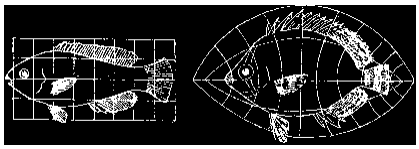
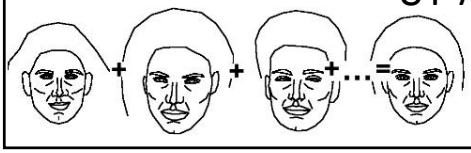


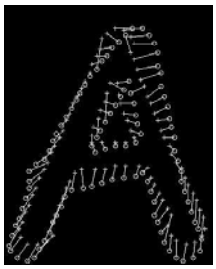
Shape Matching

Tuesday, Nov 17

Kristen Grauman

UT-Austin



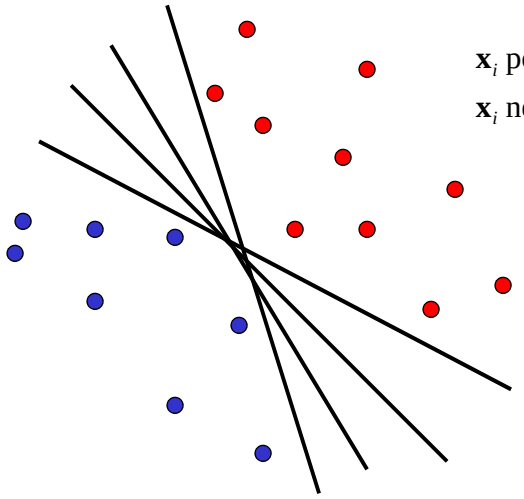


Previously

- Discriminative classifiers applied to object detection / categorization problems.
  - Boosting
  - Nearest neighbors
  - Support vector machines
    - Application to pedestrian detection
    - Application to gender classification

Linear classifiers

- Find linear function to separate positive and negative examples

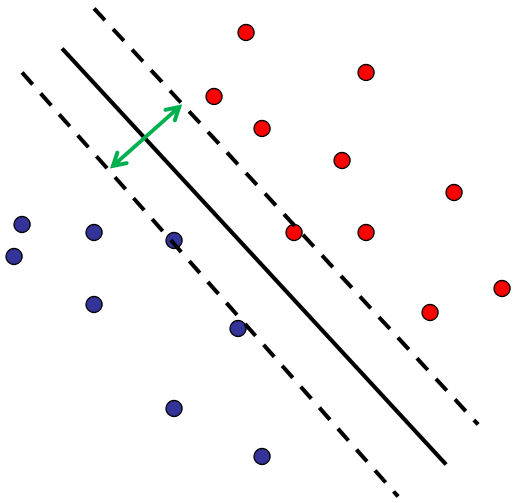


$\mathbf{x}_i$ positive: $\mathbf{x}_i \cdot \mathbf{w} + b \geq 0$

$\mathbf{x}_i$ negative: $\mathbf{x}_i \cdot \mathbf{w} + b < 0$

Which line is best?

Support Vector Machines (SVMs)



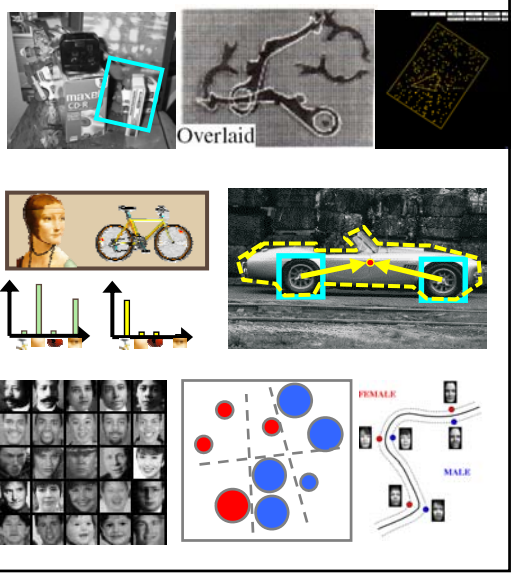
- Discriminative classifier based on *optimal separating line (for 2d case)*
- Maximize the *margin* between the positive and negative training examples

SVMs: Pros and cons

- Pros
  - Many publicly available SVM packages:
  - Kernel-based framework is very powerful, flexible
  - Often a sparse set of support vectors – compact at test time
  - Work very well in practice, even with very small training sample sizes
- Cons
  - No "direct" multi-class SVM, must combine two-class SVMs
  - Can be tricky to select best kernel function for a problem
  - Computation, memory

Recap: genres of recognition approaches

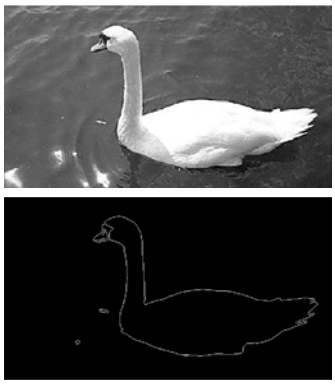
- Alignment: hypothesize and test
  - Pose clustering with object instances
  - Indexing invariant features + verification
- Local features: as parts or words
  - Part-based models
  - Bags of words models
- Global appearance: "texture templates"
  - Classification methods
  - Sliding windows or holistic



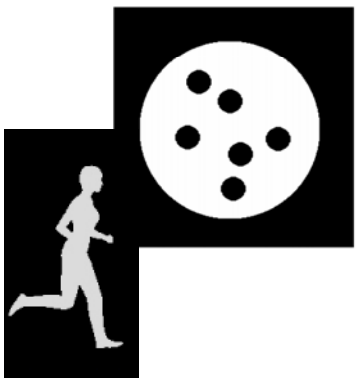
Today

- Shape matching
- *Where have we seen shape before?*

Low-level features



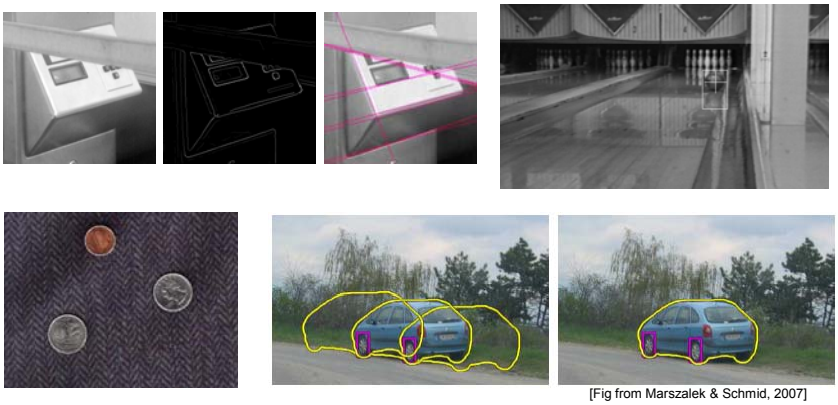
Edges



Silhouettes

Fitting

- Want to associate a model with observed features



[Fig from Marszalek & Schmid, 2007]

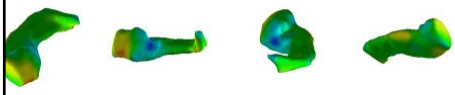
For example, the model could be a line, a circle, or an arbitrary shape.

Deformable contours

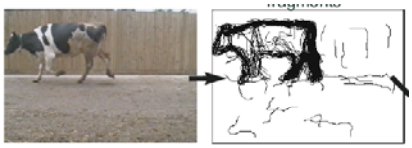
Visual Dynamics Group, Dept. Engineering Science, University of Oxford.

Applications:
- Traffic monitoring
- Human-computer interaction
- Animation
- Surveillance
- Computer Assisted Diagnosis in medical imaging

Role of shape



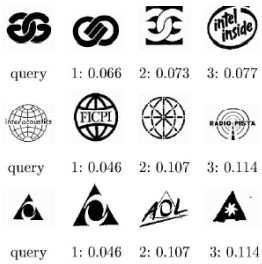
Analysis of anatomical structures
Figure from Grimson & Golland



Recognition, detection
Fig from Opelt et al.

Pose

Morphology
http://usuarios.lycos.es/lawebdelosfosiles/i



Characteristic feature
Fig from Belongie et al.

Shape in recognition

Questions and issues

- Interaction of shape and appearance
- What features?
- How to compare shapes?

Interaction of shape & appearance

- Position specific features are meaningful for objects aligned according to their shape…

Mean: μ

Interaction of shape & appearance

…but unaligned shapes are a problem.



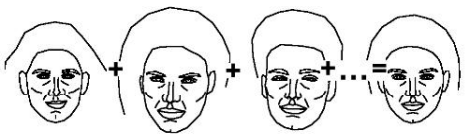
We must include appearance AND shape to construct a prototype.

Prototype faces in shape *and* appearance

1

Mark coordinates of standard features

2



Compute average *shape* for a group of faces

3

Warp faces to mean shape. Blend images to provide image with average appearance of the group, normalized for shape.

Compare to faces that are blended without changing shape.

University of St. Andrews, Perception Laboratory Figures from http://perception.st-and.ac.uk/Prototyping/prototyping.htm

Using prototype faces: aging

"Facial aging": get facial prototypes from different age groups, consider the difference to get function that maps one age group to another.

University of St. Andrews, Perception Laboratory

- **http://morph.cs.st-andrews.ac.uk//Transformer/**

Burt D.M. & Perrett D.I. (1995) Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information. Proc. R. Soc. 259, 137-143.

Questions and issues

- Interaction of shape and appearance
- What features?
- How to compare shapes?

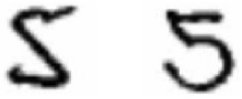


Fig. 1. Examples of two handwritten digits. In terms of pixel-to-pixel comparisons, these two images are quite different, but to the human observer, the shapes appear to be similar.

Figure from Belongie et al.

Chamfer distance

- Average distance to nearest feature

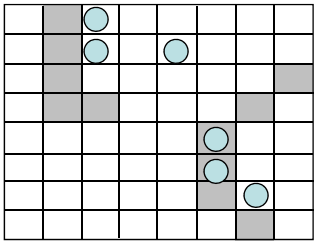
$$D_{chamfer}(T, I) \equiv \frac{1}{|T|} \sum_{t \in T} d_I(t)$$

I = Set of points in image

T = Set of points on (shifted) template

$d_I(t) =$ Minimum distance between point t and some point in I

Chamfer distance



$$D_{chamfer}(T, I) \equiv \frac{1}{|T|} \sum_{t \in T} d_I(t)$$

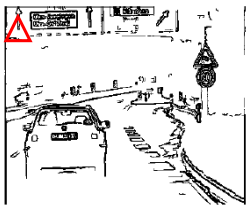
Chamfer distance

- Average distance to nearest feature

$$D_{chamfer}(T, I) \equiv \frac{1}{|T|} \sum_{t \in T} d_I(t)$$

How is the measure different than just filtering with a mask having the shape points?

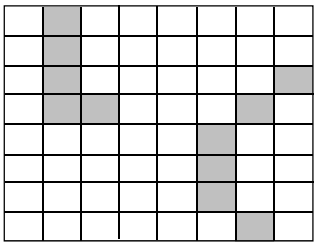
How expensive is a naïve implementation?



Edge image

Distance transform

Image features (2D)



Distance Transform

1	0	1	2	3	4	3	2
1	0	1	2	3	3	2	1
1	0	1	2	3	2	1	0
1	0	0	1	2	1	0	1
2	1	1	2	1	0	1	2
3	2	2	2	1	0	1	2
4	3	3	2	1	0	1	2
5	4	4	3	2	1	0	1

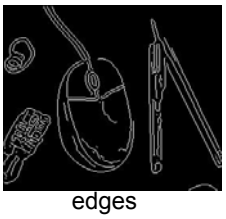
Distance Transform is a function $D(\cdot)$ that for each image pixel p assigns a non-negative number $D(p)$ corresponding to distance from p to the nearest feature in the image I

Features could be edge points, foreground points,…

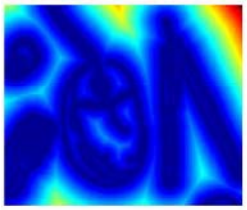
Source: Yuri Boykov

Distance transform

original



edges



distance transform

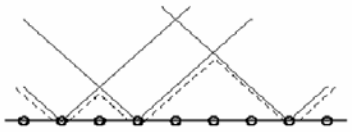
Value at (x,y) tells how far that position is from the nearest edge point (or other binary mage structure)

```
>> help bwdist
```

Distance transform (1D)

Two pass O(n) algorithm for 1D L_1 norm

1. Initialize: For all j
   $D[j] \leftarrow 1_{\mathbf{P}}[j]$ // 0 if j is in **P**, infinity otherwise
2. Forward: For j from 1 up to n-1
   $D[j] \leftarrow \min(D[j], D[j-1]+1)$
3. Backward: For j from n-2 down to 0
   $D[j] \leftarrow \min(D[j], D[j+1]+1)$

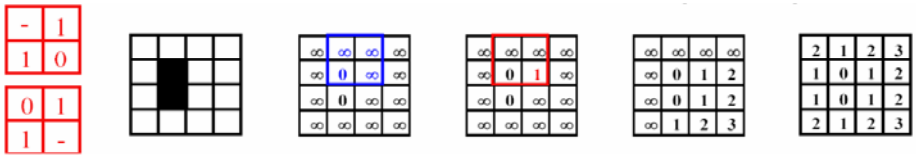


∞	0	∞	0	∞	∞	∞	0	∞
∞	0	1	0	1	2	3	0	1
1	0	1	0	1	2	1	0	1

Adapted from D. Huttenlocher

Distance Transform (2D)

- 2D case analogous to 1D
  - Initialization
  - Forward and backward pass
    - Fwd pass finds closest above and to left
    - Bwd pass finds closest below and to right

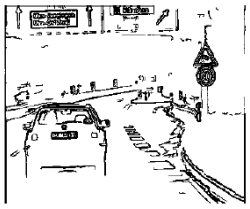


Adapted from D. Huttenlocher

Chamfer distance

- Average distance to nearest feature

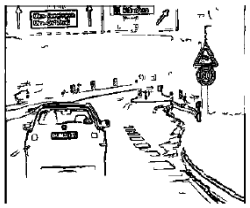
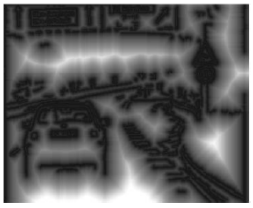
$$D_{chamfer}(T, I) \equiv \frac{1}{|T|} \sum_{t \in T} d_I(t)$$


Edge image **Distance transform image**

Chamfer distance



Edge image **Distance transform image**

Fig from D. Gavrila, DAGM 1999

Chamfer distance: properties

- Sensitive to scale and rotation
- Tolerant of small shape changes, clutter
- Need large number of template shapes
- Inexpensive way to match shapes

An aside: A limitation of active contours

- External energy: snake does not really "see" object boundaries in the image unless it gets very close to it.

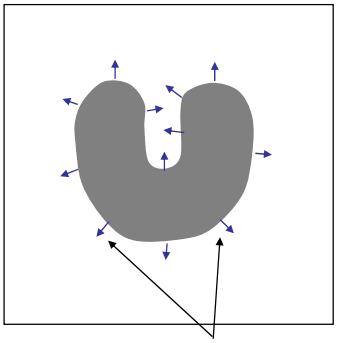
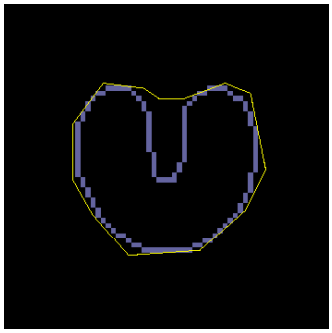


image gradients ∇I
are large only directly on the boundary

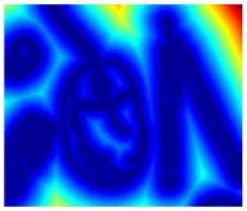


Distance transform can help

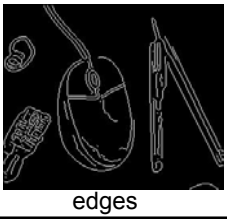
- **External image cost** can also be taken from the **distance transform** of the edge image.

original

-gradient



distance transform

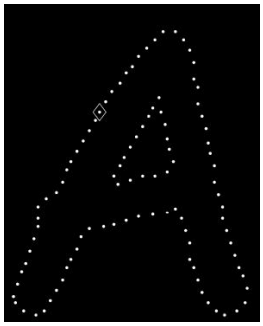


edges

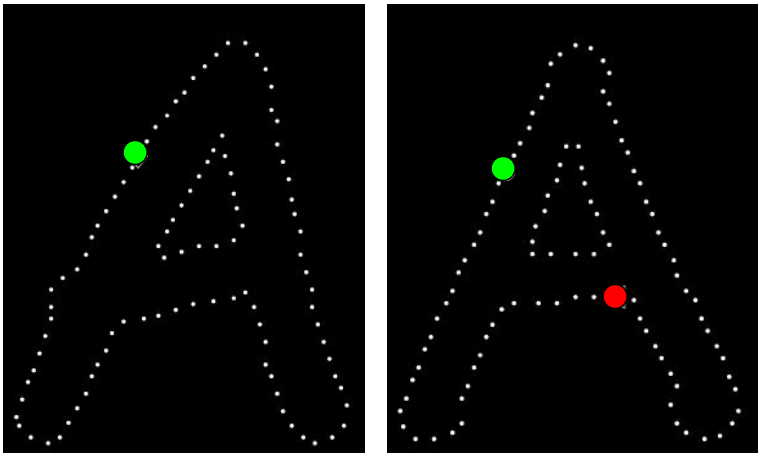
Questions and issues

- Interaction of shape and appearance
- What features?
- How to compare shapes?

- What limitations might we have using only edge points to represent a shape?
- How descriptive is a point?

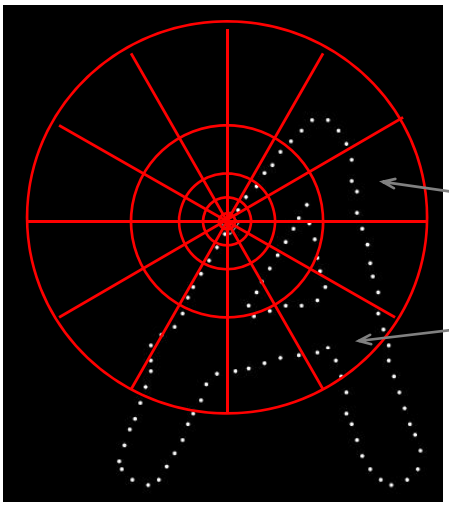


Comparing shapes



What points on these two sampled contours are most similar? How do you know?

Shape context descriptor



Count the number of points inside each bin, e.g.:

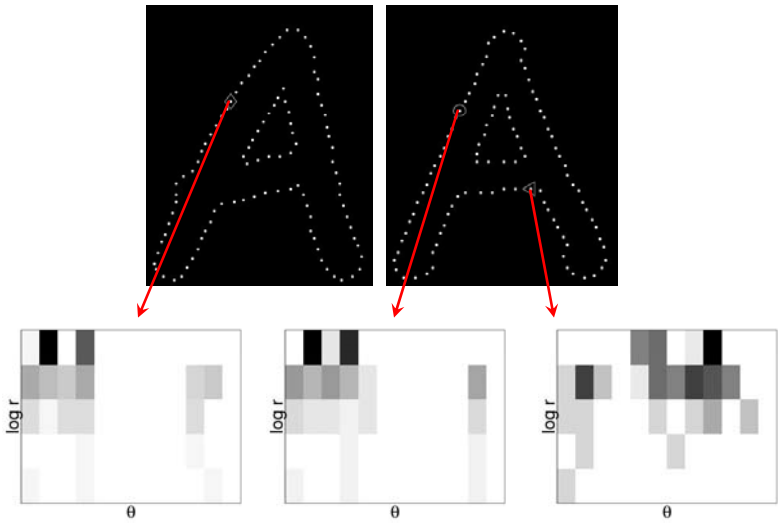
Count = 4

⋮

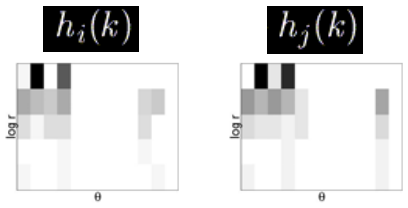
Count = 10

Compact representation of distribution of points relative to each point

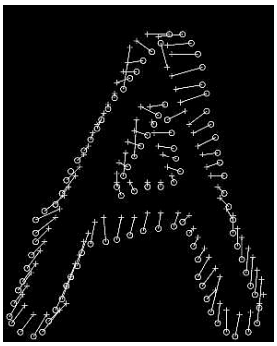
Shape context descriptor



Comparing shape contexts



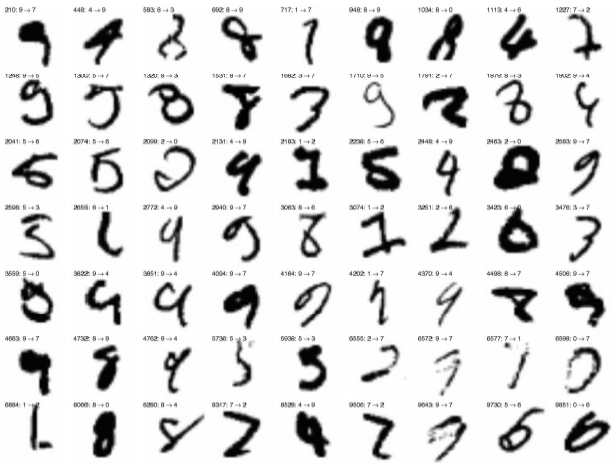
$$C_{ij} = \frac{1}{2}\sum_{k=1}^{K}\frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)}$$



Recover correspondences by solving for least cost assignment, using costs C_{ij}

(Then use a deformable template match, given the correspondences.)

Shape context matching with handwritten digits



Only errors made out of 10,000 test examples

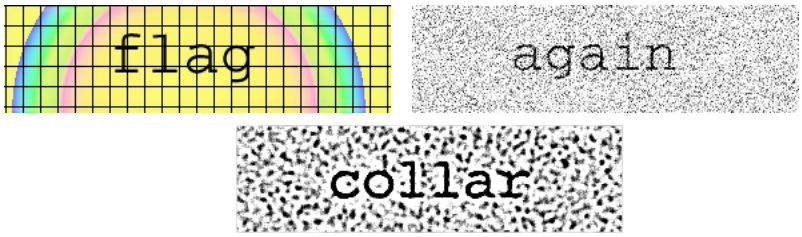
CAPTCHA's

- CAPTCHA: Completely Automated Turing Test To Tell Computers and Humans Apart
- Luis von Ahn, Manuel Blum, Nicholas Hopper and John Langford, CMU, 2000.
- www.captcha.net

Image-based CAPTCHA

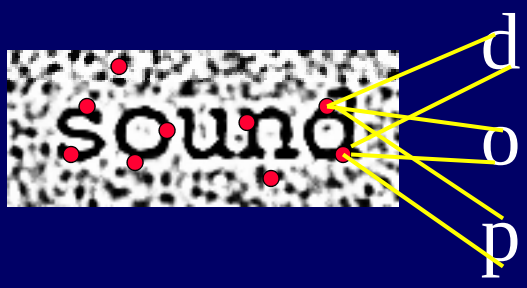
Shape matching application: breaking a visual CAPTCHA

- Use shape matching to recognize characters, words in spite of clutter, warping, etc.



Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA, by G. Mori and J. Malik, CVPR 2003

Fast Pruning: Representative Shape Contexts



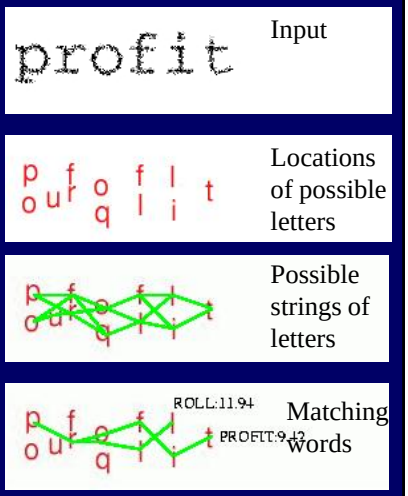
- Pick k points in the image at random
  - Compare to all shape contexts for all known letters
  - Vote for closely matching letters
- Keep all letters with scores under threshold

University of California
Berkeley

Computer Vision Group

Algorithm A: bottom-up

- Look for letters
  - Representative Shape Contexts
- Find pairs of letters that are "consistent"
  - Letters nearby in space
- Search for valid words
- Give scores to the words



University of California
Berkeley

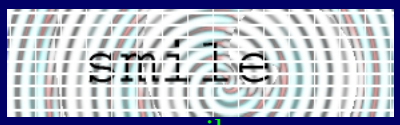
Computer Vision Group

EZ-Gimpy Results with Algorithm A

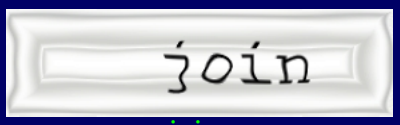
- 158 of 191 images correctly identified: 83%
  - Running time: ~10 sec. per image (MATLAB, 1 Ghz P3)

horse

spade



smile



join

canvas

here

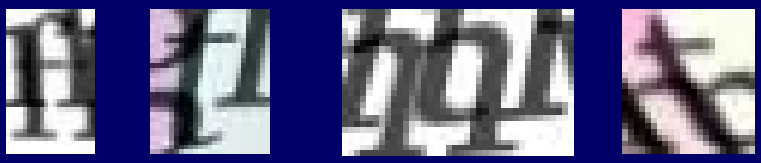
University of California
Berkeley

Computer Vision Group

Gimpy

- Multiple words, task is to find 3 words in the image
- Clutter is other objects, not texture

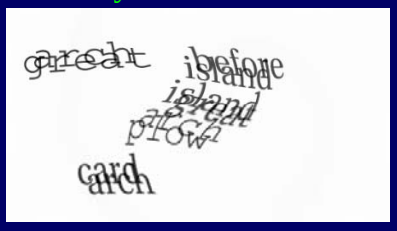
Algorithm B: Letters are not enough



- Hard to distinguish single letters with so much clutter
- Find words instead of letters
  - Use long range info over entire word
- Search problem becomes huge
  - # of words 600 vs. # of letters 26
  - Prune set of words using opening/closing bigrams

Results with Algorithm B

dry clear medical



card arch plate

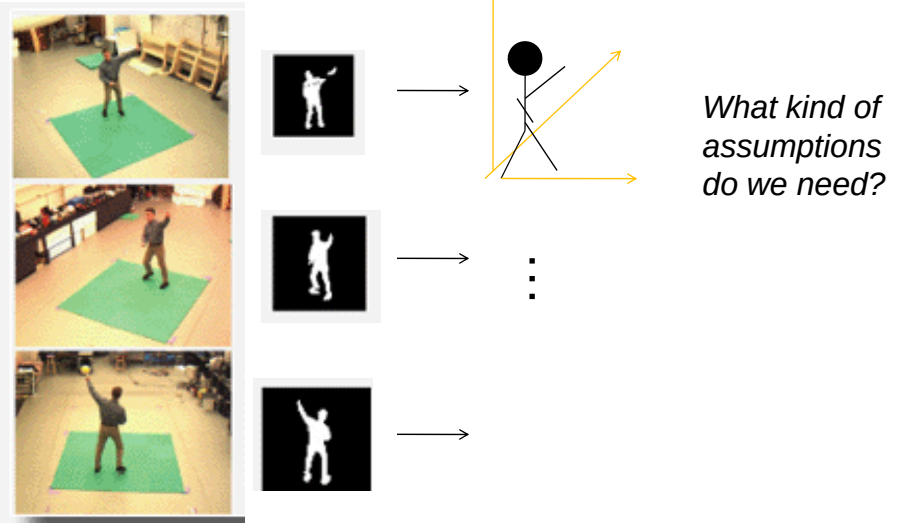
# Correct words	% tests (of 24)
1 or more	92%
2 or more	75%
3	33%
EZ-Gimpy	92%

door farm important

University of California
Berkeley

Computer Vision Group

Shape matching application II: silhouettes and body pose



What kind of assumptions do we need?

Shape-based pose estimation …and animation

- Build a two-character "motion graph" from examples of people dancing with mocap
- Populate database with synthetically generated silhouettes in poses defined by mocap (behavior specific dynamics)
- Use silhouette features to identify similar examples in database
- Retrieve the pose stored for those similar examples to estimate user's pose
- Animate user and hypothetical partner

Ren, Shakhnarovich, Hodgins, Pfister, and Viola, 2005.

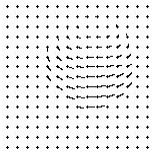
Fun with silhouettes

- Liu Ren, Gregory Shakhnarovich, Jessica Hodgins, Hanspeter Pfister and Paul Viola, Learning Silhouette Features for Control of Human Motion
- http://graphics.cs.cmu.edu/projects/swing/

Summary

- Shape can be defining feature in recognition, useful for analysis tasks
- Chamfer measure to compare edge point sets
  - Distance transform for efficiency
- Shape context : local shape neighborhood descriptor
  - Since local edge points alone are ambiguous to match
- Example applications of shape matching
  - Gimpy visual CAPTCHA: matching characters in clutter
  - Pose estimation from silhouettes

Coming up: Motion and tracking



Tomas Izo