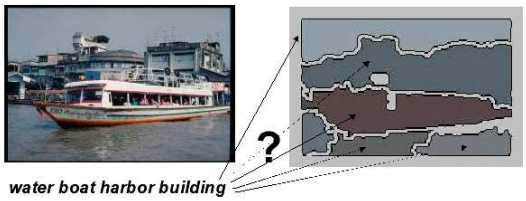
# How to run LDA on mislabeled images ?




**1) LDA should calulate for each feature $x$ and word $w_k$ the discriminant power:**

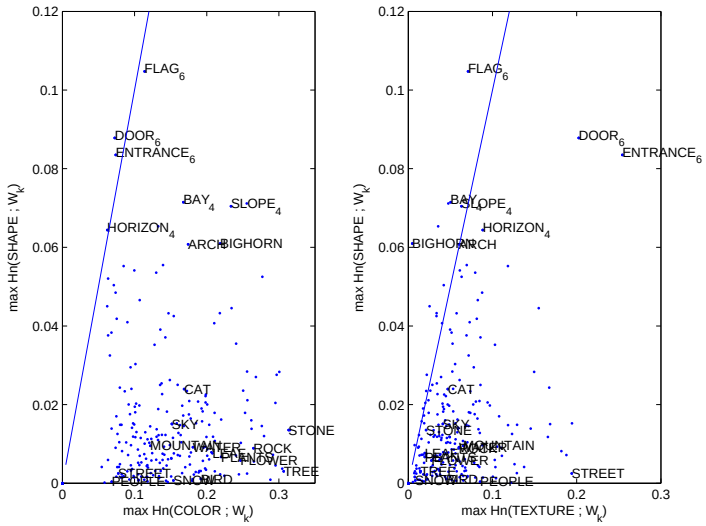
$$F(x;w_k) = \frac{1}{1+V(x;w_k)} \textbf{ where } V(x;w_k) = \frac{Within\ variance}{Between\ variance}. \tag{1}$$

We show that on large enougth image database:

$$\hat{V}(x;w_k) = \frac{\hat{W}}{\hat{B}} = A(w_k).V(x;w_k) + B(w_k).\Big(1 - C(x;w_k)\Big) \tag{2}$$

where $A > 0$ and $B > 0$ are *independent of* $x$, and $C(x;w_k) << 1$. Thus we estimate discriminant power by ranking $\hat{V}$.

**2) Estimation of features discriminant power:**




**3) Application: reduction of 90% of features space size and +37% relative gain on image classification**