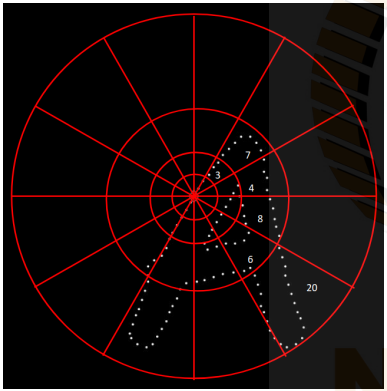
Deep Learning for Computer Vision

Other Feature Spaces

Vineeth N Balasubramanian

Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

Shape Context

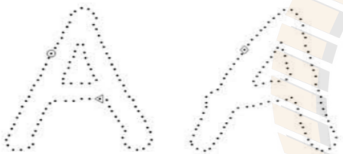


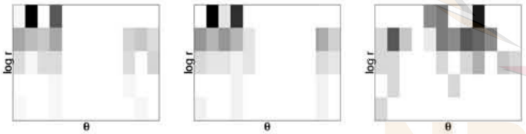
Log-polar binning over 12×5 bins

- **Log-polar representation:** Co-ordinate system in two dimensions parametrized by logarithmic distance from origin and angle.
- For each point taken on edge of shape, count number of points in each log-polar bin.
- More precision for nearby points, more flexibility for farther points.
- It is translation- and scale-invariant

Credit: Derek Hoiem, UIUC

Shape Context





Histograms over Log-polar binning for points ○,◇, ◁ respectively

- As graphs of ○ and ◇ points match, they are correspondence points between the two As.
- Two points said to be **in correspondence** if they minimize value of C_{ij} given by:

$$C_{ij} = \frac{1}{2}\sum_{k=1}^{K}\frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)}$$

where h_i and h_j are histogram representations of two points i and j across figures and K is total number of bins $(12 \times 5 = 60)$

Shape Context

- One-to-one point matching is done sequentially w.r.t. correspondence point
- Distance less than a threshold implies shape match

MSER: Maximally Stable Extremal Regions

- Method for blob detection in images, based on Watershed segmentation algorithm
- Identify regions in image that stay nearly the same through wide range of gray-level thresholds
- Sweep threshold of intensity from black to white, performing a simple luminance thresholding of image
- Extract connected components (**Extremal Regions**)
- Region descriptors serve as features

MSER: Maximally Stable Extremal Regions

- Method for blob detection in images, based on Watershed segmentation algorithm
- Identify regions in image that stay nearly the same through wide range of gray-level thresholds
- Sweep threshold of intensity from black to white, performing a simple luminance thresholding of image
- Extract connected components (**Extremal Regions**)
- Region descriptors serve as features

MSER: Methodology and Notation

- As we start thresholding the image about a grey-scale intensity level (g), regions (or blobs, a collection of pixels) appear

MSER: Methodology and Notation

- As we start thresholding the image about a grey-scale intensity level (g), regions (or blobs, a collection of pixels) appear
- As we decrease value of g, new regions appear, or regions at higher g values coalesce. Such regions can be depicted in a tree-like structure

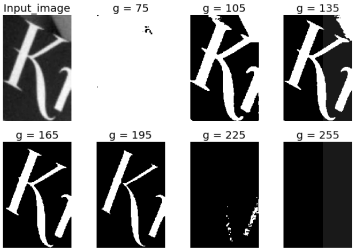
MSER: Methodology and Notation

- As we start thresholding the image about a grey-scale intensity level (g), regions (or blobs, a collection of pixels) appear
- As we decrease value of g, new regions appear, or regions at higher g values coalesce. Such regions can be depicted in a tree-like structure
- Regions at a particular g level denoted as: $R_1^g, R_2^g, ..., R_n^g$ where $|R_i^g| =$ total number of pixels in R_i^g
- Define $\Psi(\cdot)$ as:

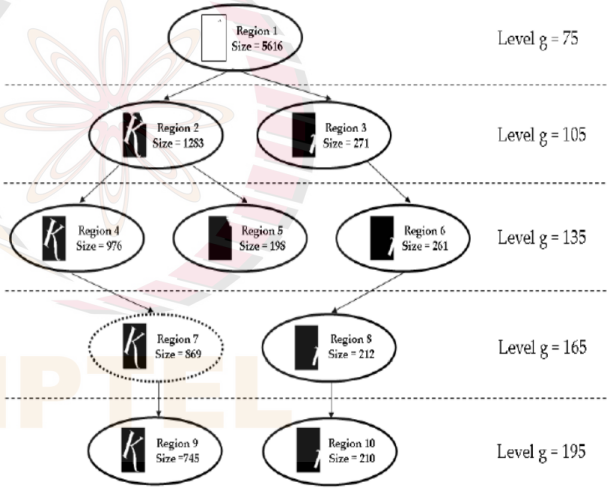
$$\Psi(R_i^g) = \frac{|R_j^{g-\Delta}| - |R_k^{g+\Delta}|}{|R_i^g|}$$

  where $\Delta =$ manually chosen intensity buffer, R_j and R_k are parent and child regions at levels $g - \Delta$ and $g + \Delta$ in the tree
- **MSER regions:** Regions where $\Psi(\cdot)$ is below a user-defined threshold

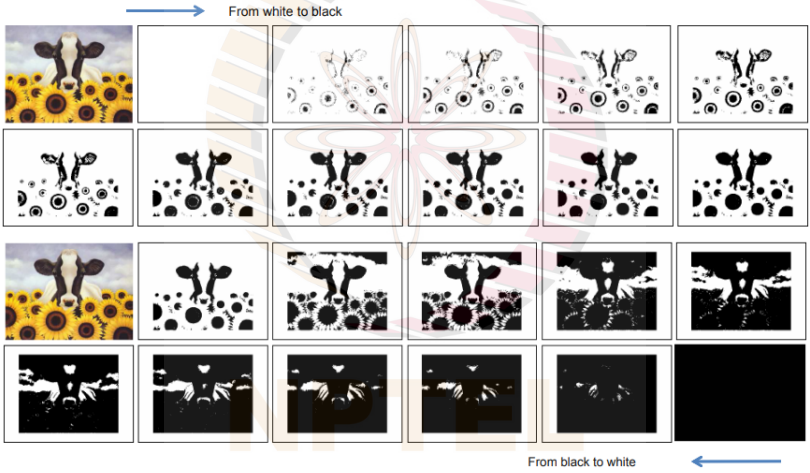
MSER: Example



Credit: Fred.A. Hamprecht

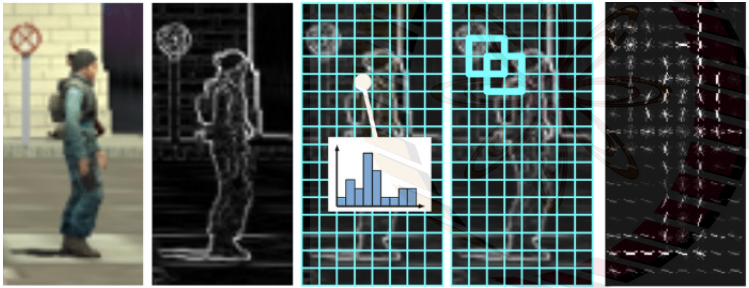


MSER: Example



Credit: Alberto Del Bimbo

Histogram of Oriented Gradients



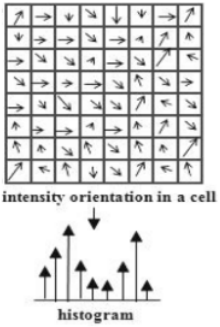
detection window slides over an image

at each location where the window is applied, gradients are computed

window is evently partitioned into cells and each pixel of the cell contirbutes to cell gradient orientation histogram

orientation histograms for overlapping 2x2 blocks of cells are normalized and collected to form the final descriptor
This is called Contrast Normalization

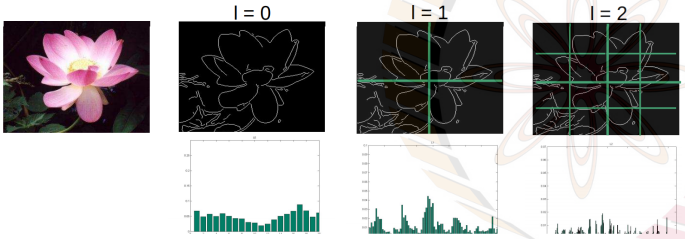
Final descriptor



Binning and weighted voting of magnitude.

Credit: Michał Olejniczak, Marek Kraft

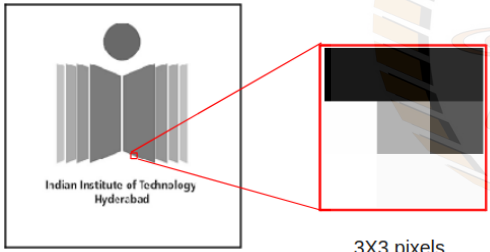
Pyramidal HoG (PHoG)



Credit: Bosch, Zisserman, Munoz 2007

- Divide image into $2^l \times 2^l$ cells at each pyramid level, l
- HOG descriptors with same dimension are extracted over each cell
- Final PHOG descriptor is concatenation of the HOGs at different pyramid levels
- Captures spatial relationship of oriented gradients better than HOG

Local Binary Patterns



3X3 pixels

116	116	128
254	209	155
255	252	252

Threshold 209

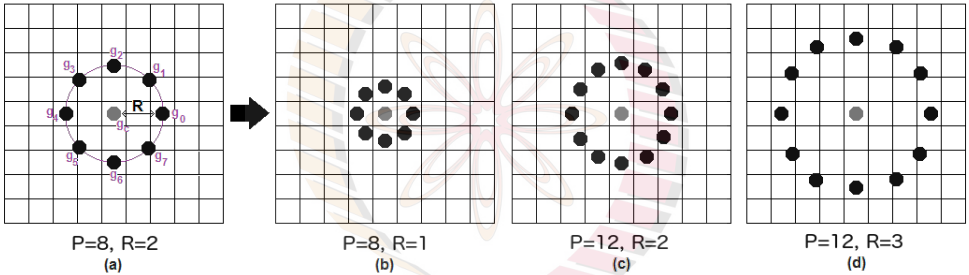
0	0	0
1	209	0
1	1	1

Binary 00001111

150	90	80
30	15	...
...	...	...

Decimal 15

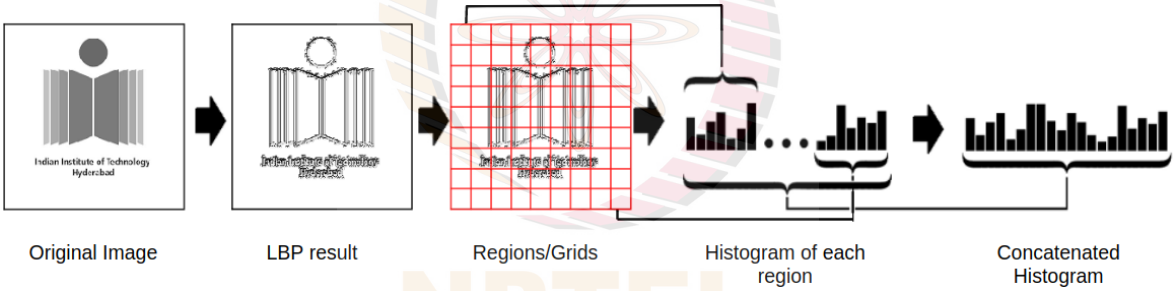
Local Binary Patterns



- Different radii (R) and neighbours (P) can be considered for per-pixel binary encoding
- Bilinear interpolation done if point shares pixels

Credit: Ojala, Pietikainen, Harwood 2007

Local Binary Patterns



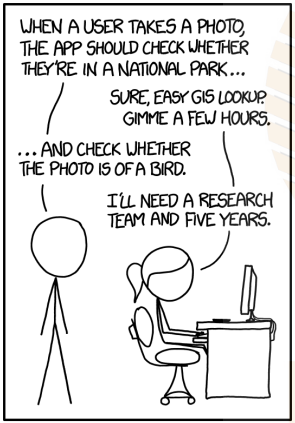
Comparison of Feature Detectors

Table 7.1 Overview of feature detectors.

Feature Detector	Corner	Blob	Region	Rotation invariant	Scale invariant	Affine invariant	Repeatability	Localization accuracy	Robustness	Efficiency
Harris	√			√			+++	+++	+++	++
Hessian		√		√			++	++	++	+
SUSAN	√			√			++	++	++	+++
Harris-Laplace	√	(√)		√	√		+++	+++	++	+
Hessian-Laplace	(√)	√		√	√		+++	+++	+++	+
DoG	(√)	√		√	√		++	++	++	++
SURF	(√)	√		√	√		++	++	++	+++
Harris-Affine	√	(√)		√	√	√	+++	+++	++	++
Hessian-Affine	(√)	√		√	√	√	+++	+++	+++	++
Salient Regions	(√)	√		√	√	(√)	+	+	++	+
Edge-based	√			√	√	√	+++	+++	+	+
MSER			√	√	√	√	+++	+++	++	+++
Intensity-based			√	√	√	√	++	++	++	++
Superpixels			√	√	(√)	(√)	+	+	+	+

Credit: Tuytelaars, Mikolajczyk 2008

...and along came Deep Learning



In the 60s, Marvin Minsky assigned a couple of undergrads to spend the summer programming a computer to use a camera to identify objects in a scene. He figured they'd have the problem solved by the end of the summer.

Over 45 years later, came along deep learning...

Credit: xkcd comics, https://xkcd.com/1425/

Homework

Readings

- Chapter 4.1.1, 4.1.2, Szeliski, *Computer Vision: Algorithms and Applications*

Further Readings

- Feature Detectors in Computer Vision: Wikipedia
- Shape Context: Wikipedia
- MSER: Wikipedia
- HoG: Wikipedia

References I

T. Ojala, M. Pietikainen, and D. Harwood. "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions". In: *Proceedings of 12th International Conference on Pattern Recognition*. Vol. 1. 1994, 582–585 vol.1.

Serge Belongie, Jitendra Malik, and Jan Puzicha. "Shape matching and object recognition using shape contexts". In: *IEEE transactions on pattern analysis and machine intelligence* 24.4 (2002), pp. 509–522.

Jiri Matas et al. "Robust wide-baseline stereo from maximally stable extremal regions". In: *Image and vision computing* 22.10 (2004), pp. 761–767.

Navneet Dalal and Bill Triggs. "Histograms of Oriented Gradients for Human Detection". In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*. CVPR '05. USA: IEEE Computer Society, 2005, 886–893.

Anna Bosch, Andrew Zisserman, and Xavier Munoz. "Representing Shape with a Spatial Pyramid Kernel". In: *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*. CIVR '07. Amsterdam, The Netherlands: Association for Computing Machinery, 2007, 401–408.

References II

Tinne Tuytelaars and Krystian Mikolajczyk. *Local invariant feature detectors: a survey*. Now Publishers Inc, 2008.

Richard Szeliski. *Computer Vision: Algorithms and Applications*. Texts in Computer Science. London: Springer-Verlag, 2011.