***Report***

*Basic image retrieval system based on colour histogram representation.*

**Histogram Intersection approach**

Histogram can be generated from the images very quickly and the comparison of these histograms is computationally fast and efficient. The histogram intersection can also handle partial matches if the area of the two histograms is different. But images retrieved by using color histogram may not be always logically related because they take into consideration just the color distribution.

Histogram intersection calculates the similarity of two discretized probability distributions or histograms, with possible value of the intersection lying between 0 (no overlap) and 1 (identical distributions).

Utilizing the effective performance of histogram intersection, I would like to try including additional features such as object location, texture, and shape to test the improvement in image retrieval.



**Earth Mover’s Distance approach**

EMD computation is very slow with image color histograms. But its performance is better than histogram intersection and is similar to human perception. EMD can match images irrespective of blurring or other local deformations. The algorithm tries to match the total amount of colors in images, at pixel to pixel level , and provides a distance value as output which physically represents the amount of color mass displaced.

Hence, when fast and rough classification is required histogram intersection might be better choice, but when exact matching is required earth mover algorithm is the winner.

Computation time for EMD is linear with respect to number of bins in the histogram, but it grows exponentially with the dimensions. Thus, to fully reap the benefits that earth mover’s distance approach provides, I would like to try its implementation of image retrieval using distributed computing.

EMD useful in detecting patterns, using summaries of data a.k.a. signatures. Signature is a list of feature -mass pairs, where feature can be anything like letters in text, words in a page or colors in an image, and mass is the frequency of occurrence of that feature. To compare signatures EMD needs to define a distance between features, which is also interpreted as the cost of turning one unit mass of feature into the other feature. EMD then equals to the minimum cost of turning one signature into the other. 