An exploration of feature detector performance in the thermal-infrared modality

Stephen Vidas$^{1,2}$, Ruan Lakemond$^1$, Simon Denman$^1$, Clinton Fookes$^1$, Sridha Sridharan$^1$, Tim Wark$^2$

$^1$ SAIVT Laboratory, QUT, 2 George St, Brisbane 4000
$^2$ Autonomous Systems Laboratory, CSIRO QCAT, 1 Technology Ct, Pullenvale 4069

\{stephen.vidas, r.lakemond, s.denman, c.fookes, s.sridharan\}@qut.edu.au
$^2$ steve.vidas, tim.wark}@csiro.au

Abstract—Thermal-infrared images have superior statistical properties compared with visible-spectrum images in many low-light or no-light scenarios. However, a detailed understanding of feature detector performance in the thermal modality lags behind that of the visible modality. To address this, the first comprehensive study on feature detector performance on thermal-infrared images is conducted. A dataset is presented which explores a total of ten different environments with a range of statistical properties. An investigation is conducted into the effects of several digital and physical image transformations on detector repeatability in these environments. The effect of non-uniformity noise, unique to the thermal modality, is analyzed. The accumulation of sensor non-uniformities beyond the minimum possible level was found to have only a small negative effect. A limiting of feature counts was found to improve the repeatability performance of several detectors. Most other image transformations had predictable effects on feature stability. The best-performing detector varied considerably depending on the nature of the scene and the test.

Keywords—thermal-infrared, feature detectors, evaluation

I. INTRODUCTION

The lack of good lighting is a major cause of poor feature detector performance in the visible modality [15]. The thermal-infrared modality is more robust to poor lighting conditions because it depends largely on emitted rather than reflected radiation [19]. Figure 1 demonstrates the potential of a thermal-infrared image for feature extraction under conditions unsuitable for visible imagery.

Accurate local feature detection forms a critical component of many algorithms used in robotics and computer vision. These include algorithms for structure from motion and SLAM (Simultaneous Localization And Mapping) [6], wide-baseline matching [14] and image classification [10]. Because feature detectors are generally designed and tested on visible images, their performance characteristics when applied to thermal-infrared images are unclear. Limited experiments such as [11] have found thermal-infrared images to be challenging to work with for problems involving feature detection. This is contributed to by relatively low SNR (Signal-to-Noise Ratio) [12]. There is therefore a need to better understand feature detection in this modality.

The widely accepted state-of-the-art analysis of local feature detectors in the visible modality can be found at [18]. Although a seminal work, the protocol presented has several limitations that make it particularly unsuitable for application in the thermal-infrared modality. Perhaps most significantly, the analysis of [18] uses default parameters in tuning the sensitivity thresholds of each feature detector. These default parameters were initially selected by the respective developers of each feature detector for perceived optimal performance in the visible modality. When these parameters are used for detection in thermal-infrared images, a negligible number of features is typically returned. The fixing of these parameters has two additional drawbacks. First, it neglects the possibility to analyse the performance of detectors as sensitivity thresholds are varied. Second, there is a bias in the analysis towards dense responses. Another key limitation noted by the authors is that their dataset is biased towards detector-friendly problems involving scenes that are texture-rich. Furthermore, since the time of its publication in 2005, there have been several new feature detectors proposed which call for this analysis to be re-examined.

The purpose of this research is to explore feature detector performance in the thermal-infrared modality. This paper proposes a new dataset of thermal-infrared images for feature detector evaluation, made publicly available at [1]. Also proposed is a modified evaluation protocol designed to utilize this dataset and avoid several of the limitations
found in [18]. A key improvement of the proposed protocol is its focus on controlling for detector sensitivity. This allows detector performance to be compared over a range of sensitivity thresholds, and without a bias towards dense responses. Experiments are then undertaken to explore thermal-infrared feature detector performance with a view to enabling new and existing computer vision algorithms to be implemented more effectively outside the visible spectrum.

II. DATASET

In order to perform the feature detector evaluation, a dataset of thermal-infrared image sequences was captured. The dataset has been made publicly available to the research community and can be found at [1]. The dataset explores the impacts of both different environments and different image transformations on detector performance.

The low SNR common in thermal images is largely due to the presence of non-uniformities in the sensor. These are unique to the thermal modality, and are mainly attributed to the difference in photo-response of each detector in the focal plane array [13]. Thermal-infrared images taken in environments with a low thermal contrast will have lower SNR, because the non-uniformities will be more dominant. The effect of non-uniformities on the SNR can be reduced by periodic NUC (Non-Uniformity Correction) operations. However, even immediately after such an operation is performed, a minimum amount of this noise is still present.

A. Data capture and preparation

The thermal-infrared camera used to capture the dataset was a Thermosteknix Miricle 307K. It has a spatial resolution of 640×480 pixels, and a pixel depth of 14 bits in raw image format. Images were captured from the camera using “yavta” [3] and DDX [9].

Raw images captured from the device have a significant amount of lens distortion. The effects of lens distortion invalidate several of the assumptions required by standard feature detector evaluations. An implementation of Zhang’s method [21] tailored for the thermal-infrared modality was used to calibrate the camera, and remove the effects of lens distortion. It should be noted that spatially remapping the image to correct for lens distortion also has the effect of remapping noise within the image.

For the purpose of several analyses, images were captured in pairs for each transformation level within each sequence. These pairs of images have slight rotations relative to one another. This is in order to vary the true locations of features within the image and include the effects of spatial quantization.

Homographies were computed to map each test image to its reference image to compensate for the effects of camera motion. Where necessary, an initial homography was provided by manually specifying point correspondences. All homographies were refined using inverse compositional image alignment [7].

Table I: Dataset histogram statistics. $\Delta I_{m\rightarrow n}$ represents the intensity range of the original pre-normalized image between the $m\%$ and $n\%$ quantiles. skewness refers to the standard measure of asymmetry in a probability distribution [2].

<table>
<thead>
<tr>
<th>Environment</th>
<th>$\Delta I_{0\rightarrow 100}$</th>
<th>$\Delta I_{1\rightarrow 99}$</th>
<th>$\Delta I_{2\rightarrow 95}$</th>
<th>skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building</td>
<td>156</td>
<td>148</td>
<td>115</td>
<td>-2.1</td>
</tr>
<tr>
<td>Car</td>
<td>148</td>
<td>138</td>
<td>77</td>
<td>-1.3</td>
</tr>
<tr>
<td>Electronics</td>
<td>68</td>
<td>63</td>
<td>36</td>
<td>1.7</td>
</tr>
<tr>
<td>Grate</td>
<td>70</td>
<td>61</td>
<td>51</td>
<td>-0.3</td>
</tr>
<tr>
<td>Kitchen</td>
<td>158</td>
<td>151</td>
<td>133</td>
<td>0.8</td>
</tr>
<tr>
<td>Night</td>
<td>479</td>
<td>45</td>
<td>37</td>
<td>10.7</td>
</tr>
<tr>
<td>Outside</td>
<td>833</td>
<td>507</td>
<td>338</td>
<td>0.5</td>
</tr>
<tr>
<td>Pavement</td>
<td>82</td>
<td>74</td>
<td>54</td>
<td>-0.2</td>
</tr>
<tr>
<td>Pipes</td>
<td>40</td>
<td>29</td>
<td>23</td>
<td>-0.1</td>
</tr>
<tr>
<td>Soil</td>
<td>36</td>
<td>26</td>
<td>17</td>
<td>0.2</td>
</tr>
</tbody>
</table>

B. Environments

The proposed dataset consists of image sequences captured from ten environments, illustrated in Figures 2, 3 and 4. Table I shows a summary of the statistical properties relating to each environment’s image histogram distribution. In thermal images, it is not unusual for the raw intensity range to be well below 255, to which it is typically normalized for display and analysis purposes.

The environments used in the dataset contain both “structured” and “textured” regions [18]. The choices of environments were made with the intention of covering a variety of typical, natural scenes with a significant variation in SNR. In the “Night” environment, the presence of a hot light generates a strong positive skew and artificially increases the intensity range. These environments with particularly low SNR have proven to be difficult to work with in the past [11].

C. Transformations

A total of eight image transformations are explored in the proposed dataset. The effect of these transformations on feature stability as their severity is increased is of particular interest. For this paper, the effect of each transformation is investigated within two different environments. Each consists of pairs of images ranging from ideal to severely transformed.

Four of the transformations can be described as digital transformations, which can be applied to a single image to varying degrees after capture. These digital transformations are:

- JPEG compression.
- Gaussian noise.
- Quantization noise.
- Histogram expansion.

Examples of the effects of these transforms are shown in Figure 3.
For the JPEG compression sequences, OpenCV [4] was used to progressively compress the images down to a quality level of 0.05 (defined as a compression level of 0.95), which corresponded to a reduction in file size of approximately 100 times. For the Gaussian noise sequences, the standard deviation was varied as Gaussian-distributed noise was added to each pixel in the original captured image. This addition of noise occurred before histogram expansion was performed on the image. For the quantization noise sequences, the original image was sub-sampled at increasing factors.

For the histogram expansion sequences, a lower and upper intensity cut-off was determined for a range of predefined percentiles. For example, when the threshold is set at 0.1, the upper threshold is set at an intensity greater than exactly 95% of pixel values, and the lower threshold is set at the 5% value. All intensity values from the original image are then linearly mapped to the [0, 255] range using these limits.

The other four image transforms are physical in nature,
and were controlled as the data is being captured. These physical transformations are:

- Non-uniformity noise.
- Out-of-focus blurring.
- Change in viewpoint.
- Time of day.

Examples of the effects of these transformations are shown in Figure 4.

For the non-uniformity noise sequences, the ideal image was taken immediately after a NUC (Non-Uniformity Correction) was performed on the camera. This operation involves a physical shutter of uniform intensity being used to radiometrically calibrate the imaging sensor. Each subsequent image in the sequence was taken after a 30 second delay, so that the final image experiences five minutes of accumulation of additional non-uniformity noise beyond the minimum level.

For the out-of-focus blurring sequences, the camera focus was manually varied incrementally between the near and far field extremes. For the change in viewpoint sequences, images were captured by varying the angle of the camera relative to the scene over a range of angles. For the time of day sequences, images were captured from a fixed location on the hour throughout the day.

III. EVALUATION

Several modifications have been made to the protocol in [18]. These changes are in order to improve the protocol’s effectiveness for the task of exploring feature detector performance in the thermal-infrared modality. However, the changes would also be useful for improving the effectiveness of evaluations in the visible modality. Some of the modifications address the following weaknesses (discussed in Section I), which are acknowledged by the original authors.

- Bias towards dense responses.
- Bias towards detector-friendly problems.

To remove the bias towards dense responses, the number of features returned by each detector is fixed for comparison. This way, all detectors can be compared on a more even footing for a range of feature counts. To address the bias towards detector-friendly problems, the proposed dataset covers a larger variety of environments which are not restricted to texture-rich scenes. Several of these environments could be described as difficult problems for feature detection. This is due to factors such as low SNR, large regions of uniformity and low texture content.

Several popular feature detectors were selected for the evaluation:

- Speeded Up Robust Features (SURF) [8].
- Hessian-affine (Hes) [17].
- Star (Star) [5].
- Maximally Stable Extremal Regions (MSER) [16].
- Features from Accelerated Segment Test (FAST) [20].

When possible, existing implementations of detectors provided by the respective authors or by libraries such as OpenCV [4] were used. However, many provided implementations did not allow re-tuning of the detector sensitivity thresholds. This made these implementations ineffective for the analyses, and ineffective for use on thermal-infrared images in general. In some cases, particular detectors were unable to return a sufficient number of features for a test. In these instances, the detector was left out of the corresponding plot in the experimental results.

A subset representative of the most interesting results has been selected for inclusion due to requirements for brevity.

A. Metrics

This paper uses the repeatability measure from [18] as the metric for measuring the stability of detected features between images. In implementing this measure, a homography is first used to map the second image to the first. The number of features in the common region of the two images is the maximum number of correspondences which can theoretically be made between the two sets of extracted features. The number of actual corresponding regions is then used to calculate repeatability.

\[
1 - \frac{R_{\mu_a} \cap R_{(H^T \mu_a)H}}{(R_{\mu_a} \cup R_{H^T \mu_a}H)} < \epsilon_0.
\]  

Here \(R_{\mu_a}\) represents the elliptic region (fitted to the size and shape of the feature) defined by \(x^T H x = 1\). \(H\) is the homography relating the two images. The union of the regions is \(R_{\mu_a} \cup R_{(H^T \mu_a)H}\), and their intersection is \(R_{\mu_a} \cap R_{(H^T \mu_a)H}\). The areas of the union and intersection of the regions are computed numerically.

In this paper, an overlap error of less than 20% is required to accept a match. This is in contrast to [18] which prefers to use a much more lenient maximum error of 40%. This lower limit is chosen to increase the standard for precision of the detectors, as should be expected as the field of computer vision advances.

B. Sensitivity Investigation

The number of features returned by a detector can often be changed by altering one parameter. In this paper, this parameter is referred to as the sensitivity threshold. In order to return a sufficient number of features from a thermal-infrared image, it was generally found that that threshold needed to be lowered considerably compared with visible-spectrum images.

As an experiment, increasingly more features were extracted from pairs of ideal images within each environment, by decreasing the sensitivity threshold. These images were identical except for a small change in time (allowing a change in non-uniformity distribution) and a small rotation. Repeatability was then calculated between the images within each pair. Figure 5 shows the effect of varying the sensitivity...
threshold on repeatability within two of the environments. These two environments were specifically chosen because of their very different statistical properties.

Under ideal conditions, a repeatability score of 1.00 would be maintained regardless of the number of features returned. However, the effect of spatial quantization and unavoidable non-uniformity noise on thermal-infrared images means that this is not the case.

The comparatively poorer performance of all detectors on the “Night” sequence can be attributed to the low SNR of this environment. In fact, in this environment, the absolute maximum feature count for several of the detectors was less than 500.

Most detectors’ performances degraded as increasing numbers of features were retained. However, in high SNR environments the Hessian-affine [17] and FAST [20] detectors maintained a relatively stable repeatability score. These detectors were designed to produce large numbers of features, which is perhaps why they do not easily reach a point where they become overly sensitive. Nevertheless, it should be noted that when the area of the image is saturated by a large number of features, the proportion of “accidental” correspondences will increase.

Table II shows a summary of the performances of each detector in each environment.

The performance of detectors varied considerably between environments, although the SURF [8] detector most consistently performed near the top. The Hessian-affine [17] detector tended to show better performance in low SNR environments, where non-uniformity noise had a greater impact.

The performance of the FAST [20] and Star [5] detectors was good in high SNR environments, but tended to be poor when SNR was low.

C. Digital Transformation Evaluation

The effect of a number of digital transformations (listed in Section II-C) was explored.

These image sequences were generated using the approaches described in Section II-C. The transformations were applied equally to the images in each pair. The top

Table II: Detector performance in each environment. Each detector was tuned to return the maximum number of features whilst maintaining a repeatability at least 90% of its best possible score. The reduced repeatability is recorded, along with the corresponding feature count in brackets. The highest repeatability achieved for each environment is highlighted in bold. Missing entries correspond to cases where the detector failed to return a sufficient number of features for that particular test.

<table>
<thead>
<tr>
<th>Environment</th>
<th>SURF</th>
<th>Hessian</th>
<th>Star</th>
<th>FAST</th>
<th>MSER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building</td>
<td>0.71 (400)</td>
<td>0.54 (400)</td>
<td>0.64 (400)</td>
<td>0.68 (46)</td>
<td>0.50 (150)</td>
</tr>
<tr>
<td>Car</td>
<td>0.75 (175)</td>
<td>0.59 (75)</td>
<td>0.72 (147)</td>
<td>0.66 (356)</td>
<td>0.53 (176)</td>
</tr>
<tr>
<td>Electronics</td>
<td>0.74 (100)</td>
<td>0.55 (250)</td>
<td>0.71 (100)</td>
<td>0.77 (70)</td>
<td>0.61 (101)</td>
</tr>
<tr>
<td>Grate</td>
<td>0.68 (250)</td>
<td>0.56 (400)</td>
<td>0.83 (25)</td>
<td>0.78 (289)</td>
<td>0.53 (75)</td>
</tr>
<tr>
<td>Kitchen</td>
<td>0.77 (200)</td>
<td>0.61 (350)</td>
<td>0.73 (99)</td>
<td>0.74 (331)</td>
<td>0.53 (75)</td>
</tr>
<tr>
<td>Night</td>
<td>0.65 (25)</td>
<td>0.56 (125)</td>
<td>0.64 (50)</td>
<td>0.62 (40)</td>
<td>0.47 (21)</td>
</tr>
<tr>
<td>Outside</td>
<td>0.85 (225)</td>
<td>0.61 (50)</td>
<td>0.74 (497)</td>
<td>0.75 (478)</td>
<td>0.68 (250)</td>
</tr>
<tr>
<td>Pavement</td>
<td>0.76 (75)</td>
<td>0.65 (125)</td>
<td>0.65 (99)</td>
<td>0.56 (62)</td>
<td>0.55 (25)</td>
</tr>
<tr>
<td>Pipes</td>
<td>0.52 (25)</td>
<td>0.39 (25)</td>
<td>0.47 (98)</td>
<td>0.40 (33)</td>
<td>-</td>
</tr>
<tr>
<td>Soil</td>
<td>0.65 (100)</td>
<td>0.60 (275)</td>
<td>0.63 (24)</td>
<td>0.63 (14)</td>
<td>-</td>
</tr>
</tbody>
</table>

100 features using each detector were then returned from each image. Repeatability was measured between the images within each pair.

Figure 6 shows the effect of different levels of JPEG compression on repeatability between pairs of images in two sequences. JPEG compression had a slight negative effect on detector performance up to relatively high levels of compression. In environments with high SNR, the degradation was less significant at first but became abrupt at high compression levels. This is likely because these images are more resilient to the quantization inherent in the JPEG compression algorithm. The reason for the upward trend of the FAST [20] detector in the “Night” environment is that quantization has artificially distorted each image to be even more similar than the originals. Overall, the SURF [8] detector performed best in the high SNR “Outside” environment, while the Hessian-affine [17] detector performed better in the lower SNR “Night” environment.

Figure 7 shows the effect of introducing Gaussian noise. All environments exhibited reduced repeatability under the
presence of Gaussian noise. Environments with high SNR were much more resilient to the presence of noise. Again, the SURF [8] detector performed better in the “Outside” environment and the Hessian-affine [17] detector performed better in the “Night” environment.

Figure 8 shows the effect of varying the histogram expansion thresholds. The effect of different levels of thresholding varied considerably between environments. Environments with intensity histograms with large tails (both positively and negatively) were most susceptible to performance degradation as thresholds varied. Most environments responded best to a threshold of zero, although in some cases the effect of changing the threshold was barely noticeable.

Figure 9 shows the effect of quantization noise on the image. Quantization noise had a relatively consistent effect on repeatability between environments. Low quantization noise had only a very small effect on detector performance. However, when the quantization factor increased beyond a factor of three or four, degradation became significant.

D. Physical Transformation Evaluation

For the out-of-focus blur and time-of-day transformations, an approach similar to that outlined in Section III-C was used.

Figure 10 shows the effect of out-of-focus blur. Each environment contained objects within a different range of distances, and therefore different focus levels achieved the best result. Reducing or increasing the focus distance relative to the optimal level causes a rapid degradation in feature stability. This is perhaps exacerbated by the narrow depth of field typical of many thermal-infrared cameras. The Hessian-affine [17] detector was shown to be the most resilient to loss of focus, while the FAST [20] detector was least resilient.

Figure 11 shows the effect of changing the time of day. The different environments were found to be affected very differently by the passage of time. For the man-made “Building” environment, repeatability was relatively constant for most detectors, although the FAST [20] detector showed a significant amount of fluctuation. For the natural “Soil” environment, the performance of all detectors except for the Hessian-affine [17] varied considerably with time of day. The effect of clouds and the change in ambient temperature in this instance made the times of 9.00am and 1.00pm particularly difficult for these detectors.

For the non-uniformity and viewpoint transformations, a
different approach was used for the evaluation. This involved determining the repeatability score between increasingly transformed images relative to an original or ideal image, rather than between two equally transformed images. This approach is more similar to that used in the protocol of [18]. This was considered a more interesting question for these transformations, in order to determine how non-uniformity noise affect the feature extraction process, and how viewpoint-covariant feature detectors behave in the thermal-infrared modality.

Figure 12 shows the effect of an accumulation of non-uniformity noise beyond the normal level on detected feature stability. A very slight decay in performance over several minutes is evident in the results, although the MSER [16] detector is more adversely affected than other detectors. Near-optimal feature stability may therefore be maintained with less frequent NUC (Non-Uniformity Correction) operations than usually applied (often as frequent as every 15 seconds). However, as the time since the last NUC operation increases, the image intensities collectively drift from their true values. Therefore the accumulation of non-uniformity noise may have a more significant influence on later stages of computer vision algorithms such as feature description and matching.

Figure 13 shows the effect of viewpoint change on repeatability. Repeatability decreased rapidly as the angle relative to the original image was increased. However, the Hessian-affine [17] detector exhibited the best affine-covariance of all of the detectors. It is unclear how affine-covariant the MSER [16] detector is in the thermal-infrared modality, since both the environments for this test were too challenging for it. It is difficult to find planar surfaces with sufficient SNR for the MSER [16] detector to be effectively evaluated.

IV. CONCLUSION & FUTURE WORK

Feature detection in the thermal-infrared modality was argued to be a research area of great interest. This was largely because of the potential performance improvements of many algorithms by using thermal-infrared cameras. This paper set out to explore the performance of several popular feature detectors in this alternative modality.

Thermal-infrared environments used to comprise the captured dataset were found to vary considerably in their basic statistical properties. These differences correlated strongly with differences in feature detector performance within these environments. Repeatability was found to degrade significantly faster as more features were retained in environments with low SNR. The results of the analyses showed that the Hessian-affine [17] detector was more resilient in these difficult environments. However, on average the SURF [8] detector achieved the highest repeatability scores.

The effect of JPEG compression and Gaussian noise was less for high SNR environments, up to extreme levels. Varying the thresholds for histogram expansion to reasonable levels had minimal negative effect, except on the FAST [20] detector. Analyses of the effects of quantization, out-of-focus blurring and viewpoint change had expected results. An increase in non-uniformity noise beyond the minimum possible level was found to have only a slight negative effect on detector performance. The Hessian-affine [17] detector achieved the most consistent results in outdoor environments throughout different times of day.

Future work planned includes an expansion and refinement of the dataset to improve the depth of the investigation into the effects of the different transformations. This dataset will also include visible images, so that the differences between modalities can be investigated. The statistical robustness of the evaluation could also be improved by considering repeatability amongst groups of images, rather than just pairs.

ACKNOWLEDGMENT

This project has been funded in part by a Queensland Government Smart Futures PhD Scholarship and a CSIRO (Commonwealth Scientific and Industrial Research Organisation) OCE (Office of the Chief Executive) Scholarship. Equipment as well as manufacturing and computing facilities were provided by both CSIRO and QUT (Queensland University of Technology).
REFERENCES


