

Object Class Recognition at a Glance

John Winn and Antonio Criminisi
Microsoft Research Ltd., Cambridge, UK

<http://research.microsoft.com/vision/cambridge/recognition>



Fig. 1

FRAMES FROM OUR VIDEO SHOWING REAL-TIME MULTI-OBJECT SEGMENTATION AND RECOGNITION.

DESCRIPTION

This video shows our real-time object class recognition system at work.

Object class recognition is a very challenging problem. The difficulty lies in capturing the variability of appearance and shape of different objects belonging to the same class, while avoiding confusing objects from different classes. However, state of the art algorithms such as [2] are capable of delivering high classification accuracy at interactive speed when dealing with a limited number of classes (around ten).

Following the texton-based modeling approach in [2] we have developed an application for real-time segmentation and recognition of objects placed on a table top (figure 1). The system comprises two steps: object segmentation and classification.

First, each object region is separated from the table top. This happens by running a patch-based classifier which discriminates between the class “table” and everything else. This technique is very different from more conventional background subtraction and is robust with respect to shadows, light changes and camera shake or motion.

Second, once all the non-table connected regions have been extracted, they are classified as belonging to one of fifteen object classes using the same discriminative technique.

Each classifier is a random forests discriminative model [1], [3] using pixel difference features. The classifier is learned similarly to [2] to achieve maximum generalisation with high efficiency and is designed to be invariant both to rotation and to small changes in scale.

Our features are computed on both the RGB image (thus providing information about appearance) and on the binary segmentation mask (thus capturing information about object shape). In the learnt random decision trees each node is hence associated to either appearance or shape. Figure 2

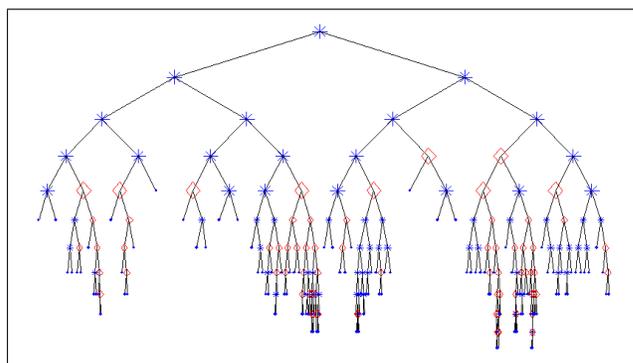


Fig. 2

A TYPICAL DECISION TREE. BLUE STARS INDICATE SHAPE NODES, WHILE RED DIAMONDS INDICATE APPEARANCE NODES.

illustrates such an example. The use of shape features is a key component of the recognition classifier since the shape information provided significantly boosts the accuracy (by more than 10%).

Our algorithm runs on 320×240 images at up to 20 frames per second, with an overall accuracy of around 90%. Training our discriminative class models for 15 classes from 600 training images takes only about ten minutes.

CONCLUSION

This video demonstrates that accurate and efficient object class recognition can be achieved by a combination of simple visual features and efficient class models. Foreground/background segmentation is achieved via the same classification framework. The simultaneous use of appearance and shape information helps improve the classification accuracy. Currently the system has been tested with fewer than twenty object classes. In the future we would like to tackle a much larger number of classes while maintaining unchanged accuracy and efficiency.

ACKNOWLEDGEMENTS

The authors would like to thank J. Shotton and C. Rother for inspiring discussion.

REFERENCES

- [1] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In *Proc. CVPR*, pages 775–781, 2005.
- [2] J. Winn, A. Criminisi, and T. Minka. Object Categorization by Learned Universal Visual Dictionary. *ICCV*, 2005.
- [3] J. Winn and J. Shotton. The Layout Consistent Random Field for recognizing and segmenting partially occluded objects. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2006.