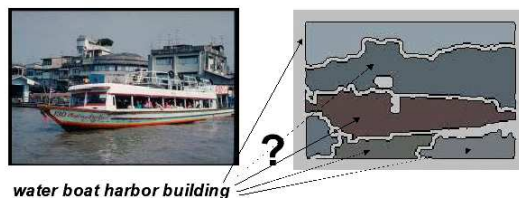


# How to run LDA on mislabeled images ?



water boat harbor building

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

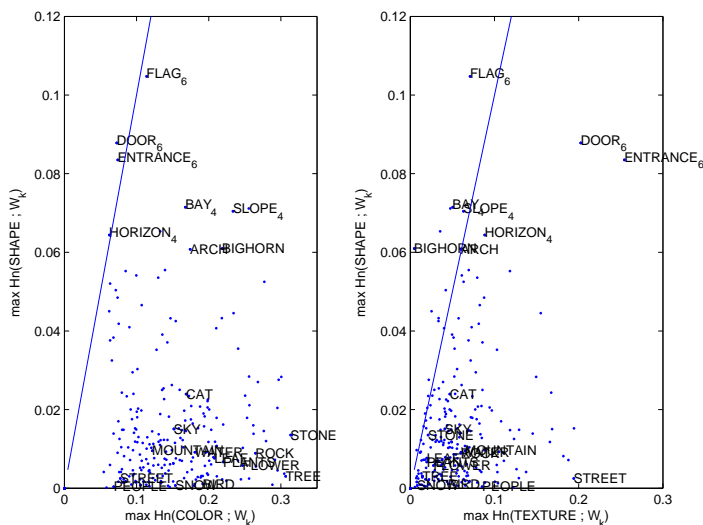
$$F(x; w_k) = \frac{1}{1 + V(x; w_k)} \text{ where } V(x; w_k) = \frac{\text{Within variance}}{\text{Between variance}}. \quad (1)$$

We show that on large enough image database:

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

where  $A > 0$  and  $B > 0$  are independent of  $x$ , and  $C(x; w_k) \ll 1$ . 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

