Robust Optical Flow in low SNR Thermal-Infrared

Stephen Vidas, *Member, IEEE,* Vu Hoang Minh, *Member?, IEEE,* and C.C. Cheah, *Senior Member?, IEEE*

***Abstract*—...**

I. Introduction

[Motivation] Thermal-infrared imaging has been proven to have advantageous properties in difficult lighting conditions such as darkness, dust, smoke and fog; especially for the tasks of detecting humans, animals, powered equipment and other anomalies. Because of this, a thermal-infrared camera is already an appealing choice to be part of a robot- or human-mounted sensor configuration for modern search and rescue, firefighting or exploration tasks. However,more can be done with the thermal-infrared video than simply 2D visualization and detection, in that it is possible to implement a SLAM (Simultaneous Localization And Mapping) system to enable a 3D understanding of the environment, including the position of the human or robot within it. This information is crucial for navigation purposes, as well as planning purposes for emergency response.

[Applications] Firefighting [1], Search & Rescue [?], Robot navigation [2], [3], Energy auditing [cite Steve's journal paper if it's published in time] and Exploration [?].

Examples of some recent investigations into these areas (can we find relevant citations for these kinds of work?)

- 3D "Building Reconstruction and Thermal Mapping in Fire Brigade Operations Categories and Subject Descriptors" [4]. Link http://www.youtube.com/watch?v=QtNvHBT3jMs
- http://youtu.be/BB8jojwl7ws?t=6m30s

[Contributions] These are what we are aiming for, for this initial paper:

- Continuous real-time operation on a standard PC (at least 10 frames per second, aiming more for 30)
- Appropriate default parameters (or adaptive parameters) for a range of typical and atypical environments and data sequences
- Tracking that can maintain a large number of stable and well-distributed features under the following challenges:
  - sudden movements
  - forwards and backwards motion
  - fairly low SNR
- Well-documented, cross-platform, OpenCV/C++ based library, integrated with ROS, that can be used for sparse optical flow on thermal-infrared

[Structure]

- motivation
- contributions

All authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798 e-mail: hoangmin001@e.ntu.edu.sg,{svidas,ecccheah}@ntu.edu.sg.

Fig. 1: ...

- intro to structure

Placeholder citation [5].

II. Background

Placeholder Figure 1.

A. Thermal-Infrared Computer Vision

Some information on calibration [6].

NUCs and other useful information for researchers to get stuck into solving similar problems with thermal-IR cameras. A lot of this can be sourced from Steve's thesis (especially Chapter 2).

B. Sparse Optical Flow

...

III. Related Work

A. Image Denoising

[7], [8], [9]

B. Feature Detection

[5], [10]

C. Feature Tracking

[11]

Can include some references to paper/s on object tracking in thermal-ir, with disclaimer that by tracking living or powered things like people or vehicles, the images are by definition high SNR, so the problem is much easier

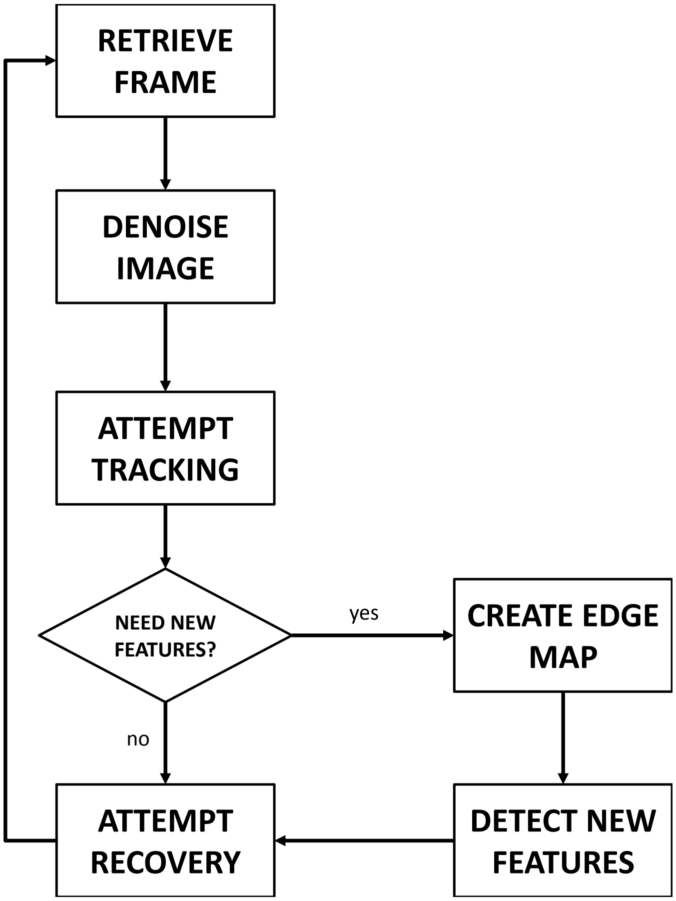


Fig. 2: Flow chart demonstrating high-level optical flow procedure.

IV. Proposed Method

Aims:

- enable detection and tracking of features in low SNR regions
- ensure a good distribution of tracked features
- default values should perform well for all kinds of thermal images

Flowchart, Figure 2.

A. Feature Tracking

Some contributions that were not included in previous paper:

- LDA-based matching for interruption handling
- initial warp for minimizing search windows based on feature velocity
- adaptive tracking window, based on no. of features and feature velocity:
  - this appropriately limits the spatial size of the search space for each new projection in order to reduce processing time while simultaneously decrease chance of incorrect tracking, without preventing good tracking results
  - no. of features relative to maximum and predefined frac defines maximum distance
  - 2 times the expected required distance defines the minimum (well, 3 is the actual minimum)
- scalable feature track management method enabling loop closure
  - limits memory consumption to a fixed, finite upper limit
  - still provides opportunity to recover lost features and potentially enables loop closure

V. Experiments

Timing tests were performed on a Intel Core I5-3317 CPU 1.7GHz, 8GB RAM.

Experiments have been designed to assess the effectiveness of the proposed method compared to existing alternatives, as well as determine which aspects of the method have the most positive effect, and at what cost (in terms of processing time). (Each part of the evaluation will include FAST and/or other conventional methods as a baseline).

First, a new set of data sequences captured for the experiments is presented in Subsection V-A. Second, the evaluation framework for the subsequent three subsections is defined in Subsection V-B. Third, several alternative denoising methods are compared for front-end image processing, and evaluated using both the proposed and conventional methods in Subsection V-C. Fourth, the edge detection and processing stage is evaluated in terms of performance and processing time in Subsection V-D. Fifth, the corner selection procedure based on the edge map is evaluated in Subsection V-E. Sixth, tracking improvements are demonstrated in Subsection V-F. Finally, the proposed method is evaluated on the same dataset as used by [11], allowing a direct comparison with the best known past performance of a thermal-infrared sparse optical flow system in Subsection V-G.

A. Datasets

The evaluation is done on thermal images collected and labeled by thermographic camera, Optris PI450, equipped with 80Hz measurement speed and 382 x 288 pixels optical resolution [1]. With high speed capturing ability, Optris PI450 is sufficient for real-time Simultaneous Localization and Mapping (SLAM).

Four datasets with different signal-to-noise (SNR) ratios are used as follows.

- High SNR indoor sequence
- Low SNR indoor sequence
- High SNR outdoor sequence
- Low SNR outdoor sequence

Thermal images not in our database can be easily added to the training set without affecting any algorithms.

Figure 3 shows samples of datasets:

B. Framework

This is where you explain the basic experiments you will be doing. When it comes to judging the performance of the

[1] http://www.optris.com/thermal-imager-pi400

(a) Sequence 1 (b) Sequence 2

(c) Sequence 3 (d) Sequence 4

(e) Sequence 5 (f) Sequence 6

(g) Sequence 7 (h) Sequence 8

Fig. 3: Datasets for evaluation

feature detection algorithms (and implicitly, the image processing algorithms), I recommend emphasizing the following characteristics:

- trackability - looking at for example the average lifetime of detected features, and how far they ”‘drift”’ throughout the sequence. Can use my reversal technique here (looping from 0 to 100 and back to 0 without doubling up, and comparing the start and end location of the tracked feature)
- precision - how accurately the features are detected on valid points-of-interest, as reflected in the repeatability of detections (whether the exact same points are detected in the real world even when the camera has shifted to a different view). You can compare each new detection with how the features have been tracked using Lucas-Kanade, but this will probably also require manual (human) verification.
- quality assessment - looking at how effective the feature ranking (strength metrics) are. You can probably use some qualitative analysis here, with some good figures, but also you can consider looking at the relationship between feature quality, and other parameters such as the precision and trackability of the feature, to determine how well the quality metric works.

You should show some comparative images from FAST and the proposed detector (perhaps with a low SNR and high SNR image each) which show the top N features from each detector. The top N from the proposed detector should seem much more logical/intuitive. Later experiments will demonstrate a good relationship between the feature ”‘strength”’ and its usefulness/effectiveness.

The framework should make clear that you want to investigate changing parameters such as:

- denoising method
- feature detection method
- feature ranking/sorting method
- thresholds and sensitivity levels

C. Image Processing

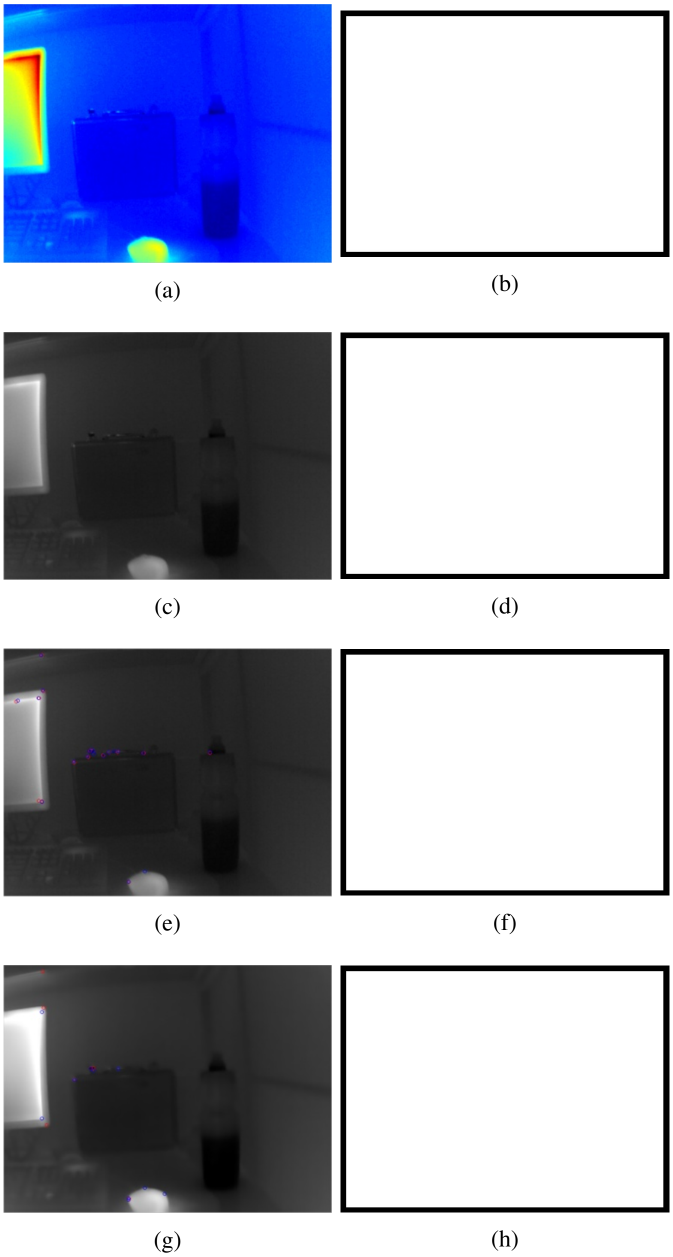
In this section, authors would like to establish the effectiveness of the two proposed denoising methods, namely Non-local algorithm [7] and Transform-domain collaborative filtering [9]. To summarize, first method analyzes noise for local smoothing filters followed by computing nonlocal means whilst second approach groups 2D image fragments into 3D data arrays to unveil the finest details based on aggregation and collaborative Wiener filtering.

Compare to Transform-domain collaborative filtering in terms of processing time, Non-local means is 5 time faster. However, in terms of output quality, Transform-domain collaborative filtering is still state-of-the-art method for denoising target. As a result, we propose to employ first denoising technique in case of low noise level. Otherwise, second approach should be used to ameliorate the following steps. To classify whether an image is noisy or not, we profound applying an estimation of noise level based on Bayesian MAP inference [12] and analogizing to a predefined threshold of 5 (for 0-255 grayscale) (need to find this threshold!?).

To evaluate the effectiveness of two proposed denoising methods, two thermal images with different level of SNR were introduced. Next, feature detectors, namely FAST [13] and HARRIS [14], were deployed in noisy and denoised frames for collation. The results are shown in Fig. 4.

D. Edge Detection and Processing

This is where you will assess the effectiveness of different thresholds and/or settings for the edge detection, and also establish which (if any) of your proposed filters and other methods actually improve performance, and how fast they are.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4: Image processing evaluation

E. Corner Selection

This section should be where you establish the effectiveness of the two denoising methods, including their processing time, by using existing (conventional) feature detectors such as FAST and HARRIS, and a conventional tracking algorithm such as Lucas-Kanade.

Also, you should look at the effect of different methods of scaling the 14-bit data to 8-bit format. The advantage of 8-bit format is that it requires less processing time, and also allows more modularity with existing functions and libraries. Furthermore, there should generally be very little (if any) information lost in this conversion, and when information is lost, it means it is a high SNR image and so the loss of the information is not a problem (there is still sufficient information left over).

F. Tracking Improvements

Want to demonstrate the following:

- adaptive windows can reduce processing time without sacrificing performance, and simultaneously reduce chance of incorrect matching
- scalable feature track management method limits memory consumption, prevents lagging, maintains ability to recover lost features

G. Motion Estimation Stability

Gain access to the exact same sequences used by Steve for his previous paper. Run the 3D motion estimation / monocular SLAM system (when ready) on these sequences, using your proposed methods. Superimpose results on the same plot as used in the previous paper, to show that error has been much improved.

Mean time until failure can be defined as the metric.

VI. Conclusion

...

Acknowledgment

(Will need to check if this is necessary with Prof Cheah) The authors would like to thank the funding from A*STAR, Singapore's National Research funding agency, for the support of this project.

References

[1] C. Schönauer, E. Vonach, and G. Gerstweiler, "3D Building Reconstruction and Thermal Mapping in Fire Brigade Operations Categories and Subject Descriptors," pp. 202–205.

[2] W. Maddern and S. Vidas, "Towards Robust Night and Day Place Recognition using Visible and Thermal Imaging," pp. 1–6.

[3] A. Gonzalez, "SLAM visuel monoculaire par caméra infrarouge Méthode Présentation du système," 2011.

[4] C. Schönauer, E. Vonach, G. Gerstweiler, and H. Kaufmann, "3d building reconstruction and thermal mapping in fire brigade operations," in *Proceedings of the 4th Augmented Human International Conference*. ACM, 2013, pp. 202–205.

[5] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, and T. Wark, "An Exploration of Feature Detector Performance in the Thermal-Infrared Modality," *2011 International Conference on Digital Image Computing: Techniques and Applications*, pp. 217–224, Dec. 2011.

[6] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, S. Member, and T. Wark, "A Mask-Based Approach for the Geometric Calibration of Thermal-Infrared Cameras," pp. 1–11, 2012.

[7] A. Buades, B. Coll, D. Matem, C. V. Km, P. D. Mallorca, J.-m. Morel, and E. N. S. Cachan, "A non-local algorithm for image denoising," no. 0, pp. 0–5.

[8] Y. Hou, C. Zhao, D. Yang, and Y. Cheng, "Comments on "Image denoising by sparse 3-D transform-domain collaborative filtering"." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 20, no. 1, pp. 268–70, Jan. 2011.

[9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, 2007.

[10] N. H. C. Yung, "Corner detector based on global and local curvature properties," *Optical Engineering*, vol. 47, no. 5, p. 057008, May 2008.

[11] S. Vidas and S. Sridharan, "Hand-held Monocular SLAM in Thermal-infrared," no. December, pp. 5–7, 2012.

[12] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang, "Noise estimation from a single image," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 901–908.

[13] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 430–443.

[14] C. Harris and M. Stephens, "A combined corner and edge detector." in *Alvey vision conference*, vol. 15. Manchester, UK, 1988, p. 50.