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 to help differentiate between similar superpixels representing different objects in different locations in the image (for example, sky and water or trees and grass), and the average gradient in the superpixel **CONTINUE HERE.**

Human vision however does not only take into consideration the colour and position of a single object at a time - the surrounding neighbours of an object are also factored in to make an accurate judgement about the object type. To simulate this, the algorithm should compare the differences between the superpixel's immediate neighbours and itself to intuitively approximate the outlines of objects, and help the classifier decide under what label superpixels are more likely to fall when in close proximity to superpixels of another specific type.

3 Results

Description	Testing accuracy	Training accuracy
Fully featured algorithm: colour features, position and neighbours taken into consideration,	52.33%	57.99%
As above, with a greater decision tree threshold cutoff of 2000	52.24%	57.93%
Only colour features taken into consideration	50.42%	53.77%
Only colour features and position taken into consideration	54.13%	54.81%
As above, with a greater decision tree threshold cutoff of 2000	54.74%	54.74%

Pictures inserted here from threshold cutoff of 2000, good and bad classification examples.

4 Discussion

Accuracy discussion - what it's good at and fails at

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 classifers 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: Improving neighbour algorithms, considering a neighbourhood wider than a superpixel's immediate neighbours (to help differentiate buildings and sky, faces/flowers and surrounds etc), taking into account a more accurate representation of texture (for example applying applying a gaussian filter to statistically analyse the texture of the superpixel) and features allowing to differentiate between objects of similar colour and contrast.

5 References

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.

S. Gould. DARWIN: A Framework for Machine Learning and Computer Vision Research and Development. In *Journal of Machine Learning Research (JMLR)*, 2012.