EXPLOITING SHAPE PROPERTIES FOR IMPROVED RETRIEVAL, DISCRIMINATION AND RECOGNITION

Vittal Premachandran

Thesis Defense Presentation

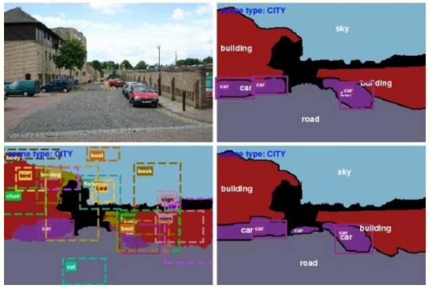
February 2014

Outline

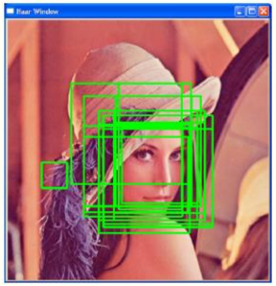
- Introduction
- Background
- Perceptually Motivated Shape Context
- What parts of a shape are discriminative?
- Shape similarity propagation on manifolds
- Object detection in images using shapes
- Summary

Introduction

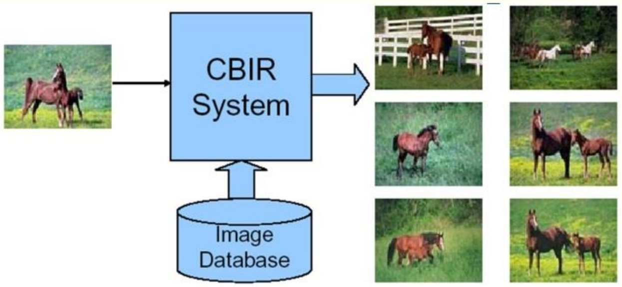
- Longstanding Goals of Computer Vision



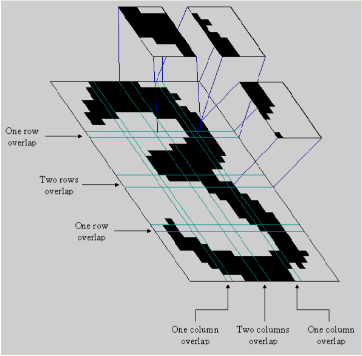
Scene Understanding



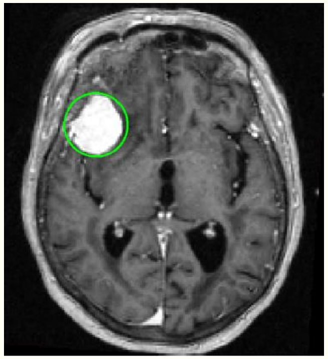
Face Recognition



Content-Based Image Retrieval



Optical Character Recognition

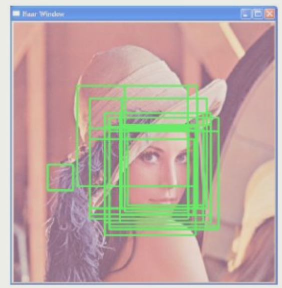


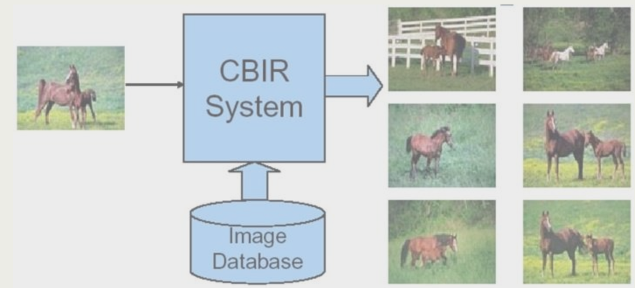
Tumor Detection

Common Task?

Introduction

- Longstanding Goals of Computer Vision



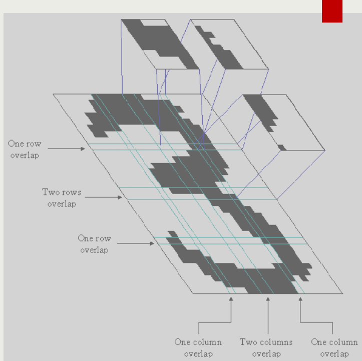


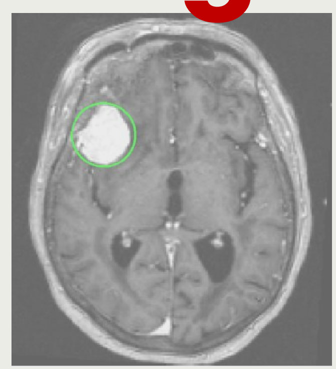
Scene Understanding

Face Recognition

Content Based Image Retrieval

Object Recognition!





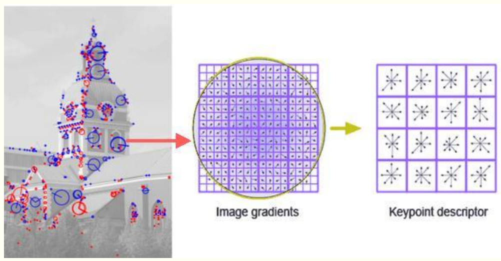
Common Task?

Optical Character Recognition

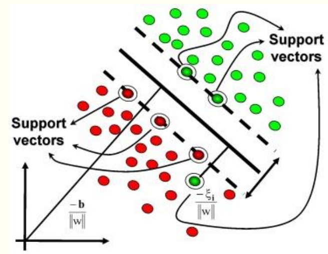
Tumor Detection

Object Recognition

- How can we recognize objects?



Extract Features



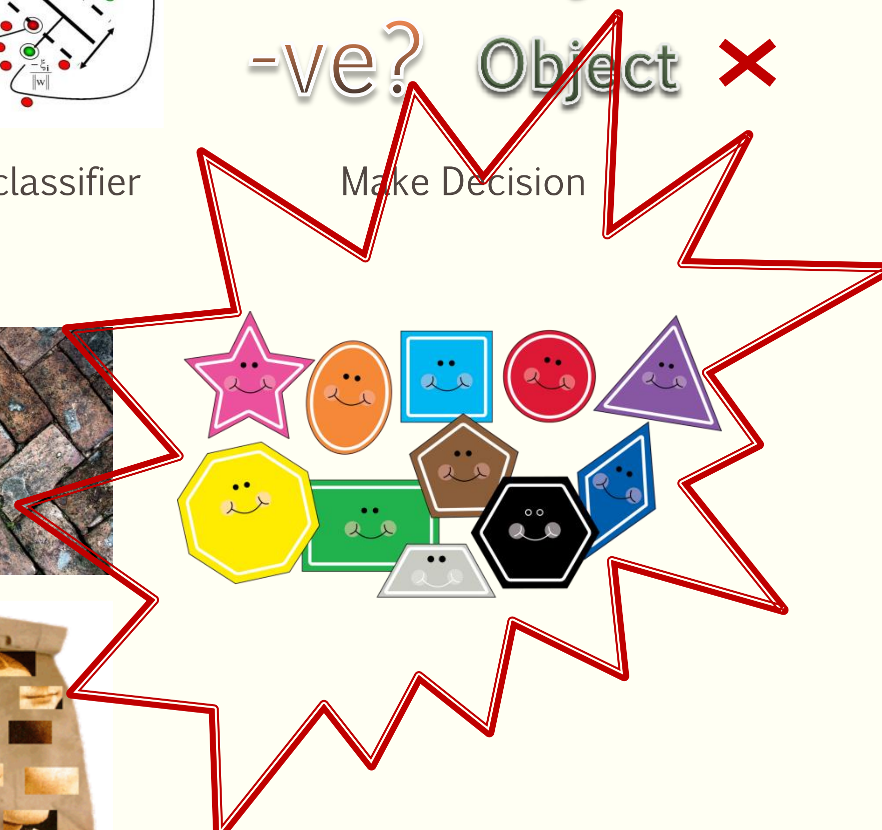
Train a classifier

+ve? Object ✓

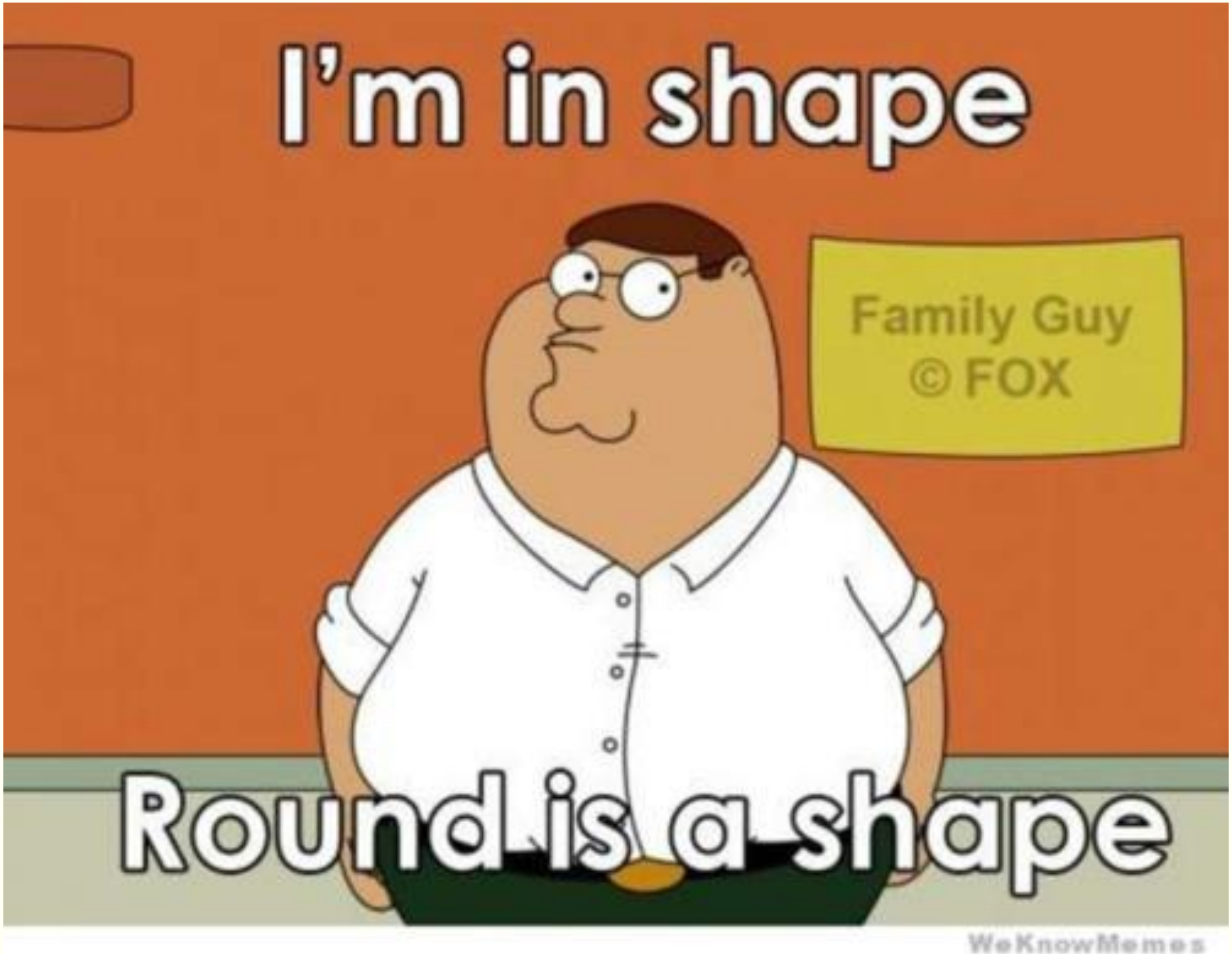
-ve? Object ✗

Make Decision

- What features to extract?



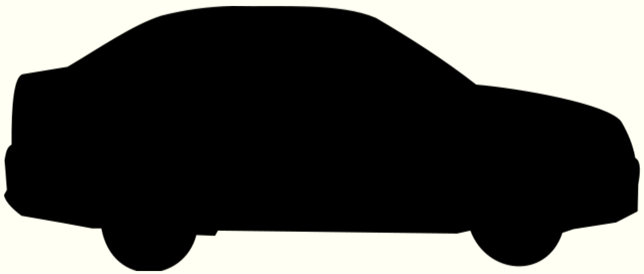
What is Shape?


I'm in shape
Family Guy
© FOX
Round is a shape
WeKnowMemes

“We here define ‘shape’ informally as ‘all the geometrical information that remains when location, scale and rotational effects are filtered out from an object.’”

– David George Kendall (1918-2007)
Statistician and Mathematician

Why use shapes?



- No colour, no texture; yet recognizable.
- Shape enables class-level recognition

- How to represent shapes?

Outline

- Introduction
- **Background**
- Perceptually Motivated Shape Context
- What parts of a shape are discriminative?
- Shape similarity propagation on manifolds
- Object detection in images using shapes
- Summary

Modern Computer Vision

Shape Matching and Object Recognition Using Shape Contexts

Serge Belongie, *Member*, *IEEE*, Jitendra Malik, *Member*, *IEEE*, and Jan Puzicha

Abstract—We present a novel approach to measuring similarity between shapes and exploit it for object recognition. In our framework, the measurement of similarity is preceded by 1) solving for correspondences between points on the two shapes, 2) using the correspondences to estimate an aligning transform. In order to solve the correspondence problem, we attach a descriptor, the *shape context*, to each point. The shape context at a reference point captures the distribution of the remaining points relative to it, thus offering a globally discriminative characterization. Corresponding points on two similar shapes will have similar shape contexts, enabling us to solve for correspondences as an optimal assignment problem. Given the point correspondences, we estimate the transformation that best aligns the two shapes; regularized thin-plate splines provide a flexible class of transformation maps for this purpose. The dissimilarity between the two shapes is computed as a sum of matching errors between corresponding points, together with a term measuring the magnitude of the aligning transform. We treat recognition in a nearest-neighbor classification framework as the problem of finding the stored prototype shape that is maximally similar to that in the image. Results are presented for silhouettes, trademarks, handwritten digits, and the COIL data set.

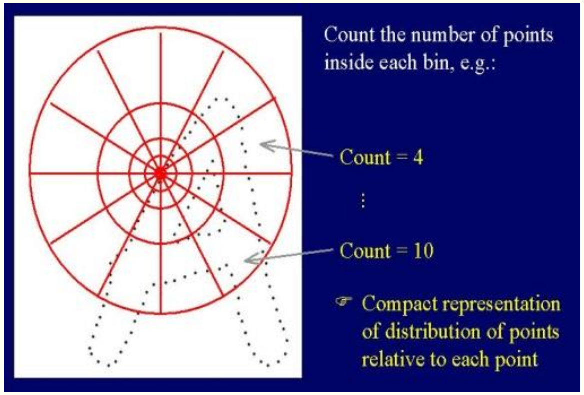
Index Terms—Shape, object recognition, digit recognition, correspondence problem, MPEG7, image registration, deformable templates.

Shape Classification Using the Inner-Distance

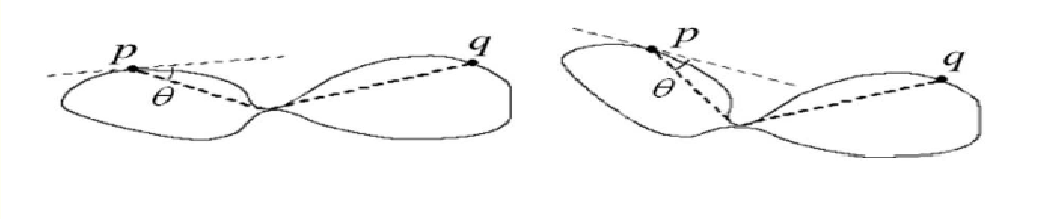
Haibin Ling David W. Jacobs
Center for Automation Research and Computer Science Department
University of Maryland, College Park
{hbling, djacobs}@ umiacs.umd.edu

Existing Methods

- Belongie, S.; Malik, J.; Puzicha, J.; , "Shape matching and object recognition using shape contexts," *PAMI 2002.*

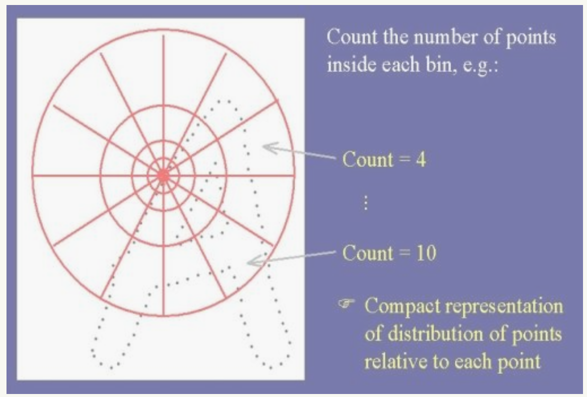


- Haibin Ling, David W. Jacobs, "Shape Classification Using the Inner-Distance", *PAMI, 2007.*
  - IDSC



Existing Methods

- Belongie, S.; Malik, J.; Puzicha, J.; , "Shape matching and object recognition using shape contexts," *PAMI 2002.*



Reliance on perfect contours!

- *PAMI, 2007.*
  - IDSC

Issues with Contour-Based Techniques

- Two assumptions of contour-based techniques
  - Noise-free contours
  - Closed contours

Contours have indentations and noise

Contours have breaks

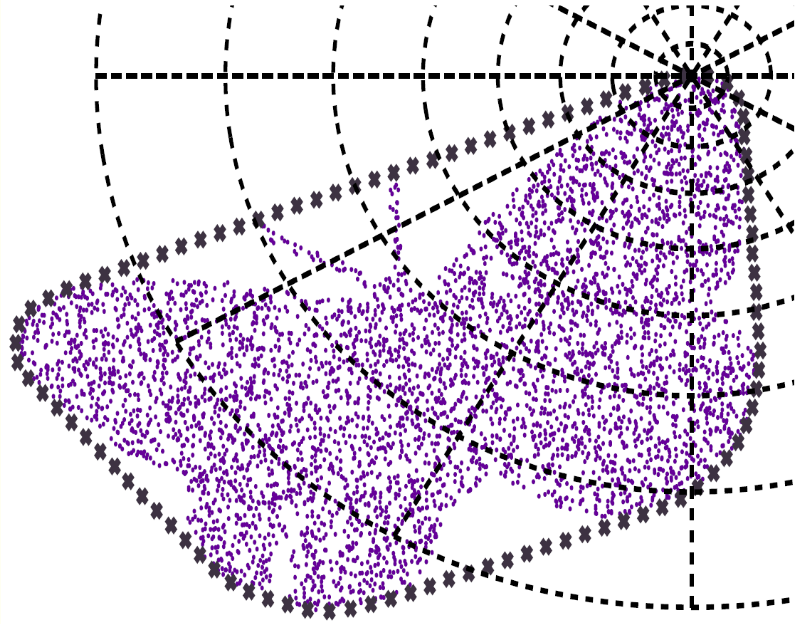
- Contours are important, but not always
- Need a way to capture entire shape; Gestalt Psychology
- Interior properties of a shape play an important role

Outline

- Introduction
- Background
- **Perceptually Motivated Shape Context**
- What parts of a shape are discriminative?
- Shape similarity propagation on manifolds
- Object detection in images using shapes
- Summary

Solid Shape Context

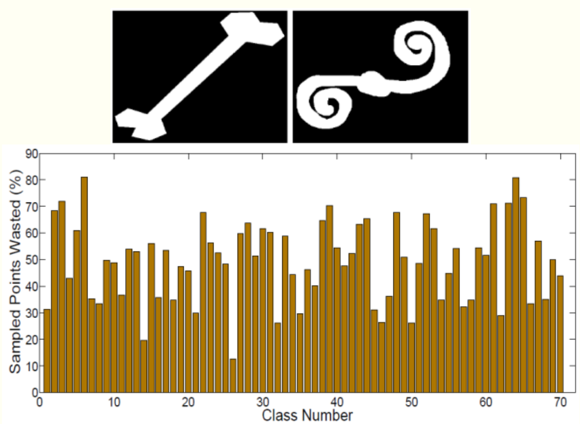
- Perceptually motivated variant of the Shape Context Descriptor
- Captures the interior properties of the shape (*Dense Points*)
- A sparse set of feature locations for easy shape-matching (*Sparse Points*)

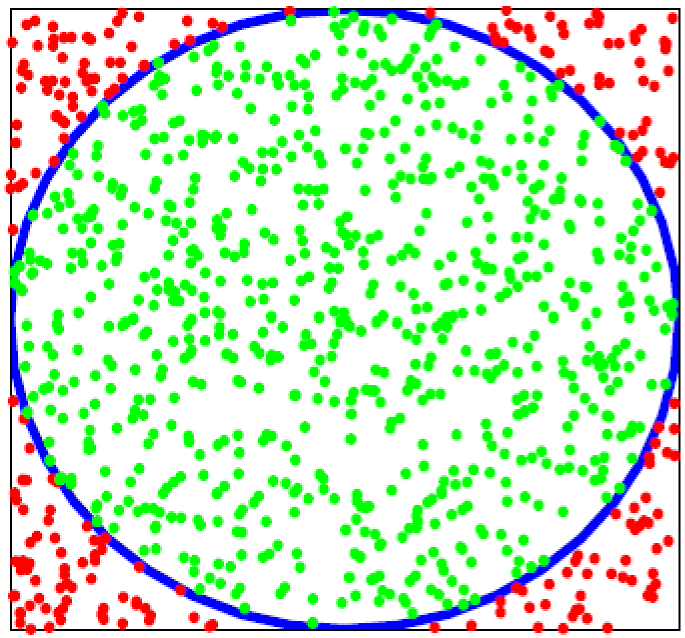


Interior Properties

- Contour Layout: Sample contour :: Interior Layout :: Sample from inside!

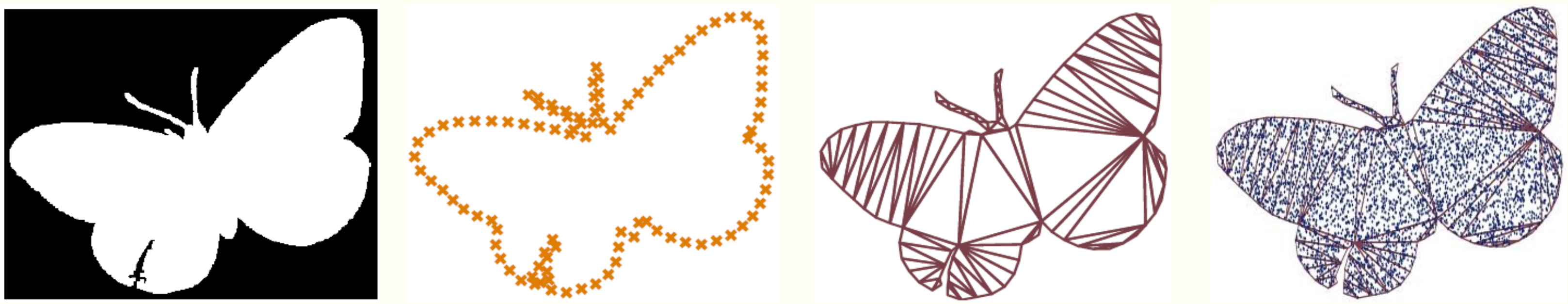
- How to sample points?
- A: Accept/Reject technique





Constrained Delaunay Triangulation - Sampling

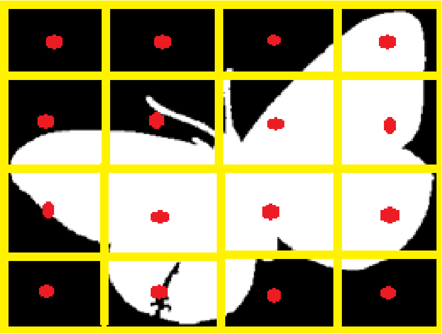
- Extract contour and sample a set of points
- Perform a Constrained Delaunay Triangulation
  on the sampled points
- Sample points from within each triangle

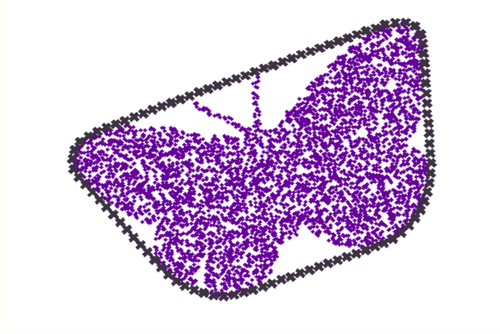


Sparse Points

- *Dense Points* ~ 2500
- Matching features are each *Dense Point* is computationally infeasible
  - (and unnecessary)
- So, need a sparser set of points to compute the descriptor

- What are a good set of sparse points?





Centers of grids?
- We lose continuity constraints

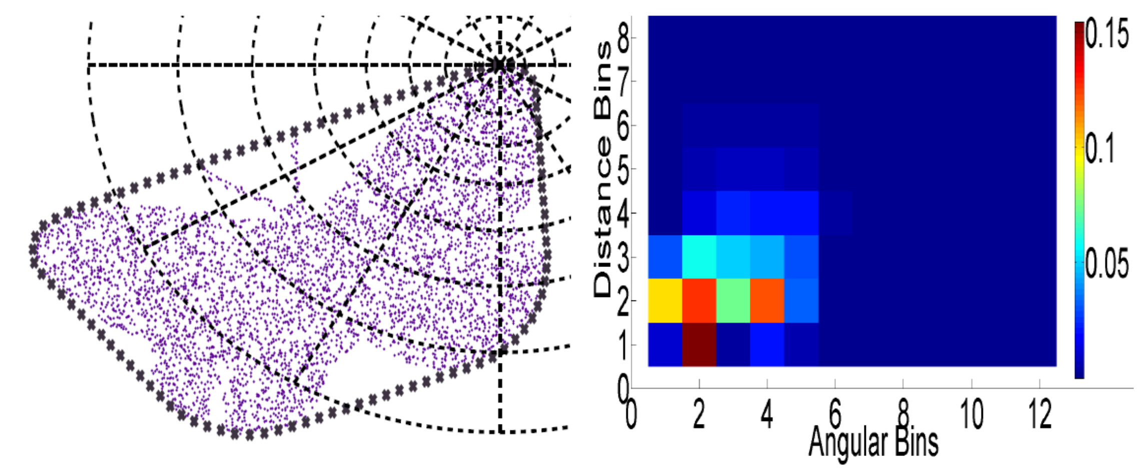
Uniform points along contours?
- Susceptible to indentations in contour

Our method:
Sample points along the ***convex hull*** of the shape

Solid Shape Context Descriptor

- SSC descriptor is a 2-D histogram of distances and angles
- The histogram is computed at each *Sparse Point*

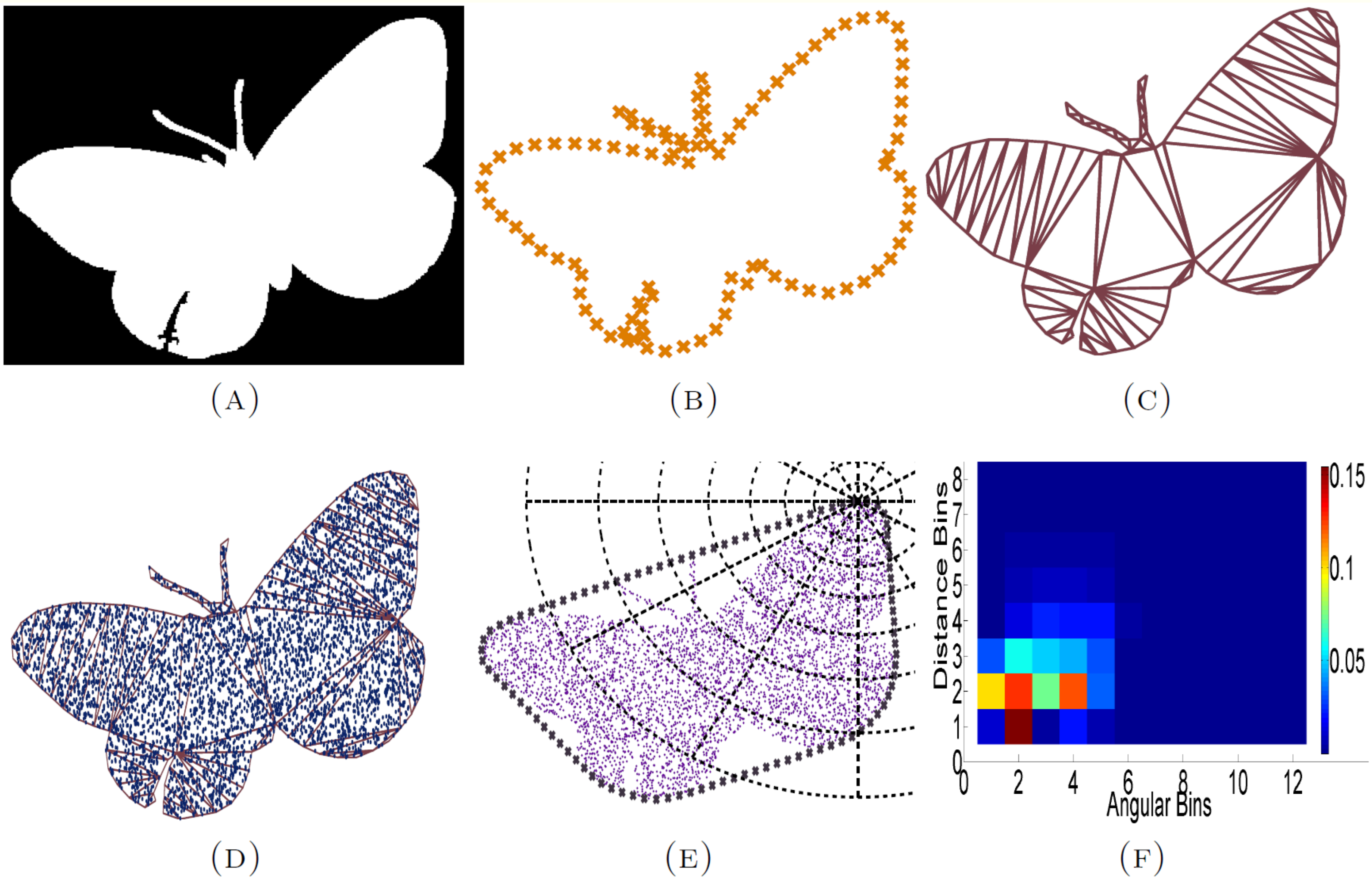
$$\mathcal{H}_i^S(k) = \#\{\mathcal{DP}_j^S : \mathcal{DP}_j^S \in bin(k)\}$$



- A shape is now described by a set of histograms

$$\mathcal{SSC}^S = \{\mathcal{H}_1^S, \mathcal{H}_2^S, \dots, \mathcal{H}_{N_{SP}}^S\}$$

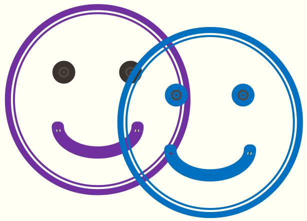
Putting it together



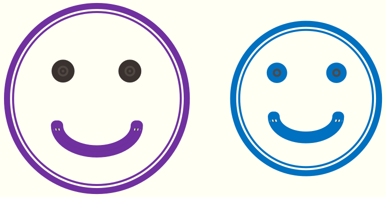
Properties of SSC

- Robust to indentations and contour noise

- Invariant to translation



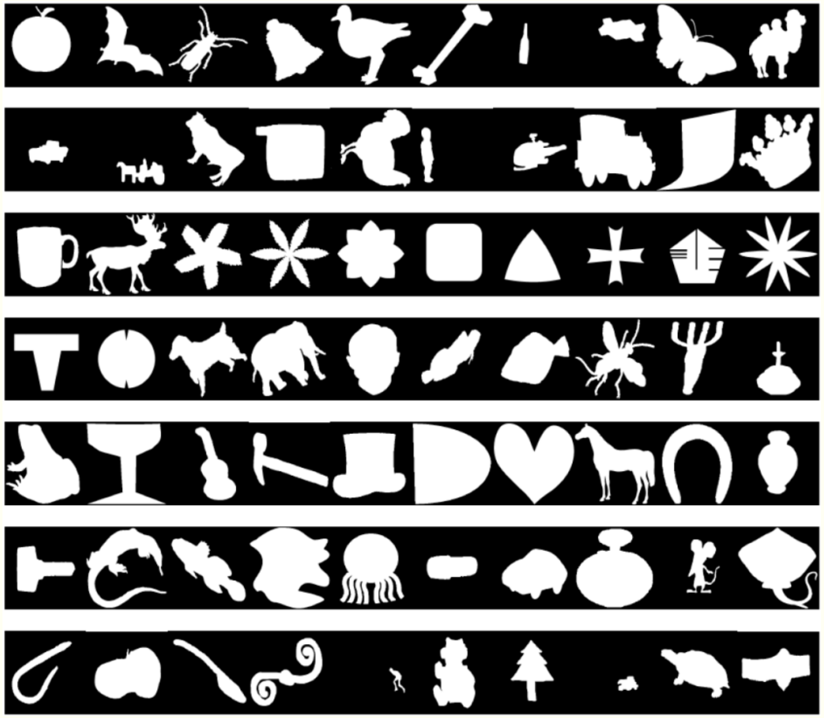
- Invariant to scale



- Invariant to rotation

Experiments and Results

- MPEG-7 Database
- 1400 images, 70 classes, 20 objects per class
- Objects vary in scale, rotation, translation, articulation

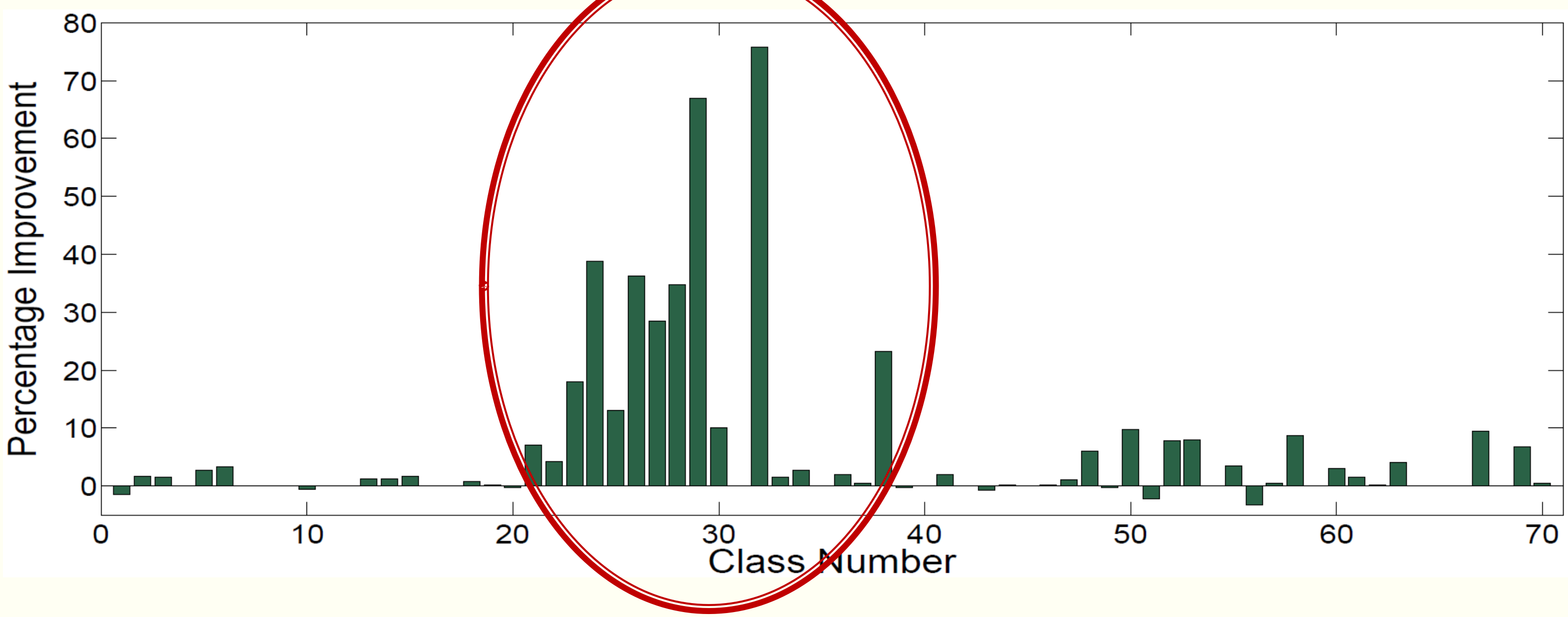


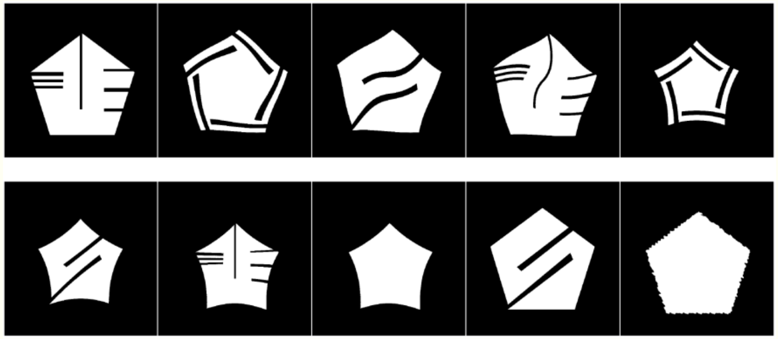
Experiments and Results

- Performance is measured using Bullseye score

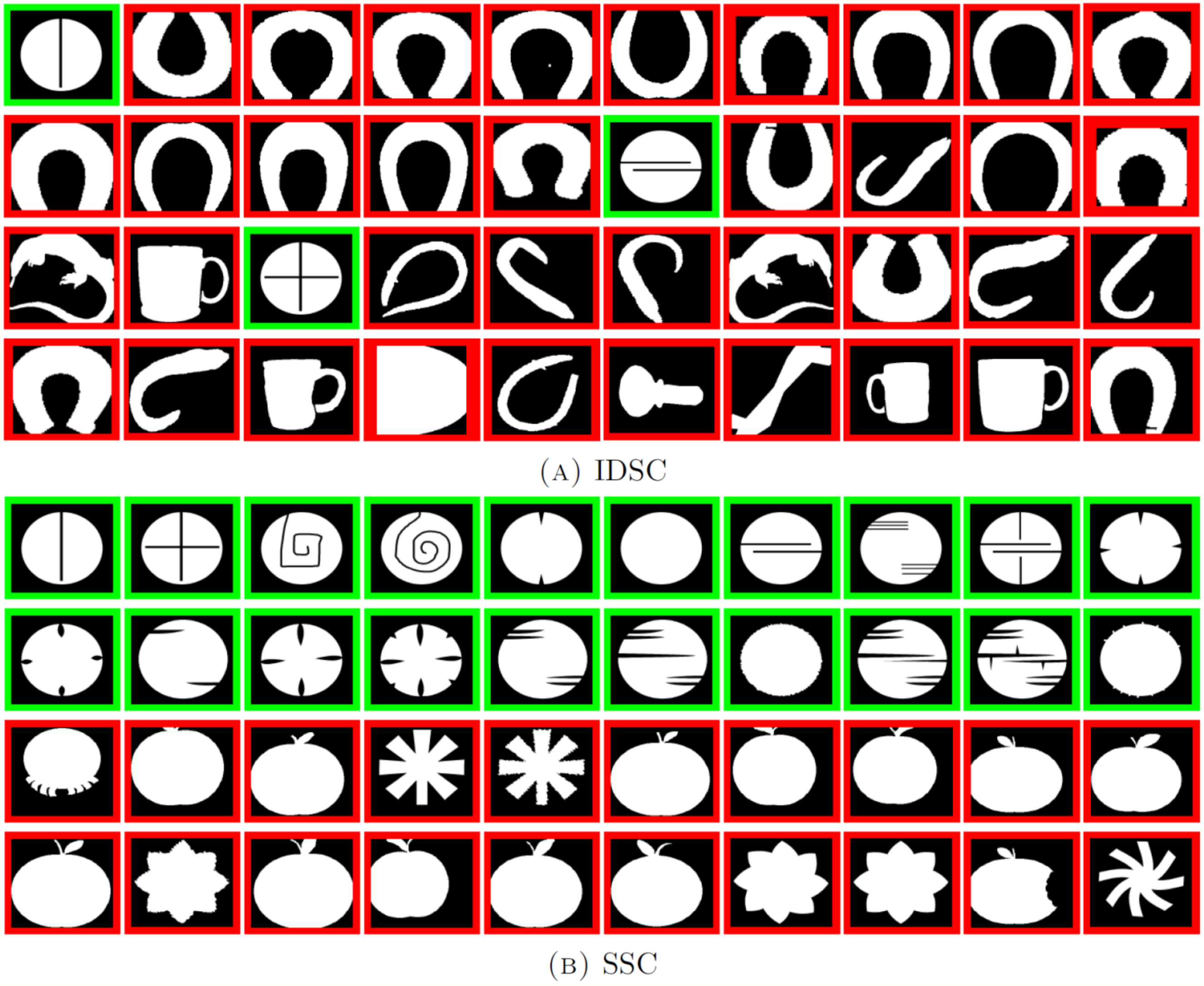
Algorithm	Bullseye Score
Visual Parts [52]	76.45%
SC+TPS [13]	76.51%
Generative Model [96]	80.03%
Curvature Scale Space [65]	81.12%
SSC	82.39%
Polygonal Multiresolution [5]	84.33%
Multiscale Representation [1]	84.93%
IDSC [55]	85.40%
Symbolic Representation [26]	85.92%
Hierarchical Procrustes Matching [64]	86.35%
IDSC(EMD) [56]	86.53%
Triangle Area [2]	87.23%
Shape Tree [30]	87.70%
ASC [57]	88.30%
IDSC+AspectNorm.+StrandRemoval [92]	88.39%
Contour Flexibility [104]	89.31%
IDSC+PMMS [39]	90.18%
IDSC+LP [106]	91.00%
IDSC+SSC	**91.65%**
IDSC+AspectNorm.+SSC	**91.83%**
IDSC+LCDP [107]	92.36%
IDSC+Affine Normalization [37]	93.67%
IDSC+AspectNorm.+StrandRemoval+LCDP [92][107]	95.60%
ASC+LCDP [57][107]	95.96%
IDSC+PMMS+LCDP [39][107]	98.56%
IDSC+SSC+LCDP	**98.85%**
IDSC+Affine Normalization+TPG [108]	99.99%

Improvement Over IDSC





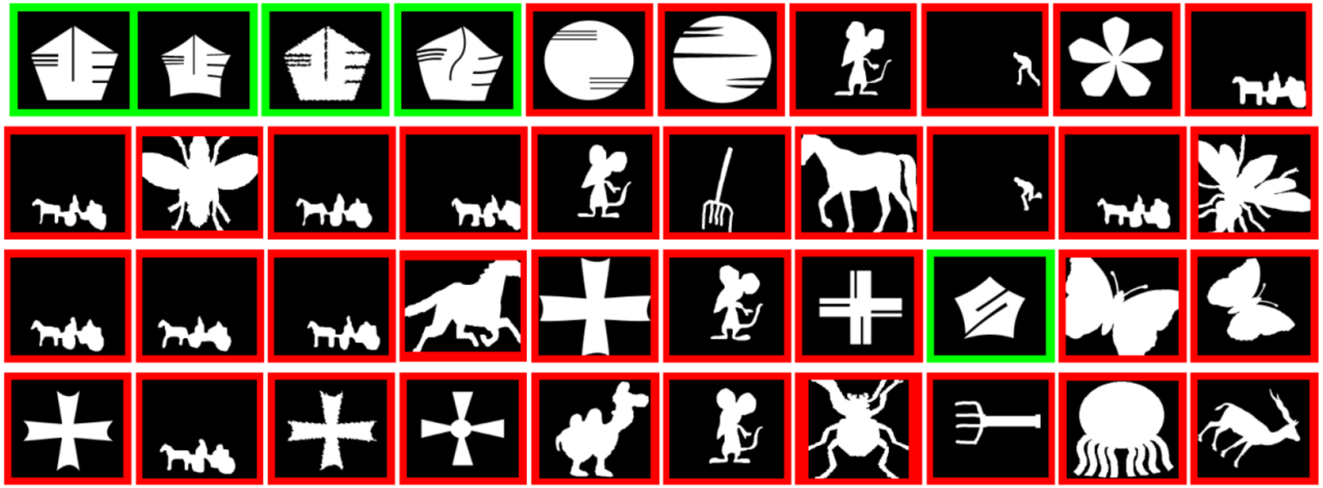
Qualitative Results



(A) IDSC

(B) SSC

Qualitative Results - 2



(c) IDSC

(d) SSC

Perceptually Motivated Shape Context - Recap

- Interior properties of the shape are important
- Proposed the Solid Shape Context descriptor to capture interior properties.

- Vittal Premachandran, and Ramakrishna Kakarala. "Perceptually motivated shape context which uses shape interiors." *Pattern Recognition* (2013).
- Ramakrishna Kakarala, Prabhu Kaliamoorthi*, and Vittal Premachandran*. "Three-dimensional bilateral symmetry plane estimation in the phase domain." *Computer Vision and Pattern Recognition (CVPR), 2013* (* indicates equal contribution)

- We assume that all parts of the shape are equally important.
- Is it so?

Are all parts of the shape uniformly discriminative?

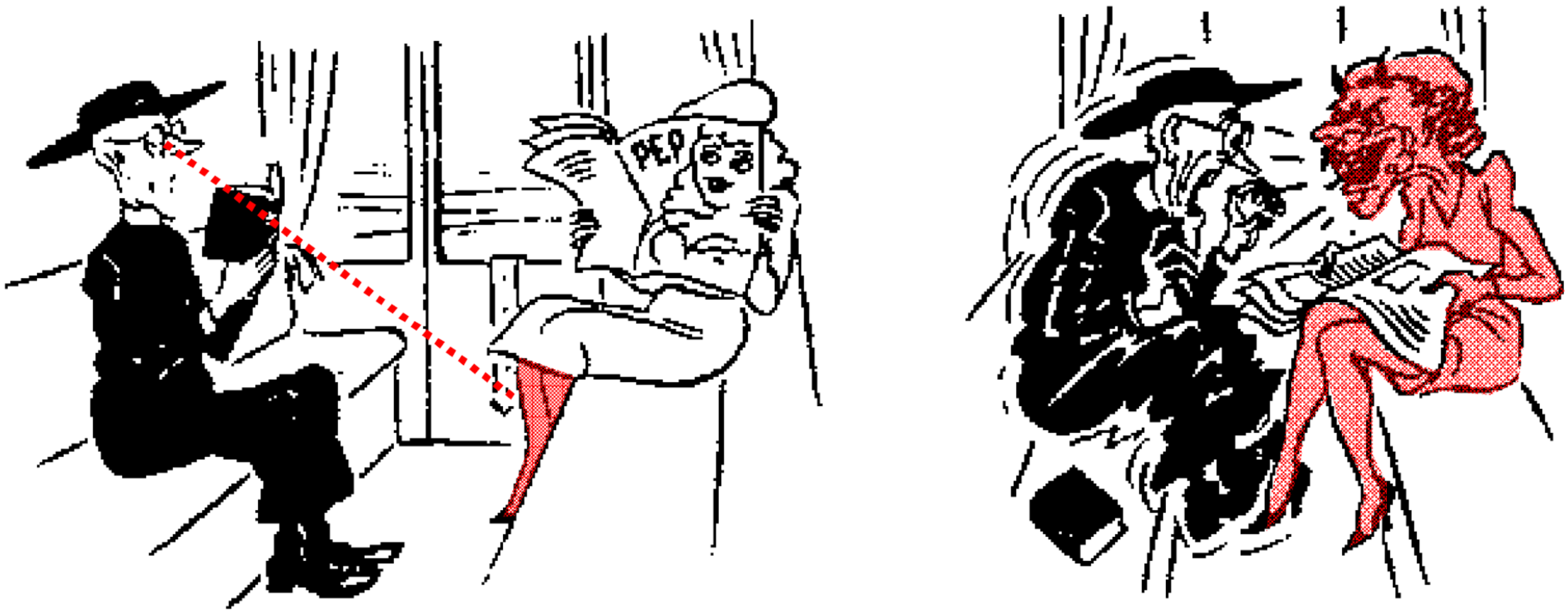
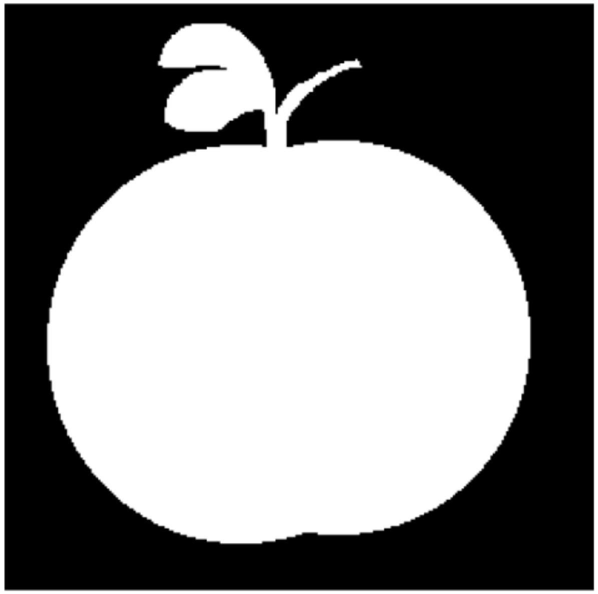


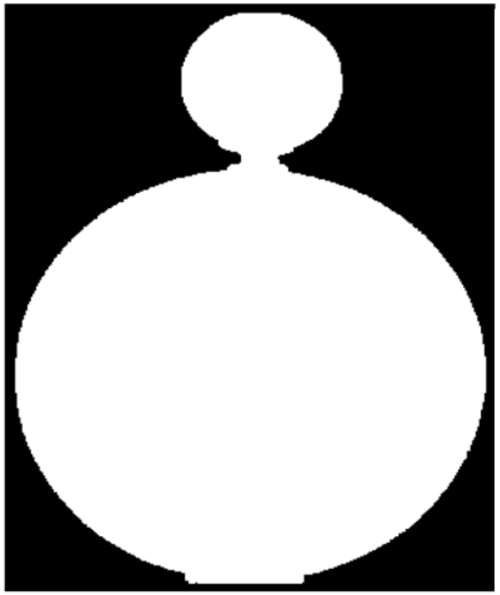
Illustration: Herluf Bidstrup

Outline

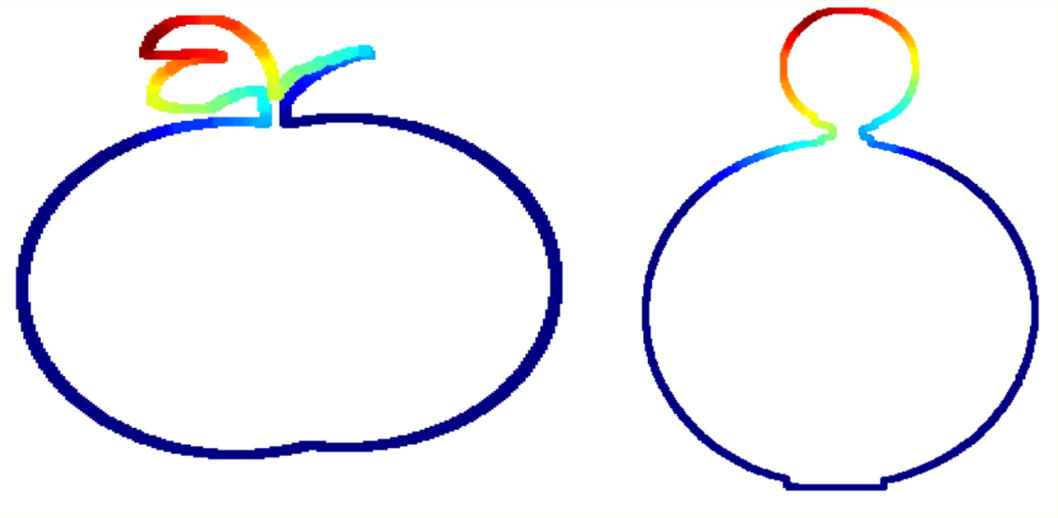
- Introduction
- Background
- Perceptually Motivated Shape Context
- **What parts of a shape are discriminative?**
- Shape similarity propagation on manifolds
- Object detection in images using shapes
- Summary

Motivation





Motivation



Discriminative Parts

- Definition: Parts of an object that are most different from parts of other objects belonging to similar-looking classes.

- Identifying discriminative parts:
  - Divide the shape into multiple parts
  - Find how unique each part is, compared to the negative examples
  - Assign an importance score to the part, which is proportional to its uniqueness

Dividing the Shape

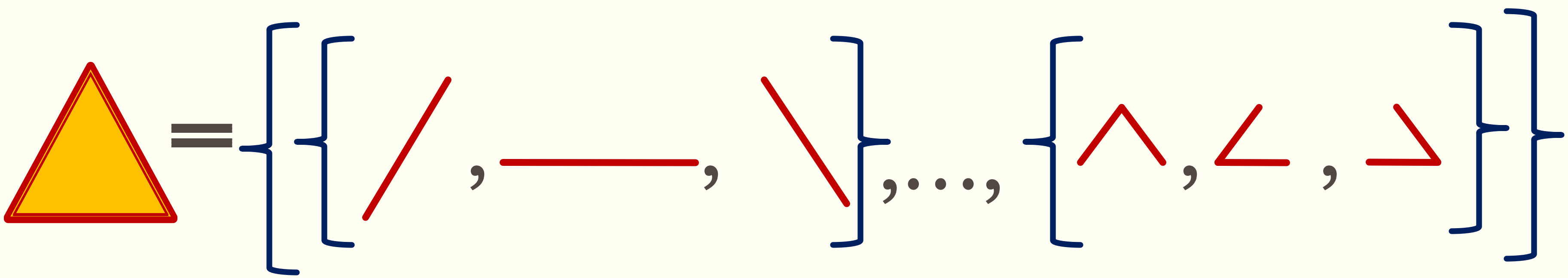
- One way: Make cuts at points of extremum (i.e. corner points) on the contour
- But, identifying extrema points is difficult and subjective.
- Our approach:
  - *Randomly* split the contour into parts
  - Randomization does not introduce subjective bias
  - Bypass the use of not-so-perfect corner detectors

Dividing the Shape - Detail

- Given a shape S_i, extract its contour C_i
- Randomly split the contour into p contour segments

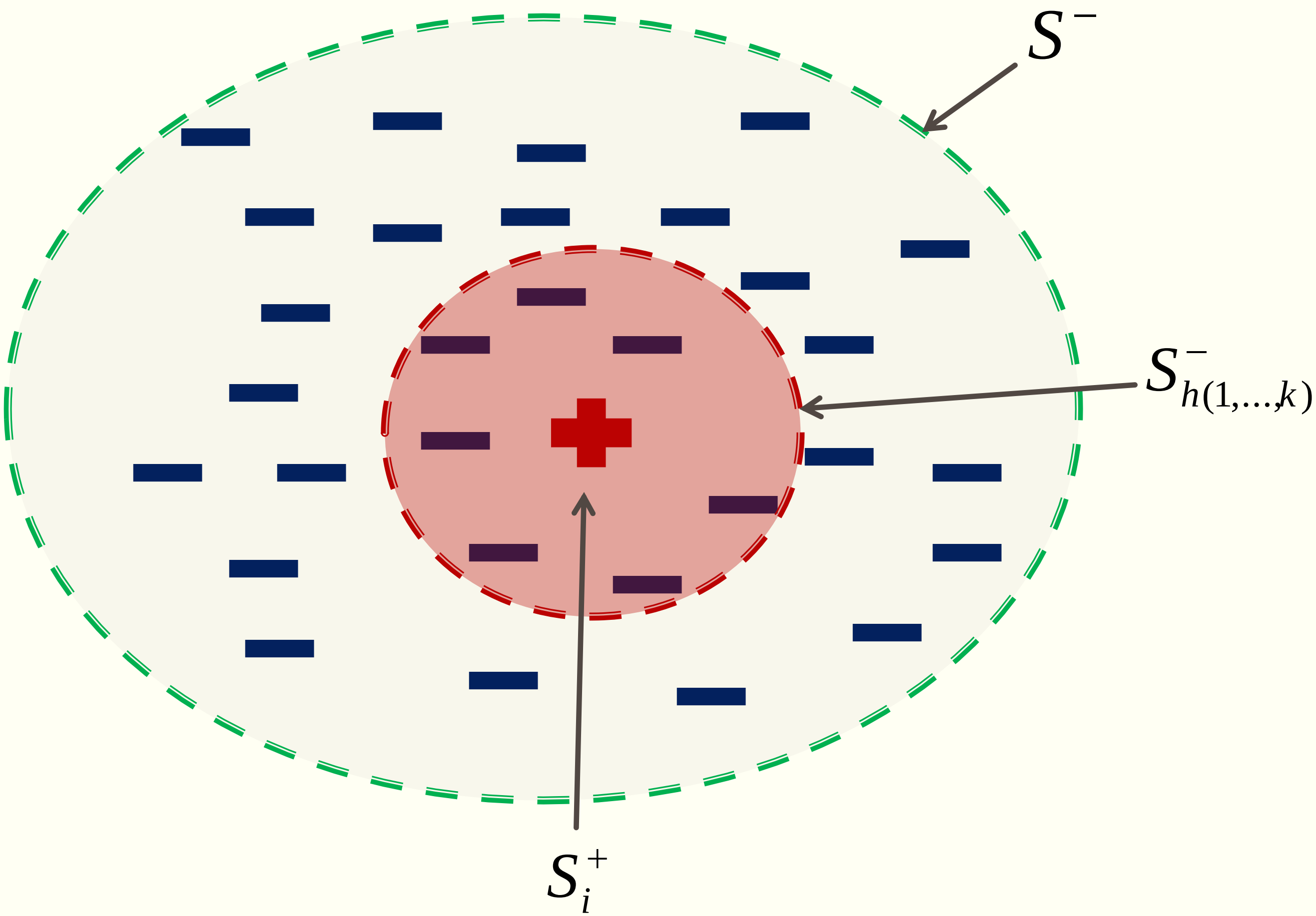
$$C_i^{seg} = \{C_i^{seg_1}, C_i^{seg_2}, \ldots, C_i^{seg_p}\}$$

- Repeat the procedure T times to get T sets of p random parts



- Next step: Compute the uniqueness by comparing to parts from negative shapes

Compare Parts to Negative Shapes



$$S_h^- = \arg\min \Psi_{IDSC}(S_i^+, S_j^-)$$

Shape Descriptor

- Requirements:
  - Invariant to translations
  - Invariant to rotation
  - Invariant to scale
  - Should be a *partial* shape descriptor! (Most important requirement)

- SC, IDSC and SSC are all global shape descriptors

Shape Descriptor - 2

- Donoser et al. "Efficient partial shape matching of outer contours", ACCV 2009.

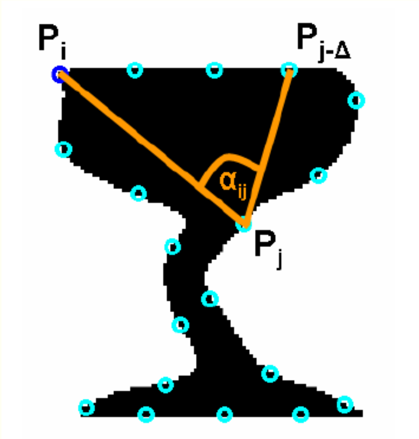


Figure from Donoser et al. ACCV '09

$$\alpha_{ij} = \angle\left(\overline{P_i P_j}, \overline{P_j P_{j-\Delta}}\right)$$

$$A = \begin{pmatrix} \alpha_{11} & . & . & . & \alpha_{1N} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ \alpha_{N1} & . & . & . & \alpha_{NN} \end{pmatrix}$$

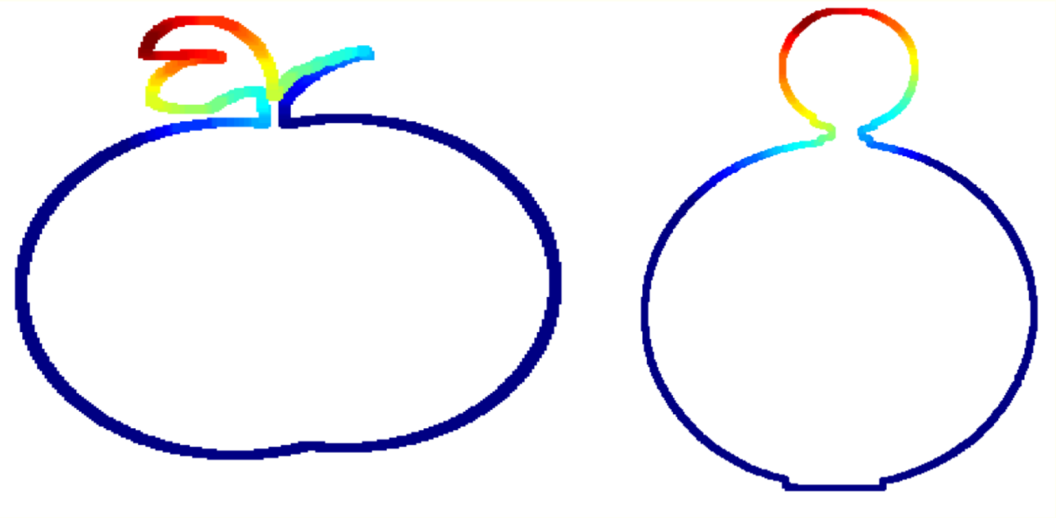
- Any sub matrix of the above matrix, is a partial shape descriptor.

- **Distance: Distinctive**

- Distance: Distinctive

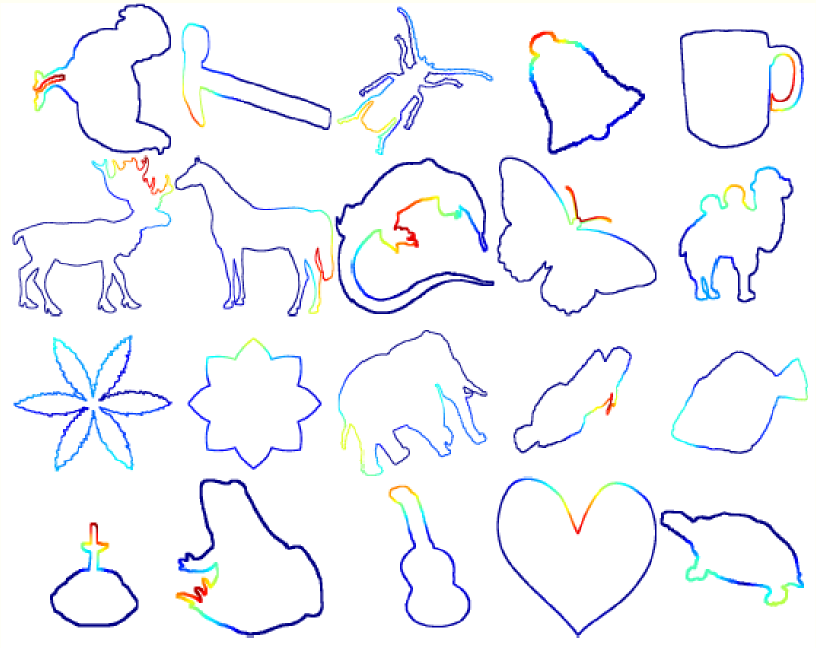
Example

- Learnt distinctiveness for two confusing classes

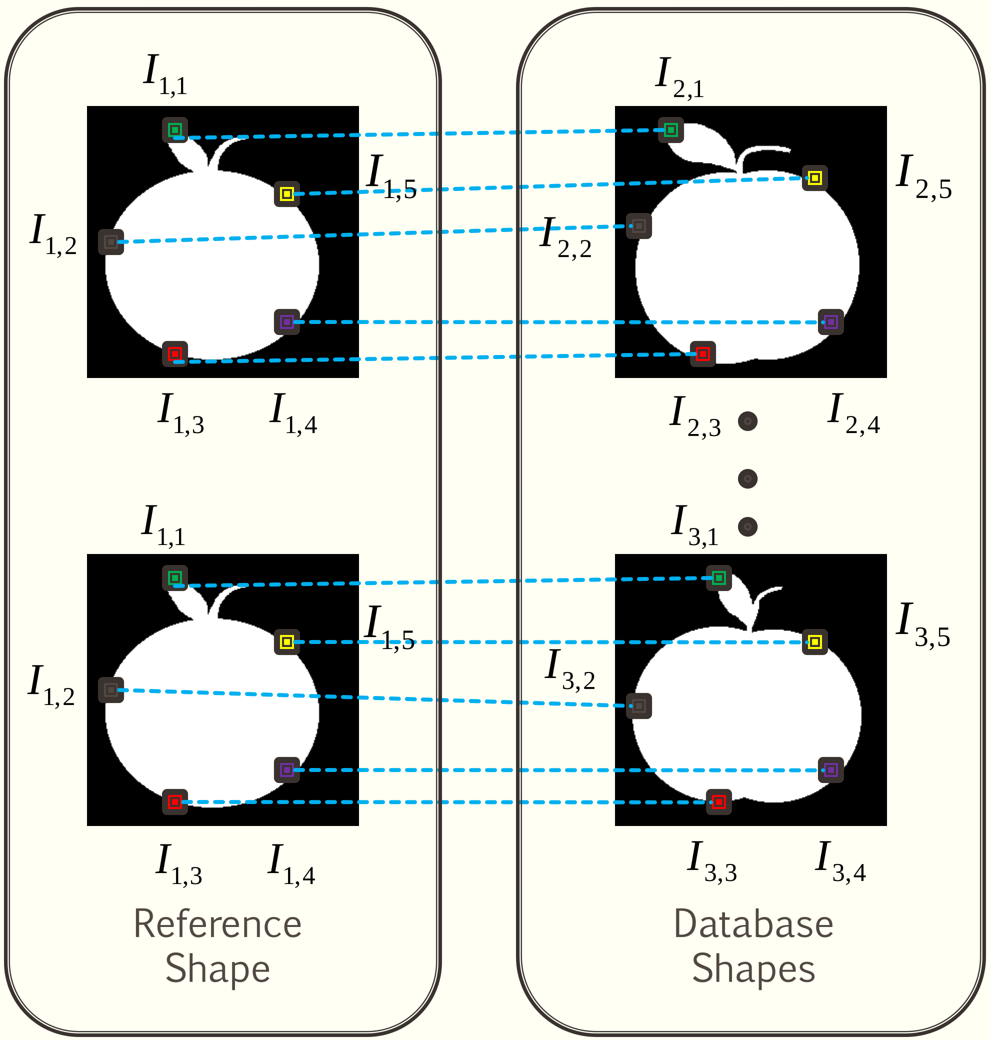


More Examples

- Example heat maps from various classes.
- Semantically meaningful parts are identified as being distinctive.

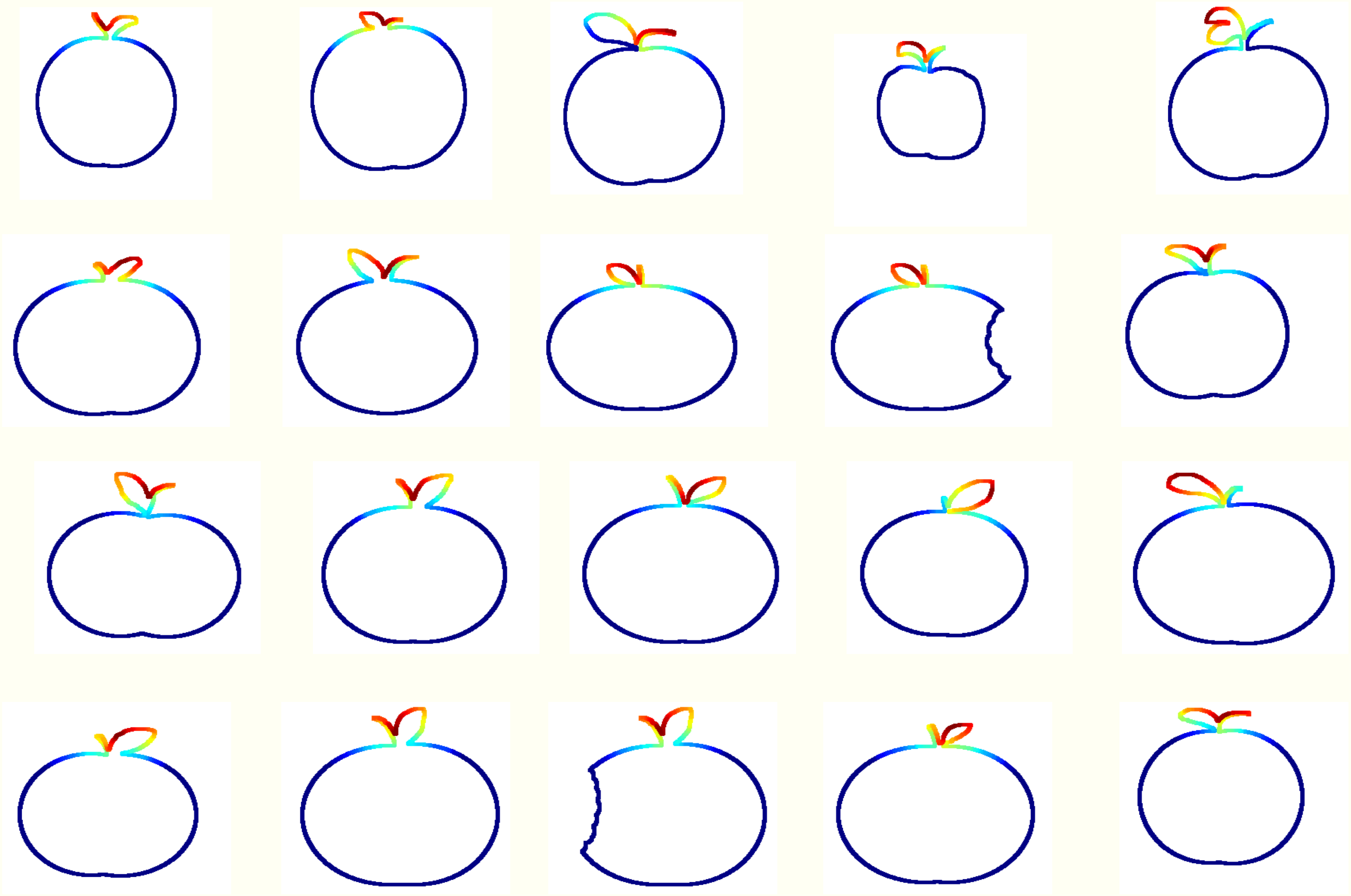


Part Consistency Matrix

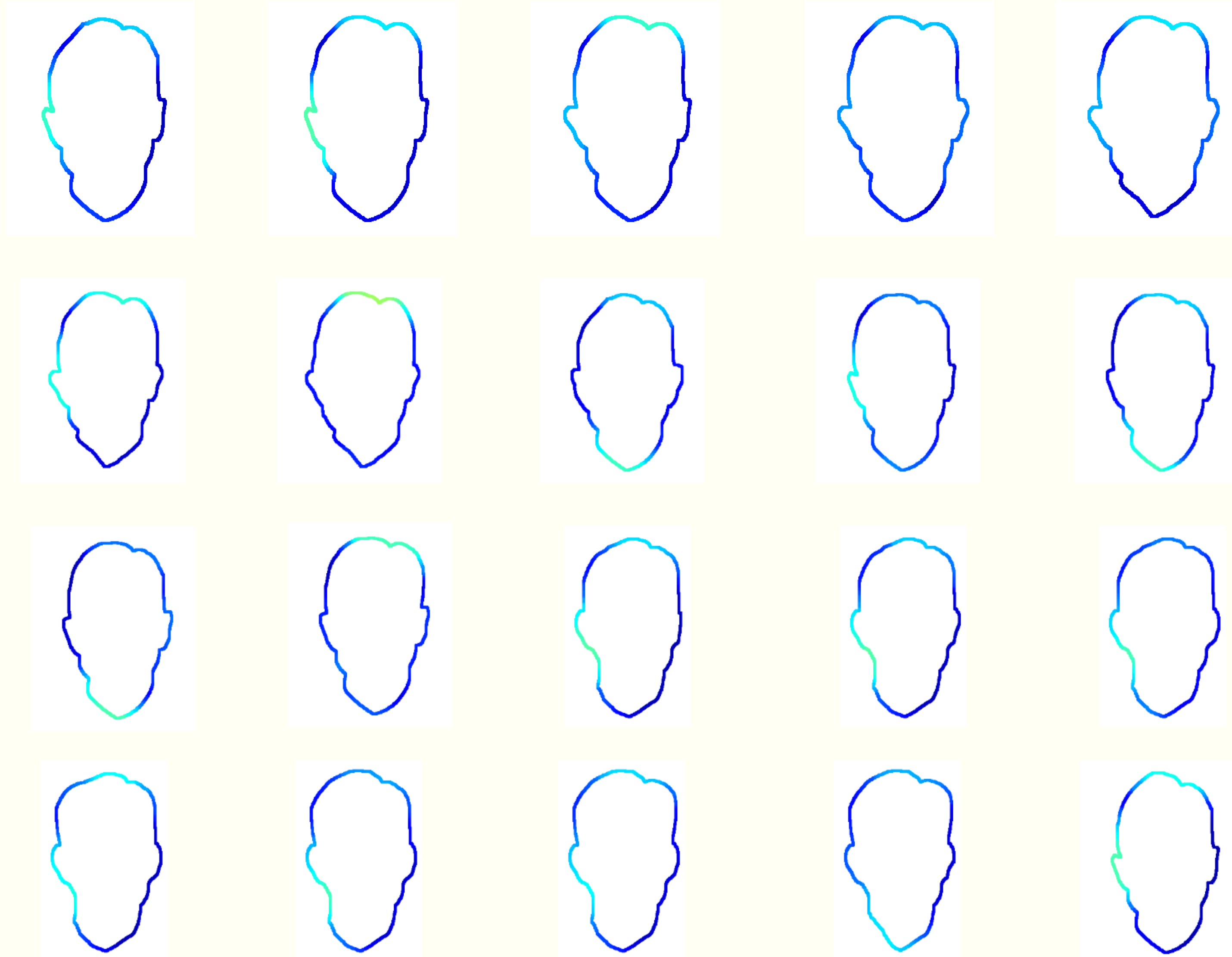


$$Q = \begin{bmatrix} I_{1,1} & I_{2,1} & I_{3,1} \\ I_{1,2} & I_{2,2} & I_{3,2} \\ I_{1,3} & I_{2,3} & I_{3,3} \\ I_{1,4} & I_{2,4} & I_{3,4} \\ I_{1,5} & I_{2,5} & I_{3,5} \end{bmatrix}$$

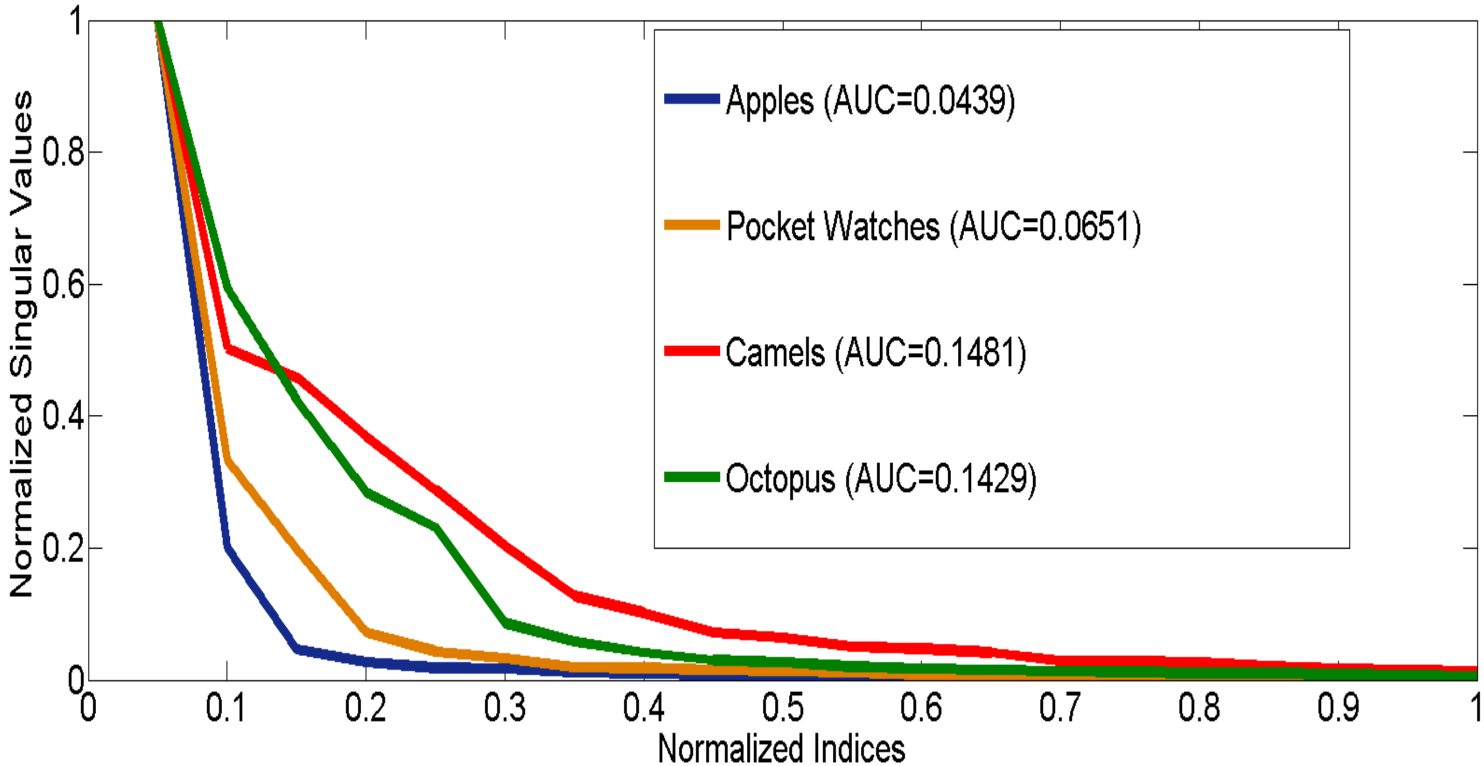
Example Class with Low AUC - Apples



Example Class with High AUC – Face Silhouettes



Singular Value Curves



Summary

- Introduced the concept of distinctiveness with a strong motivating example and provided a simple way to extract discriminative parts.

- Vittal Premachandran, and Ramakrishna Kakarala. "What parts of a shape are discriminative?" *IEEE International Conference on Image Processing,* 2013. (Best Student Paper Award!)

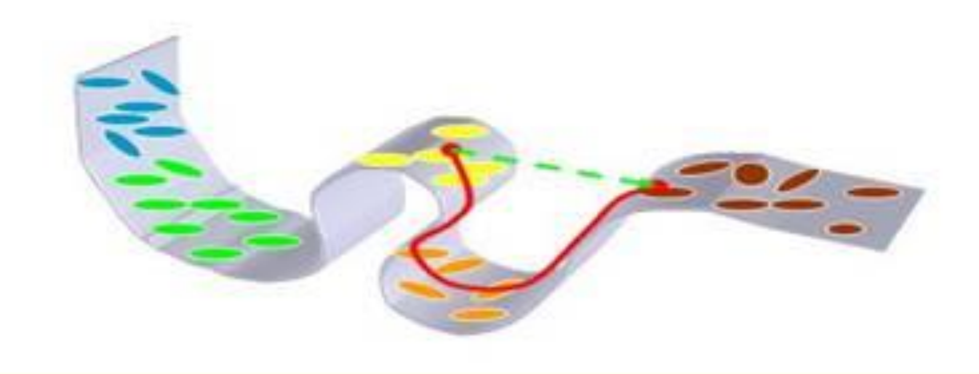
- Pairwise comparison.
- Can we go Beyond Pairwise comparison of shapes?

Outline

- Introduction
- Background
- Perceptually Motivated Shape Context
- What parts of a shape are discriminative?
- **Shape similarity propagation on manifolds**
- Object detection in images using shapes
- Summary

Shape Retrieval Using Shape Manifolds

- Shapes lie in a high dimensional space
- However, the shape manifolds are usually of lower dimensionality
- Euclidean distance is a bad metric
- Need to learn the “true” geodesic distances



Diffusion on Sparse Graphs

- Diffusion is a process to learn the affinity between shapes
- Sparse graphs are used for diffusion instead of fully-connected graphs
- Fully connected graphs have lots of noisy edges;

- Ways for graph sparsification:
  - K-NN graph neighborhood
  - Mutual K-NN graph neighborhood
  - Epsilon-NN graph neighborhood
  - Dominant neighborhood

Related work

- Yang, Xingwei, Suzan Koknar-Tezel, and Longin Jan Latecki. "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval." *CVPR 2009.*
  - LCDP for short
  - Uses k-NN for graph sparsification
  - Improves retrieval performance by ~6%

- Xingwei Yang; Prasad, L.; Latecki, L.J., "Affinity Learning with Diffusion on Tensor Product Graph," *IEEE Trans. PAMI 2013*
  - TPG for short
  - Uses Dominant Neighborhoods (DN) for graph sparsification
  - Improves retrieval performance by ~6%

- However, graph sparsification is performed in a supervised manner
- For arbitrary neighborhood sizes, their performance drops (sometimes, drastically!)

Consensus of k-NNs

- Intuition behind Consensus of k-NNs
  - Nodes that appear consistently in neighborhoods of other nodes have a high chance of being similar to each other
  - So, edges in the sparse graph should be between nodes that appear consistently
  - Prune edges between inconsistent nodes
- Higher the value of 'k', higher the probability of adding noisy edges
- Our approach is more robust at higher values of 'k'

Consensus Information

- Collecting consensus information is this simple!

```
C = 0;
for i = 1 : N do
    S_i = k-NN(v_i);
    for p = 1 : N do
        for q = p + 1 : N do
            if p ∈ S_i and q ∈ S_i then
                C(p, q) = C(p, q) + 1;
                C(q, p) = C(q, p) + 1;
            end
        end
    end
end
```

- Finally, retain the edge between p and q if C(p,q) > τ

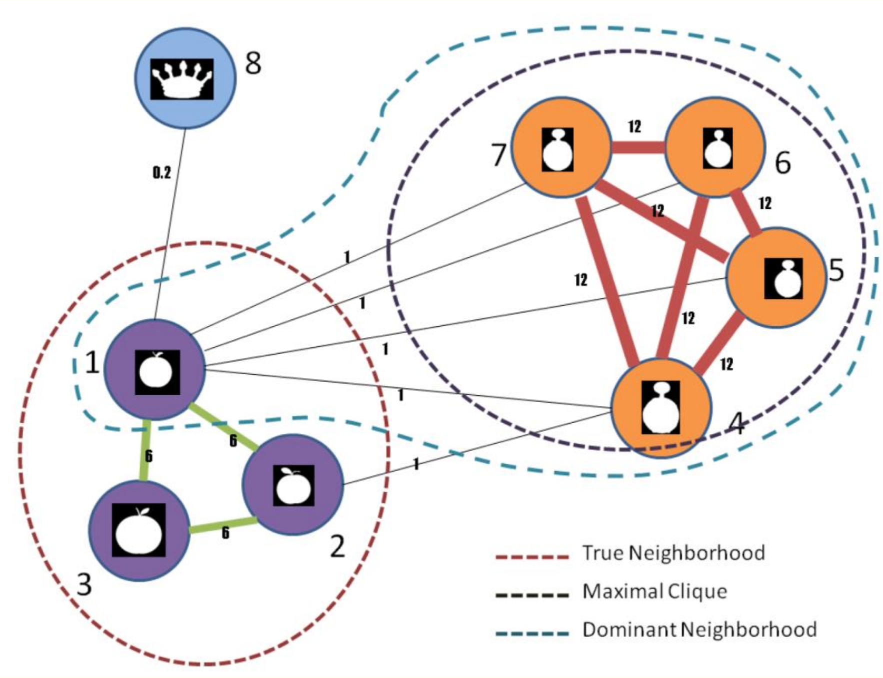
Dominant Neighborhoods

- Dominant neighbors are maximal cliques

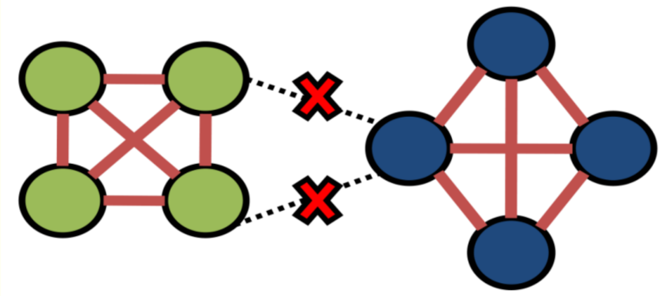
$$\begin{array}{rc} maximize & f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} \\ subject\ to & \mathbf{x} \in \Delta \end{array}$$

$$\Delta = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x} \geq 0 \ and\ \mathbf{1}^T\mathbf{x} = 1\}$$

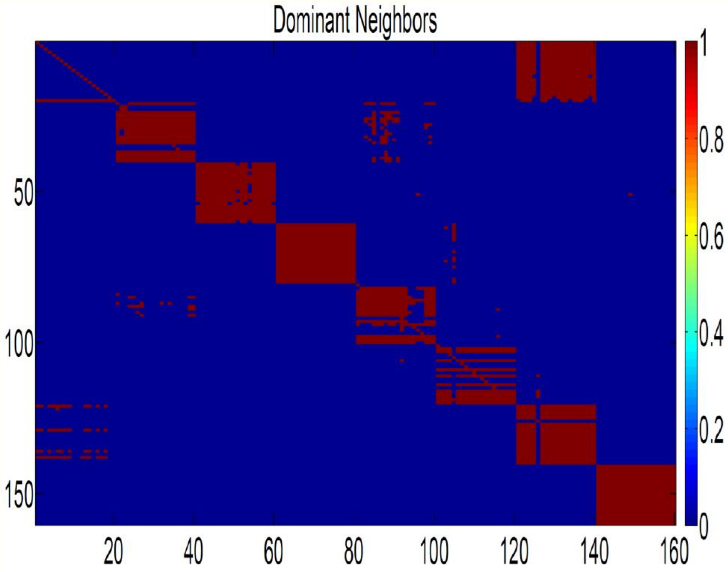
- Issues with DN:
- False Neighborhoods



Disconnected Graphs

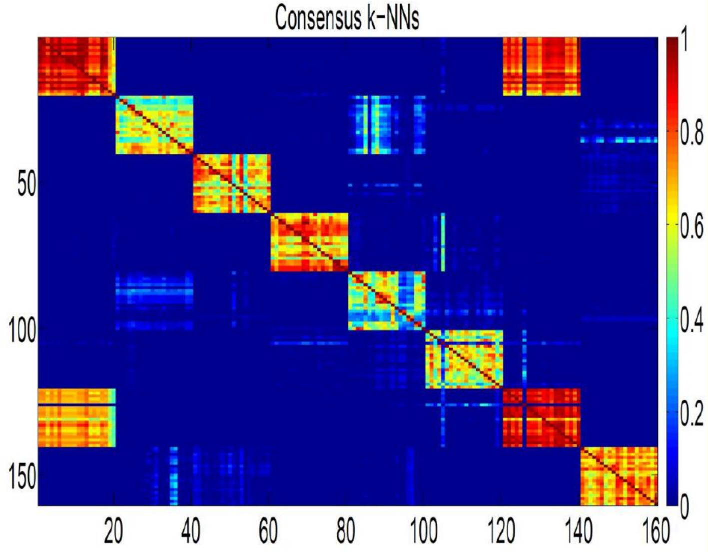


Neighborhood Comparison



- Binary neighborhood
- Class #1 gets assigned instances from class #7 as its Dominant Neighbors.

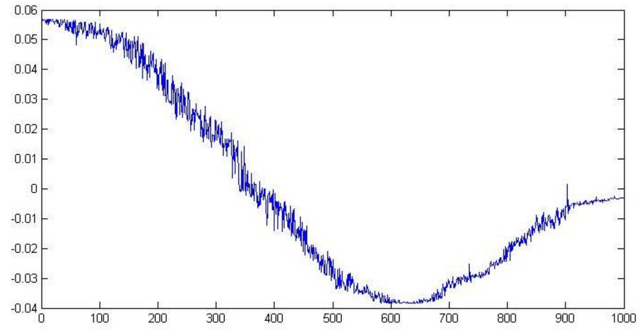
- Soft, probabilistic measures for deciding on neighborhoods.
- Class #1's neighbors remain in class #1 with high probability.

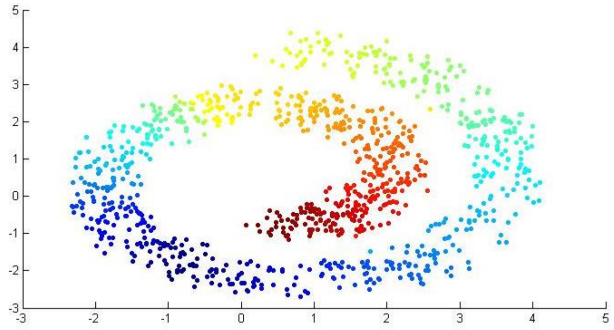


Experiments

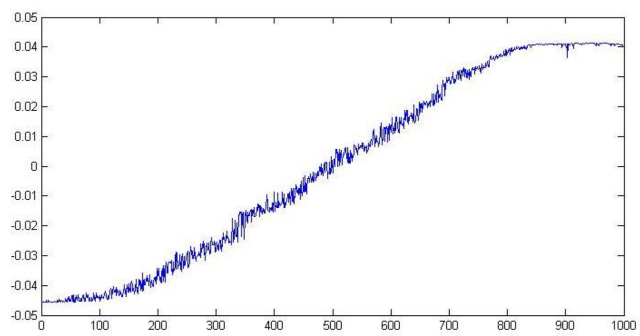
- Toy Dataset: Spiral

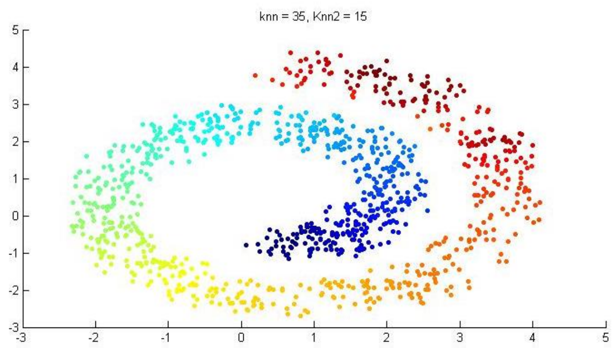
Simple k-NN



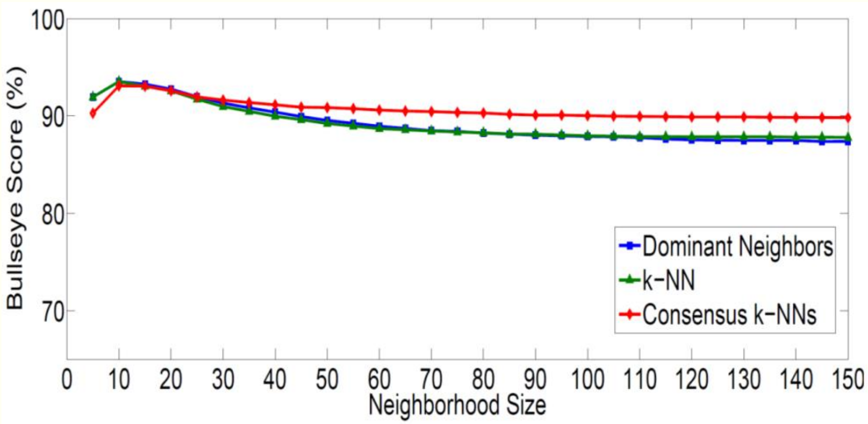


Consensus of k-NNs

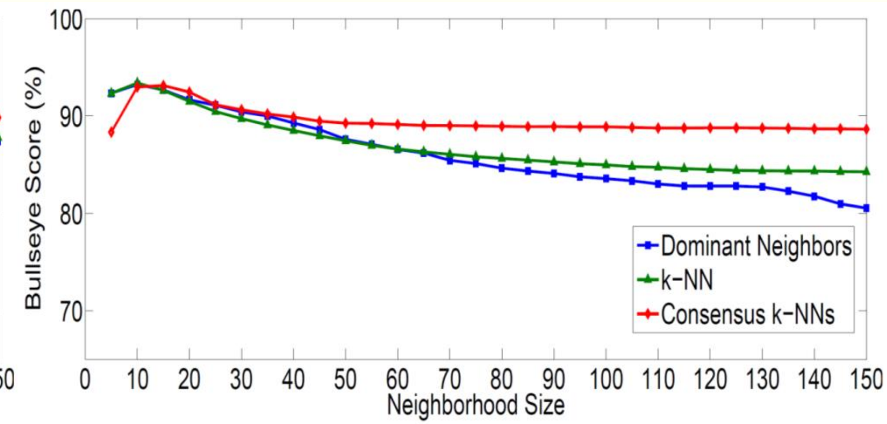




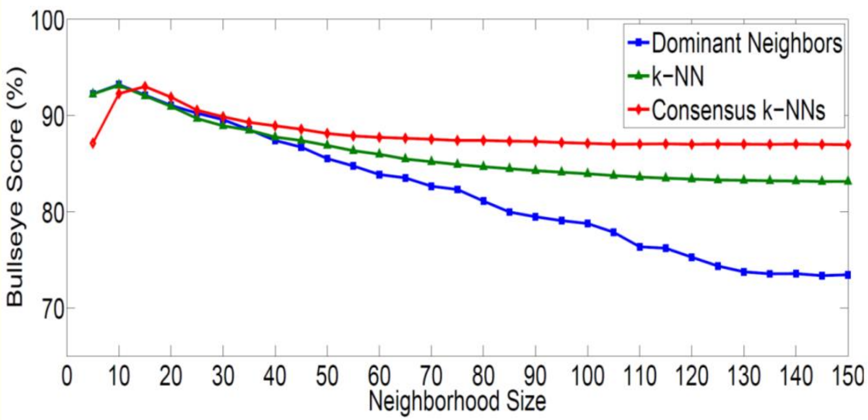
Experiments on MPEG-7 Database



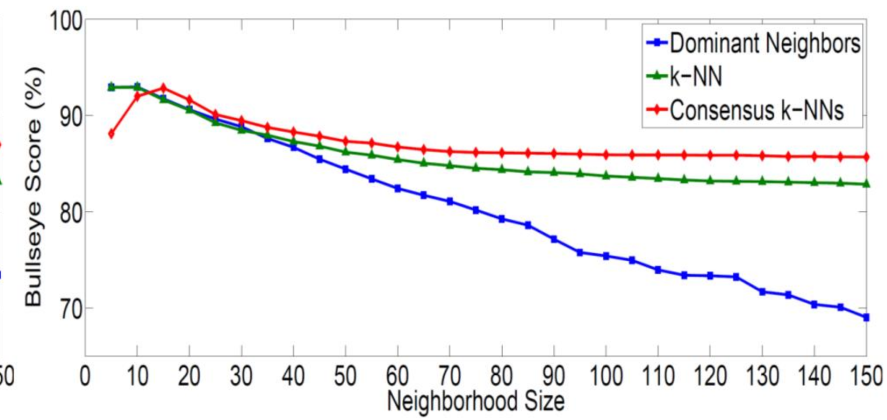
(a) K=25



(b) K=50



(c) K=75



(d) K=100

Recap

- Graph sparsification is important preprocessing step for diffusion
- Consensus of k-NNs is a much more robust way of selecting neighborhoods

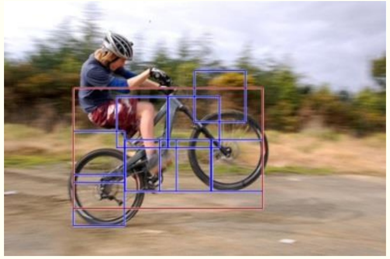
- Vittal Premachandran, and Ramakrishna Kakarala. "Consensus of k-nns for robust neighborhood selection on graph-based manifolds." *Computer Vision and Pattern Recognition (CVPR), 2013*

Outline

- Introduction
- Background
- Perceptually Motivated Shape Context
- What parts of a shape are discriminative?
- Shape similarity propagation on manifolds
- **Object detection in images using shapes**
- Summary

Object Detection Using Shapes

- Object detection is a highly challenging task
- Detection = recognition + localization



- Two main approaches for object detection:

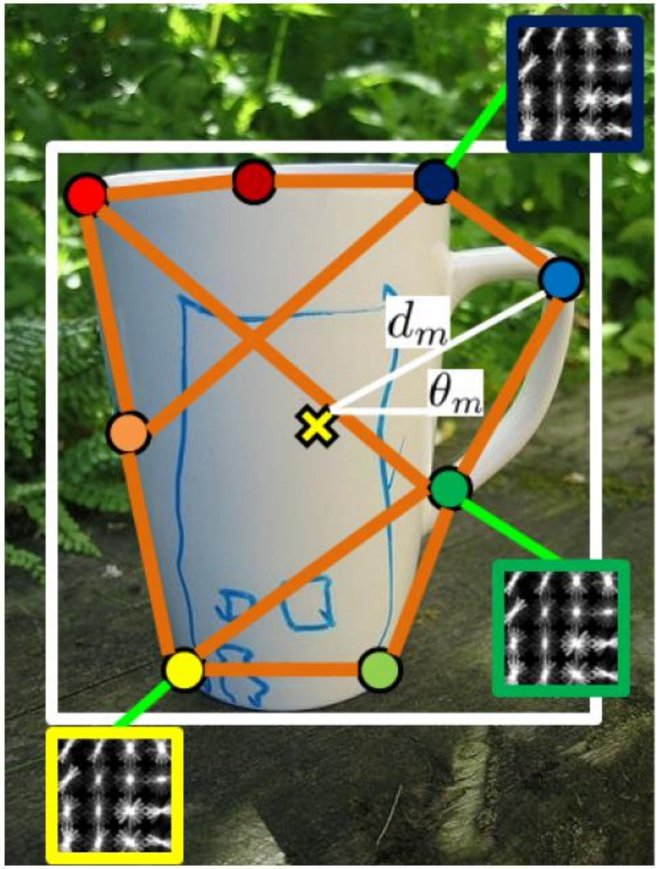
HOG-based approach;
strong discriminative capability

Contour-fragment based approach;
strong localization ability

- We aim to bridge the gap between the two approaches

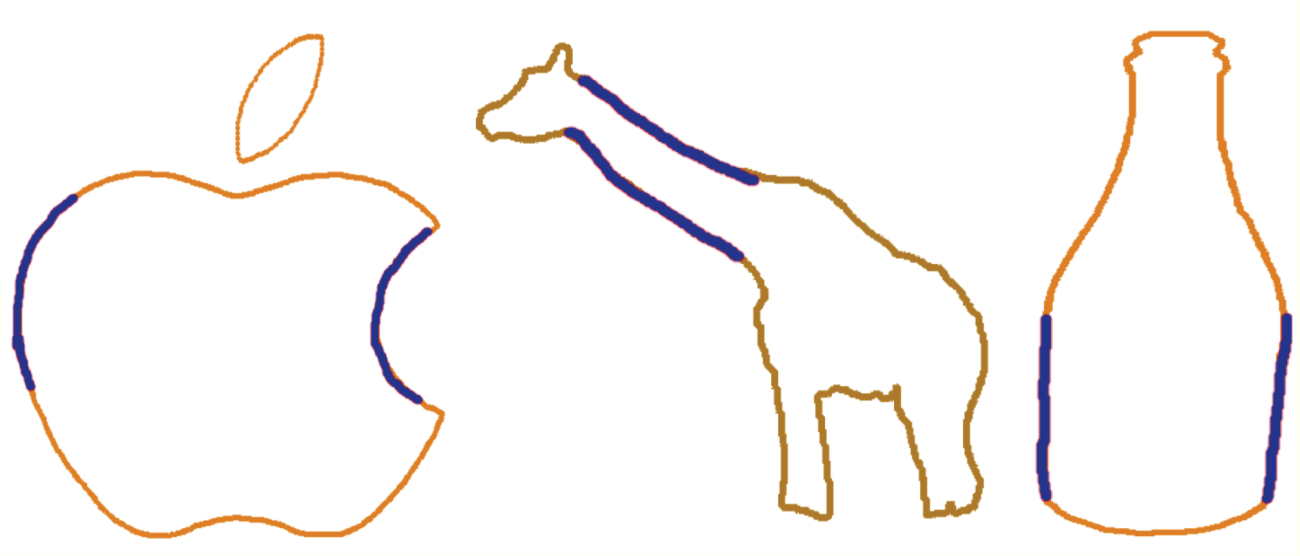
Our Model

- Parts-based model
- Each part denoted by two types of features:
  - Local features – HOG
  - Non-local features – Distance and Angle



Need For Non-Local Parts

- Locally similar looking parts are highlighted in blue
- Without non-local features, they can generate false matches



The Approach

$$E(I, Y; \ell) = \boxed{\sum_{p_i \in \mathcal{V}} \Phi(.)} + \boxed{\sum_{(p_i, p_j) \in \mathcal{E}} \Psi(.)} + \boxed{\sum_{(p_i, p_j) \in \mathcal{E}} \Omega(.)}$$

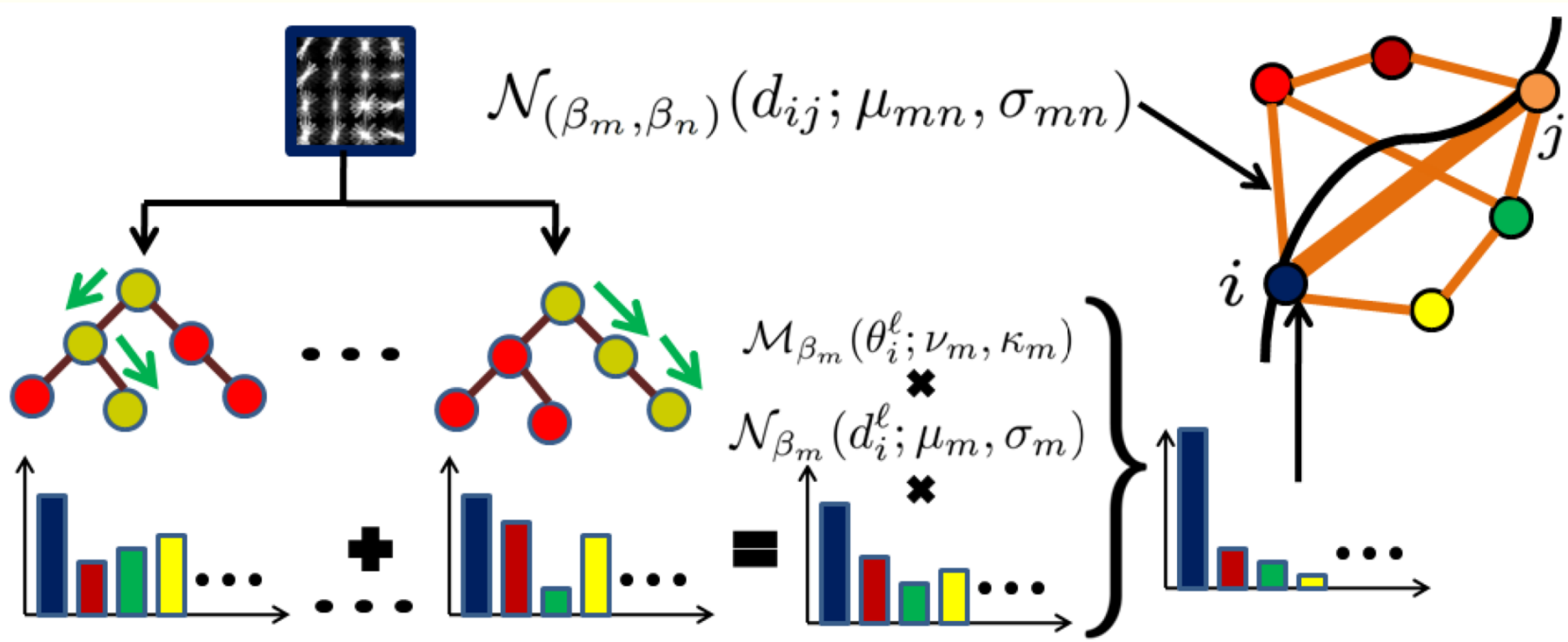
Unary Potential

Pairwise Potential

Potts Potential

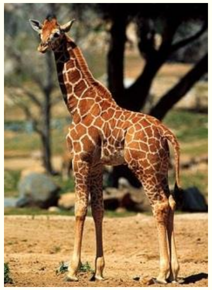
- TRW-S for optimization

Overall Procedure



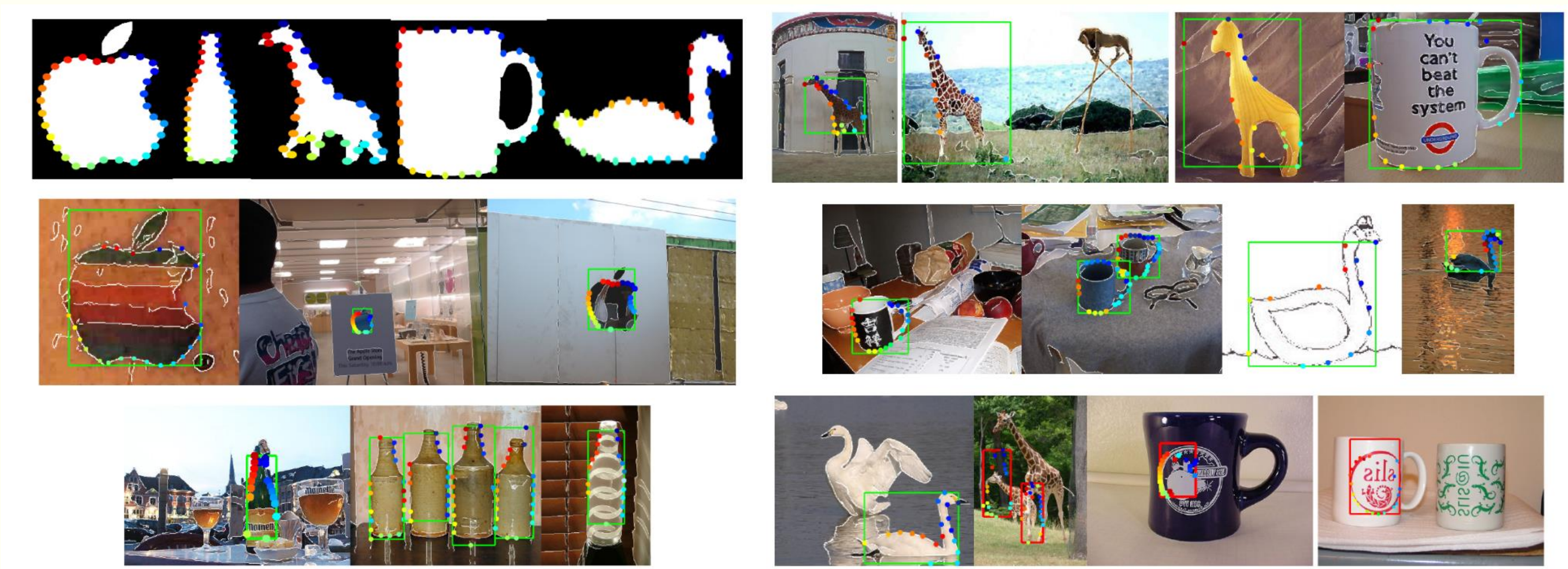
Experiments and Results

- ETH-Z Shape Database
  - 5 classes



  - 255 Images
  - PASCAL 50% overlap criteria
  - Performance measured by mean Average Precision and Detection Rate at 0.3/0.4 FPPI

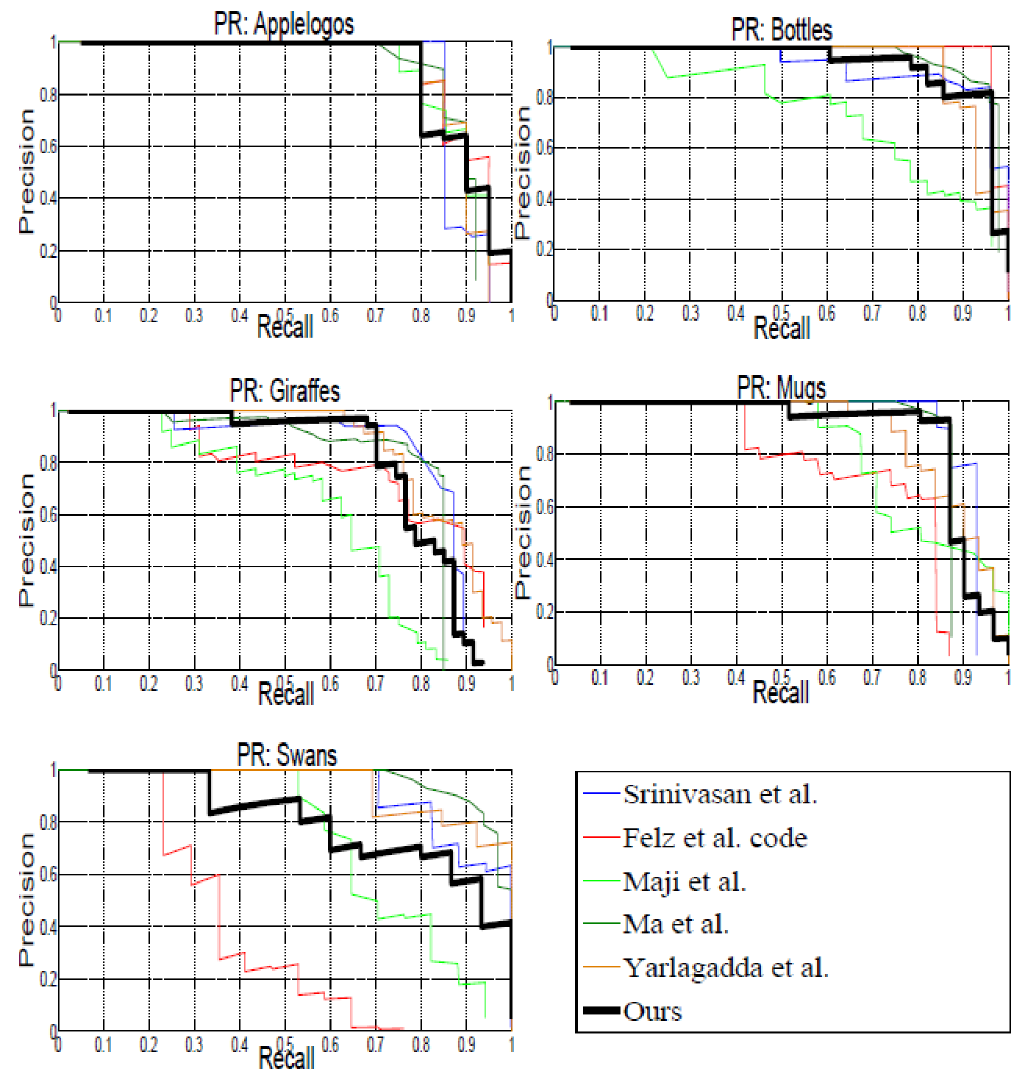
Detection Examples



Mean Average Precision

	Applelogos	Bottles	Giraffes	Mugs	Swans	Mean
Ours	**0.897**	0.939	0.800	**0.889**	0.819	0.868
Srinivasan et al. [91]	0.845	0.916	0.787	0.888	0.922	0.872
Felzenswalb et al. [31]	0.891	0.950	0.608	0.721	0.391	0.712
Wang et al. [100]	0.866	**0.975**	**0.832**	0.843	0.828	0.869
Ma et al. [61]	0.881	0.920	0.756	0.868	**0.959**	**0.877**
Lin et al. [53]	**0.909**	0.898	0.811	0.893	**0.964**	**0.895**
Ours Random	0.908	**0.951**	**0.821**	**0.894**	0.847	0.884

Precision Recall Curves



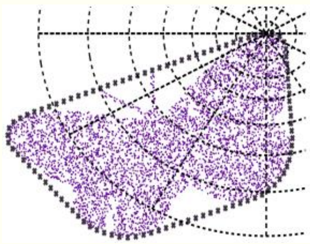
Outline

- Introduction
- Background
- Perceptually Motivated Shape Context
- What parts of a shape are discriminative?
- Shape similarity propagation on manifolds
- Object detection in images using shapes
- Summary

Summary

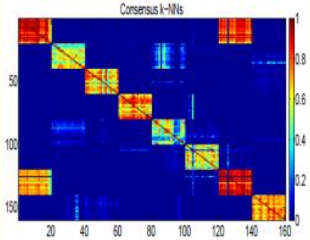
- Presented a strong motivation to use shapes

- Perceptually motivated technique to match shapes



- Presented a novel technique for identifying discriminative parts.

- Consensus of k-NNs for robust neighborhood selection



- Bridge the gap between HOG-based and contour-based object detection

Future Work

- Need other types of perceptually motivated techniques
- Use discriminativeness of parts for generating better shape descriptors
- Bigger shape databases are needed for object detection

Thank you!