Object Class Recognition Using Discriminative Local Features

Gyuri Dorko and Cordelia Schmid

Introduction

- This method is a two step approach to develop a discriminative feature selection for object part recognition and detection.
- The first step extracts scale and affine invariant local features.
- The second generates and trains a model using the features in a "weakly supervised" approach.

Local Descriptors

- Detectors
  - Harris-Laplace
  - Harris-Affine
  - Entropy (Kadir & Brady)
- Descriptors
  - SIFT (Scale Invariant Feature Transform)

Learning

- This is also a two step process
  - Part Classifier
    - EM clustering in the descriptor space
  - Part Selection
    - Ranking by classification likelihood
    - Ranking by mutual information criterion

Learning the part classifiers

With the clustering set positive descriptors are obtained to estimate a Gaussian Mixture Model (GMM). It is a parametric estimation of the of the probability distribution of the local descriptors.

$$p(\mathrm{x}) = \sum_{i=1}^{K} p(\mathrm{x}|C_i)P(C_i),$$

Where K is the number of Gaussian components and:

$$\textstyle\sum_i^K P(C_i) = 1.$$

$$p(\mathrm{x}|C_i) = \mathcal{N}(\boldsymbol{\mu}_i, |\Sigma_i)$$

The dimension of the vectors x is 128 corresponding to the dimensions of the SIFT features.

Learning the part classifiers

The model parameters μ_i, Σ_i and $P(C_i)$ are computed with the expectation-maximization (EM) algorithm. The EM is initialized with the output of the K-means algorithm. This are the equations to update the parameters at the *jth* maximization (M) step.

$$\mu_i^j = \frac{\sum_{n=1}^{N} P^{j-1}(C_i|\mathrm{x}^n)\mathrm{x}^n}{\sum_{n=1}^{N} P^{j-1}(C_i|\mathrm{x}^n)}$$

$$\Sigma_i^j = \frac{\sum\limits_{n=1}^{N} P^{j-1}(C_i|\mathrm{x}^n)(\mathrm{x}^n - \mu_i^j)(\mathrm{x}^n - \mu_i^j)^T}{\sum_{n=1}^{N} P^{j-1}(C_i|\mathrm{x}^n)}$$

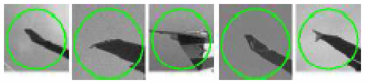
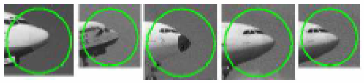
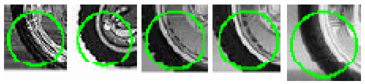
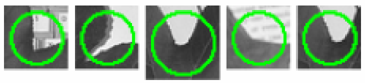
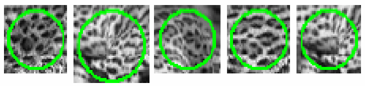
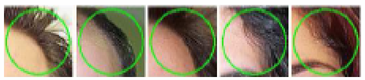
$$P^j(C_i) = \frac{1}{N}\sum_{n=1}^{N} P^{j-1}(C_i|\mathrm{x}^n),$$

Learning the part classifiers

The clusters are obtained from assigning each descriptor to its closest component. The clusters typically contain representative object parts or textures.

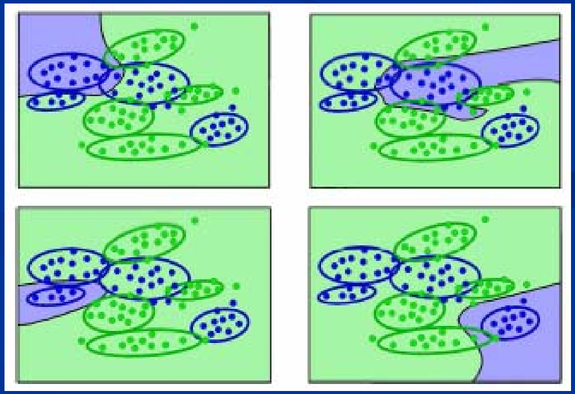
Here we see some characteristic clusters of each database.

With the mixture model a boundary is defined for each component to form *K part classifiers*. Each classifier is associated with one Gaussian

Database	Sample cluster #1	Sample cluster #2
Airplanes		
Motorbikes		
Leaves		
Wild Cats		
Faces		

$$i^* = \underset{i}{argmax}\; p(\mathbf{y}|C_i)P(C_i)$$

A test feature y is assigned to the component i* having the highest probability.



Selection

- The selection ranks the components according to its ability to discriminate between the object-class and the background.
  - By classification likelihood. Promotes having high true positives and low false positives.
  - By mutual information. Selects part classifiers based on the information content to separate background from the objects-class.

Ranking by classification likelihood

- The ranking is computed as follows:

$$R_{\mathcal{L}}(C_i) = \frac{\sum_j^{V^{(u)}} P(C_i|\mathbf{v}_j^{(u)})}{\sum_j^{V^{(n)}} P(C_i|\mathbf{v}_j^{(n)})},$$

Where $V^{(u)}$ and $V^{(n)}$ are the unlabeled (potentially positive) descriptors $v_j^{(u)}$ and negative descriptors $v_j^{(n)}$ from the *validation set*. Performs selection by classification rate. This component hay have very low recall rates. Even though this parts are individually rare, combinations of them provide sufficient recall with excellent precision.

Recall: true features/(true features + true negatives)

Ranking by mutual information

- Best to select a few discriminative general part classifiers.
- Ranks parts classifiers based on their information content for separating the background from the object-class.
- The mutual information of component C_i and object-class O is:

$$
\begin{aligned}
R_I(C_i) = {} & P(\bar{C}_i, \bar{O}) \log \frac{P(\bar{C}_i, \bar{O})}{P(\bar{C}_i)P(\bar{O})} \\
& + P(C_i, \bar{O}) \log \frac{P(C_i, \bar{O})}{P(C_i)P(\bar{O})} \\
& + P(\bar{C}_i, O) \log \frac{P(\bar{C}_i, O)}{P(\bar{C}_i)P(O)} \\
& + P(C_i, O) \log \frac{P(C_i, O)}{P(C_i)P(O)} \\
= {} & \sum_{\substack{k=\{C_i, \bar{C}_i\} \\ l=\{O_i, \bar{O}_i\}}} P(k,l) \log \frac{P(k,l)}{P(k)P(l)}.
\end{aligned}
$$

$$P(\bar{C}_i, \bar{O}) = \frac{\sum_j^{V^{(n)}} P(\bar{C}_i | \mathbf{v}_j^{(n)})}{V^{(u)} + V^{(n)}}$$

$$P(C_i, \bar{O}) = \frac{\sum_j^{V^{(n)}} P(C_i | \mathbf{v}_j^{(n)})}{V^{(u)} + V^{(n)}}$$

$$P(\bar{C}_i, O) = \frac{\sum_j^{V^{(u)}} P(\bar{C}_i | \mathbf{v}_j^{(u)})}{V^{(u)} + V^{(n)}}$$

$$P(C_i, O) = \frac{\sum_j^{V^{(u)}} P(C_i | \mathbf{v}_j^{(u)})}{V^{(u)} + V^{(n)}}$$

$$P(\bar{C}_i) = P(\bar{C}_i, \bar{O}) + P(\bar{C}_i, O)$$

$$P(C_i) = P(C_i, \bar{O}) + P(C_i, O)$$

$$P(O) = \frac{V^{(u)}}{V^{(u)} + V^{(n)}}$$

$$P(\bar{O}) = \frac{V^{(n)}}{V^{(u)} + V^{(n)}}$$

Naively assumes all unlabeled as the object

Final feature Classifier

- Based on the ranking, the n part classifiers of the highest rank are chosen and marked as positive.
- The rest are marked as negative, the true negative and the non-discriminative positive ones.
- Note that each part classifier is based on a Gaussian component, thus the MAP criterion only activates one part classifier per descriptor.

Applications

- Initial step for localization within images. The output is not binary but a ranking of the part classification.
- Classification of the presence or absence of an object in an image. Here is required an additional criterion of *how many p positive* classified descriptors are required to mark the presence of an object. The authors uses this because it is easier to compare.

Experimental Results

Precision by detector and ranking

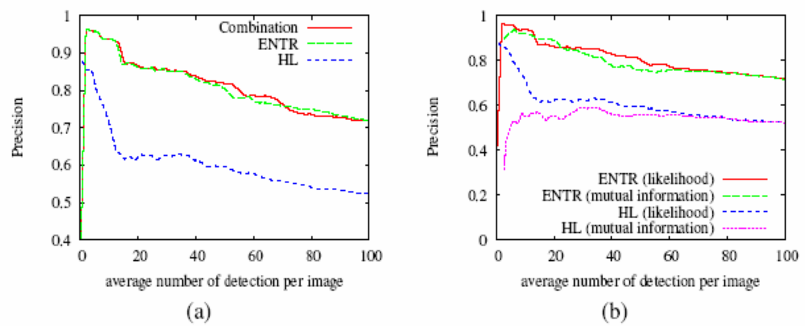


Fig. 6. The precision of the detected features on the bicycle database. (a) evaluates the two detectors and their combination with the ranking method $R_{\mathcal{L}}$. (b) compares the two different ranking methods for the individual detectors.

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}},$$

Feature selection with increasing n

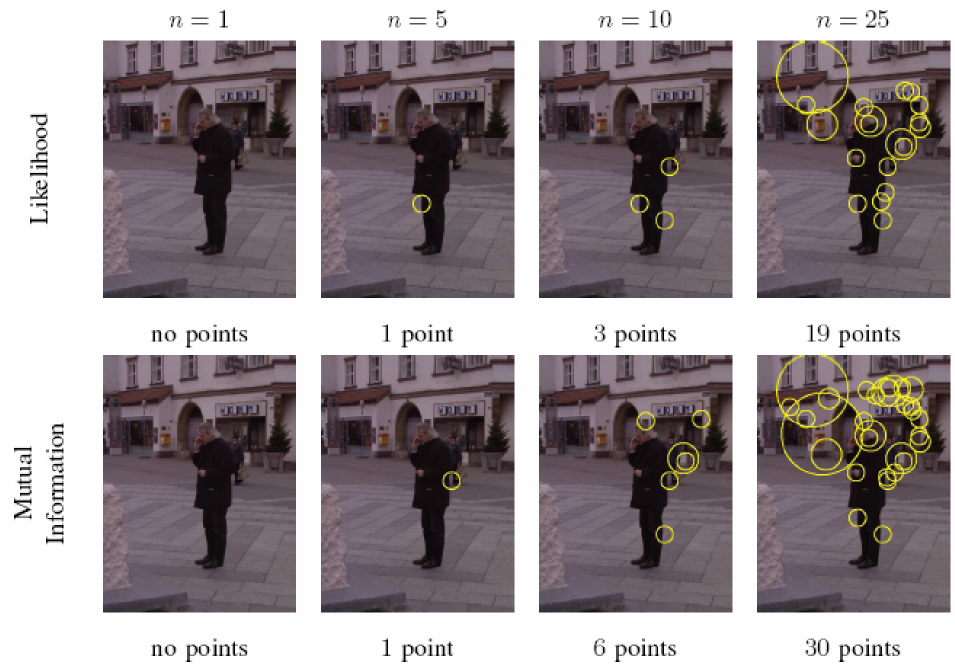


Fig. 8. Feature selection results with increasing n on a sample from the people database. This is one of the most challenging databases as the appearance of the people is very variable. In this case likelihood and mutual information focused on different *part classifiers*, there were no "very special" or "very general" clusters.

Experimental Results

<u>ROC (Receiver Operating Characteristic)</u>

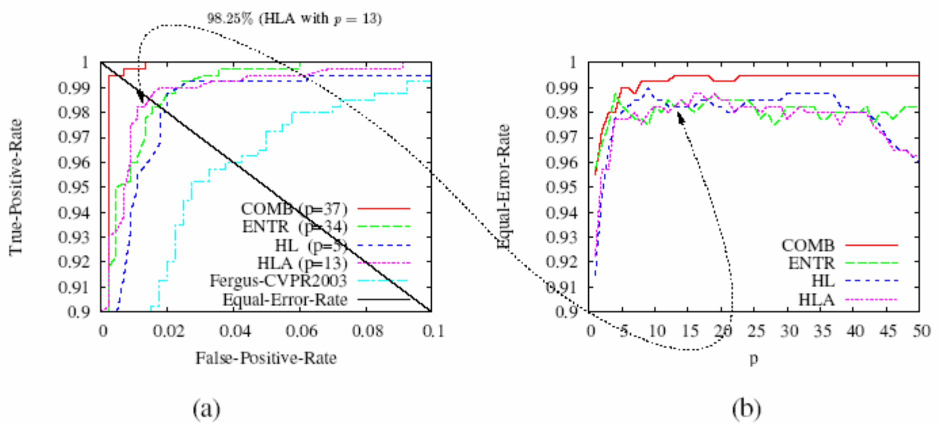


Fig. 9. On the left, the ROC curves for image classification on the motorbikes database using different detectors and estimated p parameters. On the right the corresponding equal error rate curves. The dotted line with arrows shows the connection between the two curves. See the text for an explanation.

<u>True positives on equal-error rate</u>

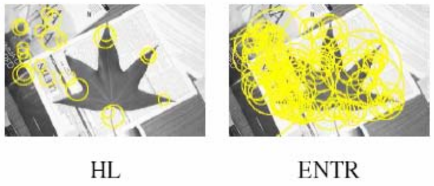


Fig. 10. The output of the HL and ENTR operators on the leaves database.

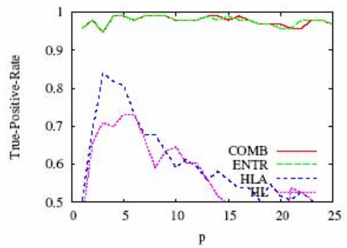


Fig. 11. Equal-error-rate results of image classification on the leaves database.

Experimental Results

TABLE I

Equal-Error-Rate results on image classification using the combination of HL and ENTR detectors (COMB) and $R_{\mathcal{L}}$ ranking. The last column shows the best results reported by *other groups* on the same datasets.

Database	This paper				Others
	Ideal p		Estimated p		
	p	%	p	%	%
Airplanes 	25	**98.75**	28	98.5	94.0 [11]
Faces 	45	**99.54**	33	99.08	96.8 [11]
Motorbikes 	37	**99.5**	37	**99.5**	96.0 [11]
Wild Cats 	7	**91.0**	13	87.0	90.0 [11]
Leaves 	8	**98.92**	8	**98.92**	84 [27]
Bikes 	26	**92.0**	14	88.0	86.5 [15]
People 	13	**88.0**	13	**88.0**	80.8 [15]

TABLE II

Equal-Error-Rate results on image classification with different databases, detectors.

Database	Detector	Ideal p		Estimated p		Others
		p	%	p	%	%
Airplanes	ENTR	18	97.0	8	96.00	94.0
	HL	14	97.75	9	96.25	
	HLA	8	96.75	8	96.75	
Faces	ENTR	12	97.70	19	96.77	96.8
	HL	11	99.54	11	99.54	
	HLA	21	**100.0**	21	**100.0**	
Motorbikes	ENTR	4	98.75	11	98.0	96.0
	HL	9	99.0	5	98.0	
	HLA	16	98.75	13	98.25	
Wild Cats	ENTR	7	83.0	25	82.0	90.0
	HL	12	**93.0**	10	91.0	
	HLA	12	**92.0**	68	89.0	
Leaves	ENTR	8	**98.92**	8	**98.92**	
	HL	5	73.12	2	65.59	84
	HLA	3	83.87	2	68.82	
Bikes	ENTR	29	**92.0**	19	90.0	
	HL	24	84.0	24	84.0	
	HLA	32	70.0	12	64.0	86.5
People	ENTR	12	**88.0**	29	80.0	
	HL	27	78.0	30	76.0	
	HLA	21	76.0	17	74.0	80.8

Experimental Results

Selection of the entropy detector

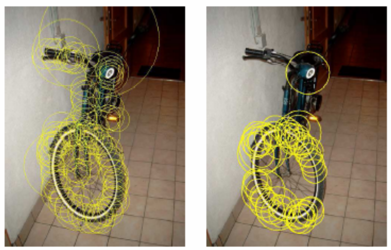


Fig. 12. Selection results on the bicycle database. The ENTR detector output is shown on the left, and the selected discriminative features are shown on the right.

Selection results of different feature detectors

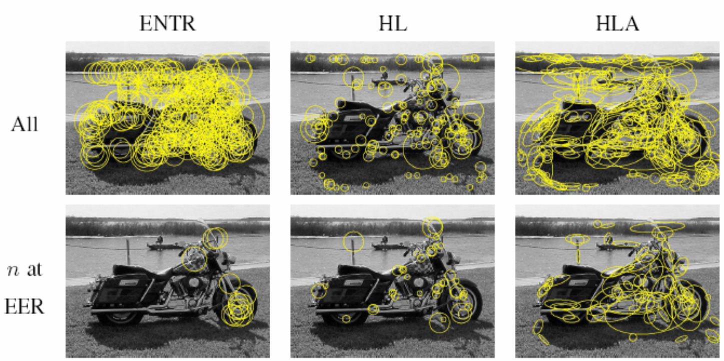


Fig. 13. Selection results using different feature detectors: Entropy of region histograms (ENTR) [8], Harris-Laplace (HL) [19], Harris-Affine (HLA) [20]. The top row shows the output of the interest point detectors, i.e the input to our selection method. In the bottom row we mark only the n best ranked features. For this example we set our parameter n according to the equall error rate operating point from our ROC curves.

TABLE III

EQUAL-ERROR-RATE RESULTS ON IMAGE CLASSIFICATION USING LIKELIHOOD AND MUTUAL INFORMATION AS RANKING METHODS.

Database	$R_{\mathcal{L}}$		$R_{\mathcal{I}}$	
	p	%	p	%
Airplanes	25	98.75	37	98.5
Faces	45	99.54	16	99.54
Motorbikes	37	99.5	49	99.0
Wild Cats	7	91.0	41	90.0
Leaves	8	98.92	9	97.85
Bikes	26	92.0	14	90.0
People	13	88.0	12	82.0

Thanks!