Object Recognition Showcase

Nicu Sebe

University of Amsterdam, The Netherlands

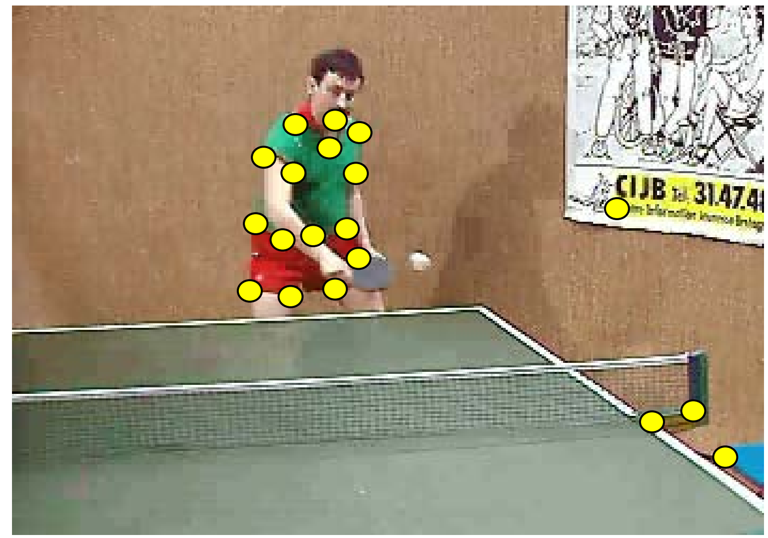
Allan Hanbury, Julian Stoettinger

TU Wien - PRIP

Jaume Amores

INRIA - IMEDIA

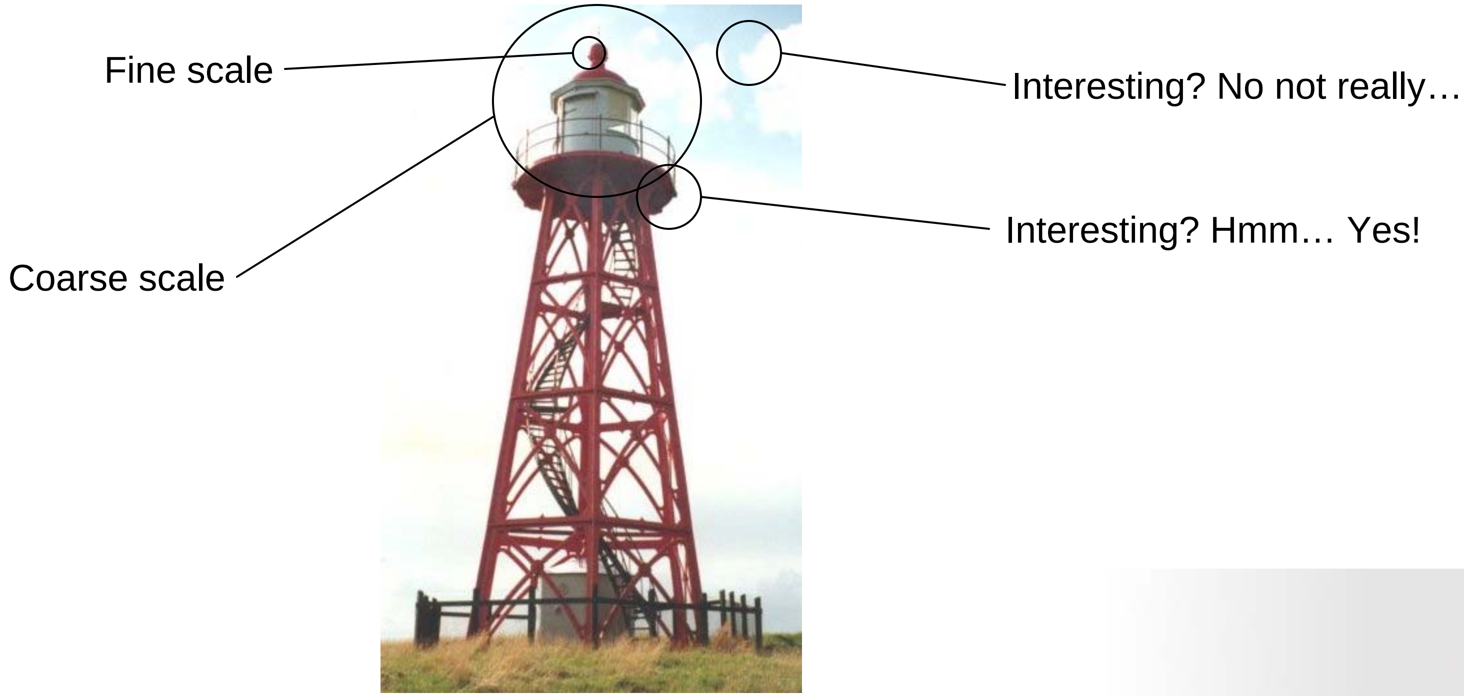
Salient Points



luminance-based points
color-based points?

Salient Points

- **Capture visual "interesting" parts of an image**
- **All points should summarize the image content**
  - Multiple scales: coarse … fine

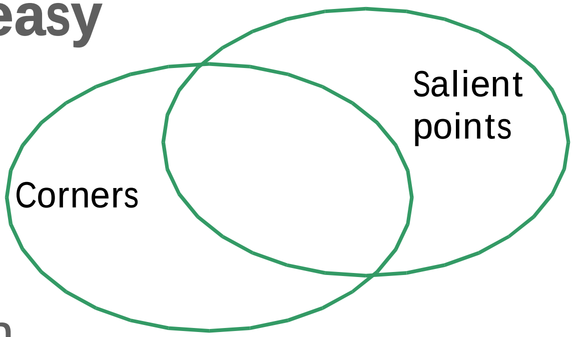


Salient Points - Usage

- **Matching them!**
  - Compare detected salient points
    - Detect points in different images
    - Describe these points and compare using a similarity measure
    - Derive relations between images:
      - i.e.: Same scene with different viewpoint; common object(s); etc.
- **For example:**
    - Object recognition
      - Different scales: Hierarchical object model
    - Object tracking
    - Content Based Image Retrieval (CBIR)

Existing Research

- **Finding visual "interesting" points is not easy**
  - Mathematical definition?



- **Local Image Descriptors**
  - Harris [Harris88], Multi-scale point detection [Mikolajczyk01], Local gray value invariants [Schmid97], Edge-based region detector [Tuytelaars04], SUSAN [Smith97], Wavelets [Sebe 03], etc.
- **Local Region Descriptors**
  - SIFT [Lowe04], shape context [Belongie02], moment invariants [Gool96], N-jet [Koenderink87], etc.

Existing Research - Issues

- **Images are mostly color**
  - Why are the existing salient point techniques luminance-based?
  - They typically focus on shape saliency rather than color saliency
  - They cannot distinguish between black-and-white corners (low salient) and red-green corners (high-salient)
- **Few existing salient point algorithms that use color [Montesions98][Itti98][Heidenman04]**
  - Their results do not differ greatly from the intensity-based methods
  - Difficulties in combining the information available from the color channels
  - Many possible color spaces

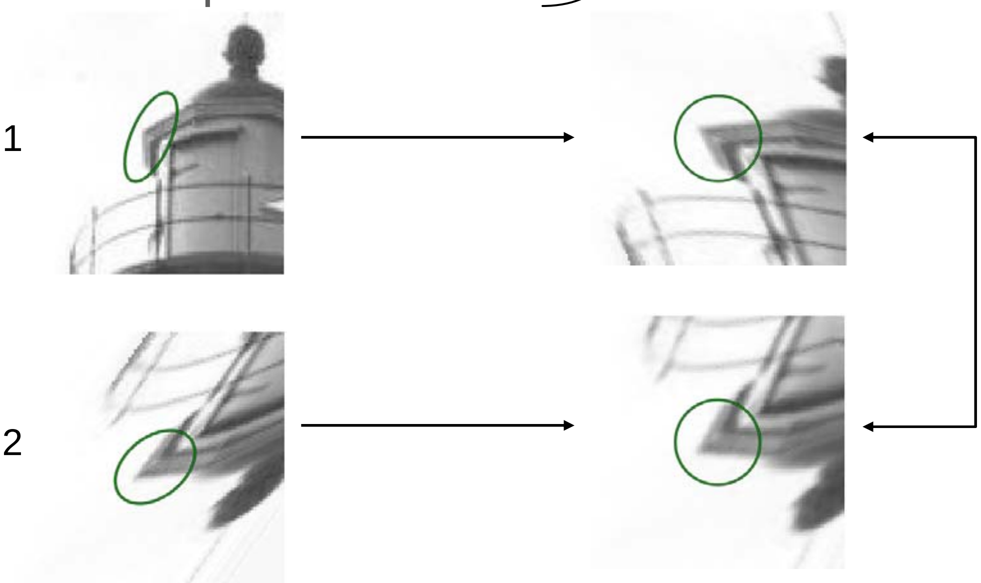
Research – Affine invariance

- **Detect regions under common transformations**
  - Translation
  - Rotation
  - Scaling
  - Viewpoint

Affine invariance !

Viewpoint 1

Viewpoint 2



Related by rotation

Detected regions

Normalized detected regions

Research - Framework

- **Existing method by Mikolajczyk**
  - Iterative affine invariant point detector
    - Multi-scale Harris corner detector
    - Laplacian characteristic scale selection
    - Second moment matrix shape determination

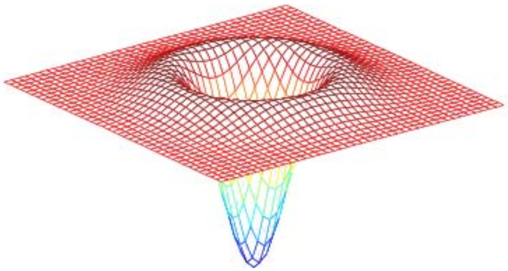
Initial region based on initial scale and location

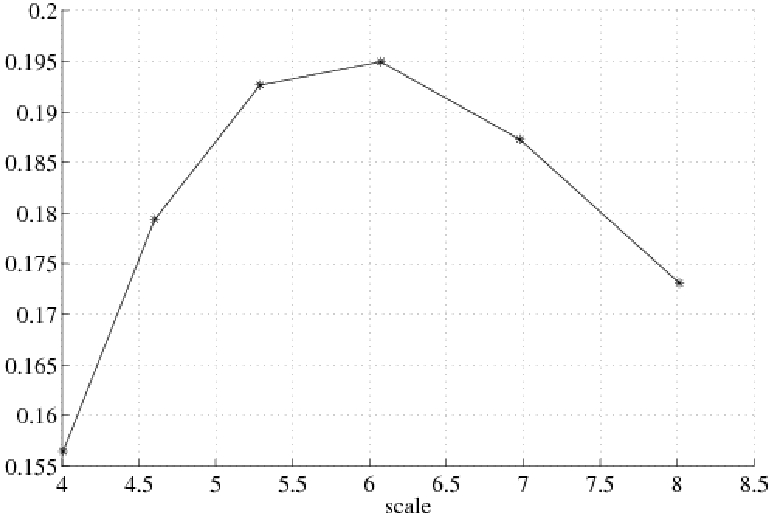
Iteratively adjust scale, position and shape of region

final region

Research – Framework

- **Characteristic scale**
  - Convolve with multiple Laplacian of Gaussian kernels: scale trace.
  - Select maximum



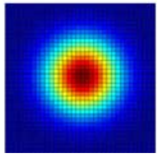


Research – Framework

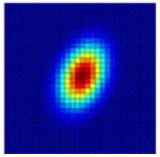
- **Affine deformation**
  - Second moment matrix

$$\mu(\mathbf{x},\sigma_I,\sigma_D) = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} = \sigma_D^2 g(\sigma_I) \cdot \begin{bmatrix} L_x^2(\mathbf{x},\sigma_D) & L_x L_y(\mathbf{x},\sigma_D) \\ L_x L_y(\mathbf{x},\sigma_D) & L_y^2(\mathbf{x},\sigma_D) \end{bmatrix}$$

    - Suppress noise without suppressing the anisotropic shape of a structure.
    - Eigenvalues represent two principal curvatures of a point: shape normalization!
    - Calculated using (affine) Gaussian kernels



Uniform kernel



Affine kernel

  - Affine invariance
    - Detect regions that comply to:

$$\mu(\mathbf{x},\Sigma_I,\Sigma_D) = M$$
$$\Sigma_I = \sigma_I M^{-1}$$
$$\Sigma_D = \sigma_D M^{-1}$$

Color-based salient points

- **Color Harris (Weijer04)**
  - Extend calculation of second moment matrix to color
    - Sum gradients of the channels

color-based points?
luminance-based points

What's the problem?

Evaluation Criteria [Schmid98]

- **Repeatability**
  - Salient point detection should be stable under varying viewing conditions
- **Distinctiveness**
  - Salient points should focus on events with a low probability of occurrence

Idea: Incorporate color distinctiveness into the design of salient point detectors!!!!!!

Color-based salient points

- The efficiency of the salient point detection depends on distinctiveness of the extracted points
- At the salient points' positions, local neighborhoods are extracted and described by local image descriptors
- The distinctiveness of a descriptor describes the conciseness of the representation and the discriminative power of the salient points
- The distinctiveness is measured from the information content
- the information content of an event, v, **is equal to :**

$$I(v) = -\log(p(v))$$

Color-based salient points

- For luminance-based descriptors the information content is measured by the local two-jet of the local structure [Schmid00]
- Due to extra information available in color images, the local one-jet is sufficient

$$v = \begin{pmatrix} R & G & B & R_x & G_x & B_x & R_y & G_y & B_y \end{pmatrix}$$

- Assuming independent probabilities of the 0^{th} order signal and the 1^{st} derivatives, the information content is:

$$I(v) = -\log(p(v)) = -\log\left(p(\mathbf{f})\,p(\mathbf{f}_x)\,p(\mathbf{f}_y)\right) \qquad \mathbf{f} = (R, G, B)$$

- By adapting the saliency map to focus on rare color derivatives, the color distinctiveness of the detector is improved!!!!

Color-based salient points

Saliency boosting

- Image derivatives that occur equally often should contribute equally to the saliency measure
- Vectors with equal information content should have equal influence on the saliency map
- Find a transformation *g* for which it holds:

$$\boxed{\textit{Color Boosting Saliency:}\quad p(\mathbf{f}_x) = p(\mathbf{f}'_x) \leftrightarrow |g(\mathbf{f}_x)| = |g(\mathbf{f}'_x)|}$$

Color-based salient points

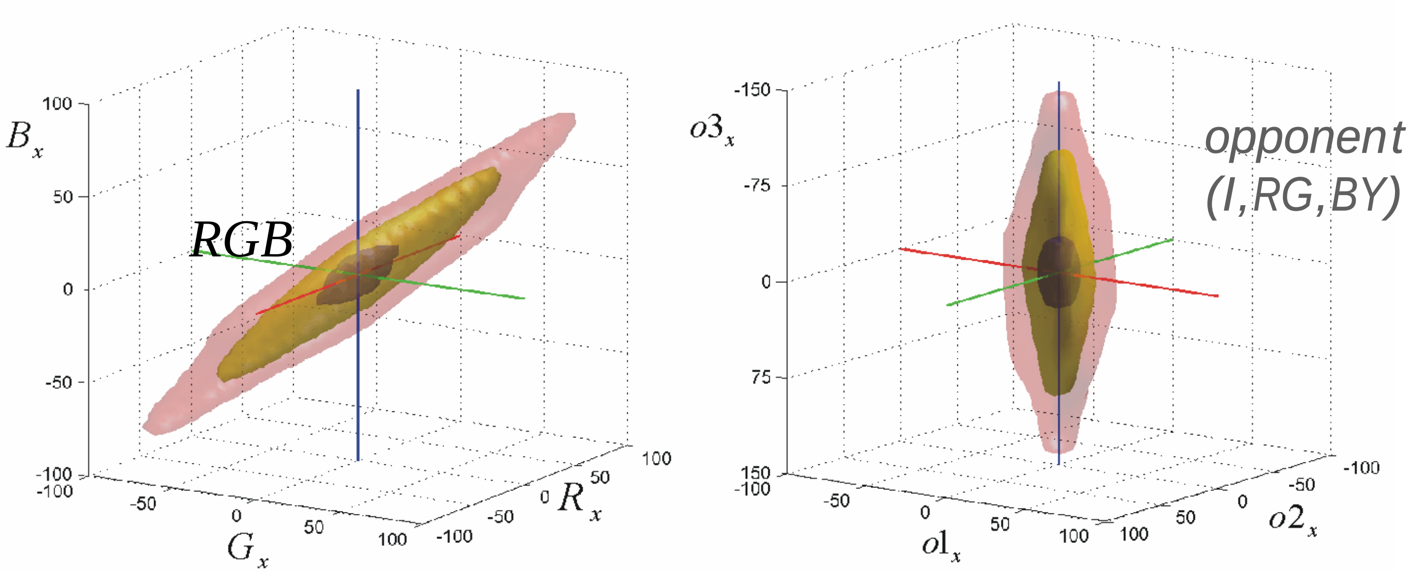
Invariance ⇔ distinctiveness

- The channels of $\mathbf{f}_x$ are correlated!!! Shadows, shading, and specularities will have a great influence
- There is a need to use different color spaces which will eliminate the influence of these perturbations

	shadows	shading	highlights	ill. intensity	ill. Colour
I	-	-	-	-	-
RGB	-	-	-	-	-
rgb	+	+	-	+	-
Ratios	+	+	-	+	+

Statistics of color images

• The statistics of $\mathbf{f}_x$ are computed by looking at the 40.000 images of the Corel database.

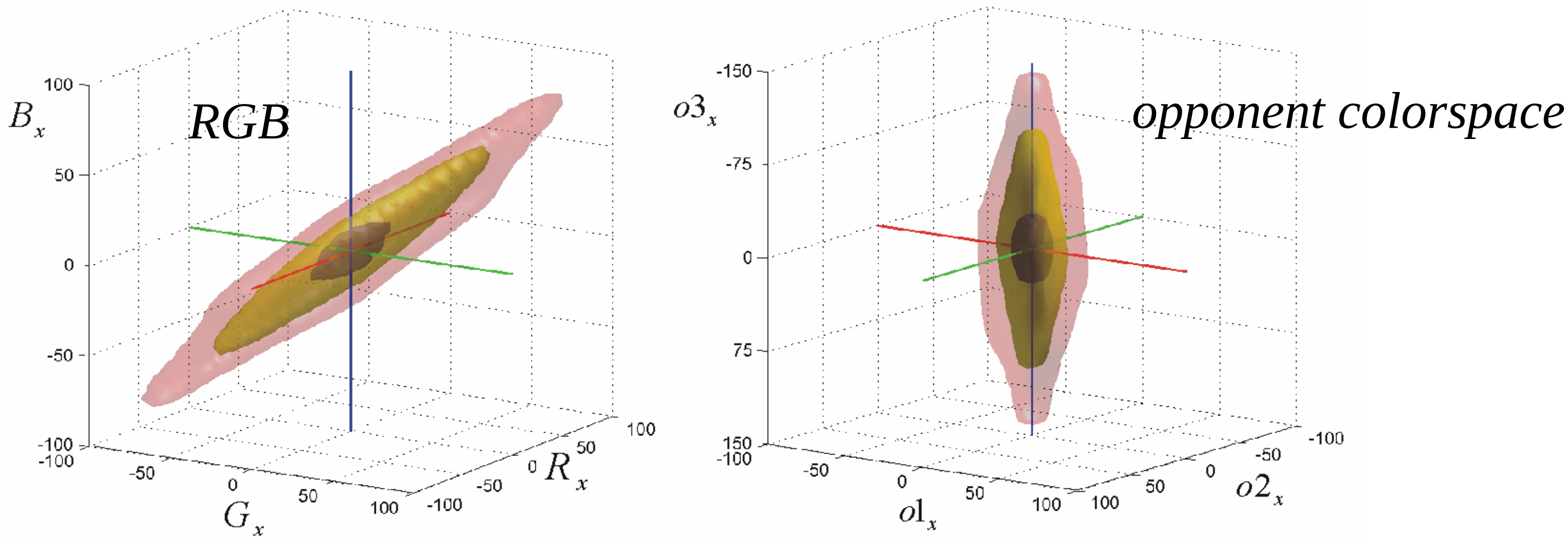


opponent colorspace (I,RG,BY)

• Isosalient surfaces can be approximated by aligned ellipsoids in decorrelated color spaces.

Statistics of color images

Color Boosting Saliency: $p(\mathbf{f}_x) = p(\mathbf{f}'_x) \leftrightarrow |g(\mathbf{f}_x)| = |g(\mathbf{f}'_x)|$



Color Boosting function: $$g(\mathbf{f}_x) = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} h(\mathbf{f}_x)$$

Statistics of color images

	spherical	opponent	HSI
λ_1	0.85	0.85	0.86
λ_2	0.52	0.52	0.51
λ_3	0.10	0.065	0.066

- **Opponent color space was to perform best [vdWeijer04]**
  - One of the components is still the intensity (although, with a very low weight, i.e., 0.065)
- **Investigate a more invariant color space which has no intensity information anymore: color ratios**
- **The goal is to analyse the tradeoff between invariance and distinctiveness**

Color constancy: Color Ratios

$$m_1 = \frac{R^{x_1} G^{x_2}}{R^{x_2} G^{x_1}}, m_2 = \frac{R^{x_1} B^{x_2}}{R^{x_2} B^{x_1}}, m_3 = \frac{G^{x_1} B^{x_2}}{G^{x_2} B^{x_1}}$$

Taking the natural logarithm of both sides results for m_1 in :

$$\ln m_1 = \ln\left(\frac{R^{x_1} G^{x_2}}{R^{x_2} G^{x_1}}\right) = \ln R^{x_1} + \ln G^{x_2} - \ln R^{x_2} - \ln G^{x_2} =$$

$$\ln R^{x_1} + \ln G^{x_2} - (\ln R^{x_2} + \ln G^{x_2}) =$$

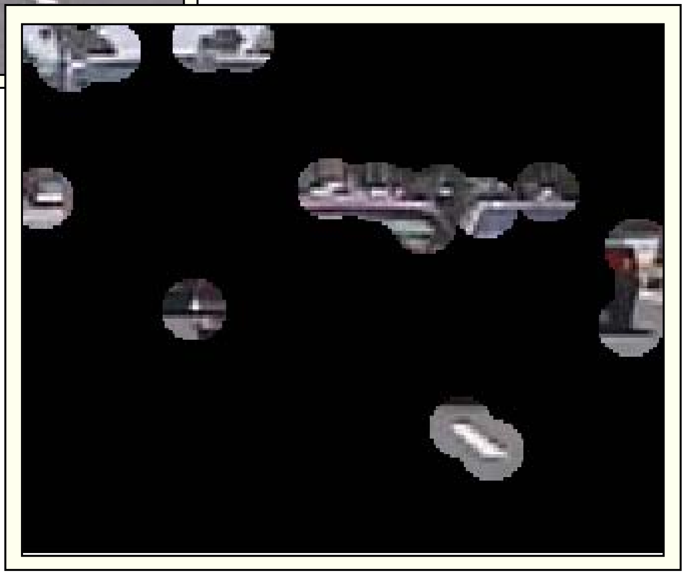
$$\ln\left(\frac{R^{x_1}}{G^{x_1}}\right) - \ln\left(\frac{R^{x_2}}{G^{x_2}}\right) = \ln\left(\frac{R}{G}\right)^{x_1} - \ln\left(\frac{R}{G}\right)^{x_2} = \frac{\partial}{\partial x} \ln\left(\frac{R}{G}\right)$$

Color constancy: Derivatives

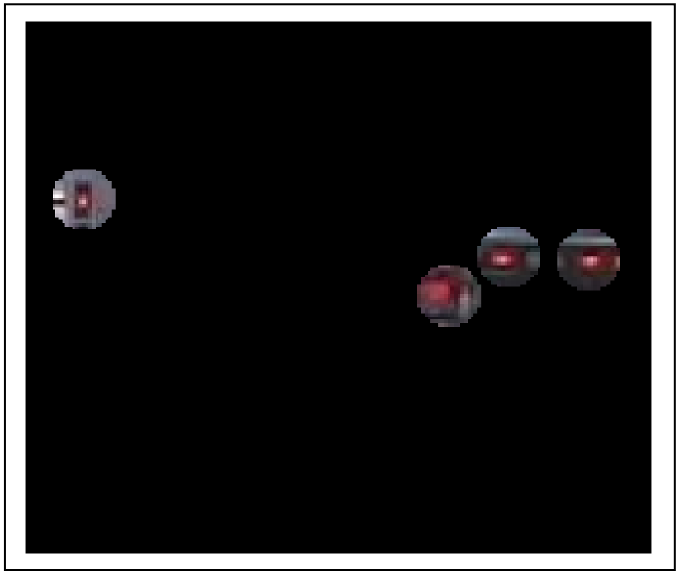
Funt and Finlayson (Mondrian-world)	Gevers and Smeulders (3D world)
$\begin{pmatrix} \frac{\partial}{\partial x}\ln R \\ \frac{\partial}{\partial x}\ln G \\ \frac{\partial}{\partial x}\ln B \end{pmatrix}$	$\begin{pmatrix} \frac{\partial}{\partial x}\ln\left(\frac{R}{G}\right) \\ \frac{\partial}{\partial x}\ln\left(\frac{R}{B}\right) \\ \frac{\partial}{\partial x}\ln\left(\frac{G}{B}\right) \end{pmatrix}$

Saliency boosted points

input car-image

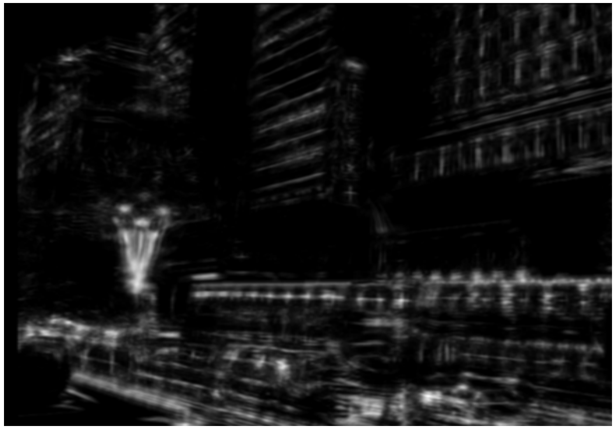


RGB-based (first 20 points)

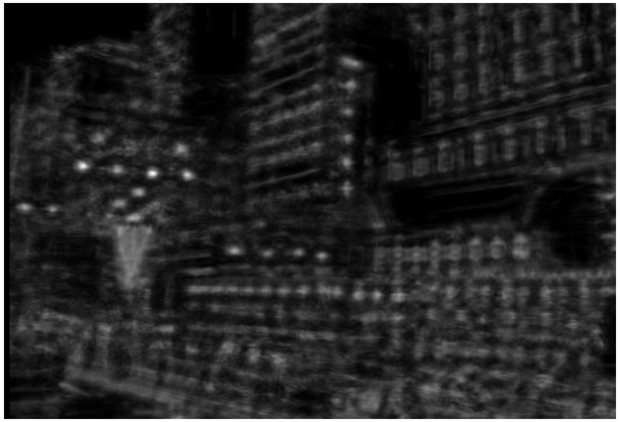


saliency boosting (first 4 points)

Saliency boosted points



RGB-based



saliency boosting

Research - Approach

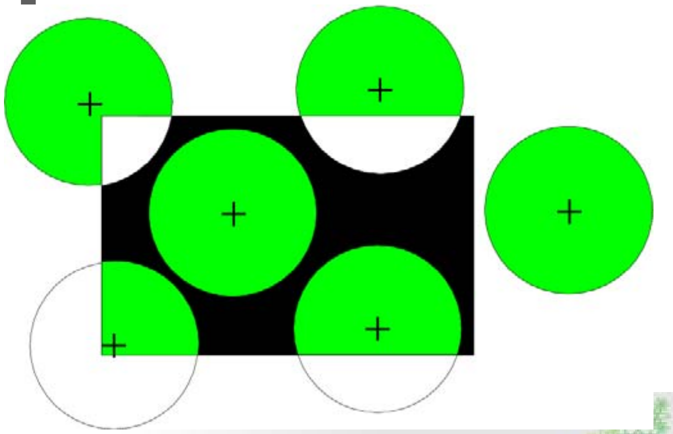
- **Use different corner detectors in the framework**
  - Intensity: Harris, SUSAN
  - Color: 2 colorHarris variants (colOppHarris, colRatHarris)
- **Evaluation**
  - Repeatability under common transformations (invariance)
    - Test sets for different common variations in imaging conditions
      - Blur, Lighting, Rotation/Scaling, viewing angle, JPEG compression
  - Information content of the detected regions (distinctiveness)
    - Detect lots of regions, estimate entropy from them.
  - Complexity

Intensity-based corner detectors

- **Harris corner detector**
  - Second moment matrix (SMM) at certain scale
  - Eigenvalues of SMM represent principal curvatures
    - Detect regions with high gradient in different directions

$$C_H(\mu) = \det(\mu) - \alpha \mathrm{trace}^2(\mu)$$

- **Discrete low-level corner detector [Smith 97]**
  - Fundamentally different from Harris detector
  - Circular mask
  - Determine the area of the mask with a similar value as the center
    - Derive cornerness measure from it



Experimental results

- **Repeatability**
  - For each common transformation a number of test sets
  - Each test sets contains 6 images
    - Gradual increase of transformation between images
    - Related by homography to establish a ground truth
      - Detected regions are projected onto the first image of set
  - Next: 4 test sets and repeatability results
    - Impression of overall performance

Experimental results - Repeatability/blur

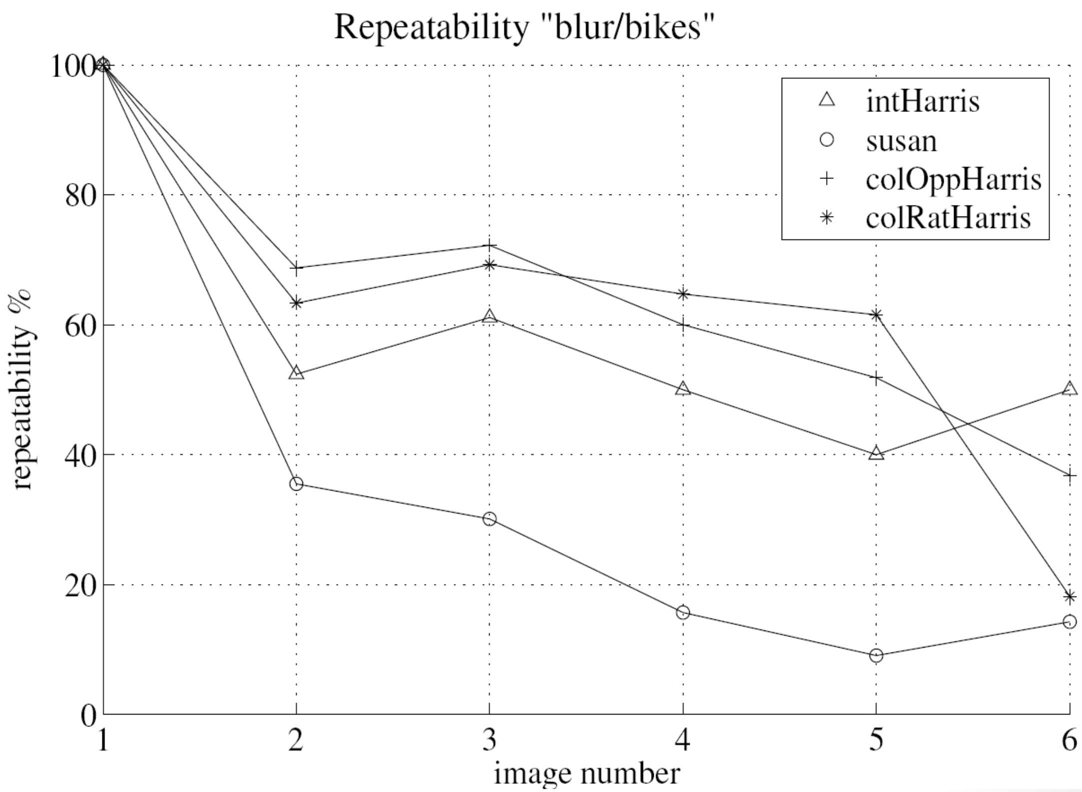
1

...

4

...

6



Experimental results - Repeatability/light

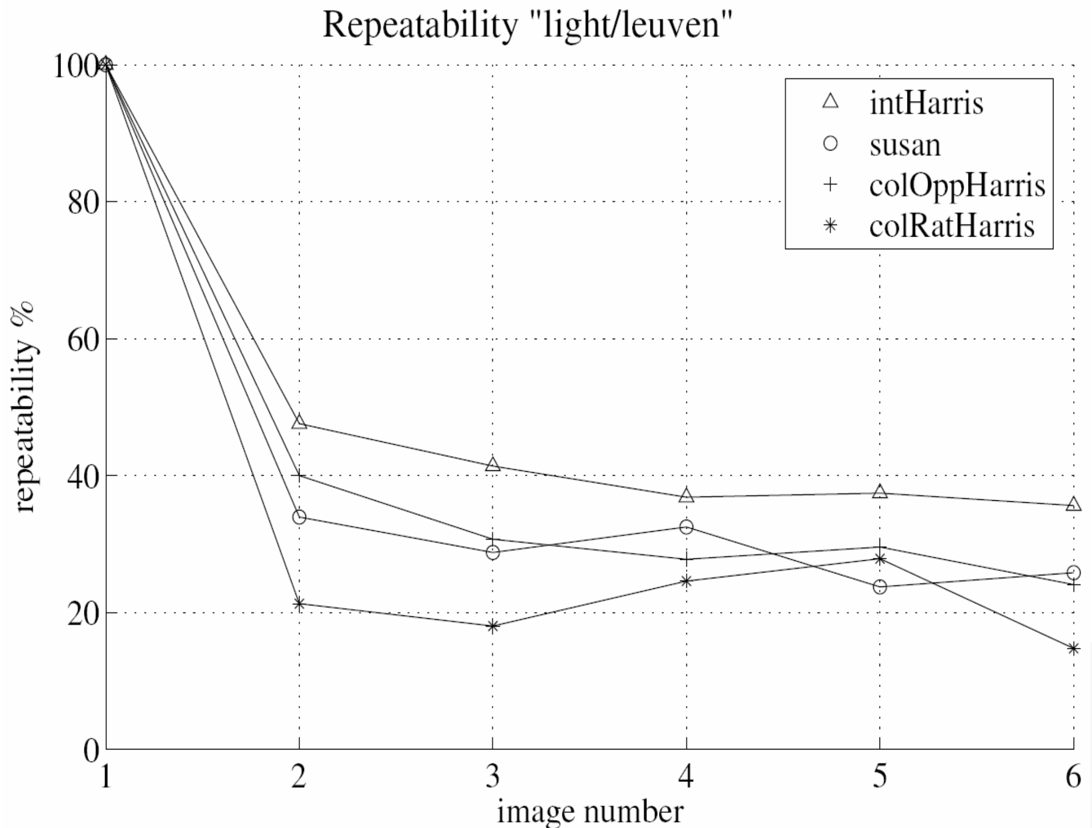
1

...

4

...

6



Experimental results - Repeatability/viewpoint

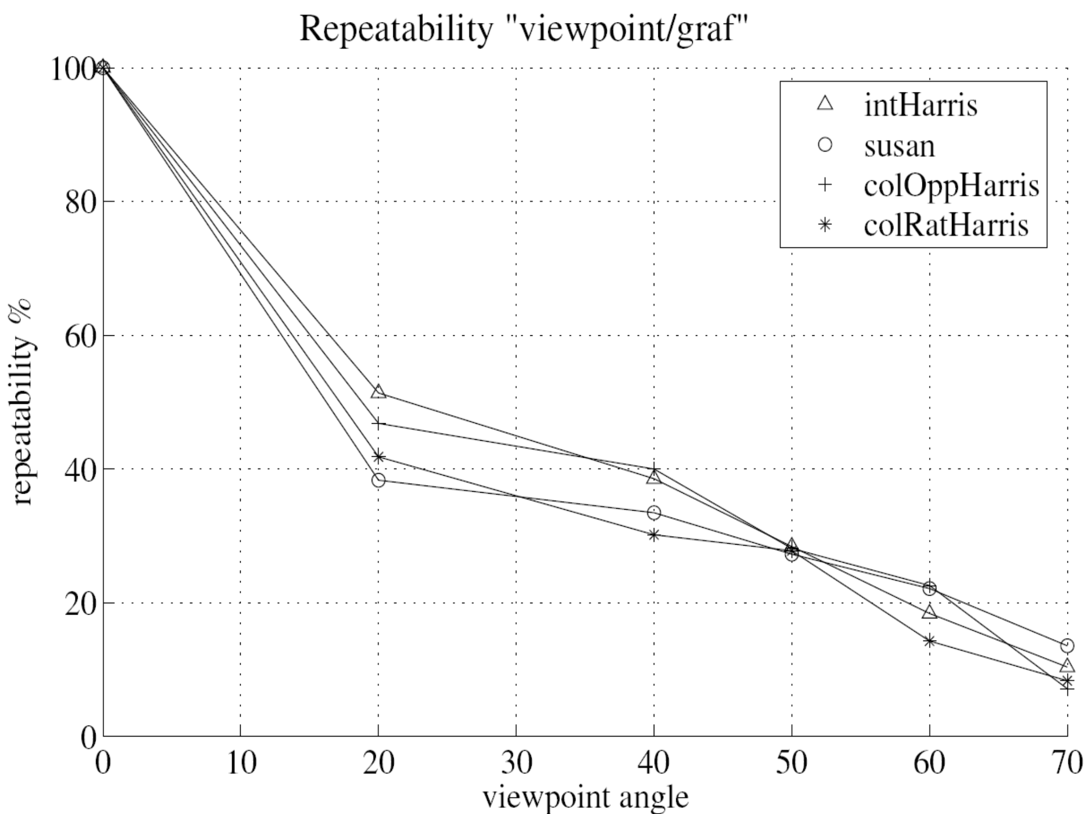
1

...

4

...

6



Experimental results - Repeatability/JPEG

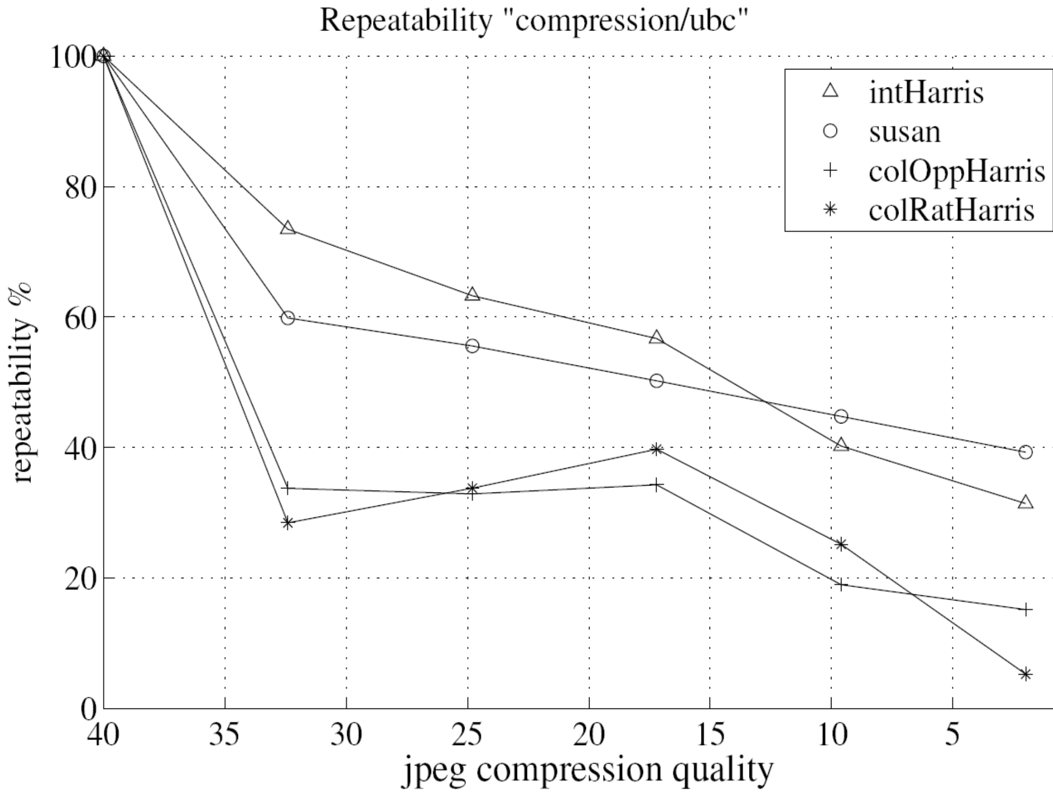
1

...

4

...

6



Experimental results - Information content

- **Distinctiveness of the regions detected**
  - Create descriptors and estimate entropy

Detector	Entropy
intensity Harris	11.41
SUSAN	11.23
colOppHarris	13.41
colRatHarris	13.96
random	9.24

- **Probability to produce a collision when matching is 7.4 times higher for intensity than for color.**

Experimental results - Complexity

- **Detection complexity**
  - Intensity Harris and SUSAN approximately equal
  - colorHarris using *n* color channels
    - *n* times more expensive compared to intensity only
- **Matching complexity**
  - colorHarris tends to need less regions to perform optimal
    - Lower matching complexity

Experimental results

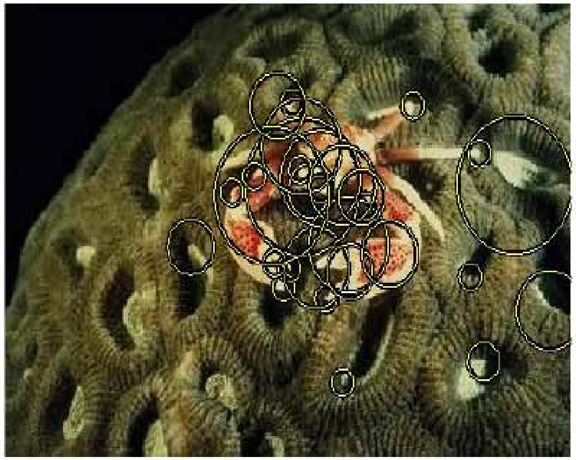
(a) *RGB* (b) *rgb* (c) Spherical colour space (d) *OCS* (e) *HSI* (f) colour boosted *OCS*

Experimental results

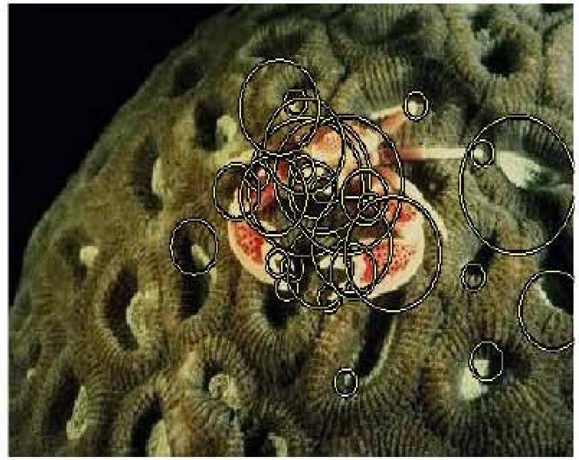
(a) Illumination (b) HSI

Figure 3: 30 extracted regions based on luminance and HSI information with $t = 1.2; s = 10; \sigma_I = 0.7$. The regions shift towards colour differences, specular, and shading changes are not regarded anymore. The parrot is therefore highly prioritized.

Experimental results

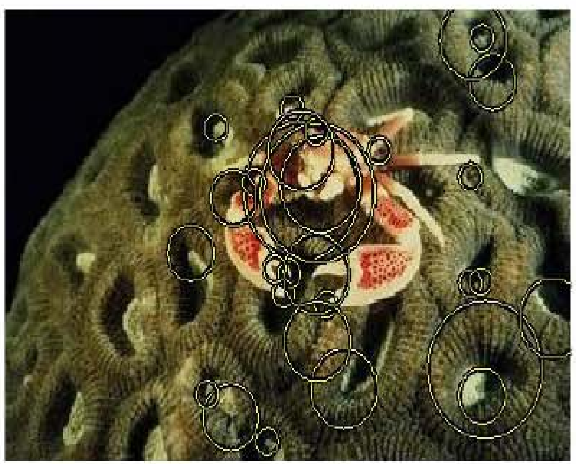


(a) Illumination

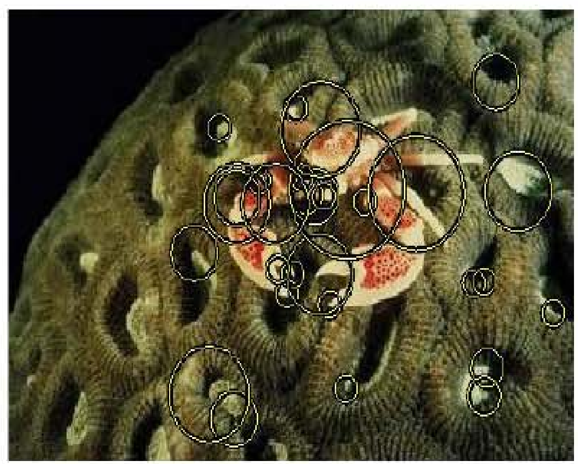


(b) *RGB*

(c) *rgb*



(d) *OCS*



(e) colour boosted *OCS*

(f) Quasi invariant *HSI*

What's Next? Using Context Information

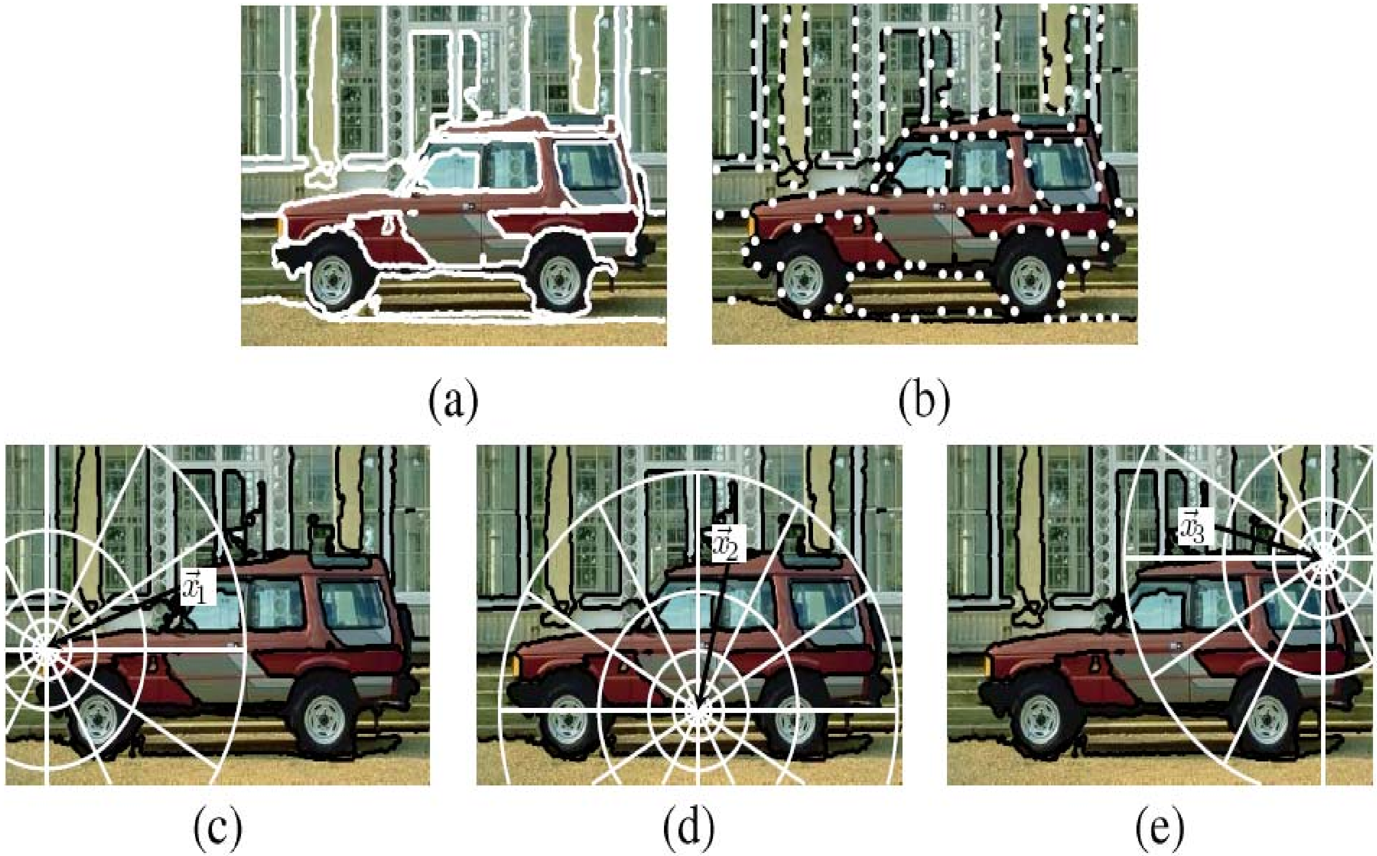


Fig. 1. (a) Dense cloud of points at contours of the image (in black). (b) Sampled set of points taken as reference (in white). (c)-(e) Log-polar spatial quantization of our descriptor given three different references $\vec{x}_1$, $\vec{x}_2$, $\vec{x}_3$. The image representation is a set of descriptors, one for each reference point $\vec{x}_i$

What's Next? Using Context Information

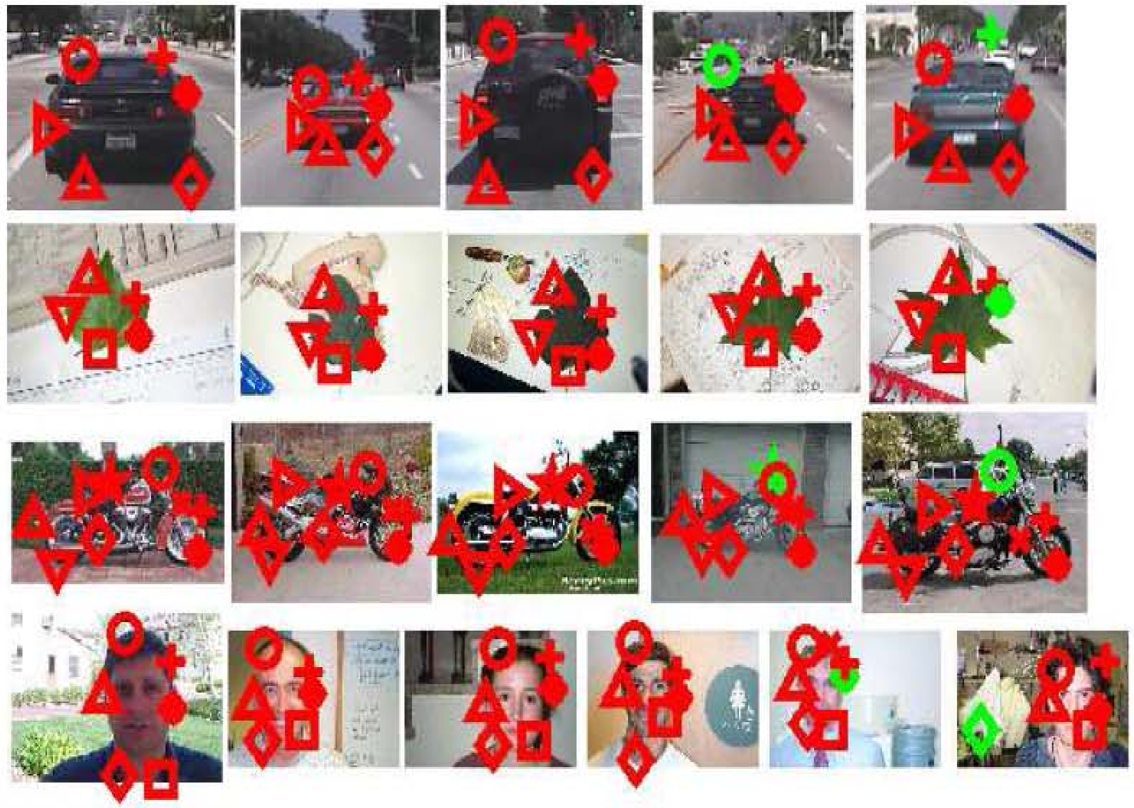


Fig. 5. Matching model parts across images. Different parts are shown with different symbols, where we sampled a few matchings for clarity.

Publications

- **Context-based object class recognition and retrieval by generalized correlograms**
  - J. Amores, P. Radeva, N. Sebe, IEEE Trans. PAMI (to appear)
- **Color interest points for image retrieval**
  - J. Stottinger, N. Sebe, A. Hanbury, T. Gevers, Computer Vision Winter Workshop, Feb 2007
- **Do color interest points help image retrieval?**
  - J. Stottinger, N. Sebe, A. Hanbury, T. Gevers, submitted to ICIP
- **Object retrieval and recognition with color interest points**
  - J. Stottinger, N. Sebe, A. Hanbury, T. Gevers, submitted to ICCV