Fang Chen*, Jan Müller, Jens Müller, and Ronald Tetzlaff

Efficient feature-based motion estimation in neurosurgery using non-maximum suppression

https://doi.org/10.1515/cdbme-2018-0133

Abstract: In this contribution we propose a feature-based method for motion estimation and correction in intraoperative thermal imaging during brain surgery. The motion is estimated from co-registered white-light images in order to perform a robust motion correction on the thermographic data. To ensure real-time performance of an intraoperative application, we optimise the processing time which essentially depends on the number of key points found by our algorithm. For this purpose we evaluate the effect of applying an non-maximum suppression (NMS) to improve the feature detection efficiency. Furthermore we propose an adaptive method to determine the size of the suppression area, resulting in a trade-off between accuracy and processing time.

Keywords: brain motion, Harris corner detection, non-maximum suppression, Normalised Cross Correlation.

1 Introduction

Time-resolved thermal imaging is a novel technique to monitor cortical temperature variations during neurosurgery. The goal of our work is to apply thermography to identify normal and anomalous brain tissue intraoperatively. Previously, Steiner et. al. [1] demonstrated an approach to cortical perfusion imaging by analysing local temperature changes caused by an ice-cold saline solution injected via a central vein. Hollmach et. al. [2] investigated thermography with multivariate statistical data analysis to detect and demarcate tumour borders. Hoffmann et. al. [3] proposed a new classifier to locate the cerebral cortex in thermal images.

However, brain motion occurring during surgery can reduce the accuracy of the automated identification and has to be firstly removed. Furthermore, motion detection directly from thermographic sequences may cause errors because temperature shifts over time may not be related to brain motion, but to other brain activities we want to analyse. In a surgery system, a co-registered white-light camera is utilised for observation. In our approach we estimate the brain motion from white-light images, since in this modality the brain motion is more unambiguous compared to thermography.

For the intraoperative application, a real-time motion correction is required. Since the processing time of feature-based methods mostly depends on the number of tracked features, we investigate the effect of non-maximum suppression (NMS) methods to reduce the number of key points and hence to improve the efficiency of feature detection.

2 Methods

Feature-based motion estimation establishes correspondences between image features such as points, contours, and blobs. These correspondences can be constructed to a dense displacement field applying interpolation. Generally, we divide the process into five steps: pre-processing, feature detection, feature description, feature matching, and interpolation.

I) The task of *pre-processing* is to reduce noise and to enhance the local contrast, such that the features can be better distinguished. For noise reduction, we employ the non-linear Perona and Malik diffusion filter [4]. Furthermore, we perform the so-called Contrast Limited Adaptive Histogram Equalisation (CLAHE) [5] to improve the local contrast.

II) We apply the Harris corner detector [6] to *identify features* in the image. The brain cortex is segmented manually as a Region of Interest (ROI) for the *feature detector*.

III) The *feature descriptor* characterises local features. Inspired by [7], a small patch (31×31) of neighbouring pixels is employed as the *feature descriptor*.

IV) The *matching feature* pairs are those that show very similar patches. In the matching stage, the small patches are compared by applying the similarity measure NCC:

$$NCC = \frac{\sum_{i=1}^{n}(I_i^{\mathrm{r}} - \overline{I^{\mathrm{r}}})(I_i^{\mathrm{t}} - \overline{I^{\mathrm{t}}})}{\sqrt{\sum_{i=1}^{n}(I_i^{\mathrm{r}} - \overline{I^{\mathrm{r}}})^2 \sum_{i=1}^{n}(I_i^{\mathrm{t}} - \overline{I^{\mathrm{t}}})^2}}, \qquad (1)$$

where I^{r} and I^{t} denote intensities of pixels located at the same position in the reference and target images/patches. $\overline{I}$ is the average of all pixel intensities in an image/patch; both images/patches have the same number of pixels n. The value of

***Corresponding author: Fang Chen,** Institute of Circuits and Systems, Faculty of Electrical and Computer Engineering, TU Dresden, 01062 Dresden, Germany, e-mail: fang.chen@tu-dresden.de

NCC varies from '0' for totally different, to '1' for identical images/patches.

V) A 2D cubic B-spline *interpolation* is applied to construct the displacement vector field. Before the interpolation, a mean value filter is employed to remove outliers and to smooth the displacement field.

The Harris detector in II) calculates a corner response which expresses intensity changes within a neighbourhood. A point is assumed to be a corner when its corner response is larger than a threshold. In our work, the image intensities for the calculation of corner responses are 8-bit grey values.

NMS is the process of searching for the local maximum within a certain neighbourhood and suppressing the rest. It is usually applied as an edge thinning technique. Here the method is applied to suppress points that do not exhibit the *locally maximal* corner response. We implemented an NMS using a dilation operator: We calculate the corner response for every pixel, compose a new image whose pixel values are the corner responses, and then dilate this image with a rectangular structuring element. By comparing the dilated image with the corner response image, all locally maximal responses are retained. The size of the structuring element of dilation therefore must be equal to the size of the suppression area.

Essentially, an application of an NMS can improve the efficiency of feature matching because it suppresses redundant points and selects unique points representing corners. On these grounds, we test the effect of NMS as well as the influence of the neighbourhood size.

3 Experiments and results

3.1 Dataset

The visual-light images were recorded using a white-light camera (Basler acA1920-155us) with a resolution of 1920×1200 px in RGB colour channels. During an intraoperative application, the camera is operated with a frame rate of 7.5 fps. Here we analyse image sequences from five clinical cases (S1 - S5). The patient of S1 and the patient of S2 have metastasis, the patient of S3 has glioblastoma, the patient of S4 has reactive inflammatory changes, and the patient of S5 has anaplastic astrocytoma. Each image sequence has a duration of 10 seconds; no surgical procedure was recorded.

3.2 Performance metrics

We employ the similarity measure NCC according eq 1, which is usually used to evaluate the alignment of images, as a performance metric . The first frame of every image sequence is selected to be the reference image and the others are target images. The target images are aligned to the reference image with the obtained displacement fields. If a displacement field represents exactly the brain motion, the aligned image will have the maximal similarity to the reference image with NCC≈1.

For an additional comparison we measured the processing time. The programs are implemented in Python, the Harris corner detector is implemented using the OpenCV library. All computations, the metric included, are performed on the HPC system at the Center for Information Services and High Performance Computing at TU Dresden.

To assess the NMS method, we selected five sizes of suppression neighbourhood: 10×10, 15×15, 20×20, 25×25, and 30×30 px. For comparison we applied the method without NMS, which determines the features by thresholding the Harris corner response. We calculated the Harris corner responses of all pixels in the ROI of the reference images (see Figure 1). The features are pixels with large Harris corner responses, and we assume that they amount to less than 10% of pixels. Therefore we selected three thresholds: 10^4, $5 \cdot 10^4$ and 10^5. In the NMS method we also eliminated negative Harris corner responses that indicate other features like edges and areas.

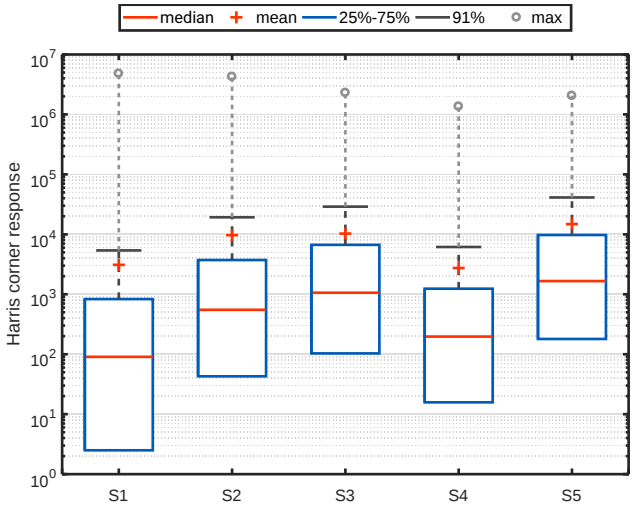

Fig. 1: Harris corner responses of pixels in the ROI of the reference images in S1-S5.

3.3 Results

In Figure 2, the resulting NCC and processing time are plotted against the number of features of every frame in the five sequences. The ranges of NCC values are almost the same with or without NMS. Nevertheless, some frames have a clearly

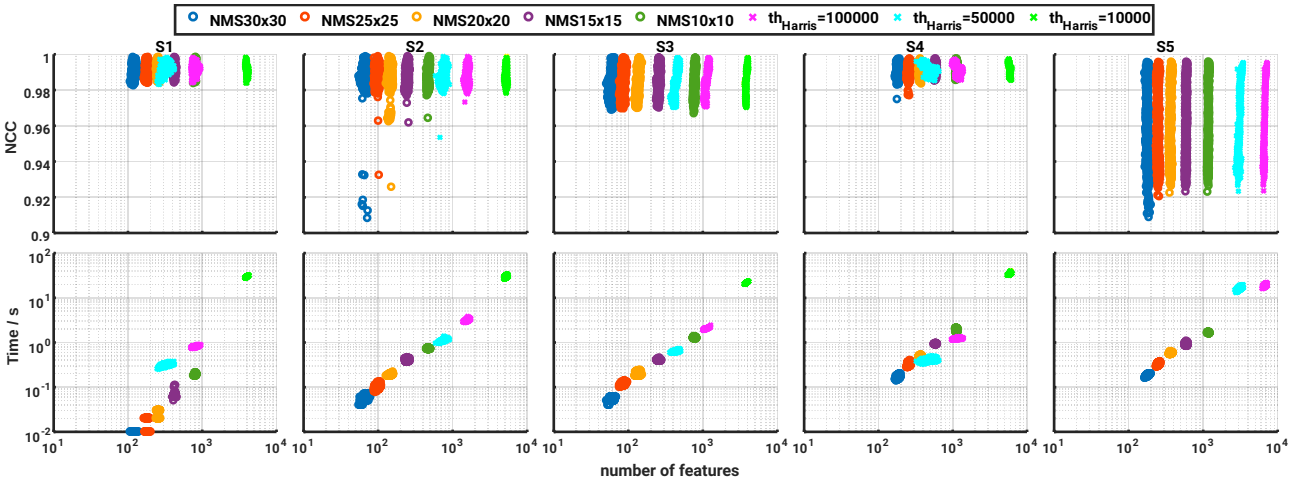

Fig. 2: NCC and processing time against the number of features of every frame without (magenta/cyan/green) and with (other colours) NMS in five image sequences S1-S5

smaller NCC after applying NMS when the size of the suppression area enlarges. The reason is that some of the relevant feature points are suppressed. Furthermore, if there is a false match, the error will become very large because of the large suppression area (see Figure 3a).

Although the NCC values of the tests without NMS are stable and slightly higher than with NMS, the computation takes much longer. As shown in Figure 2, the processing time is roughly proportional to the number of features. Furthermore, many features gather in the areas with high contrast (see Figure 3b). These features do not contribute to a better motion estimation because they are located in the same area and provide the same information about brain motion.

Repeatability is another problem of the simple thresholding technique. As shown in Figure 1, the Harris corner responses of five image sequences are not generally similar. In S5, the vessels are much denser than in S1-S4. When we use the same threshold of 10^4, significantly more features are detected in S5 ($> 2 \cdot 10^4$) (see Figure 4). Table 1 shows the mean number of features in five sequences with the three thresholds.

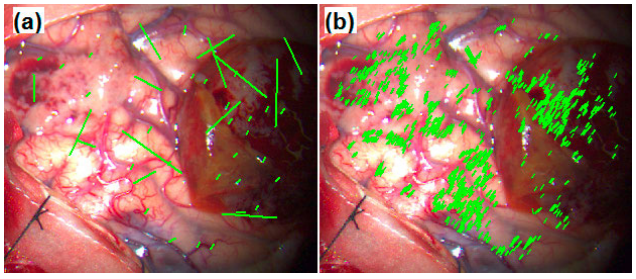

Fig. 3: Putative matches (green lines) in S3-frame 512 using (a) NMS30×30, (b) without NMS, threshold 10^4.

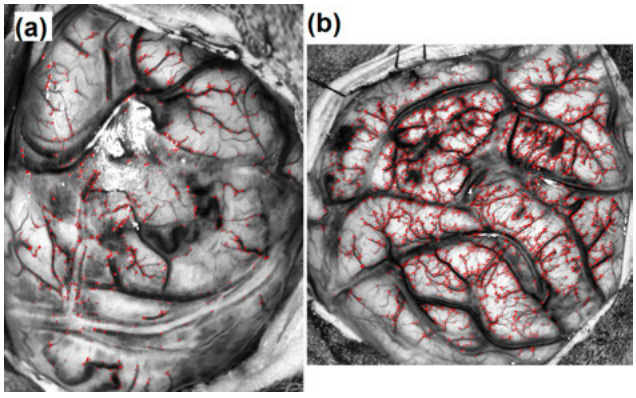

Fig. 4: Features (red points) in (a) S4-frame 1 and (b) S5-frame 1, without NMS and with threshold 10^4.

Tab. 1: Number (mean) of features detected with thresholds 10^4, $5 \cdot 10^4$ and 10^5 in five image sequences S1-S5.

thresh.	S1	S2	S3	S4	S5
$\mathbf{10^4}$	3,973	5,193	3,817	5,864	20,378
$\mathbf{5 \cdot 10^4}$	805	1,558	1,106	1,172	6,666
$\mathbf{10^5}$	326	745	430	490	3,071

Furthermore, the computation of a single frame in S5 by using a threshold of 10^4 takes longer than 40 minutes in contrast to maximally 40 seconds in the other sequences. Because of this large time consumption we omitted the test with a threshold of 10^4 throughout the whole sequence S5.

Since our project focuses on an intraoperative application, it is inconvenient to adjust the threshold during surgery. Even though we do not have to adjust the threshold in the NMS method, the size of the suppression area must be determined. Based on Figure 2, the number of features of around 10^3 is a

good trade-off between accuracy and processing time. We propose an adaptive method to calculate an appropriate size $z \times z$ of the suppression area with the help of the ROI size:

$$z = \lfloor \sqrt{A_{\mathrm{ROI}}/4000} \rfloor , \quad (2)$$

A_{ROI} being the number of pixels in the ROI. Table 2 shows the number of features and the processing time with the suppression area calculated by eq 2. The numbers of features in S4 and S5 are larger than in S1-S3. This difference is reasonable because of the floor function in eq 2. Additionally, areas with specular reflections are masked, such that features in these areas are eliminated.

Tab. 2: Number (mean) of features and processing time (mean) by applying an NMS. The sizes of the suppression areas $z \times z$ are calculated by eq 2.

	S1	S2	S3	S4	S5
$\mathbf{z \times z}$	8×8	6×6	5×5	9×9	9×9
# of features	1,012	1,069	1,076	1,305	1,359
Time/s	0.4664	0.5534	0.5909	0.5086	0.4220

4 Conclusion

We evaluated the performance of feature-based motion estimation methods with and without NMS by calculating the similarity measure NCC and the processing time. For comparison reasons, we reduced the number of features without NMS by thresholding the Harris corner response. This method is not efficient since a large number of redundant features are detected. The large amount of features also accounts for a long processing time. Moreover, this simple thresholding technique lacks of repeatability because of the varying image quality in different datasets. In contrast, we are able to control the number of features and the processing time in the NMS method through variation of the suppression area. However, large errors occurred in some images because of the large suppression area. We therefore propose an adaptive method to determine the size of the suppression area.

Acknowledgment: The authors would like to thank Prof. Matthias Kirsch of the University Hospital in Dresden, Germany, for supporting the intraoperative measurements.

Author Statement

Research funding: This work is supported by the European Social Fund (grant no. 100270108) and the Free State of Saxony. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] G. Steiner, S. B. Sobottka, E. Koch, G. Schackert, and M. Kirsch, "Intraoperative imaging of cortical cerebral perfusion by time-resolved thermography and multivariate data analysis," *Journal of biomedical optics*, vol. 16, no. 1, p. 016001, 2011.

[2] J. Hollmach, N. Hoffmann, C. Schnabel, S. Küchler, S. Sobottka, M. Kirsch, G. Schackert, E. Koch, and G. Steiner, "Highly sensitive time-resolved thermography and multivariate image analysis of the cerebral cortex for intrasurgical diagnostics," in *SPIE BiOS*. International Society for Optics and Photonics, 2013, p. 856550.

[3] N. Hoffmann, E. Koch, U. Petersohn, M. Kirsch, and G. Steiner, "Cerebral cortex classification by conditional random fields applied to intraoperative thermal imaging," *Current Directions in Biomedical Engineering*, vol. 2, no. 1, pp. 475–478, 2016.

[4] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[5] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[6] C. Harris and M. Stephens, "A combined corner and edge detector." in *Alvey vision conference*, vol. 15, no. 50. Citeseer, 1988, pp. 10–5244.

[7] C. Faria, O. Sadowsky, E. Bicho, G. Ferrigno, L. Joskowicz, M. Shoham, R. Vivanti, and E. De Momi, "Validation of a stereo camera system to quantify brain deformation due to breathing and pulsatility," *Medical physics*, vol. 41, no. 11, 2014.