8 Conclusions and further work

This thesis considered two problems: the general problem of classifier learning, and the computer vision problem of object detection. In the field of classifier learning, the meta-classification technique of virtual attribute subsetting was developed and tested. In object detection, the Haar Classifier Cascade algorithm was considered in detail; its behaviour on rotated object detection was studied and it was extended to report numeric confidence instead of binary detections. Finally, methods of combining detections representing separate object segments were created and compared.

The object detection experiments were partially driven by applications in fish and seahorse detection in underwater images, as discussed in sections 4.2 and 4.3. To offer some comparison with existing methods, the common application of face detection was also used, by running an existing face detection cascade on a standard set of frontal face images.

8.1 Rotated object detection

Haar Classifier Cascades are not rotation-invariant, so detection of widely rotated objects requires either image rotation or detection by multiple rotated cascades. Chapter 5 reports tests on the fish and seahorse datasets, varying three parameters: rotation type (images or cascades), rotation steps (15°, 30° or 45°) and training sample random angle ranges (0°..90°). These were evaluated by plotting and comparing ROC curves to find the most accurate detectors.

The results in section 5.4.6 show that image rotation was consistently more accurate than combining rotated cascades; this held for all three detection problems (fish, seahorse heads and seahorse bodies). Meanwhile, the best angle step of those tested was, as predicted in section 5.3 and shown experimentally in section 5.4.5, the smallest: 15°.
8.1.1 Training angle ranges

A single cascade will only detect objects in the orientations presented during training. Randomly varying these angles by small amounts will make a cascade more robust, but large amounts may make a cascade needlessly complex and a poor detector. This was considered in previous research (Jones & Viola, 2003a; Kölsch & Turk, 2004a), but only for angle ranges from $0^\circ$ to the angle step of $15^\circ$. In this thesis, larger angle ranges were also tested – up to $90^\circ$, which results in large overlap between the cascade coverages.

The results for binary detection showed no clear pattern: section 5.4.2 found that the best angle ranges for fish detection equalled or exceeded the angle step, while section 5.4.3 found that in equivalent experiments in seahorse detection the best angle ranges were usually less than the angle step.

The confidence mapping introduced in chapter 6 had its own angle range behaviour, as discussed in section 6.4.6 and shown in appendix B. While the best angle ranges were once again both over and under the angle step, they were usually closer to it than the best binary cascade angle range. ROC curves for confidence mapping with similar but suboptimal angle ranges were also closer to the ‘best’ curve, showing that poorly chosen angle ranges had less impact on confidence mapping than on binary detection.

8.2 Confidence measures

Haar Classifier Cascades usually return a binary output: object or non-object. In chapter 6, they were modified to return a numeric measure of their confidence. Two means of using this measure were developed in section 6.1: hill-climbing the binary detections and confidence mapping. They were then tested on all three image sets.

In section 6.4.5, confidence mapping consistently improved upon the accuracy of binary cascades; hill-climbing sometimes also achieved improvements, though not as significantly. The confidence mapping detections were based on local maxima from the confidence map, with the confidences from nearby non-maxima added; section 6.4.1 shows that this improves its accuracy. On the fish images, it was also best to evaluate each cascade beyond its first stage failure, as found in section 6.4.3.
Confidence mapping remained effective when combining the output from multiple cascades, as shown by the fish and seahorse detection by rotated cascades in figs. 6.12(b), 6.13(b) and 6.13(d). Combining results from multiple rounds of virtual attribute subsetting yielded a small additional increase in accuracy, as seen in section 6.4.4, although the cost in classification time was large.

Those classification times were measured in section 6.4.8, which showed that the confidence mapping time is equivalent to binary detection time unless stage failures are tolerated.

### 8.3 Seahorse segment matching

The seahorse detection cascades trained in chapter 5 were trained to detect seahorse head segments and seahorse body segments. The hill-climbing and confidence mapping in chapter 6 used the same cascades, so also only found individual segments. In chapter 7 the detected segments were matched to create whole seahorse detections, using a formula to give the cost of matching a given head and body. The results in section 7.5 were best when the segments were matched by a greedy matching algorithm, using a cost formula learnt from detections on the training images.

For low false positive counts, the resulting whole seahorse detections were more accurate than either of the component segment detections.

### 8.4 Virtual attribute subsetting

Virtual attribute subsetting was developed in chapter 3 as a generic ensemble classification technique. Section 3.4.1 shows that it produces the same results as attribute subsetting using Naïve Bayes as the base classifier while reducing training time and storage. It also frequently increased the accuracy of J4.8 decision trees and PART rule sets, as shown in sections 3.4.2 and 3.4.3, for no change in the training process. The subset choice in each case was found to be significant. On most datasets, balancing attributes improved accuracy and balancing both attributes and classifiers using the algorithm described in section 3.2.1.4 increased it further. For maximum accuracy across most datasets, the proportion of attributes per subset must be high – around 0.9.
An unusual feature of virtual attribute subsetting is that it is entirely applied after its base classifier is learnt, so can be applied to a classifier that was learnt without virtual attribute subsetting in mind. Most ensemble classification techniques must modify the training data to the base classifiers (e.g. bagging, boosting and attribute subsetting). Stacking does not, but must take the time to run the base classifiers on some training data to evaluate their accuracy. This also means that virtual attribute subsetting with a new base classifier needs no training time beyond that used to train the base classifier. The implementation used in chapter 3 chooses subsets at training time, but even this could be left until classification time if necessary.

Confidence mapping also has no special training requirements, so virtual attribute subsetting was applied to confidence mapping using a face detection cascade trained in 2002 or earlier (Lienhart & Maydt, 2002), before either virtual attribute subsetting (Horton et al., 2006) or confidence mapping (Horton et al., 2007) were even proposed.

### 8.5 Recommendations

The results above lead to the following recommendations for Haar Classifier Cascade training and detection for other types of objects:

1. If detecting rotated objects is necessary, the best random angle ranges to use in cascade training are problem-dependent, but setting the angle range equal to the angle step is a good start. Ideally, multiple cascades or cascade sets should be trained and tested, with ranges both above and below the angle step.

2. Also for rotated objects, the angle step between successive images (or between cascade training angles, for rotated cascades) should be around $15^\circ$. If speed is essential, steps up to $30^\circ$ may be sufficient, especially for if the images are rotated.

3. Confidence mapping instead of binary detection will probably improve accuracy, as long as only local maxima are returned with neighbouring confidences added. It may also reduce classification time unless stage failure tolerance is necessary.

4. If high accuracy is essential and computation time permits, virtual attribute subsetting can help.
Virtual Attribute Subsetting was tested on three base classifiers, only two of which behaved differently compared with standard attribute subsetting. Any classifier which can return a classification for an instance with unknown values can be used as the base classifier, but not all are likely to be effective. Further work could be undertaken to test the effectiveness of virtual attribute subsetting on top of other base classifiers and possibly modifying them to better suit the algorithm.

In object detection, it would be useful to find some aspect of the images or the cascades trained on them which could predict the best angle range to use – however, investigating this would involve experiments with many image sets and cascades.

Also, no satisfactory explanation was found here for the poor performance of rotated cascades relative to rotated images. This is a problem, since rotated cascade detection has the potential to be much faster than rotated image detection, especially if used in a two-stage classification process such as that in (Jones & Viola, 2003a). Further work to investigate this might therefore be useful.

While the confidence measure introduced in chapter 6 was effective on cascades that were trained for binary detection, it could possibly be improved by training with the confidence measure in mind. In particular, it may be viewed as a simple form of margin measurement. Section 2.3.4.1 noted that increasing the margin can improve classifier accuracy even after training data error is zero, and that some boosting algorithms specifically try to maximise the margin. It is therefore possible that altering the cascade training process to maximise the margin would make object detection by confidence measurement more accurate. More complex ways to combine the margins instead of simply summing them may also deserve investigation.