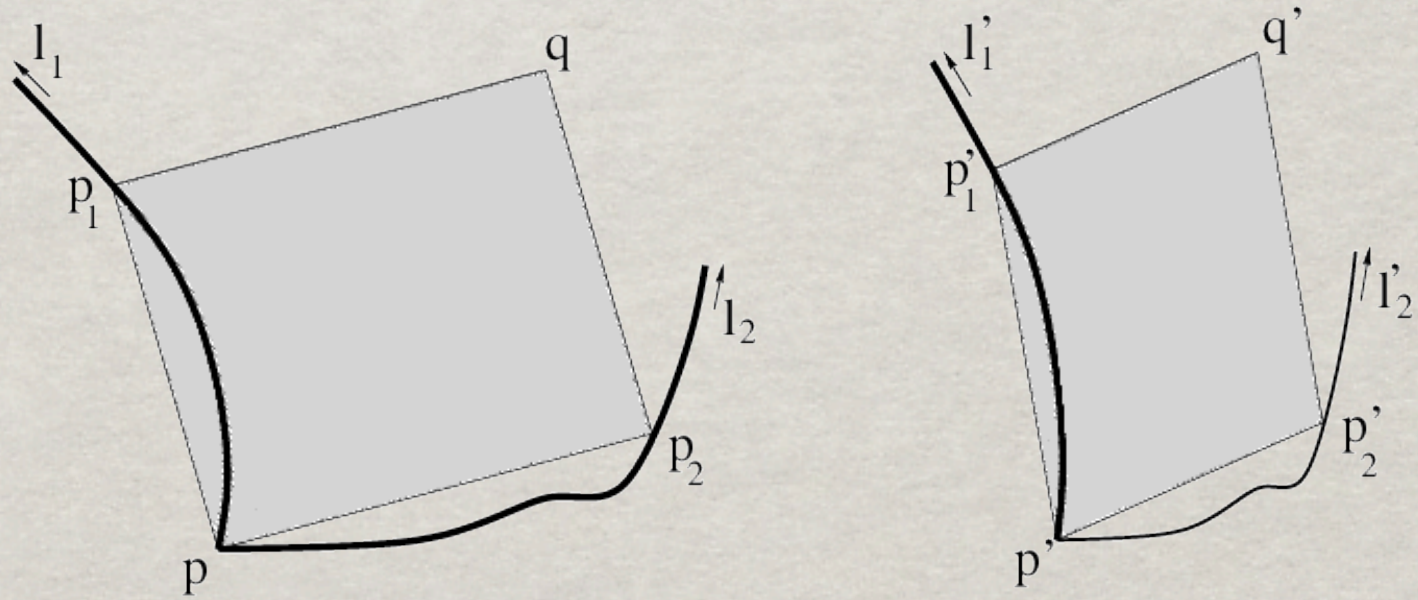
MSER+ and MSER-

MSCR

- ~3x more computationally expensive.

EBR

- Edge Based Regions (Tuytelaars&vanGool ICVIS'99)
- Start in a Harris point (Harris&Stephens AVC'88) situated on a Canny contour (Canny PAMI'86).
- Move in both directions in an affine invariant manner.

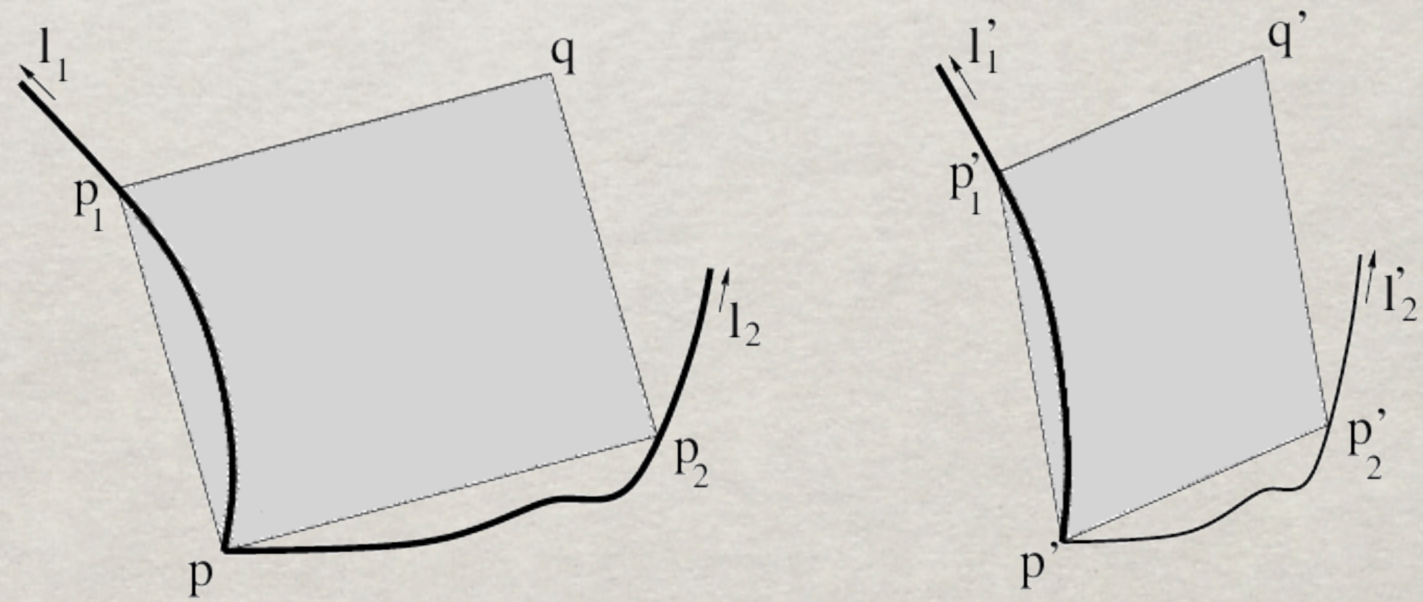


EBR

- Affine invariance by maintaining equality of two integrals $l_1 = l_2$

$$l_i = \int (|\mathbf{p}'_i(s_i) \quad \mathbf{p} - \mathbf{p}_i(s_i)|) ds_i$$

- Proportional to two areas and affinities preserve area ratios.

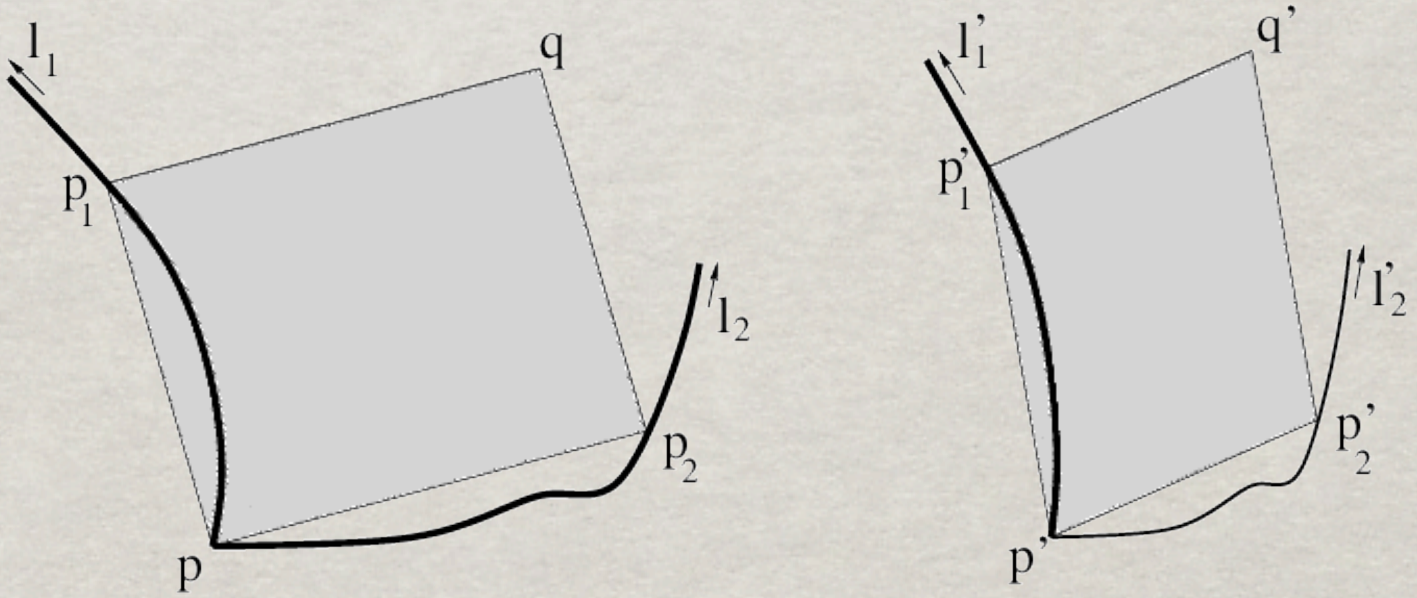


EBR

- All values of s_i give affine invariant regions. Which should be selected?

  A. Local max of mean intensity in parallellogram

  B. Two other choices in today's paper.



Discussion

- Questions/comments on paper and lecture.