DETECTING STRUCTURED LIGHT PATTERNS IN COLOUR IMAGES USING A SUPPORT VECTOR MACHINE

Tom Botterill[1], Richard Green[1], Steven Mills[2]

[1] Department of Computer Science, University of Canterbury, Christchurch, New Zealand
[2] Department of Computer Science, University of Otago, Dunedin, New Zealand

ABSTRACT

3D reconstruction from multiple cameras is challenging in some environments because of ambiguous matches between similar-looking features. These ambiguities can be resolved by projecting a structured light pattern into the scene, and detecting points in the light pattern in each image. Robust detection of the structured light pattern is hard because of variations in object colour and lighting within the scene, however for specific applications, training data can easily be collected and labelled, enabling the detection problem to be solved using machine learning techniques. We demonstrate the application of a Support Vector Machine (SVM) to detect laser light patterns projected into images of vines, using Feature Subset Selection to design a feature descriptor. A descriptor is computed for every candidate pixel, and the SVM determines if each descriptor is part of the laser line pattern. On test images, the proposed detector achieves 99.4% precision at 90% recall, outperforming a detector which uses only one pixel's colour.

***Index Terms*—** Structured light, Feature detection, Support Vector Machine, Machine learning, Feature Subset Selection

1. INTRODUCTION

Complex scenes can be reconstructed in 3D by capturing multiple images, and using a multi-view 3D reconstruction algorithm, such as bundle adjustment or a dense stereo algorithm. These algorithms struggle to reconstruct some scenes however, often due to ambiguities in matching similar-looking objects between multiple views. To improve the robustness of these reconstruction algorithms, additional 3D information captured with a structured light 3D imaging system can be used. Structured light systems work by projecting a pattern of light into the scene, identifying the light pattern in images of the scene, and then using triangulation to reconstruct the 3D positions of points in the pattern [1].

Structured light has long been used as a stand-alone 3D imaging technique, with either laser-generated patterns, or patterns projected from a digital projector [2]. Systems using laser patterns (such as Microsoft's Kinect [3]) use a camera

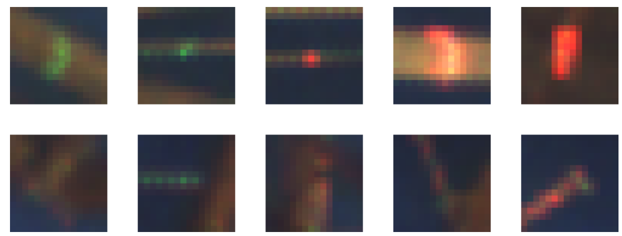

Figure 1: Examples of points on a red and green laser light pattern projected on vines and wires (top), and image points with similar colour because of image noise, variation in the scene, and demosaicing artefacts (bottom).

with a narrow bandpass filter to image only the laser light reflected from the scene. Projector-based systems require the objects to be illuminated only (or predominately) by the projector, and generally make strong assumptions about the colours or reflectance properties of objects [4, 5]. For augmenting a vision-based multi-view 3D reconstruction system, it is desirable to detect the structured light pattern in the same images that are captured for the 3D reconstruction. Detecting the structured light pattern directly from these images avoids the need for additional cameras, and the additional bandwidth, synchronisation and calibration which these would require.

To augment an images based 3D reconstruction system in this way, the structured light pattern should be visible in the images even in the presence of ambient light or machine vision lighting. In many environments, planes of coloured light projected by line lasers are ideal for this purpose.

One application for the combination of vision and structured light is for building 3D models of complex branching plants. Our particular application is a tractor-mounted robot which images vines [6]. Three high resolution cameras image the vines and the wire trellis over which they grow, and a model-based bundle adjustment framework is used to reconstruct their 3D structure. The 3D reconstruction requires vines and wires to be matched between different images, however different vines and wires appear similar, so these matches are often ambiguous. To address these ambiguities, red and green laser lines are projected into the scene. Every pixel where the laser line intersects a vine or wire provides a 3D point, which enables matching ambiguities to be resolved, and also aids the 3D reconstruction [6].

The challenge addressed in this paper is how to detect the structured light points in the colour images. A good detector must be robust to the variation in colours and reflectance properties of objects, and variation in laser intensity and light levels across the scene [7]. This is a hard problem in general, but for specific applications, large amounts of training data can easily be collected. The problem of detecting which pixels show the laser line can then be framed as a machine learning problem. In this paper we demonstrate that this problem can be solved efficiently and effectively by training a Support Vector Machine (SVM) to classify pixels. A set of candidate pixel locations (i.e. red or green pixels) is selected from each image; a descriptor (a set of nearby pixel values) is extracted from the region around each candidate pixel; and an SVM is applied to determine whether each candidate pixel is part of the laser line pattern. On test images, this approach finds 90% of detected laser line points with a false positive rate of 0.5%. The methods described could easily be applied in many other application domains.

2. DETECTING STRUCTURED LIGHT PATTERNS

Most contemporary structured light systems use a dedicated camera and bandpass filter [8, 1, 9] to image only a laser light pattern. Finding the pattern consists of simply finding maxima in the image [10]. More recently, systems have been developed which project complex "coded light" patterns into the scene, and these patterns are detected from colour images. These patterns include patterns of continuously varying hue, and patterns of dots or lines from a discrete set of colours. If the only illumination source is the projector, and the objects which are imaged have uniform colour and approximately Lambertian reflectance, then these patterns can be identified from the pixel's colour [11, 2]. Patterns of dots or stripes can be detected more robustly by using dynamic programming to use knowledge of the pattern to resolve ambiguous pixels [5], however this requires the order of dots or stripes in images to match (locally) the order in which they are projected, which is not true for complex 3D scenes. Systems which capture multiple images with different patterns are more robust, and have higher resolution, but are unsuitable for dynamic scenes [2].

A line laser is used in combination with colour vision for obstacle detection for mobile robots by Ta et al. [12] and by Chang et al. [13]. In both cases the line is detected by applying a threshold to the pixel colours. The laser lines are clearly detected on many indoor objects.

The vine imaging robot previously used a threshold-based approach to estimate laser line points [14], but the method was not sufficiently robust to small changes in vine colour and light levels, and often missed laser line points on smaller or darker vines. The aim of this research is to improve the previous system by designing a detector which uses the region around the laser line, in addition to the laser line colour.

Figure 2: Part of a training image (left) with red (marked in magenta) and green (marked in cyan) laser line points labelled by hand (right). The pattern on the background is irrelevent so is masked out.

3. MACHINE LEARNING FOR STRUCTURED LIGHT PATTERN DETECTION

Our goal is to design a binary classifier for testing whether a pixel in an image of vines is part of a laser line structured light pattern, based on the pixel's colour, and the colours of pixels in its neighbourhood (Figure 1). This is a classic supervised machine learning problem, where large numbers of labelled examples of structured light pixels are given by manually labelling images (Figure 2). We use a Support Vector Machine (SVM), although many alternative machine learning algorithms could be used instead. Across a range of standard datasets, SVMs show classification performance comparable with other leading machine learning algorithms, including Neural Network and Random Forests [15, 16]. SVMs are widely used for object detection and feature classification in computer vision, for example by entrants in the PASCAL Visual Object Class challenge for deciding whether a descriptor of a region of an image shows an object [17], and for classifying land cover in satellite images [18].

3.1 Support Vector Machines

Given a set of feature vectors divided into two classes, an SVM finds a splitting surface which maximises the separation between the classes. The linear SVM finds the planar surface which maximises the separation between the classes. To make the SVM robust to mislabelled training data, and overlap between classes, the splitting surface is chosen so that some fraction of examples may be misclassified. For the ν-SVM formulation we use, the hyperparameter ν controls the fraction of the training data which may be misclassified [19].

More complex splitting surfaces can be found by transforming the feature vectors to a higher dimensional space using a nonlinear kernel function, with a linear SVM

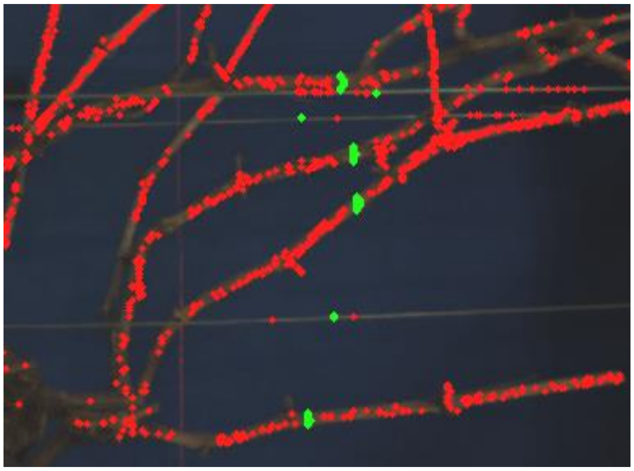

Figure 3: Candidate points for red line laser pattern. Positive examples (as identified in manually labelled images) are marked with green dots; negative examples (all other candidate points) are marked in red.

applied in this higher dimensional space. While many different kernel functions can be used, the Gaussian Radial Basis Function (RBF) kernel often performs well in practice [15, 20], as it can approximate a wide range of smooth splitting surfaces. The curvature (complexity) of the splitting surface is controlled by setting the hyperparameter γ.

To avoid overfitting the training data, cross-validation should be used to set the hyperparameters [20]. In two-fold cross-validation, the SVM is trained on half of the training data, and validated on the other half. This is repeated with the two halves of the training data swapped, and the validation scores are averaged. A brute-force search over the space of hyperparameters is used to find a parameterisation which maximises the cross-validation score, i.e. where the SVM generalises well within the training set. The SVM is then retrained on the entire training set.

3.2. Features for laser line detection

The SVM is used to identify which of a set of candidate pixels from each image are part of the laser line pattern. It is unnecessary and infeasible to apply an SVM to classify every pixel in an image. Instead, a fast check is first applied to identify candidate pixels which could possibly be on the laser line pattern. In our system, the lasers are aligned so that the laser plane is perpendicular to the scanlines. For the red laser, only extrema in the red colour channel along each scanline are considered as candidates pixels. In 12 labelled training images, there are 828 662 extrema in the red channel, of which 5806 are labelled as laser line. By considering only pixels where the red component is 25 greylevels greater than the green component, 94% of the negative examples, but only 0.4% of the positive examples can be excluded. This criterion enables us to select a subset of approximately 4500 candidate pixels per image to evaluate using the SVM, of which about 500 are laser line points. In the labelled training images, candidate pixels which have been labelled as laser line (Figure 2) are used as positive training examples, and other candidate pixels are used as negative training examples (Figure 3).

The feature vectors used are based on the pixel values in a region around each candidate pixel. A patch of the image could be used, but high dimensional patch descriptors are unlikely to give good performance, because the quantity of training data required to train a high-dimensional classifier is too large [21]. We are interested in detecting very small regions of laser light, e.g. one or two pixels high on wires (Figure 1), and intuitively, pixels far from the laser line point should not affect the classification, so we restrict the descriptors to a row of 9 pixels centred on the candidate pixel. As there are significant levels of noise in the image (with standard deviation approximately 5 greylevels), the same row in images convolved (blurred) with Gaussian filters of size 1, 2 and 4 are included.

SVMs perform best when every feature is normalised to a range of about zero to one [20]. It is also desirable that the detector is invariant to changes in illumination (due to position with respect to the lasers and lights, from shadows, and from variations in camera settings and configuration). To address both of these criteria, the descriptors are normalised by dividing by the extreme value in the red channel.

3.3. Feature subset selection

The proposed feature vectors still have considerable redundancy, because the colour values are interpolated from a Bayer image, and pixels in blurred images are a linear combination of other pixels. In addition, they still have high dimensionality (80D for a 9 pixel long row). Feature Subset Selection methods aim to find a subset of input features so that a classifier trained on this subset will generalise better than a classifier trained on all of the features, and will be more computationally efficient.

Sequential Forward Selection [22] is a simple and popular feature subset selection method: the method starts from an empty feature set, then on each iteration adds the feature which increases the cross-validation score the most. Alternative feature selection methods can be applied, e.g. using heuristics such as genetic algorithms, or a brute-force search over all subsets, however these methods are considerably more costly (or computationally infeasible), and can lead to additional overfitting problems [21].

As the laser line detector should be symmetric, features are added in pairs—each pixel location is added along with the same pixel location from the opposite side of the candidate pixel.

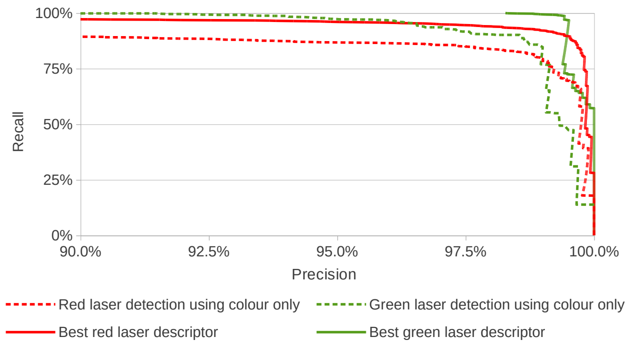

Figure 5: Empirical precision-recall curves from test data (the irregular profiles are caused by clusters of misclassified features). The descriptors found by feature subset selection outperform classifiers based only on one pixel's colour.

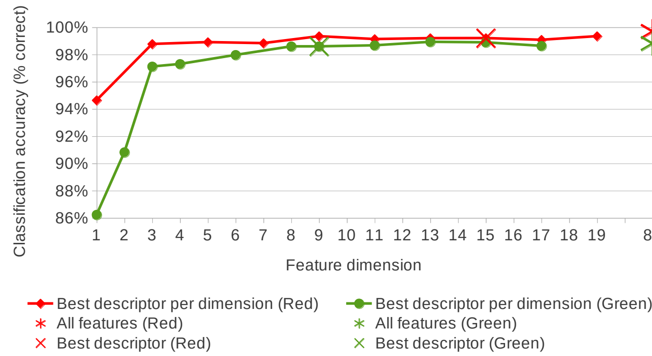

Figure 4: Classification accuracy of descriptors found by feature subset selection, evaluated on test data. Short descriptors perform as well as a descriptor including all 80 features.

4. EXPERIMENTAL RESULTS

Red and green laser light points are hand-labelled in 18 images of vines, to give two training sets of 6 images each (for two-fold cross validation), and one test set. The images include vines, wires, metal and wooden trellising, calibration patterns, and junk such as dead leaves and string. Training images are also reflected horizontally to generate more training samples, and ensure that the detector is symmetric. The training data contains 81 076 red candidate pixels (of which 12% are positive examples) and 29 560 green candidate pixels (21% positive examples).

The hyperparameters ν and γ, and the best feature subset (by Sequential Forward Selection) are found by brute-force search to maximise the two-fold cross-validation score. The classification accuracy on the test set for the best descriptor found for each dimension, the best descriptors overall, and a descriptor using all 80 features are compared in Figure 4. Descriptors using more than 9 features offer no significant improvement in classification accuracy. A 15D descriptor for the red laser was selected by cross-validation score, however a shorter descriptor could be used with a negligible drop in performance. Using all 80 features gives similar test performance to the best feature subsets, indicating that the SVM does not overfit the data even for this long descriptor, but that the long descriptor is not necessary.

The SVM finds a classifier which minimises classification error, but in practice false positive detections are a more serious problem than missing positives, as false positives lead to incorrect 3D measurements, which cause feature registration to fail. The classifier can be tuned to give a higher precision at lower recall by using a different threshold for the SVMs decision function [23]. Precision-recall curves for the best classifiers found are shown in Figure 5. The descriptors chosen by feature subset selection clearly outperform descriptors based on colour alone—the red laser detector can achieve 99.4% precision at 90% recall, whereas the colour-based classifier achieves only 88.0% precision.

Sometimes the nonlinear RBF kernel is unnecessary, and a linear SVM will provide good results. We also trained a linear SVM and descriptor for each colour, however the best descriptors found have comparatively poor performance, with 70% and 130% more misclassified candidate pixels on the test set for the red and green laser patterns respectively.

4.1. Results on a vine imaging robot

We integrated the proposed laser line detector into the vine imaging robot's vision system. The geometry of the laser line can be used to restrict the search space for laser line pixels, and to eliminate some false detections. Laser line detection takes about 40ms per frame on one core of an Intel I7 2.93GHz processor. On the first 10 sets of three stereo frames of a video sequence there are 220 red or green laser line points visible on wires. 193 of these are detected by the new detectors, but only 132 with the old detector based on one pixel's colour alone. Neither detector has any false detections on wires. Matching wires between views is challenging, as they all appear similar [6]. These results indicate that the new laser line detector will improve the robustness of the 3D reconstruction system.

5. CONCLUSIONS

This paper has described how an SVM can be used to detect a laser line structured light pattern in colour images. A short descriptor based on pixel values around candidate pixels is designed by Feature Subset Selection, and is classified using an SVM. In test images, the proposed detector returns 90% of laser line points with 99.4% precision. The detector will be used on a vine imaging robot, where the laser line points will improve the robustness and accuracy of a multi-view 3D reconstruction system. The proposed method could also be applied to other problems where additional 3D information will improve robustness and accuracy.

6. REFERENCES

[1] P. Besl, "Active, optical range imaging sensors," *Machine vision and applications*, vol. 1, no. 2, pp. 127–152, 1988.

[2] J. Geng, "Structured-light 3d surface imaging: a tutorial," *Advances in Optics and Photonics*, vol. 3, no. 2, pp. 128–160, 2011.

[3] Kinect; Wikipedia, January 2013. [Online]. Available: http://en.wikipedia.org/wiki/Kinect

[4] C. Je, S. Lee, and R. Park, "High-contrast color-stripe pattern for rapid structured-light range imaging," in *ECCV*, 2004, pp. 95–107.

[5] L. Zhang, B. Curless, and S. Seitz, "Rapid shape acquisition using color structured light and multi-pass dynamic programming," in *International Symposium on 3D Data Processing Visualization and Transmission*, 2002, pp. 24–36.

[6] T. Botterill, R. Green, and S. Mills, "Reconstructing partially visible models using stereo vision, structured light, and the g2o framework," in *In Proc. Image and Vision Computing New Zealand*, 2012.

[7] T. Botterill, S. Mills, R. Green, and T. Lotz, "Optimising light source positions to minimise illumination variation for 3D vision," in *In Proc. 3DIMPVT*, 2012, pp. 1–8.

[8] Y. Shirai, "Recognition of polyhedrons with a range finder," *Pattern Recognition*, vol. 4, no. 3, pp. 243–244, 1972.

[9] F. DePiero and M. Trivedi, "3-d computer vision using structured light: Design, calibration, and implementation issues," *Advances in Computers*, vol. 43, pp. 243–278, 1996.

[10] D. Naidu and R. Fisher, "A comparative analysis of algorithms for determining the peak position of a stripe to sub-pixel accuracy," in *Proc. British Machine Vision Conf*, 1991, pp. 217–225.

[11] C. Rocchini, P. Cignoni, C. Montani, P. Pingi, and R. Scopigno, "A low cost 3d scanner based on structured light," in *Computer Graphics Forum*, vol. 20, no. 3. Wiley Online Library, 2001, pp. 299–308.

[12] H. Ta, D. Kim, and S. Lee, "A novel laser line detection algorithm for robot application," in *International Conference on Control, Automation and Systems (ICCAS)*, 2011, pp. 361–365.

[13] W. Chang, V. Nguyen, and P. Chu, "Reconstruction of 3D contour with an active laser-vision robotic system," *Asian Journal of Control*, 2012.

[14] T. Botterill, S. Mills, and R. Green, "Design and calibration of a hybrid computer vision and structured light 3D imaging system," in *Proc. International Conference on Automation, Robotics, and Applications (ICARA)*, 2011.

[15] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, no. 1, pp. 169–186, 2003.

[16] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 161–168.

[17] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 3, pp. 3–338, 2010.

[18] C. Huang, L. Davis, and J. Townshend, "An assessment of support vector machines for land cover classification," *International Journal of Remote Sensing*, vol. 23, no. 4, pp. 725–749, 2002.

[19] B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett, "New support vector algorithms," *Neural computation*, vol. 12, no. 5, pp. 1207–1245, 2000.

[20] C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification," 2003.

[21] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153–158, 1997.

[22] A. Whitney, "A direct method of nonparametric measurement selection," *IEEE Transactions on Computers*, vol. 100, no. 9, pp. 1100–1103, 1971.

[23] P. Lingras and C. Butz, "Precision and recall in rough support vector machines," in *International Conference on Granular Computing*, 2007, pp. 654–654.