Brandon Lucia
CSE 590v
Review   Learning Globally-Consistent Local Distance Functions for Shape-Based Image Retrieval and Classification

The key contribution of this paper is very straightforward. Using a technique based on prior work, and leveraging such previously developed techniques as SIFT and arithmetic blur feature detection, this work determines the distance in feature space of images from other images in a set of test images. What makes this contribution especially interesting is the fact that while many previous attempts on this goal have used methods which learn a single distance function for all images in the training set, this work develops a local distance function for each. Further, the training process uses triples of images, each a positive, negative, or reference for the others in each triple in which it appears   this causes the constraints imposed by the training to be strongly inter-related, and the set of constraints to be globally consistent across all images. This fact is a very novel aspect, as it is an extension of the technique which they seem to heavily draw upon from ([6]   Frome, et al), and they claim explicitly that due to the fact that they employ local distance functions w.r.t. single images, instead of global distance functions, the technique is a departure from the body of metric learning work .

For the most part, given that this paper is just 7 pages long, and probably so by necessity, not option, it is as well motivated as can be expected. While the SIFT and arithmetic blur techniques on their own, and the image recognition techniques which draw on them were not explained in great detail in this work, they are all well cited, providing an interested competent graduate student enough background and ancillary information to comprehend and analyze the ideas presented. In addition to this, section 5 even provides a more detailed view of some points from the technical section with which I was left unsatisfied (feature selection, what each of the triplets contains, etc). Technically, the paper was thorough, though their mathematics, to the inexperienced could appear a bit oblique. These details are necessary, however, as they constitute the important, novel and reproducible aspects of the technique. At times, the explanations accompanying the math seemed to be somewhat detached from the mathematics at work, and while throughout the derivation of their dual-method optimization it was clear what they were doing, it was not always completely clear what directed them to make this decision (though some of what they were doing was explained by the fact that [6] and [18] used a similar or the same derivation).

The experimental results presented were the point on which I was least satisfied. This recognition technique was trained with a certain set of images, and was set then to recognize related images to those it learned. In the paper the authors actually indicate that their approach (rightfully) exploits the artifacts in the dataset , but then that it is not tuned to the ... data set . This seems a little inconsistent, and leads me to wonder how general this approach actually is, and what results would have been obtained using a data set beside the Caltech101 data set.

Overall the material within was novel and enlightening, and fairly approachable, even with little background. It would be interesting to consider how this technique applies to data sets, and how it reacts to data sets of a variety of sizes. Also, since the confusion matrix presented in figure 6 is included, it would have been informative to consider what especially confounding factors appear in the data set, and in general data sets, and what this technique does to avoid them, how it reacts to them, and what effect they have on this techniques ability to make accurate recognitions.