Developing Task-Specific RBF Hand Gesture Recognition

A. Jonathan Howell, Kingsley Sage, and Hilary Buxton

School of Cognitive and Computing Sciences,
University of Sussex, Brighton BN1 9QH

{jonh,khs20,hilaryb}@cogs.susx.ac.uk

Abstract. In this paper we develop hand gesture learning and recognition techniques to be used in advanced vision applications, such as the ActIPret system for understanding the activities of expert operators for education and training. Radial Basis Function (RBF) networks have been developed for reactive vision tasks and work well, exhibiting fast learning and classification. Specific extensions of our existing work to al-

Our work on appearance-based approaches using RBF nets suggests they are very learnable and robust in comparison with structural approaches for general object categorisation on real-world tasks such as face recognition [9, 11]. Natural deformable objects are difficult to specify and so are their movements and actions, so adaptive methods are required. At the heart of a visual learning system is the ability to find the relevant mapping from observable or derivable attributes of image(s) onto the visual categories we require for real-world tasks.

manifold formed by the example views of objects in a space of all possible views of that object [17].

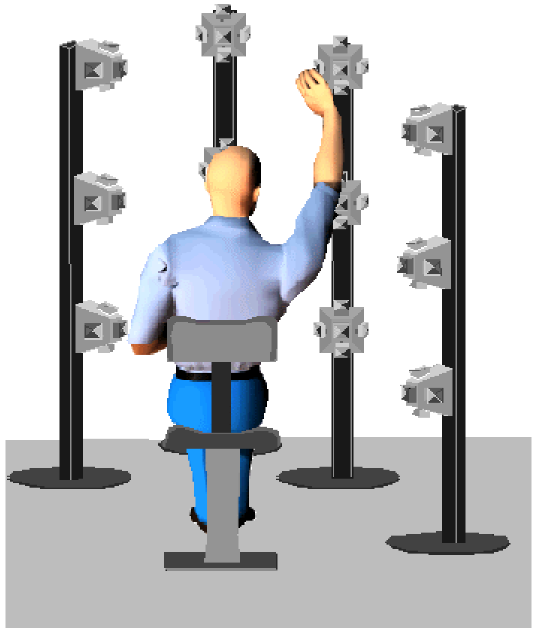
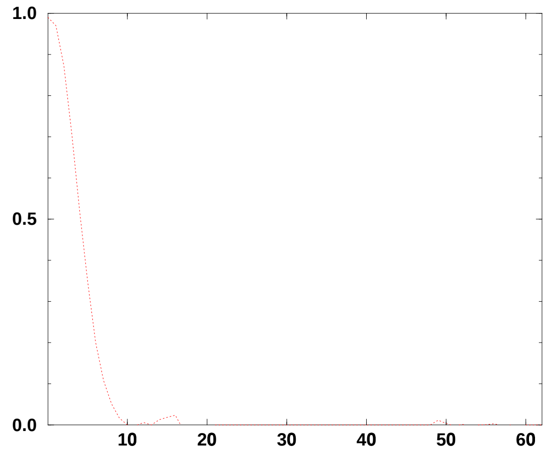


Fig. 1. The target system for the HUMOSIM hand trajectory data.

- Start with a static hand placed at a 'home location' on the subject's leg, followed by:
- A movement toward the target, 2 nt to



1.0
0.5
0.0
10
20
30
40
50
60

Table 1. Generalisation over hand trajectory angle (around y-axis) for TDRBF networks trained with a range of tower data, from Tower 0 (45° left) to Tower 3 (90° right). The '% Correct' values show the proportion of test trajectories where gesture phases were correctly interpreted at every time step of the entire trajectory.

Training Towers		Test Tower, % Correct			
		0	1	2	3
Single	0	100	66	0	0
	1	100	100	0	0
	2	0	0	94	15
	3	0	0	41	94
Consecutive	0 + 1	100	100	0	0

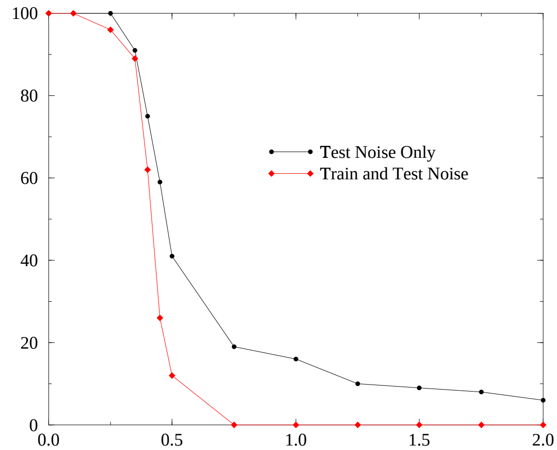


Fig. 3. Classification performance for TDRBF network trained and tested with targets in Tower 0 (45° left), with varying amounts of RMS noise added to the trajectory positions (x-axis, values in cm). The y-axis shows the proportion of test trajectories where gesture phases Tag stt owe A tin B k ae A pe k r e k a

Table 2. Performance for TDRBF networks trained for multiple tasks: 'Gesture' has six gesture phase classes, 'Tower Position' has four position classes (from 45° left to 90° right) and 'Pod Position' has three (from 45° above to 45° below). The '% Correct' values show the proportion of test trajectories where combinations of gesture phase, tower and pod positions were correctly interpreted at every time step of the entire trajectory.

Trained Tasks	Classes	Test Tower, % Correct			
		0	1	2	3
Gesture + Tower Position	24	100	100	82	84
Tower + Pod Position	12	100	100	100	100
Gesture + Tower + Pod	72	100	88	82	89

square (RMS) value, for example, a noise level of 1.0cm RMS produced random values between about ±1.2cm. To produce a smoother variation of values, each vector of random values had individual values averaged with its neighbour.

Fig. 3 shows how classification performance deteriorates as noise increases. Two test arrangements are shown, each with a separate line on the graph. The first trains the TDRBF network without noise, and tests with varying noise. The second both trains and test with an equal level of noise. The TDRBF network performs slightly better when trained without noise, but overall the limit for useful performance would appear to be around 0.5cm RMS noise (on every axis, every time step).

5.3 Multiple Tasks

In this section, we consider how to learn multiple tasks, such as 'which gesture *and* which tower is the hand aiming for?' In previous work, we have shown that separate RBF networks can learn different tasks (face identity, expression, head pose) from the same training data through altering the training signal [5], and that one TDRBF network could learn both gesture and identity by giving different classes to gestures from different individuals [10].

Three tasks can be learnt from the HUMOSIM hand trajectory data:

- *'Which gesture phase?'*, using six gesture phase classes,
- *'Which tower position is the hand aiming for?*, using four position classes (from 45° left to 90° right),
- *'Which pod position is the hand aiming for?*, using three position classes (from 45° above to 45° below).

As an example of combining these tasks, in order to learn both gesture and tower position, we train a network with individual phase classes for each tower. This uses six phases for each of the four towers, 24 classes in all. The results for networks trained on three combinations of these tasks are shown in Table 2, including one trained with all three tasks, which required 72 classes.

Table 2 shows that minimal reduction in performance is observed, compared the the network trained with all towers in Table 1, whilst useful extra information is provided alongside the gesture output.

6 Summary

In this paper we have shown:

- The TDRBF network can learn individual gesture phases from 3-D hand trajectories collected from a magnetic sensor.
- An efficient method for parsing network output and measuring correct classification over an entire hand trajectory file has been developed.
- The 3-D coordinate representation limits trajectory angle generalisation due to values moving from one axis to another as the angle is varied, but this can be overcome by explicit training for several target positions.
- Although the m

Acknowledgements

The authors gratefully acknowledge the invaluable help provided by the Laboratory for Human Motion Simulation (HUMOSIM) at the University of Michigan, USA in allowing us access to their 'Terminal Hand Orientation and Effort Reach Study, 2000' hand trajectory database; also framework concepts and funding from the EU ActIPret project.

References

1. R. Basri. Recognition by prototypes. *International Journal of Computer Vision*, 19:147–168, 1996.

18. D. A. Pomerleau. ALVINN: An autonomous land vehicle in a neural network. In D. S. Touretzky, editor,