
COMP2550/COMP3130 ANU

Warmup Project Report

Christopher Claoué-Long (u5183532@anu.edu.au)

Jimmy Lin (u5223173@anu.edu.au)

1 Introduction and Overview

The field of computer vision is often thought of as a diverse topic, however in reality there are three major paths to which each project can be linked. The first two – scene categorisation and object detection – aim to provide a summary of a scene using tags or bounding boxes around discrete objects, however do not deliver an accurate object outline nor consider the image as a whole like what happens in human vision.

The task for this warmup project was to further the third path of research: annotating an entire image at the pixel level to describe the entire scene. This would be done by considering a unified cluster of pixels (a superpixel) and extracting a set of useful features to check against using the open source Darwin framework for machine learning and the MSRC dataset. The desired outcome was a feature vector that allowed machine learning algorithms to classify superpixels in images with at least 50% accuracy.

2 Method

The first few features implemented were the average discrete red, green and blue values of the superpixel as well as their standard deviation, since this would begin to differentiate between extremely different superpixels such as those describing sky and grass. Calculating the difference and absolute difference between the red, green and blue average values across the superpixel to further bring out features dependent on the superpixel composition could also help the classifier decide which label to apply as it brings out major differences in colour composition. Knowing the centre of the superpixel in order to work out its average position in the image is taken into account Other features tested include the average gradient in the superpixel, its centre to know the general location within the image (to help differentiate between sky and water

for example), and taking into consideration the differences between the superpixel's immediate neighbours and itself using a hash set for speed to intuitively approximate the outlines of objects to help the classifier decide what superpixels are more likely to be when in close proximity to superpixels of another type.

3 Results

Pictures, maybe accuracy tables with different features implemented?

4 Discussion

While the accuracy is above the desired outcome, the efficiency of the algorithm is quite low when considering neighbours since in memory this is represented as a linear list. A way to speed this up is to use a hash set to index the neighbours in $O(1)$ instead of $O(n)$ time.

Accuracy discussion

Where the algorithm fails - animals and objects with outlines of similar appearance especially!

Performance on training set compared to test set, is this difference expected?

Changes when modifying depth of decision tree

Why is it important to evaluate pixelwise accuracy instead of accuracy on the superpixels? Checking accuracy of a superpixel is difficult, since it may overlap multiple different may overlap onto different label areas, making it impossible to check consistently whether the superpixel was labelled correctly. Pixelwise accuracy has no possible overlap, ensuring that the overall accuracy of the machine learning algorithm is measured correctly.

What do you think is more important, the features or the machine learning classifier? Machine learning classifiers can be built in multiple ways - vector distance tests, binary search trees, hash

sets, cosine similarity, comparing individual features, considering features as a whole, and so on. However, these classifiers are only as good as the features they are given – it is therefore crucial to discover a set of features that allows the classifier to differentiate superpixels accurately.

Possible improvements include considering a neighbourhood wider than a superpixel's immediate neighbours (to help differentiate buildings and sky, faces/flowers and surrounds etc), taking into account the texture

5 References

- S. Gould. DARWIN: A Framework for Machine Learning and Computer Vision Research and Development. In *Journal of Machine Learning Research (JMLR)*, 2012.
- K. Park and S. Gould. On Learning Higher-Order Consistency Potentials for Multi-class Pixel Labeling. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2012.