

Experiments in Binning Image Statistics

Nimar S. Arora

December 6, 2007

1 Introduction

Various vision tasks require the computation of image statistics and aggregating them into histograms. These histograms are usually compared using the χ^2 distance which gives a rough idea of the similarity of the two image patches. For example, in [2] a histogram of key-points is collected on a rectangular grid for the category recognition task. Other researchers, [3], have used local image statistics around a point to find corresponding points.

In general, image statistics produce a vast array of number of varying magnitudes. Some form of aggregation is required to ensure robustness. Although k-means clustering is a commonly used device in this setting, the Euclidean distance on which it is based is not always the appropriate measure of distance. A pre-processing step where the statistics are aggregated into histograms is quite helpful. In this study, we study the issues surrounding binning these statistics by feature location and feature magnitude. Where, a feature can be interpreted as the response of an arbitrary filter or key-point detection algorithm.

We claim that the optimal binning strategy depends on the particular feature being computed. Towards this end, we try various strategies on a fixed set of features and evaluate their performance on a specific task. The task that we consider is object categorization using the CalTech 101 dataset [1]. The overall approach is to use a nearest neighbour classifier using the χ^2 distance between the feature histograms of the test image and the training images.

In the subsequent sections we describe the features that we collected and the performance of various binning strategies that we evaluated on these features.

2 Feature Collection

The features used in this study are very simple, however they are representative of the type of features normally collected in image statistics. We used two types of features: brightness magnitudes, and brightness gradients. For brightness gradients, we computed the difference in brightness magnitude between the pixel and its neighbour to the right as well as the neighbour below. This generates two numbers per pixel, which can, in fact, be thought of as four numbers since we preserve the signs.

No smoothing was performed before computing the gradients. We did run experiments with smoothing at various scales ($\sigma = .5, 1, \sqrt{2}$), but these produced no significant improvement. We also considered collecting color statistics by measuring the *hue* of each pixel, but this didn't help either.

Once a feature has been collected at a location on the image, it is put into an appropriate histogram bin based on its location w.r.t the center of the image and the magnitude and sign of the feature. A complete descriptor for the image is constructed by computing these counts over all the features in every location in the image.

3 Spatial Aggregation

We considered two different forms of spatial aggregation: a log-polar grid, and a rectangular grid. In the log-polar grid, the image was divided into eight angular regions at 45° intervals around the center of the image. The distance from the center was binned by taking the integral part after computing a logarithm to the base 4. The rectangular grid was constructed by dividing each image axis into 8 or 16 equally spaced intervals. The category recognition results showed that the log-polar grid handily outperformed the rectangular grid.

Categories	Log-polar grid	Rectangular Grid
10	80%	69%

The experiments above and all other validation experiments in this study were performed on the first 10 categories (including the *background* category) using 10 randomly selected training and test images.

For the log-polar grid we also considered aggregating the distances by relative distances w.r.t the size of the image. However, this performed much poorly (66%) on the same task as above.

4 Brightness Aggregation

Similar to the distances, we considered the integral part of the logarithm to the base 4 for binning the brightness on a log-scale. We also considered binning the distances using different size linear buckets ranging from 4 to 40. However, the log-scale seems to easily outperform the linear scale as shown below.

Categories	Log bins	Linear bins
10	80%	72%

5 Gradient Aggregation

The gradients (on the un-smoothed image) were subjected to the same log-scale and linear scale binning as for the features above. It was observed that linear scale binning using a step size of 8 performed the best. It turns out that the sign of the gradient does matter in the binning process.

Categories	Log bins	Linear bins	Linear bins ignoring sign
10	78%	80%	80%
20	48%	53%	52%

The experiment with 20 categories used 15 randomly generated training and test images.

6 Results

Our final results on the CalTech 101 dataset using the simple methods above using all 102 categories were as follows:

Training Images	Test Images	Category Avg.
15	15	25%
30	< 30	30%

The best performing binning that we used for the above experiments were – 8 angles, log base 4 for the square of the distance, gradients in steps of 8 preserving sign, and log base 4 for the brightness magnitude.

7 Conclusion

From the experiments it appears that features should be aggregated in space using a log-polar grid rather than a uniform rectangular one. Brightness magnitudes should be aggregated by log-intervals. However, brightness gradients should be aggregated by linear intervals.

In general, this study demonstrated that simple Euclidean distance is not appropriate for clustering image features. Certain features are better aggregated on a log-scale and this needs careful analysis on a per-feature basis.

References

- [1] L. FEI-FEI, R. FERGUS AND P. PERONA. *Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories*. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004
- [2] SVETLANA LAZEBNIK, CORDELIA SCHMID, AND JEAN PONCE. *Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories*. CVPR, 2006
- [3] ALEXANDER C. BERG, TAMARA L. BERG, JITENDRA MALIK. *Shape Matching and Object Recognition using Low Distortion Correspondence*. CVPR 2005