Probe-based Rapid Hybrid Hyperspectral and Tissue Surface Imaging Aided by Fully Convolutional Networks

Abstract. Tissue surface shape and reflectance spectra provide rich intra-operative information useful in surgical guidance. We propose a hybrid system which displays an endoscopic image together with a fast joint inspection of tissue surface shape using structured light (SL) and hyperspectral imaging (HSI). For SL a miniature fibre probe is used to project a random distribution of coloured spots onto the tissue surface using a dispersed supercontinuum laser. In HSI mode standard endoscopic illumination is used, with the fibre probe collecting reflected light and encoding the spatial information into a linear format that can be imaged onto the slit of a spectrograph. Correspondence between the arrangement of fibres at the distal and proximal ends of the bundle was found using spectral encoding. Then during pattern decoding, a fully convolutional network (FCN) was used for spot detection, followed by a matching propagation algorithm for spot identification. This method enabled fast reconstruction (12 frames per second) using a GPU. The hyperspectral image was combined with the white light image and the reconstructed surface, showing the spectral information of different areas. Validation of this system using phantom and ex vivo experiments has been demonstrated.

Keywords: structured light, hyperspectral imaging, endoscopy, deep learning

1 Introduction

Tissue surface shape recovery is an important tool for both surgical navigation and pathology detection. For example, the morphological appearance could assist in colorectal polyp detection [1], in cases where the lesion is not obvious from standard images. In addition intra-operative tissue surface shape can be combined with pre-operative imaging modalities like CT or MRI, aiding surgical navigation [2]. SL has shown potential in tissue reconstruction in surgical environments, due to its insensitivity to surface texture [2]. A typical SL system consists of a projector and mono or stereo camera for detection. A pattern decoding step to find the correspondences between the camera and/or projector image planes then enables reconstruction of the surface by triangulation for a specific camera-projector position.

Spectral imaging techniques including HSI and multispectral (MSI) implementations, measure the reflectance spectra at particular pixel locations in an image, and have useful clinical applications in discriminating tissues which are indistinguishable
under white light [3]. HSI can measure blood oxygen saturation by exploiting the known absorption properties of haemoglobin and reveal the hallmarks of some tumours [4]. MSI has been used to monitor perfusion and tissue oxygen saturation intraoperatively in the bowel [5], during transplant surgery, and vascular procedures [6]. The high resolution of HSI, compared to multispectral imaging (MSI), can also enable more quantitative analysis allowing extraction of structural information or identification of disease markers in tissue using machine learning algorithms [7]. Previously combination of spectral and 3D information using MSI systems has been attempted [8, 9], but acquisition time has limited the spectral resolution to tens of wavelengths or less. To the best of our knowledge, this is the first time HSI (hundreds of wavelengths) has been combined with 3D reconstruction techniques based on SL and we believe that this solution to the compromise of spectral versus spatial resolution can allow highly accurate real-time distinction between different pathologies.

Computational image analysis involves the development of a robust real-time algorithm to decode the projected pattern for 3D tissue surface reconstruction, based on deep learning and a pattern-specific feature matching method. Recently, advances in hardware have made solving large-scale problems using Convolutional Neural Networks (CNN) possible in a reasonable amount of time. Long et al. proposed to use Fully Convolutional Networks (FCN) for pixelwise semantic segmentation, exceeding the state-of-the-art segmentation [10]. FCN has the benefits of non-dependency on the manually extracted features, combining both global and local information, and fast execution time. In the field of microscopic imaging, FCN has been used to detect and segment cells successfully [11]. So in this work FCN is applied for pattern decoding to detect projected spots. We also show that even with a limited amount of training images (n=17) by data augmentation FCN still performs robustly and accurately.

2 Materials and Methods

2.1 Hybrid Structured Light and Hyperspectral Imager (SLHSI)

The system is built around a custom optical fibre assembly (Fibertech Optica, Inc., Canada) similar to that in our previous work [12, 13]. It is a 2.5 m long incoherent bundle of 171 50 μm core fibres arranged in a linear array at one end and a circular bundle at the other (Fig. 1). A 20 mm working distance GRIN lens (GRINtech GmbH, Germany) is attached distally. The probe’s outer diameter is 2.8 mm.

SL. A 4 W supercontinuum laser, dispersed by a prism, is coupled into the probe’s linear array (Fig. 1). The GRIN lens projects an image of the bundle’s end face which, due to the incoherent fibres, is a mixture of multi-coloured spots. The pattern is reflected by the object, collected by a laparoscope (Karl Storz GmbH, Germany) and imaged onto a CCD (Prosilica GX1050C; Allied Vision Technologies, Inc., USA).

HSI. With the laser switched off a white light source (Xenon 300; Karl Storz GmbH, Germany) is activated. Reflected light is collected by the probe and directed, via a 45° mirror on a motorised flipper mount (MFF101; Thorlabs Ltd., UK), towards a 250/50 mm focal length lens combination to form a demagnified image of the fibre array on the slit of a HSI camera (Nano-Hyperspec; Headwall Photonics, Inc., USA) (Fig. 1).
Correspondences between SL and HSI. The incoherent bundling of fibres in the probe meant that the mapping between an individual fibre’s HSI spectral line and its position in the object plane had to be determined. This was done by projecting the SL pattern onto a white screen and imaging it with a separate hyperspectral camera [5] to determine each spot’s mean wavelength. These could then be linked to the corresponding HSI sensor locations by sorting their wavelengths from shortest to longest.

**Fig. 1.** Left: setup in SL mode, and the physical appearance of the miniaturised probe (lower left). Right: setup in HSI mode and the captured spectrograph image (lower right).

Spectral processing. For each fibre in the HSI image its corresponding pixel columns were averaged to reduce signal noise. The wavelength-pixel row relationship was determined by illuminating the probe with laser light and recording the position of the intensity maximum in the HSI image. A straight-line fit to the pixel and known wavelength data then allowed calibration of each row. Each fibre’s spectrum was divided by the corresponding signal from a white reference target (Spectralon; Labsphere, Inc., USA) to correct for wavelength-dependent transmission characteristics of the system. Further processing to generate absorbance spectra and determine relative haemoglobin concentration in tissue was performed as described previously [5].

### 2.2 3D tissue surface reconstruction

**SL system calibration.** During calibration the projector can be regarded as an inverse camera. The calibration method has been described previously [14]. Besides, a virtual projector image (reference image) is generated from an SL pattern on a white plane using the probe intrinsic parameter and the homography [15]. This image functions like the image from the second camera in passive stereo, facilitating pattern decoding.

**Pattern decoding – spot detection.** In this system projected spots with different colour are considered as features. In surgical environments spot detection is the key challenging step for accurate pattern decoding. An algorithm based on FCN has been employed to detect the spots in endoscopic images robustly in real-time despite confounding factors such as blood, smoke or non-uniform illumination.

Our FCN model (Fig. 2 (a)) consists of two parts: the contractive and the expansive phases. The former halves image size and doubles the feature dimension at each step.
At the end of this phase, each neuron has an effective receptive field of 16×16 pixels, and the feature dimension of its output volume is 512. The expansive phase also contains four steps, where each step upsamples image, fuses the upsampled output and the convoluted output from the counterpart in the contractive phase, and then convolutes to decrease the feature dimension. This expansive phase thus combines the coarse “what” and fine “where” information, resulting in a pixel-wise prediction of the whole image. The output of this model is a 2-channel image with the same size as the input image. Each channel indicates the “probabilities” of being foreground (spots) or background. The loss function is evaluated using the softmax function.

Fig. 2. (a) The deployed FCN model; (b) Cropped SL image; (c) Density map of the cropped SL image; (d) Feature matching between the captured (left) and the reference SL image (right)

For training only 17 captured images have been used and image argumentation including resizing, flipping, and rotation, was applied to increase the training set. To increase the speed the sizes of input images were halved. During training the manual segmentations of the SL images were used as the desired label images. The training included a coarse training and a fine tuning, each with momentum 0.9 and weight decay 0.0005. The former had a learning rate of 0.0000005 until 80000 iterations, while the latter had 0.0000001 until 150000 iterations. In prediction the subtraction between the foreground and background channels was used as the “density map” of spot detection (Fig. 2 (c)). Both the training and prediction were applied using Caffe [16]. In prediction the spot centres were detected at the local maxima locations of the density map.

Pattern decoding- spot matching. Like feature matching in passive stereo, detected spots should be matched to counterparts in the reference image for 3D reconstruction. On the reference image the spots were manually segmented offline.

The spot matching algorithm is based on colour and neighbourhood information detected using Delaunay triangulation. Neighbourhood areas of spots were defined according to distances between neighbours. Then a customised feature descriptor was
defined for each spot. Taking the spot as the area centre, its feature descriptor was a 32x3 matrix, each row of which represented the colour in one direction on the area boundary. Only one feature descriptor was generated per spot in the captured SL image, while 32 descriptors starting from different directions were generated per spot in the reference image, taking rotation into consideration.

Epipolar lines were used to constrain the matching search space. The smallest distance between a spot descriptor on the captured image and all 32 descriptors of a spot on the reference image was used to describe the distance between them. The matching with closest distance smaller than a threshold was chosen. This was followed by a pruning procedure based on neighbourhood information. An iterative method was then applied: matching was propagated to other neighbouring unmatched spots, followed by pruning, until the number of matching stops changing.

**Combination of SL and HSI.** With calibration and spot matching the 3D tissue surface could be triangulated. Hyperspectral data from different fibres, corresponding to individual spots, could be projected onto the reconstructed surfaces, providing a hybrid view of both the spectral and shape information relating to the target tissue. The computation time for reconstruction from single SL image was ~80 ms on a PC (OS: Ubuntu 14.04; processor: i7-3770; graphics card: NVIDIA GTX TITAN X).

## 3 Experiments and Results

Phantom and *ex vivo* tissue experiments have been carried out to validate the SL reconstruction and demonstrate hybrid hyperspectral and surface shape imaging.

**SL Reconstruction.** In SL mode the angle between the probe and the camera is ~10-15°, baseline ~3 cm, working distance 5-9 cm, with the endoscope optical axis roughly perpendicular to the tissue surface. Previous work has shown that the projected spot patterns are robust to changes in background albedo, due to their narrow bandwidth [12]. It was also found that, given an accurate calibration and perfect feature matching, reconstruction error can reach 0.7 mm at a working distance of ~10 cm [14]. Thus, the reconstruction accuracy mainly depends on the pattern decoding results. Here we provide the validation of feature matching from a silicone heart phantom, and *ex vivo* experiments on ovine heart and liver. Each group of validation data contains 10 images chosen randomly from recorded videos. Automatic feature matching results were compared with those from manual annotation. The true positive, annotated matches, together with the matching sensitivity and precision are listed in Table 1.

**Table 1.** Validation of feature matching results. The manual annotation of feature matching is used as the ground truth.

| Object               | Annotated matches | True positive | Sensitivity   | Precision   |
|----------------------|-------------------|---------------|---------------|-------------|
| Silicone heart phantom | 170 ± 1           | 154 ± 17      | 0.907 ± 0.101 | 0.997 ± 0.004 |
| Ovine heart          | 134 ± 8           | 113 ± 10      | 0.844 ± 0.038 | 0.997 ± 0.006 |
| Ovine liver          | 128 ± 11          | 110 ± 13      | 0.861 ± 0.052 | 0.994 ± 0.005 |
Table 1 shows that the pattern decoding algorithm functions robustly with both the phantom and even some challenging ex vivo data, with feature matching sensitivity higher than 0.8, and precision 0.99. Compared with previous work [14], this shows a much higher sensitivity for feature matching due to the high spot detection accuracy that can be achieved with FCN. Meanwhile, the real-time performance of FCN guarantees its practicability in real-world applications. Furthermore, the simple structure of the feature descriptors, and its approach which takes pattern rotation into consideration, not only functions robustly but also simplifies the matching procedure, contributing to the real-time performance.

**HSI validation.** Fig. 3 shows the results of the spectral validation experiment. Spectral data detected from the spot regions indicated by the SL pattern in Fig. 3 (a) are plotted in Fig. 3(b) alongside ‘gold standard’ reflectance spectra measured by a spectrometer (USB4000HR; Ocean Optics, Inc., USA).

![Image](image.png)

**Fig. 3.** (a) Macbeth colour chart illuminated with SL. (b) Sample reflectance spectra from selected colour panels of the Macbeth chart, with R and C indicating the row and column of the chart, respectively. Data from the hybrid system (SLHSI) is plotted alongside the high resolution spectrometer results (GS).

In each Macbeth chart panel the SLHSI system’s mapping procedure returns the correct reflectance spectral data for each location. When all recorded panels are considered there is a mean error of 10% between the SLHSI and gold standard result.

To demonstrate the ability of the system to combine shape and spectral information a cylindrical three-coloured pattern was imaged (Fig. 4). Mean reflectance curves for each region show peaks in the red, green and blue, respectively. Convolution of these spectra with the spectral transmission properties of a colour camera allows RGB information to be recovered, as demonstrated in the surface mesh overlay (Fig. 4 (c)).

Reflectance spectral data from a murine abdomen, imaged post mortem, were used to calculate the absorbance spectrum for each spot location and extract relative concentrations of oxy and deoxyhaemoglobin using the method described previously [5]. Tissue oxygen saturation (StO2) at various spot locations is shown in Fig 4 (d), along with an absorbance spectrum obtained from one of the regions (Fig. 4 (e)). Oxygen saturation throughout the abdomen is low (14±10%), as expected for an ex vivo experiment. Spectra that did not match the model well (r^2<0.8) were rejected.
Fig. 4. (a) Three-coloured cylindrical target and SL illumination. (b) Mean reflectance spectra for each of the target’s regions. (c) 3D surface reconstruction coloured by RGB values generated from the mean reflectance spectra. (d) Murine abdomen with $S\text{tO}_2$ overlay. (e) Sample tissue absorbance spectrum and model fit.

4 Discussion and Conclusion

We have developed a probe-based rapid hybrid system capable of 3D depth sensing as well as hyperspectral imaging. Making use of a miniaturised probe with different fibre arrangements at its two ends, a slit hyperspectral sensor was integrated with an SL system, using a motorised flipper mirror to switch between SL and HSI. The SL mode enabled tissue surface reconstruction of up to 171 data points, using the random-coloured spot patterns; while HSI was able to measure the reflectance spectra of the same regions in one exposure. In addition, a robust, accurate, and fast 3D reconstruction algorithm was proposed, using FCN together with specific feature descriptors, resulting in a real-time measurement of tissue surface even with images of low quality. Phantom and ex vivo experiments have been applied for the validation of the algorithm. Future work will emphasise further hardware development to increase the reliability and practicability to enable acquisition of in vivo data in the clinic.

5 References

1. Schwartz, J.J., Lichtenstein, G.R.: Magnification endoscopy, chromoendoscopy and other novel techniques in evaluation of patients with IBD. Techniques in Gastrointestinal Endoscopy 6, 182-188 (2004)

2. Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., Kolb, A., Rodrigues, M., Sorger, J., Speidel, S., Stoyanov, D.: Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery. Medical Image Analysis 17, 974-996 (2013)
3. Roblyer, D., Richards-Kortum, R., Sokolov, K., El-Naggar, A.K., Williams, M.D., Kurachi, C., Gillenwater, A.M.: Multispectral optical imaging device for in vivo detection of oral neoplasia. BIOMEDO 13, 024019-024019-024011 (2008)
4. Lu, G., Fei, B.: Medical hyperspectral imaging: a review. BIOMEDO 19, 010901-010901 (2014)
5. Clancy, N.T., Arya, S., Stoyanov, D., Singh, M., Hanna, G.B., Elson, D.S.: Intraoperative measurement of bowel oxygen saturation using a multispectral imaging laparoscope. Biomedical Optics Express 6, 4179-4190 (2015)
6. Clancy, N.T., Arya, S., Corbett, R., Singh, M., Stoyanov, D., Crane, J.S., Duncan, N., Ebner, M., Caro, C.G., Hanna, G., Elson, D.S.: Optical Measurement of Anastomotic Oxygenation Dynamics. In: Biomedical Optics 2014, pp. BS3A.23. Optical Society of America. (2014)
7. Akbari, H., Halig, L.V., Schuster, D.M., Osunkoya, A., Master, V., Nieh, P.T., Chen, G.Z., Fei, B.: Hyperspectral imaging and quantitative analysis for prostate cancer detection. BIOMEDO 17, 0760051-07600510 (2012)
8. Clancy, N.T., Lin, J., Arya, S., Hanna, G.B., Elson, D.S.: Dual multispectral and 3D structured light laparoscope. pp. 93160C-93160C-93165. (2015)
9. Clancy, N.T., Stoyanov, D., James, D.R.C., Di Marco, A., Sauvage, V., Clark, J., Yang, G.-Z., Elson, D.S.: Multispectral image alignment using a three channel endoscope in vivo during minimally invasive surgery. Biomedical Optics Express 3, 2567-2578 (2012)
10. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, pp. 3431-3440. (2015)
11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, M.W., Frangi, F.A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III, pp. 234-241. Springer International Publishing, Cham (2015)
12. Clancy, N.T., Stoyanov, D., Maier-Hein, L., Groch, A., Yang, G.-Z., Elson, D.S.: Spectrally encoded fiber-based structured lighting probe for intraoperative 3D imaging. Biomedical Optics Express 2, 3119-3128 (2011)
13. Maier-Hein, L., Groch, A., Bartoli, A., Bodenstedt, S., Boissonnat, G., Chang, P.L., Clancy, N.T., Elson, D.S., Haase, S., Heim, E., Hornegger, J., Jannin, P., Kenngott, H., Kilgus, T., Muller-Stich, B., Oladokun, D., Rohl, S., dos Santos, T.R., Schlemmer, H.P., Seitel, A., Speidel, S., Wagner, M., Stoyanov, D.: Comparative Validation of Single-Shot Optical Techniques for Laparoscopic 3-D Surface Reconstruction. Medical Imaging, IEEE Transactions on 33, 1913-1930 (2014)
14. Lin, J., Clancy, N.T., Elson, D.S.: An endoscopic structured light system using multispectral detection. Int J CARS 10, 1941-1950 (2015)
15. Lin, J., Clancy, N.T., Stoyanov, D., Elson, D.S.: Tissue Surface Reconstruction Aided by Local Normal Information Using a Self-calibrated Endoscopic Structured Light System. In: Navab, N., Hornegger, J., Wells, M.W., Frangi, F.A. (eds.) Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part I, pp. 405-412. Springer International Publishing, Cham (2015)
16. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional Architecture for Fast Feature Embedding. Proceedings of the 22nd ACM international conference on Multimedia, pp. 675-678. ACM, Orlando, Florida, USA (2014)