Elliptical Specularity Detection in Endoscopy with Application to Normal Reconstruction

Karim Makki[1*], Kilian Chandelon[1] and Adrien Bartoli[1,2*]

[1]EnCoV, Institut Pascal, UMR6602 CNRS/UCA, Clermont-Ferrand, France.
[2]Direction de la Recherche Clinique et de l'Innovation, CHU de Clermont-Ferrand, France.

*Corresponding author(s). E-mail(s): {karim.makki, adrien.bartoli}@uca.fr;

Abstract

Purpose. To detect specularities as elliptical blobs in endoscopy. The rationale is that in the endoscopic setting, specularities are generally small and that knowing the ellipse coefficients allows one to reconstruct the surface normal. In contrast, previous works detect specular masks as free-form shapes and consider the specular pixels as nuisance. **Methods.** A pipeline combining deep learning with handcrafted steps for specularity detection. This pipeline is general and accurate in the context of endoscopic applications involving multiple organs and moist tissues. A fully convolutional network produces an initial mask which specifically finds specular pixels, being mainly composed of sparsely distributed blobs. Standard ellipse fitting follows for local segmentation refinement in order to only keep the blobs fulfilling the conditions for successful normal reconstruction. **Results.** Convincing results in detection and reconstruction on synthetic and real images, showing that the elliptical shape prior improves the detection itself in both colonoscopy and kidney laparoscopy. The pipeline achieved a mean Dice of 84% and 87% respectively in test data for these two use cases, and allows one to exploit the specularities as useful information for inferring sparse surface geometry. The reconstructed normals are in good quantitative agreement with external learning-based depth reconstruction methods manifested, as shown by an average angular discrepancy of $12.11° \pm 9.86°$ in colonoscopy. **Conclusion.** First fully automatic method to exploit specularities in endoscopic 3D reconstruction. Because the design of current reconstruction methods can vary considerably for

different applications, our elliptical specularity detection could be of potential interest in clinical practice thanks to its simplicity and generalisability. In particular, the obtained results are promising towards future integration with learning-based depth inference and SfM methods.

Keywords: specularity, ellipse, detection, endoscopy, 3D reconstruction.

1 Introduction

Specularities are generally considered as nuisance [1], even in conditions where they are numerous [2]. For instance, they have been detected and removed before running the Shape-from-Shading method for colonoscopic surface reconstruction in [3]. However, they have recently been shown to bring useful information in endoscopy, with as main assumptions that the light source is to some extent an isotropic point collocated with the camera's centre of projection. These assumptions were also used in Shape-from-Shading [4, 5] and recently led to the following results [6]: *in the endoscopic setting, 1) specularities are near-elliptical blobs and 2) the surface normal can be reconstructed from the ellipse coefficients*. The ability to reconstruct surface normals from a single image is a potential boost for learning-based depth inference [7], which suffers high uncertainty in specularities, and multi-image 3D reconstruction methods such as Structure-from-Motion [8]. The state of the art lacks an effective method to fully exploit the information contained in specularities. The main reason is the lack of a reliable method to detect elliptical specularities. The problem involves four main challenges: C1) the local intensity strongly varies; C2) the overlap of neighbouring blobs; C3) the small size of many blobs, which make ellipse fitting unstable; C4) the unwanted blobs off the tissues, for instance on the interventional instruments. We propose a method which addresses these challenges by leveraging deep learning to reach a global understanding of the image and ellipse fitting to achieve high accuracy. Experimental results show that our method produces stable normal estimates in a fully automatic manner from a single endoscopic image.

2 Methods

2.1 Pipeline

Our method follows the pipeline of figure 1, which combines a neural network and handcrafted steps, taking the best of both. The neural network finds relevant blobs but has limited precision whilst the handcrafted steps, exploiting the elliptical prior, bring high precision.

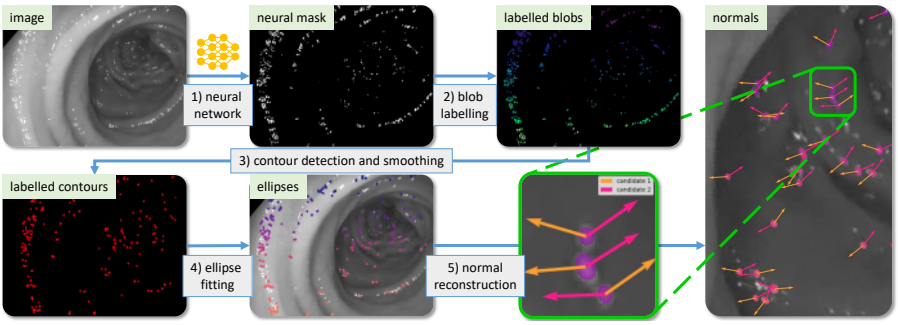

Fig. 1: Proposed pipeline for specularity detection and normal reconstruction, combining deep learning with handcrafted steps exploiting the elliptical prior.

2.2 Neural Network

We achieve *Step 1)* with a fully convolutional network to produce an initial *neural mask* M which specifically finds specular pixels on the colon tissue, dealing with challenges C1) and C4).

Architecture

Our network architecture is shown in figure 2. It has 18 convolution layers with 64 filters, a kernel size of $(3, 3)$, strides of $(1, 1)$ and an orthogonal kernel initialiser. It uses ReLU activations and batch normalisation with a momentum of 0.1 and a minimal value of 0.0001, and sigmoid activation in the last layer. Overall, the network has 594 240 trainable and 2 048 non-trainable parameters.

Training

We created a dataset of 2 240 images extracted from 13 colonoscopy procedures from the Endomapper dataset, which we manually annotated using adaptive thresholding. We split the dataset in 2 000 training, 200 validation and 40 test images. We used data augmentation with standard image transformations, namely horizontal and vertical flipping and realistic brightness adjustment within range $[-0.3, 0.3]$. We trained with Adam optimiser for 20 epochs with a batch size of 32, a learning rate of 0.001 and a binary cross-entropy loss function. We obtained a segmentation Dice score of 80.05% with standard deviation 23.06% on the test set.

2.3 Handcrafted Steps

The three handcrafted Steps 2) to 5) start from the neural mask M and reconstruct the normals. *Step 2) blob labelling.* The blobs are isolated, dealing with challenge C2), by giving a unique id to the connected components in M, with background pixels labelled as 0. We only keep those components whose area lies between 7×10^{-4} and 3×10^{-3}% of the image area. For our 1248×1080 images, the range is $[10, 40]$ px^2. This was found empirically to discard overly

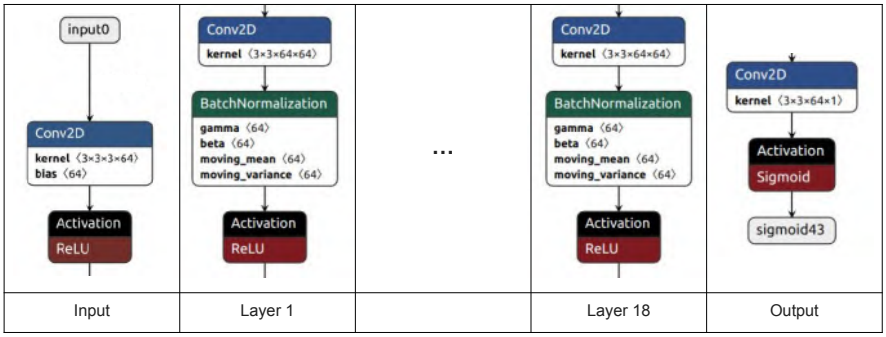

Fig. 2: Architecture of the proposed fully convolutional neural network used in Step 1 of the pipeline.

small blobs (unstable ellipses) and large blobs (breaking the local planarity assumption or consisting of overlapping specularities). *Step 3) contour detection and smoothing.* The isophote contours are extracted and smoothed. We use marching squares to detect the zero level set for the labelled blobs, leading to a labelled set of closed contours. We fit a smoothing cubic B-spline, which reduces the effect of noise, and resample the countours with 1000 points. *Step 4) ellipse fitting.* We use [9], which is a stable version of the direct least squares fitting method [10], dealing with challenge C3). *Step 5) normal reconstruction.* We use the pose from circle method [11], which takes the ellipse coefficients and the camera's intrinsic parameters as inputs.

3 Experimental Results

3.1 Simulated Data

We modified the synthetic image renderer of [6] to use a sphere instead of a plane, hence creating curvature. We use the same parameters but increase the roughness parameter to $n = 120$ for a better trade-off between specularity size and surface curvature. The average angular error on the estimated normals over 300 trials is shown in figure 3. First, we observe that the normal error remains in a reasonable range lower than 14 degrees. Second, we observe that the larger the sphere radius R, the smaller the normal error. This was expected as a larger radius flattens the surface, hence conforms it to the planar hypothesis made in normal reconstruction. Note that we took care of keeping the distance from the camera to the brightest point of the specularity unchanged. Third, we observe that the large the camera tilt θ with respect to the sphere's normal at the brightest point of the specularity, the larger the error. This was expected because a larger tilt creates more perspective, carrying the effects of curvature more strongly.

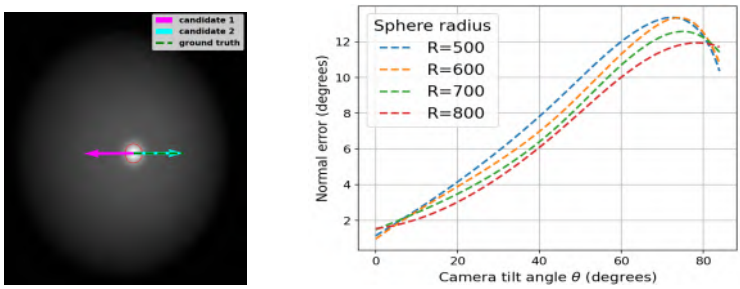

Fig. 3: Synthetic data. (left) Example image, showing a detected specular isophote (in red) and the two normal candidates estimated using the pose from circle method (in magenta and cyan). (right) Normal error vs camera tilt angle.

3.2 Real Data

Evaluation of specular blob segmentation

We selected 5 representative colonoscopic images from the Endomapper dataset and 6 laparoscopic kidney images from a partner hospital via the UroCCR #122 ancillary study, with various viewing conditions and in the presence of interventional tools, as shown in figure 4. First, we evaluated the accuracy of our neural network model relative to the accuracy of a classical specularity detector from endoscopic video frames [12] using several standard evaluation metrics. As illustrated in the first part of table 1, the segmentation accuracy is significantly improved for the two observed organs. For instance, the mean Dice is improved from 41% to 60%, and from 10% to 74% for the colon and kidney respectively. Note that at this level, specular masks are still considered as free-form shapes. In the following, we consider specular masks as sparsely distributed elliptical blobs and our evaluation is be based on this condition. For the colon and the kidney, an average of 86% and 78% of relevant elliptical blobs are respectively detected by the neural network. From these blobs, 68% and 72% respectively meet the normal reconstruction condition, which is that they have an elliptical shape, as tested by ellipse fitting. This eventually gives a total amount of about 250 and 1620 pairs of blobs and normals per image.

We perform an ablation study with highly controlled ground truth data, as follows. We manually segmented the images with ITK-SNAP to only consider the relevant elliptical blobs, creating 1252 specular blobs for the colon and 9724 blobs for the kidney, which we ensured are isolated, on the observed organ surface, and in the range of size leading to stable normal reconstruction. An example of a relevant, though incomplete for the sake of visualisation, set of exploitable elliptical blobs is shown in figure 5.a. We used this carefully labelled dataset to evaluate performance and perform an ablation study. First, we ran our pipeline without the neural network, which we replaced by mere intensity thresholding, obtaining a mean Dice of 43% and 59% for the colon

and the kidney respectively. Second, we ran our pipeline without the handcrafted steps, directly using the neural mask as final result, obtaining a mean Dice of 68% and 81% for the colon and the kidney respectively. Finally, we ran our full pipeline, obtaining a mean Dice of 84% and 87% for the colon and kidney respectively. Part 2 of table 1 reports segmentation accuracy using several evaluation metrics, showing for instance an increase in sensitivity from 51% (for the neural mask) to 71% (for the combined method) while the specificity remained almost unchanged for the colon. An illustration for a set of specularities in colonoscopy is provided in figure 5. This shows that 1) both the learning-based and the handcrafted steps contribute significant to the complete system and 2) the complete pipeline segmentation has a high coverage of the exploitable specular blob and is accurate.

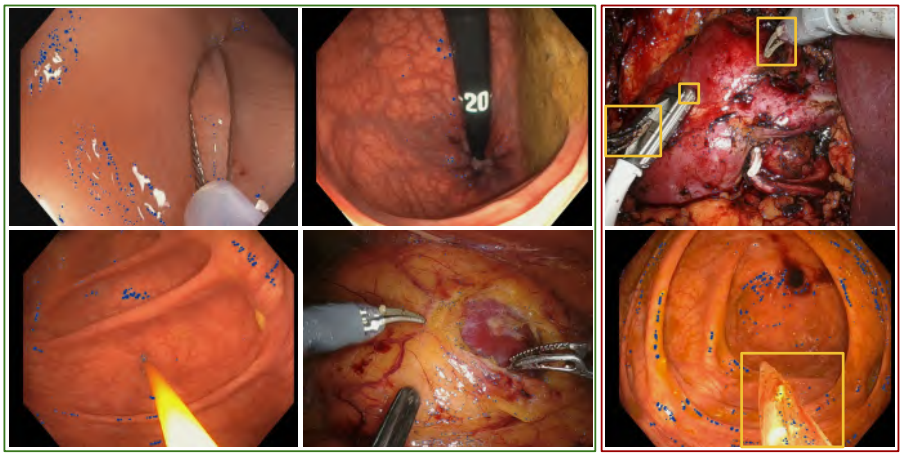

Fig. 4: Neural network capacity in dealing with challenge C4). The first two columns show successful detection of specularity on the colon and kidney surfaces. The third column shows two failure cases for the kidney and colon, with specular blobs detected on the instruments. In our evaluation dataset, failure cases are rare but hardly avoidable due to the wide variety of surgical scenes.

Evaluation of normal reconstruction

We compared the normals reconstructed by our method to the external depth reconstruction method [13]. This allowed us to perform a test on real colonoscopic images in challenging conditions, where ground truth is not available. Concretely, we used the 7 images and depthmaps showcased in [13], from which we extracted normal maps. Our method reconstructed 67 normals and we obtained an average angular discrepancy of 12.11° with standard deviation 9.86°. This discrepancy indicates a satisfying agreement between the two methods. The low image resolution of 256×256 required by [13] explains the

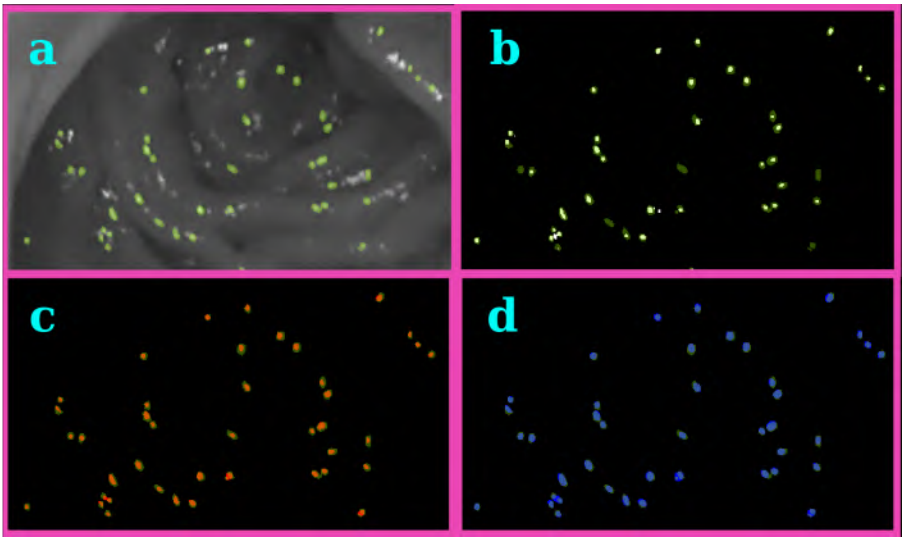

Fig. 5: Segmentation evaluation by ablation: a. shows a close-up on a surface patch example with a set of manually segmented elliptical blobs forming the ground truth in green. b. shows the specular mask obtained using mere thresholding in white vs the ground truth mask, with a Dice of 0.55 for the selected surface patch. c. shows the neural mask in red vs the ground truth mask, with a Dice of 0.61. d. shows the final mask obtained using both the neural mask and ellipse fitting for local refinement in blue vs the ground truth, with a Dice of 0.83.

lower number of normals detected by our method and possibly the angular discrepancy.

4 Conclusion

We have proposed a method that combines learning-based and handcrafted steps to achieve elliptical specularity detection, from which robust and accurate normal reconstruction follows. We plan to improve the detection capacity and accuracy of our neural network in order to extract more small size exploitable ellipses even in very challenging conditions. We also plan to combine our method with Structure-from-Motion and run advanced tests in colonoscopic navigation.

Acknowledgements. This work has been supported by the Endomapper H2020 project.

Compliance with Ethical Standards. The authors declare that they have no conflict of interest. This is a retrospective analysis, no interventional procedures were performed, and the data was already collected. All patients gave written informed consent in accordance with the UroCCR project (French

Table 1: Top: Evaluation of the proposed network and comparison to [12] in terms of segmentation performance. Bottom: evaluation of the proposed pipeline with ablation of the neural and handcrafted steps in terms of detection performance. We use the following metrics: $TPR = TP/(TP + FN)$, also known as sensitivity, $TNR = TN/(TN + FP)$, also known as specificity, $PPV = TP/(TP + FP)$ also known as precision and $ACC = (TP + TN)/(TP + TN + FP + FN)$. In segmentation, the measured quantity is the number of pixels; in detection, it is the number of blobs. All statistics are given as averages over the test datasets with standard deviation.

Part 1: segmentation accuracy

Metric	DICE (%)	FP	FN	TP	TN	TPR (%)	TNR (%)	PPV (%)	ACC (%)
Kidney									
Method [12]	10.63 ± 3.99	161905 ± 39919	1187 ± 659	9282 ± 3760	1138346 ± 39691	89.14 ± 3.35	87.55 ± 3.06	5.71 ± 2.23	87.56 ± 3.04
Neural mask	74.71 ± 7.61	1430 ± 999	3161 ± 1155	7309 ± 3246	1298821 ± 4230	69.20 ± 3.32	99.89 ± 0.08	82.30 ± 13.84	99.65 ± 0.11
Colon									
Method [12]	41.20 ± 36.24	70149 ± 121367	5359 ± 2010	8509 ± 6928	1346768 ± 186063	52.37 ± 25.02	94.77 ± 9.07	41.20 ± 40.41	94.42 ± 9.02
Neural mask	23.73 ± 6.76	2145 ± 1737	11763 ± 6319	2105 ± 778	1414772 ± 97368	17.57 ± 7.48	99.84 ± 0.13	61.25 ± 31.08	99.04 ± 0.26

Part 2: elliptical detection accuracy

Metric	DICE (%)	FP	FN	TP	TN	TPR (%)	TNR (%)	PPV (%)	ACC (%)
Kidney									
Handcrafted	59.40 ± 4.12	11211 ± 3642	1187 ± 659	9282 ± 3761	1289040 ± 7712	89.13 ± 3.35	99.13 ± 0.28	44.79 ± 4.83	99.05 ± 0.32
Neural mask	81.66 ± 2.34	16 ± 9	3161 ± 1155	7309 ± 3246	1300235 ± 4330	69.20 ± 3.31	100 ± 0.00	99.71 ± 0.25	99.76 ± 0.09
Full pipeline	87.34 ± 2.34	22 ± 4	2960 ± 1185	7450 ± 2256	1281574 ± 4330	71.31 ± 3.31	100 ± 0.00	99.72 ± 0.22	99.74 ± 0.07
Colon									
Handcrafted	42.73 ± 6.10	260 ± 71	6918 ± 421	2717 ± 541	1337945 ± 191	28.13 ± 5.26	99.98 ± 0.00	91.39 ± 0.62	99.47 ± 0.03
Neural mask	68.07 ± 0.75	19 ± 24	4403 ± 238	4705 ± 322	1338713 ± 728	51.73 ± 0.98	100 ± 0.00	99.56 ± 0.56	99.67 ± 0.03
Full pipeline	84.19 ± 4.32	581 ± 14	2273 ± 34	5373 ± 34	13450900 ± 890	71.22 ± 5.45	99.3 ± 0.68	90.16 ± 3.26	99.8 ± 0.01

network of research on kidney cancer, NCT03293563). This article does not contain any studies with animals performed by any of the authors.

References

[1] Morgand, A., Tamaazousti, M., Bartoli, A.: A geometric model for specularity prediction on planar surfaces with multiple light sources. IEEE Transactions on Visualization and Computer Graphics **24**(5), 1691–1704 (2017)

[2] Daher, R., Vasconcelos, F., Stoyanov, D.: A temporal learning approach to inpainting endoscopic specularities and its effect on image correspondence. arXiv preprint arXiv:2203.17013 (2022)

[3] Ciuti, G., Visentini-Scarzanella, M., Dore, A., Menciassi, A., Dario, P., Yang, G.-Z.: Intra-operative monocular 3d reconstruction for image-guided navigation in active locomotion capsule endoscopy. In: 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), pp. 768–774 (2012). IEEE

[4] Okatani, T., Deguchi, K.: Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. Computer vision and image understanding **66**(2), 119–131 (1997)

[5] Yeung, S., Tsui, H.-T., Yim, A.: Global shape from shading for an endoscope image. In: MICCAI, pp. 318–327 (1999). Springer

[6] Makki, K., Bartoli, A.: Normal reconstruction from specularity in the endoscopic setting. arXiv preprint arXiv:2211.05642 (2022)

[7] Rodriguez-Puigvert, J., Recasens, D., Civera, J., Martinez-Cantin, R.: On the uncertain single-view depths in colonoscopies. In: MICCAI, pp. 130–140 (2022)

[8] Sengupta, A., Bartoli, A.: Colonoscopic 3D reconstruction by tubular non-rigid structure-from-motion. International Journal of Computer Assisted Radiology and Surgery **16**(7), 1237–1241 (2021)

[9] Halır, R., Flusser, J.: Numerically stable direct least squares fitting of ellipses. In: Proc. 6th Int. Conf in Central Europe on Computer Graphics and Visualization. WSCG, vol. 98, pp. 125–132 (1998). Citeseer

[10] Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct least square fitting of ellipses. IEEE Transactions on pattern analysis and machine intelligence **21**(5), 476–480 (1999)

[11] Forsyth, D., Mundy, J.L., Zisserman, A., Coelho, C., Heller, A., Rothwell, C.: Invariant descriptors for 3D object recognition and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence **13**(10), 971–991 (1991)

[12] Tchoulack, S., Langlois, J.P., Cheriet, F.: A video stream processor for real-time detection and correction of specular reflections in endoscopic images. In: 2008 Joint 6th International IEEE Northeast Workshop on Circuits and Systems and TAISA Conference, pp. 49–52 (2008). IEEE

[13] Rau, A., Edwards, P.E., Ahmad, O.F., Riordan, P., Janatka, M., Lovat, L.B., Stoyanov, D.: Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy. International journal of computer assisted radiology and surgery **14**, 1167–1176 (2019)