Improving the Boosted Correlogram

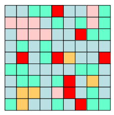
Nicholas R. Howe and Amanda Ricketson
Smith College
Northampton, Massachusetts, USA

Abstract

Introduced seven years ago, the correlogram is a simple statistical image descriptor that nevertheless performs strongly on image retrieval tasks. As a result it has found wide use as a component inside larger systems for content-based image and video retrieval. Yet few studies have examined potential variants of the correlogram or compared their performance to the original. This paper presents systematic experiments on the correlogram and several variants under different conditions, showing that the results may vary significantly depending on both the variant chosen and its mode of application. As expected, the experimental setup combining correlogram variants with boosting shows the best results of those tested. Under these prime conditions, a novel variant of the correlogram shows a higher average precision for many image categories than the form commonly used.

Correlogram Variants, Old and New

All correlograms are sets of statistics describing the average environment surrounding points with similar color in an image. A vector can be assembled containing all the chosen correlogram statistics for all possible colors. This vector, capturing useful information about the contents of the image, can be compared with vectors for other images and used for retrieval and classification. The figures below illustrate the computation of various environment statistics for a single point (the red pixel at the center of the image).

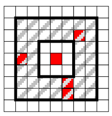


Original Sample Image

Banded Autocorrelogram

(The Old Standby)

Typically called simple the correlogram, this statistic has become widely used. It measures the probability that pixels within specified square rings around a central pixel will have the same color.

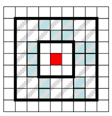


P(Red within 2-3 pixels of Red) = 4/40 = 10%

General Correlogram

(Awkward Due to Size)

This statistic measures correlations between all pairs of colors, requiring quadratic storage space compared to the autocorrelogram.

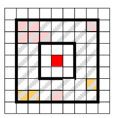


P(Blue within 2-3 pixels of Red) = 15/40 = 37.5%

Color Band Correlogram

(A Novel Combination)

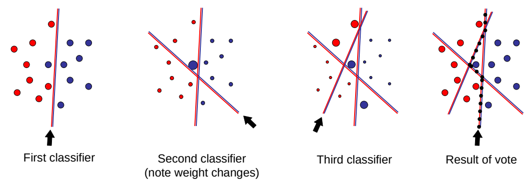
This statistic augments the autocorrelogram with correlations between a color and groups of similar colors. The experiments here use either one or two such similarity levels.



P(Red-like within 2-3 pixels of Red) = 10/40 = 25%

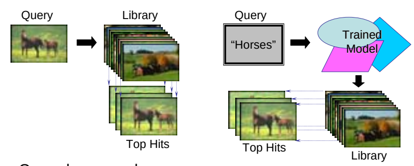
Boosting: An Introduction

Boosting repeatedly trains simple classifiers to recognize a target class. Adjustable weights focus attention on instances that have been classified incorrectly by earlier rounds. A weighted vote of all the trained classifiers together gives greater accuracy than any single classifier alone. The figures below illustrate this process for a simple task.



Experimental Setup

The traditional way to evaluate approaches to image retrieval uses queries consisting of a single image, which are compared with a library of images to determine the most similar for retrieval. Boosting motivates a paradigm shift away from query-by-example, toward natural language queries with models trained offline. Experiments 2 and 3 adopt the offline training model. For a fair comparison between boosted and unboosted methods, all techniques are allowed access to a training set containing positive and negative examples of one of 15 different image concepts. The models generated are then tested for accuracy on a previously unseen test set.



Query-by-example paradigm

Trained model paradigm

Experiment 1: Unboosted Retrieval

This experiment tests the native performance of four sets of correlogram statistics on a query-by-example task. The four descriptors used are the banded autocorrelogram (Auto), the general correlogram (GC), and the color band correlogram with two and three bands (CB2 and CB3, respectively.)

Experiment 2: Greedy Nearest Exemplars

This experiment establishes a control for the boosted classifiers using the trained model paradigm. The model consists of a set of exemplars of the target class, with new images ranked according to the distance to their nearest exemplar image. Exemplars are chosen greedily.

Experiment 3: Boosted Retrieval

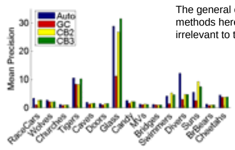
This experiment applies boosting with the four descriptors from Experiment 1. The boosted models created for each class on the training data yield a score for each unknown image, which is used to determine the retrieval rank.

Experimental Results

Query-by-example is not a particularly effective approach for image retrieval. The bars in the graphs represent the mean precision for each image category, integrated over all recall levels. The results in Experiment 1 are markedly lower than those 2 and 3. Still, all methods perform significantly above chance (0.5%).

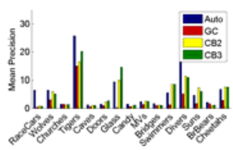
The general correlogram performs much worse than the other methods here, presumably due to the inclusion of many features irrelevant to the target class.

Experiment 1 Results

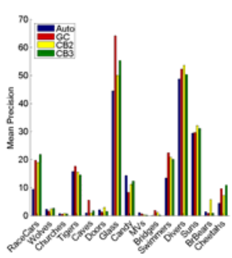


Experiment 2 Results

Greedy nearest exemplars works somewhat better than query-by-example. Again, the general correlogram does worse than the other methods.



Experiment 3 Results



Boosting produces the highest mean precision of the three experiments. In addition, the general correlogram has the best overall performance. The color band correlograms do nearly as well. The boosting process acts as a filter, selecting the statistics most relevant for identifying the target class.

Conclusions

- **Without boosting**, extra irrelevant information in the correlogram variants **lowers** performance.
- **With boosting**, the same variants show **higher** performance.
- Boosting apparently discriminates relevant features from irrelevant.
- Boosted models give the best average precision over all the experimental frameworks.
- Color-band correlogram achieves boosted performance near general correlogram, with much lower storage cost.

Smith College