1 Approximating Visual Clusters

We present our model "DIMATEX", a fast image auto-annotation system. Trying to tackle with the high-dimensional problem, fuzzy segmented and fuzzy labelled visual object, DIMATEX splits each visual feature in two classes: low (0) or high (1) value. See above 2 visual vector approximation clusters in 13 visual features (3 for shapes, 6 LAB colors, 4 textures):

\[ P(w, c_k|J) = P(w, b_i|J). \]

2 Training Joint Visuo-Textual Distributions

Assuming that for each image \( J_i \) of training set \( T \), \( P(J_i|T) = \frac{1}{|T|}. \)

\[
P(w, c_k|T) = \sum_{J_i} P(w, b_i|J_i) P(J_i|T) = \frac{1}{|T|} \sum_{J_i} P(w, b_i|J_i, T) = \frac{1}{|T|} \sum_{J_i} P(w|b_i) P(b_i|J_i, T).
\]

We set \( P(J_i|b_i, T) = 1 \) if \( w \) annotates \( J_i \), else 0. And we estimate for each of the \( m \) regions \( b_i \) of image \( J_i \) of \( T \):

\[
P(b_i|J_i, T) = \frac{\text{area}(b_i)}{\sum_{i=1}^{m} \text{area}(b_i)}, \quad (3)
\]

where \( \text{area}(b_i) \) is the number of pixels contained in \( b_i \).

3 Image Auto-annotation

Using previous joint distributions, one can estimate the most accurate words for annotating an image \( I \), by simply picking a desired number of words that have the highest probability \( P(w|I, T) \) that word \( w \) annotates \( I \). As image regions \( \{b_1, \ldots, b_m\} \) form a partition of \( I \), we estimate that:

\[
P(w|I, T) = \sum_{i=1}^{m} P(w|b_i, I) P(b_i|I, T). \quad (4)
\]

Let be \( c_k \) the cluster of region \( b_i \), then from (1):

\[
P(w, b_i|I, T) = P(w, c_k|I, T) \geq \frac{P(w, c_k|T)}{P(c_k|T)}, \quad (5)
\]

where \( P(w, c_k|T) \) is given from (2), and \( P(c_k|T) \) is simply estimated from the training data.

Gain over priors of Normalized Score (Sensi.+Specif.-1), for different auto-annotation models. Prior model uses only word frequency. DIMATEX results are given with less or more than 4 emitted words. LSA and PLSA are from [Monay2003] and Hierar. from [Barnard2003], trained and tested on the same database than DIMATEX:

4 Results

Experiments are done on COREL where a maximum of 5 'manual' words are globally labelling each image. DIMATEX is runing segmenting each image in 10 regions with Normalized Cuts algorithm [Barnard2003]. We give below images with 1: COREL manual annotation, 2: DIMATEX Auto-Annotations, 3: PL-SAWORDSFEATURES, 4: PL-SAWORDSFEATURES, 5: DIRECT, and 6: LSA. (3,..,6) are from [Monay2003 & http://www.idiap.ch/monay/acmm04]:

Gain over priors of Normalized Score (Sensi.+Specif.-1), for different auto-annotation models. Prior model uses only word frequency. DIMATEX results are given with less or more than 4 emitted words. LSA and PLSA are from [Monay2003] and Hierar. from [Barnard2003], trained and tested on the same database than DIMATEX: