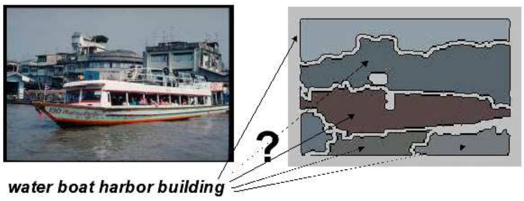
Approximation of Linear Discriminant Analysis for Word Dependent Visual Features Selection from Mislabeled Images

Hervé Glotin, Sabrina Tollari, Pascale Giraudet
LSIS UMR CNR–INCOD team - 83957 La Garde cedex, France. {glotin, tollari, giraudet@univ-tln.fr}

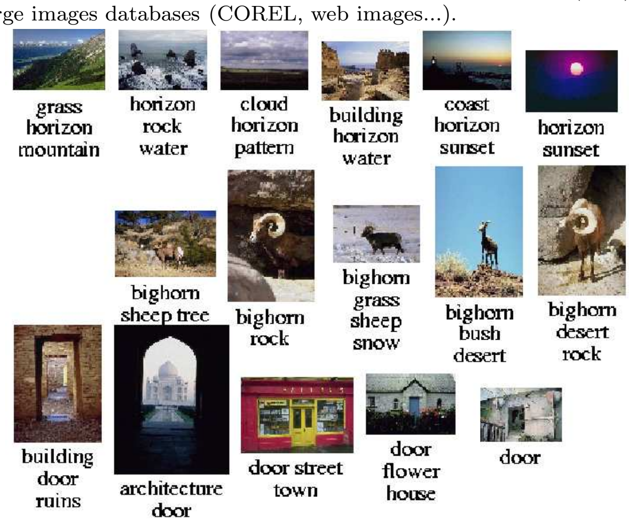
1 PCA seeks for representation but LDA for discrimination

Although PCA finds components that are useful for representing data, these components are not said to be useful for discriminating between data in different classes. Considering data for 'O' and data for 'Q', PCA might discover the gross features that characterize 'O's and 'Q's, but might ignore the tail that distinguishes an 'O' from a 'Q'. LDA would discover this discriminative feature.

2 How to run LDA on mislabeled images ?



Due to automatic segmentation errors or abstract words, there is no one to one relationship between word and image segment (blob) in large images databases (COREL, web images...).



3 Approximation of LDA

Let be:
- S the theoretical values set of one feature x for all the blobs that are *exactly* representing word w_k,
- T the features set *of all blobs included in all images labeled* by w_k,
- $T = T_G \cup S$, with $T_G \cap S = \emptyset$ (we have $c_{T_G} \neq 0$; we note for set E, c_E cardinal, μ_E average of x_i values of $x \in E$, v_E variance),
- G the set containing all values of x from all blobs contained in images that are not labeled by w_k,
- B_{DE} (resp. W_{DE}) the Between variance (resp. the Within variance) between any sets D and E.

LDA calulates for each x the discriminant power:

$$F(x; w_k) = \frac{1}{1 + V(x; w_k)} \text{ where } V(x; w_k) = \frac{W_{SG}}{B_{SG}}. \quad (1)$$

We show that, if $\mu_{T_G} = \mu_G$ *and* $v_{T_G} = v_G$ (simple assumption of context independency provided by any large enougth image database):

$$\hat{V}(x; w_k) = \frac{W_{TG}}{B_{TG}} = A(w_k).V(x; w_k) + B(w_k).\Big(1 - C(x; w_k)\Big) \quad (2)$$

where $A > 0$ and $B > 0$ are *independent of* x. Moreover we show that for discriminant x $C(x; w_k) << 1$. Thus $\hat{V}(x; w_k)$ is a linear function of $V(x; w_k)$, and order of $\hat{V}$ and V values are the same. So we estimate the most discriminant features by ranking $\hat{V}$.

4 Application to visual features selection

Experiments are conducted on COREL: 267 words, 10K images, 10 blobs by image generated with Normalized Cut method. Each blob is described by 40 visual features: 2 of positions, 6 of shapes, 16 of colors (rgb, lab...) and 16 of texture extracted by gaussian filters.

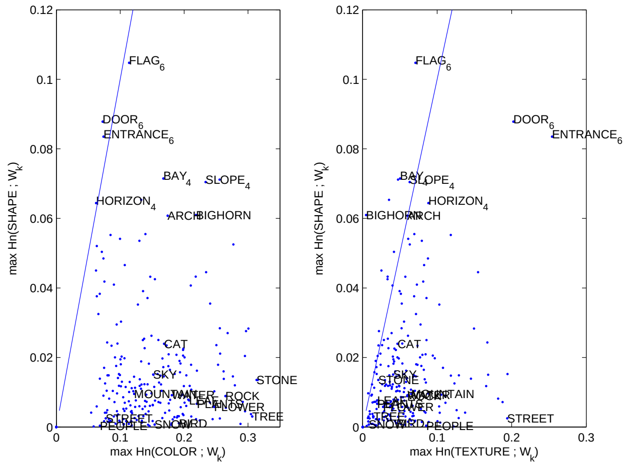


Fig.: Maximum values, in 3 feature subsets (COLOR, TEXTURE, SHAPE), of normalised approximated discrimination power $Hn(x; w_k) = \hat{F}(x; w_k) / \sum_x \hat{F}(x, w_k)$, for the 14 most frequent words. Results are intuitively correct: TREE, ROCK, FLOWER, PLANTS are only discriminated by color; while STREET, ENTRANCE are more discriminated by texture. SHAPE is in average not very competitive in comparison to COLOR or TEXTURE.

5 Conclusion: image classification gain

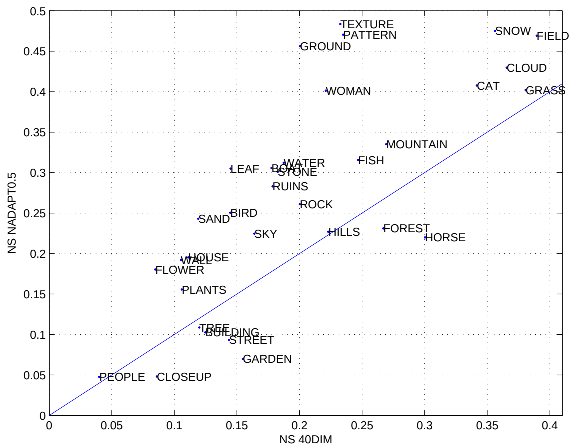


Fig.: Normalized Score (=sensi+specifi-1) of image classification using HAC with 40 features (40DIM), versus automatically selected ones (NADAPT0_5) using only 1 to 8 features by word while increasing classification (+37% relative gain). Details in Tollari & Glotin, "Keyword dependant selection of visual features and their heterogeneity for image content-based interpretation", internal report LSIS RR.2005.003-2005.