Image Retrieval and Classification using Local Distance Functions

Pushkala Iyer
03/22/2007

Introduction

- The Problem
  - Visual Categorization
- The Solution
  - Application of combined local distance functions

General Discriminative Approach

- Identify interest points
- Select a patch around interest point
- Compute fixed length feature vector (set)
- Define a function which can compare the similarity between 2 such sets
- Feed distances to a learning algorithm (SVM, Nearest neighbor classifier)

Approach

- Metric learning
- Relative importance of features is useful
- Distance function for each exemplar, thus learning a weighting over features
- Advantages
- Output of learning is a quantitiative measure of relative importance
- Ability to combine and select features of different types

Distance functions and Learning Procedure

- Abstract Patch based image features
- N training images => N learning problems
- Concepts: Focal image $\mathcal{F}$, Learning set Candidate Image $\mathcal{I}$
- Distance function is a combination of elementary patch based distances.
- M patches => M patch-to-image distances ($d^{\mathcal{F}}_j(\mathcal{I})$) to compute between $\mathcal{F}$ and $\mathcal{I}$
- $D(\mathcal{F}, \mathcal{I}) = \sum_{j=1}^{M} w^{\mathcal{F}}_j \, d^{\mathcal{F}}_j(\mathcal{I}) = \langle w^{\mathcal{F}} \cdot d^{\mathcal{F}}(\mathcal{I}) \rangle$

Learning

- Triplets of images – ($\mathcal{F}$, $\mathcal{I}^d$, $\mathcal{I}^s$)
- Ideally, using the learned distance function, we want D($\mathcal{F}$, $\mathcal{I}^d$) > D($\mathcal{F}$, $\mathcal{I}^s$)
- $\langle w^{\mathcal{F}} . d^{\mathcal{F}}(\mathcal{I}^d)\rangle > \langle w^{\mathcal{F}} . d^{\mathcal{F}}(\mathcal{I}^s)\rangle$
- If $x_i = d^{\mathcal{F}}(\mathcal{I}^d) - d^{\mathcal{F}}(\mathcal{I}^s)$, then $\langle w^{\mathcal{F}} . x_i\rangle > 0$
- For a given focal image, T triplets are chosen
- Maximal-margin formulation allowing slack for triplets that do not meet condition, while minimizing total slack
- arg min $_{w^{\mathcal{F}},\xi}$ $\frac{1}{2} \| w^{\mathcal{F}} \|^2 + C \sum_{i=1}^{T} \xi_i$

such that for all i in the set of triplets, $\langle w^{\mathcal{F}} . x_i\rangle > 1- \xi_i,\ \xi_i > 0$

Learning

- arg min $_{w^{\mathcal{F}},\xi}$ $\frac{1}{2} \| w^{\mathcal{F}} \|^2 + C \sum_{i=1}^{T} \xi_i$
- Desired margin increased to 1
- L2 regularization is robust to outliers and noise
- Generalization of Distance Metric learning by Schultz and Joachims in [7]
- Some differences
- Triplets do not share focal image
- Exemplar represented by fixed length vector & L22 distance between these vectors is used.
- Contribution: The distance metric algorithm is more widely applicable
- Primal positivity constraint, no bias term (vs SVMs)

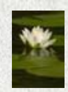
Visual Features and Elementary Distances

- Different kinds of features can be combined – shape features at 2 scales, color feature.
- Filter based patch features – geometric blur descriptors over SIFT
- Two scales of geometric blur features – patch radii - larger 72 pixels, smaller 42 pixels
- 4 oriented channels, 51 sample points = 204 dimensions
- Color features – histograms of 8 pixel radius patches
- Only features of the same type are compared.

Applications

- Image Browsing – navigating image space by visual similarity
- Image Retrieval – given a new image, return a listing of the top K training images that are similar
- Image Classification – run retrival to assign probabilities to each training image, assign the image to the class with the largest total probability.



Experiments

- Caltech101 Dataset – 101 different categories, median 50 images per class



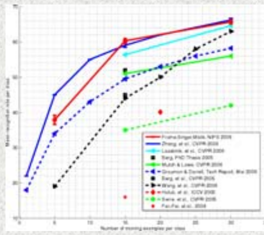


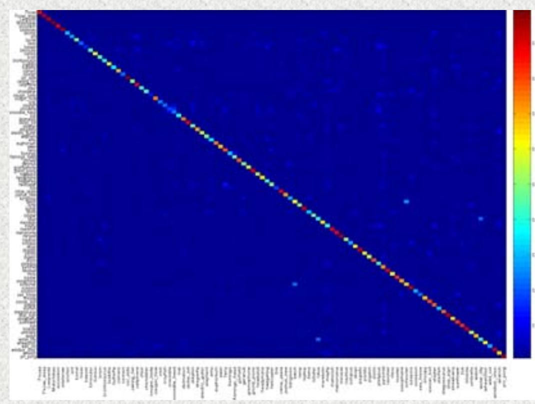
Training Data

- Images resized to 200 x 300
- 3 types of feature, 400 of each type = 1200 features per image
- Triplet choice – uses category labels
- For each M elementary patch distance measure, find top K closest images.
- 3 cases as to what is contained in the K images set (K=5)
- Both in and out of class images
- Only In class images
- Only out of class images
- Final set of triplets for focal image is the union of triplets chosen by the M measures (average 2210 triplets)

Results

- Experiments run with all features, different number of training images per category (5, 15, 30)
- 10 random splits of data into training and test images.
- Average of the mean recognition rate across splits, and standard deviation reported.
- Best value of C (.1), but recognition robust to changes in C value.
- Recognition rates
- Color only – poorest – 6% $\pm$0.8%
- Big geometric blur features – moderate – 49.6% $\pm$1.9%
- Small geometric blur features – better – 52.1% $\pm$0.8%
- Combined shape – 58.8% $\pm$0.8%
- Combined color, shape – 60.3% $\pm$0.7%
- Performance variations – combining shape and color – better on 52 categories, worse on 46, no change on 3.



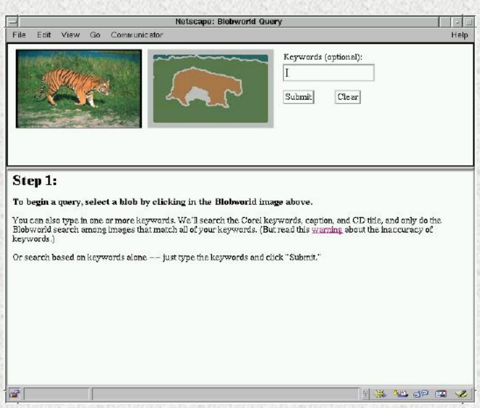


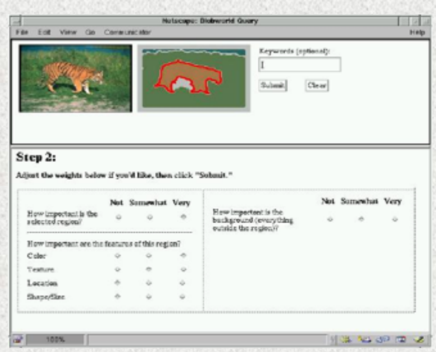
Summary

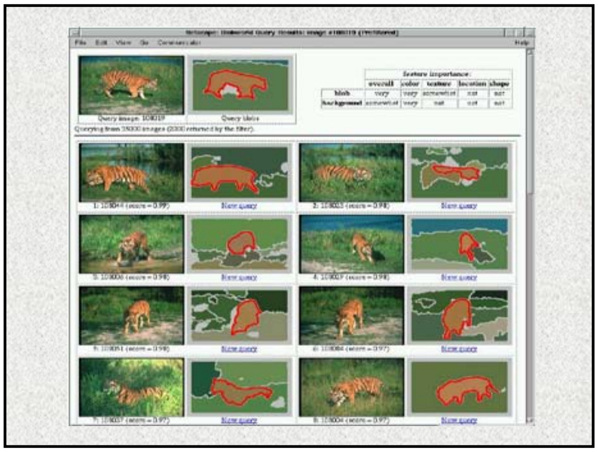
- Relative importance of features can be measured
- Different types of features can be combined
- Shows that the distance metric learning generalization (Schultz and Joachims) is more widely applicable
- Weight vectors are usually sparse(69% are 0) – reduces feature comparisons at test time.
- After comparisons, processing time for computing linear combinations and scoring is negligible – over KNN-SVM of Zhang
- 9 out of 10 worst categories were animal categories
- One possible enhancement – make use of geometric relationships between features in experiments

Blobworld

- Past Research project at UC Berkeley
- System for content based image retrieval
- Segments every image into objects they contain, allowing users to query for photographs based on objects






File Edit View Go Communicator
Help
feature importance:
overall color texture location shape
blob very very somewhat not not
background somewhat very not not not
Query blobs
New query