EEE422 (EEE6082) Computational Vision

Salient Features

Ling Shao

Overview

- Introduction
- Invariant Salient Regions Detection
  - ➢ Algorithm
  - ➢ Performance Evaluation
- Invariant Salient Regions Based Image Retrieval
  - ➢ Specific Object Retrieval
  - ➢ Object Category Retrieval

Introduction

- Good performance in many computer vision tasks often depends upon the reliable selection of a sufficient number of image regions or salient features.
- The state of the art region detectors:
  - ➢ Harris Affine (Mikolajczyk and Schmid, 2002)
  - ➢ Maximal Stable Extremal Regions (Matas et al., 2002)
  - ➢ Intensity Extrema Based Detector (Tuytelaars and Van Gool, 2000)
  - ➢ Edge Based Detector (Tuytelaars and Van Gool, 1999)

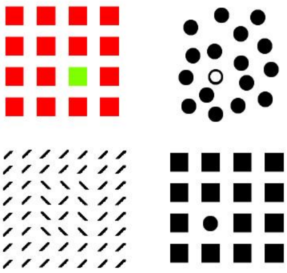
  The above region detectors do not select the most informative regions.

  They do not rank the detected regions according to their importance.
- To design a region detector that can rank the selected regions according to their saliency.

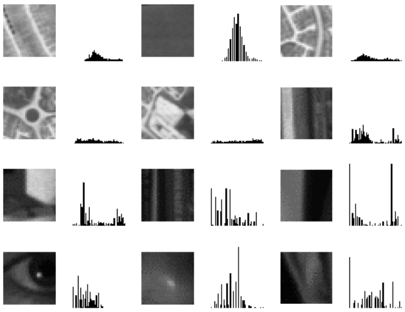
Salient Regions Detection

- The most informative regions
- A ranking algorithm
- Invariance to global changes in imaging conditions
  - – e.g. Viewpoint, photometric changes
- Robustness to local perturbations in the image function
  - – e.g. noise, motion
  - – changes in the background do not affect features
- Repeatability under intra-class variations of scene
  - – Similar scene contents produce similar results
  - – e.g. a detector firing on the door of one particular washing machine model should respond well to other models.

Saliency: “pop out”



Local histograms of intensity for some image regions

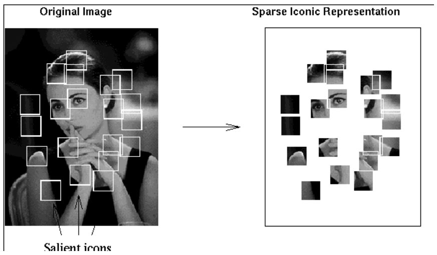


The histograms have the topology of a torus.

Uniform parts have peaked histograms

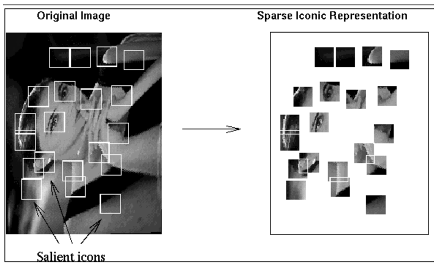
structured parts have flatter histogram, more complexity

Recognition using salient icons

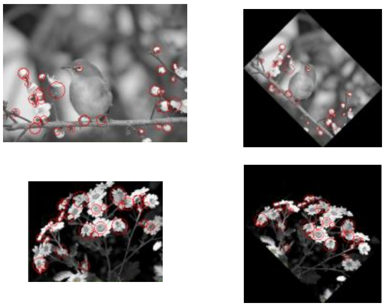


Chopped up Emanuelle B éart

Emanuelle does an affine roll



Invariant Salient Regions Detection



Invariant salient regions detection

- The algorithm considers an image region to be salient if it is simultaneously unpredictable in some feature-space and over scale.
- A region is quantized by the product of two items:

$$Y_D(s_p, \mathrm{x}) = H_D(s_p, \mathrm{x}) \times W_D(s_p, \mathrm{x})$$

Feature-space saliency:

$$H_D(s, \mathrm{x}) = \int_{d \in D} p(d, s, \mathrm{x}) \log_2 p(d, s, \mathrm{x}) \cdot \mathrm{d}d$$

Inter-scale saliency:

$$W_D(s, \mathrm{x}) = s \cdot \int_{d \in D} \left| \tfrac{\partial}{\partial s} p(d, s, \mathrm{x}) \right| \cdot \mathrm{d}d$$

Scale selection:

$$s_p = \left\{ s : \frac{\partial H_D(s, \mathrm{x})}{\partial s} = 0, \frac{\partial^2 H_D(s, \mathrm{x})}{\partial s^2} < 0 \right\}$$

Feature-space saliency - discrete

- Local entropy of a descriptor:

$$H_D(s, \mathrm{x}) = -\sum_{d \in D} p_{d,s,\mathrm{x}} \log_2 p_{d,s,\mathrm{x}}$$

The descriptor could be colour, orientation, phase, etc.

Histograms are used to approximate the PDFs.

Inter-scale saliency - discrete

$$W_D(s, \mathrm{x}) = \frac{1}{2}\left(\frac{N_s}{N_s - N_{s-ds}} \sum_{d \in D} \left(\left| p_{d,s,\mathrm{x}} - p_{d,s-ds,\mathrm{x}} \right| \right) + \frac{N_{s+ds}}{N_{s+ds} - N_s} \sum_{d \in D} \left(\left| p_{d,s+ds,\mathrm{x}} - p_{d,s,\mathrm{x}} \right| \right) \right)$$

where *ds* represents a small change in scale and *Ns* is the number of pixels in the sampling windows at scale *s*.

Pixel sampling

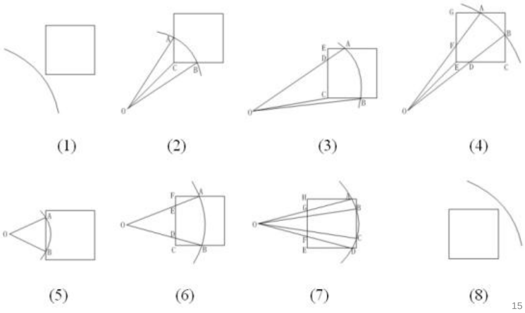
- The pixels in the central part of the sampling window are more reliable than those on the edge, hence we want to rely more on the central part.
- The difference in the number of pixels between consecutive scales is not the same at different scales. It follows that the inter-scale saliency , which is calculated from the difference of pixels between consecutive scales, is likely to be less reliable for smaller scales.
- In order to enlarge the difference between consecutive scales and to weight the central pixels over those at the edge, we use two-dimensional Gaussian functions to sample pixels within a circular region.

Partial Volume Estimation

- The discretised circular sampling window gives rise to step changes in the histogram as the scale is increased.
- Our solution to this problem is to define a smoother transition between those pixels that are included in the histogram and those that are not. We use partial volume estimation: those pixels at the sampling window edges are weighted by the area of intersection between the window and the pixel.

Partial Volume Estimation

Different alignments between the sampling window and a pixel



Parzen windowing and bin interpolation

- For a particular scale s, the PDF is estimated from N(s)=pi*sqrt(s) pixels. For an image with (say) 256 grey-scales, only a small number of intensities would have value, especially when scale s is small. Hence, in histogramming, if the horizontal axis takes values from 0 to 255 i.e. the number of bins is 256, many of the bins may be empty.
- With Parzen windowing and bin interpolation, the value of each intensity is spread among the neighbouring bins, hence fewer empty bins are left.

Performance evaluation - repeatability

- Repeatability is a standard criterion for measuring the robustness of a region detector;
- Repeatability rate is the percentage of the total observed regions that are detected in both images.
- The test sequences used include:
  - Viewpoint
  - Illumination
  - Scaling
  - Image blur
  - Compression

Repeatability Rate

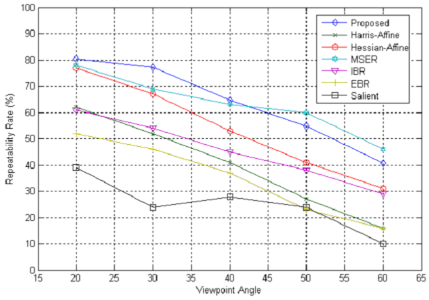
$$r = \frac{C(I_1, I_2)}{N}$$

- The detected regions are first rescaled to the same size;
- We consider two regions to be corresponding if the error of area covered by two regions is less than 40%.

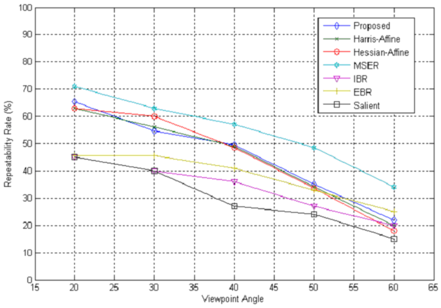
Examples of dataset for comparison of different region detectors

(a), (b) Viewpoint change; (c), (d) Scaling + rotation; (e), (f) Image blur; (g) JPEG compression; (h) Illumination change.

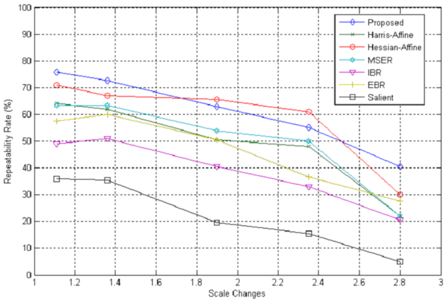
Viewpoint change for structured scene



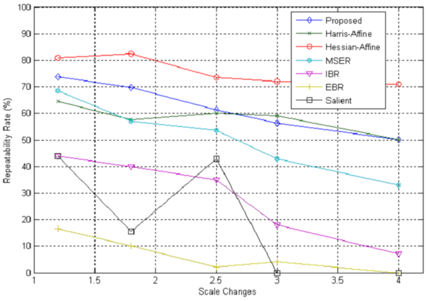
Viewpoint change for textured scene



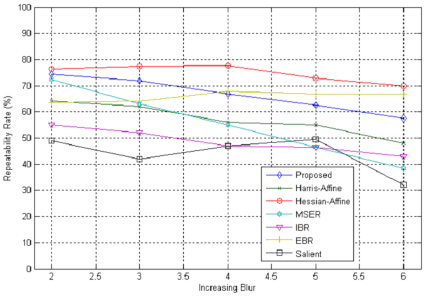
Scale change for structured scene



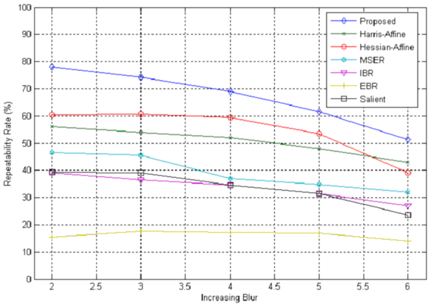
Scale change for textured scene



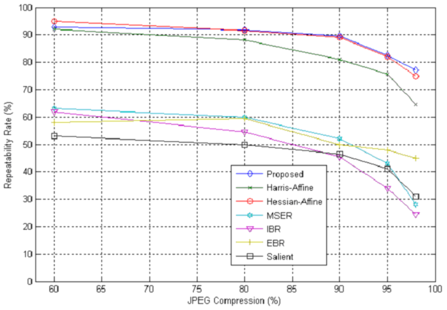
Blur for structured scene



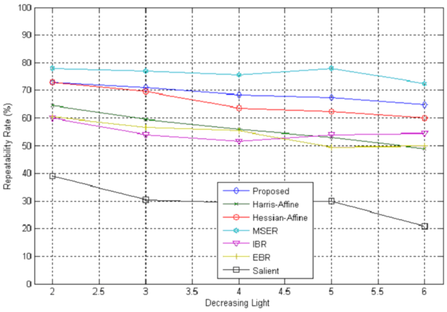
Blur for textured scene



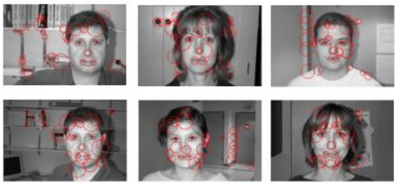
JPEG compression



Illumination change

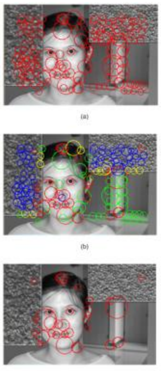


Feature Classification



- Blobs
- Edges and Lines
- Texture
- Texture Boundary

Feature classification



(a)

(b)

(c)

Content based image retrieval

With the emergence of multimedia and the advancement of Internet technologies, a huge amount of images are generated and stored every day. Content-Based Image Retrieval (CBIR) has attracted increasing attention in the computer vision and multimedia database communities. The features used by most existing image retrieval systems include:

- Colour
- Shape
- Texture
- Spatial Layout

The main drawback of such systems is that they use features that are only relevant to a narrow image domain. For the broad domain of image retrieval, we require more advanced features which are effective for generic images.

Region Segmentation

- Global features are not effective
- Most features in CBIR are extracted from segmented regions
- The region-based Image Retrieval techniques are highly dependent on the accuracy of the region segmentation
- No generic segmentation algorithm
- Segmented regions are homogeneous thus contain little information content

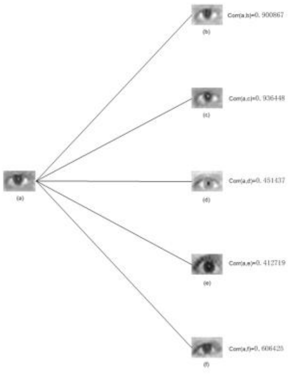
Our feature – salient regions

- The most informative regions
- Invariant to geometric and photometric transformations
- Invariant to intra-class variations
- Generic image retrieval

Specific object retrieval

- The purpose of specific object retrieval is to distinguish specific objects of the same category, e.g. human faces
- Each instance of one object category has some unique features
- Two datasets are used for experiments: the first is the Caltech Human Face Dataset, which contains 435 images of different people taken at different background with different expressions; the second is the CMU VASC Motion Dataset, which contains a large set of motion series of different objects from which 1301 images are selected.

Correlation between Regions



Specific object retrieval algorithm

- A number of salient regions are detected from the images, and normalised for scale and illumination
- Compute the correlations between the regions of each query image and the regions of each image in the database
- The Similarity Score between each query image and each image in the dataset is computed
- Images with Similarity Scores higher than a threshold are considered to contain the same object as the query

Experiment: Caltech Human Face Dataset

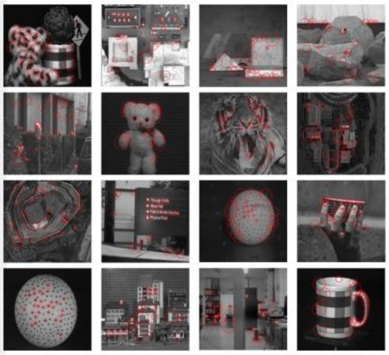
- True Positive: 80.5%
- False Positive: 3.14%

Caltech Face Dataset - Benchmarking

- Histogram intersection
- Harris affine + SIFT descriptor

Algorithms	Histogram Intersection	Harris Affine + SIFT Descriptor	Proposed
Retrieval Rates	17.9%	57.8%	80.5%

Experiment: CMU VASC Motion Dataset



- True Positive: 100%
- False Positive: 1.6%

CBIR under viewpoint and illumination changes

- Moment invariants are used as region descriptor;
- Moment invariants are invariant to geometric and photometric transformations;
- Traditional moment invariants and generalised colour moment invariants;
- Suitable for object category retrieval.

Hu’s moments

$$M_{pq} = \sum_x \sum_y x^p y^q f(x, y)$$

$$\overline{x} = \frac{M_{10}}{M_{00}} \qquad \overline{y} = \frac{M_{01}}{M_{00}}$$

Central Moments:

$$\mu_{pq} = \sum_x \sum_y (x - \overline{x})^p (y - \overline{y})^q f(x, y)$$

Normalisation to scaling:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}$$

Hu’s moment invariants

$$\phi_1 = \eta_{20} + \eta_{02}$$

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

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

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

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

Generalised colour moment invariants

$$M_{pq}^{abc} = \sum_x \sum_y x^p y^q [R(x, y)]^a [G(x, y)]^b [B(x, y)]^c$$

- GPD invariants – affine geometric and diagonal photometric transformations
- GPSO invariants – affine geometric and scaling and offset photometric transformations

Specific Scene Retrieval

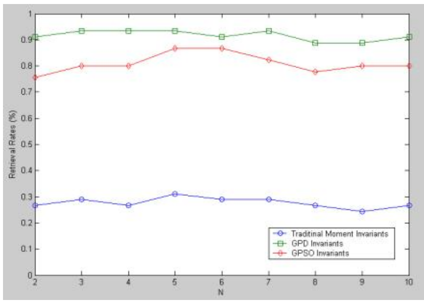
- A number of salient regions are selected for each query image and each image in the dataset using the salient region detection method;
- Moment invariants are computed for each selected region for both the query images and the test image in the database;
- The distances of invariant vectors between regions in the test image and regions in the query image are calculated and sorted; The average value of the *N* smallest distances is used as the dissimilarity measure between the test image and the query image;
- Compare the dissimilarity measures of the test image and all the query images, and the smallest dissimilarity measure indicates the test image contains the same scene as that particular query image.

Experiment: specific scene retrieval

- Traditional moment invariants: 31.1%
- GPD invariants: 93.3%
- GPSO invariants: 86.7%
- Harris affine + SIFT: 90.7%

Variations of the images

Retrieval Results with Different *N*



Object Category Retrieval

- *T* images of each category are randomly selected from the dataset as training data for that object category; the remaining images are used for retrieval;
- 100 salient regions are selected for each image in the training data and each image in the experimental dataset using the salient region detection method;
- For a particular test image in the experimental dataset, the moment invariants are computed for each region selected; The same type of moment invariants are also computed for each region of all the images in the training data;
- The distances of invariant vectors between regions in the test image and regions in each category of training images are calculated and sorted; The average value of the 10 smallest distances is used as the dissimilarity measure between the test image and that particular object category;
- Compare the dissimilarity measures of the test image and all the object categories, and the smallest dissimilarity measure indicates the test image belongs to that particular object category.

Experiment: object category retrieval

- Traditional moment invariants: 28.6%
- GPD invariants: 87.9%
- GPSO invariants: 84.2%
- Harris affine + SIFT: 78.4%

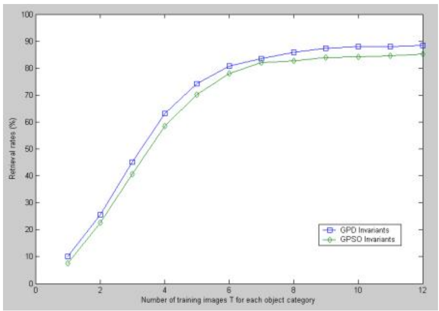
GPD invariants

	Cat	Elephant	Emu	Crocodile	Crayfish	Hawksbill	Leopard	Kangaroo
Cat	52	0	1	0	0	0	4	2
Elephant	1	47	0	1	0	0	2	3
Emu	0	1	38	2	0	1	0	1
Crocodile	0	0	1	34	2	3	0	0
Crayfish	0	0	0	1	52	7	0	0
Hawksbill	0	1	0	2	8	78	1	0
Leopard	13	0	2	1	0	0	171	3
Kangaroo	5	2	2	1	0	0	0	66

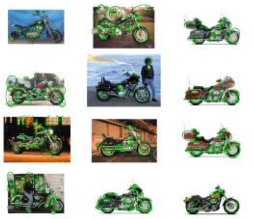
GPSO invariants

	Cat	Elephant	Emu	Crocodile	Crayfish	Hawksbill	Leopard	Kangaroo
Cat	48	2	0	0	0	0	6	3
Elephant	3	47	1	0	2	0	0	1
Emu	0	0	35	2	0	0	1	5
Crocodile	1	0	0	34	1	4	0	0
Crayfish	0	0	0	2	48	9	0	1
Hawksbill	0	0	0	1	11	75	3	0
Leopard	15	0	0	1	2	4	166	2
Kangaroo	3	2	5	0	1	0	3	62

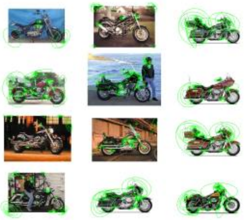
Retrieval Results with Different Numbers of Training Images



Affine invariant salient regions under intra-class variations

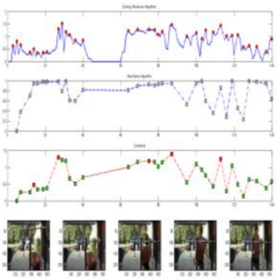


Affine invariant salient regions respond more consistently to similar scene contents e.g. Bike wheels



Affine Interest Points
Features are affected by background changes and intra-class variations.

Extension: Salient frame extraction



- Entropy maxima: *selecting frames containing the most complex motions*
- Histogram intersection: *calculating motion variations between consecutive frames*
- Combined: *salient frames*