Perception Quality Evaluation with Visual and Infrared Cameras in Challenging Environmental Conditions

Christopher Brunner and Thierry Peynot

Abstract. This work aims to contribute to the reliability and integrity of perceptual systems of unmanned ground vehicles (UGV). A method is proposed to evaluate the quality of sensor data prior to its use in a perception system by utilising a quality metric applied to heterogeneous sensor data such as visual and infrared camera images. The concept is illustrated specifically with sensor data that is evaluated prior to the use of the data in a standard SIFT feature extraction and matching technique. The method is then evaluated using various experimental data sets that were collected from a UGV in challenging environmental conditions, represented by the presence of airborne dust and smoke. In the first series of experiments, a motionless vehicle is observing a 'reference' scene, then the method is extended to the case of a moving vehicle by compensating for its motion. This paper shows that it is possible to anticipate degradation of a perception algorithm by evaluating the input data prior to any actual execution of the algorithm.

1 Introduction

Unmanned Ground Vehicles (UGVs) require high integrity perception systems in order to achieve long-term autonomous operation. Current state-of-the-art perception modules employ a range of sensors along with a variety of algorithms to interpret and fuse data to provide a meaningful Representation of the Environment (RotE) that is generally used to enable successful navigation of the UGV. However, catastrophic consequences might occur if the sensed data is misinterpreted by the perception system, generating an inaccurate RotE. A simplified perception block is shown in Fig. 1.

Christopher Brunner · Thierry Peynot
ARC Centre of Excellence for Autonomous Systems,
Australian Centre for Field Robotics (ACFR),
The University of Sydney, NSW 2006, Australia
e-mail: c.brunner@acfr.usyd.edu.au, tpeynot@acfr.usyd.edu.au

O. Khatib, V. Kumar, and G. Sukhatme (eds.), *Experimental Robotics*, Springer Tracts in Advanced Robotics 79,
DOI: 10.1007/978-3-642-28572-1_49,

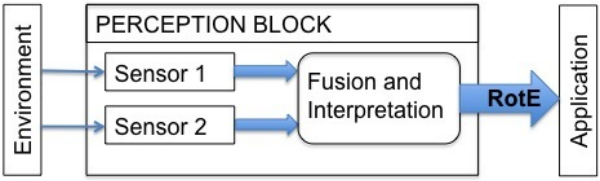


Fig. 1 Simplified Perception Block. Sensors 1 and 2 provide data from the environment. The sensed data is fused and interpreted to provide an internal Representation of the Environment (RotE) which is suitable for the required application.

Challenging Environmental Conditions (CECs) for perception can be defined as when sensors perceive elements in the environment that will result in an erroneous RotE. If this data is used without precaution, the RotE will be incorrect. Environmental phenomena such as rain, dust and smoke are currently recognised CECs for UGVs [5]. For example, a cloud of dust can appear as a large solid obstacle when using laser range finders [6, 15]. Additionally, CECs can mean that different sensors may perceive the environment in different ways. For example, smoke is usually clearly perceived by visible cameras but it may not be seen by infrared (IR) cameras, and dust blocks waves emitted by laser range scanners but it is penetrated by radar waves [12]. Direct fusion of sensor data that reflects different aspects of the environment leads to an inconsistent, therefore inaccurate, RotE.

Recent work has proposed powerful methods to compensate for the presence of CECs such as fog in images [9, 10, 14]. These techniques propose a physical model that represents the challenging condition and allows for an improvement of the quality of the image after filtering, although this quality is not specifically evaluated. However, these methods are typically computationally expensive, do not explicitly detect the condition (it is assumed, a priori, that fog is present and needs to be compensated for), and require a complicated model of the CEC in question. To the best of our knowledge such a model is not available for other types of CECs such as airborne dust or smoke, and the detection problem remains largely unsolved. Even if these known conditions were solved, it is unrealistic for every situation to be anticipated and accounted for in a long-term autonomous vehicle where the spectrum of possible CECs is unbounded.

By analysing the sensor data directly prior to fusion and interpretation in the perception module (Fig. 1), we can ensure it is suitable to contribute to the RotE. First, the quality of data from individual sensors can be measured. Then redundant characteristics between multi-modal sensors can be used to check the consistency of sensor data prior to fusion to prevent inappropriate fusion and interpretation errors. Additionally, these methods can discriminate what data from which sensor may cause misinterpretation and allow the most appropriate sensor combination to be utilised. This paper demonstrates how this can be achieved using quality metrics for visual and infrared cameras mounted on a UGV and exposed to challenging environmental conditions. The concept is illustrated specifically with sensor data that is evaluated prior to use in a standard Scale Invariant Feature Transform (SIFT) [7] feature extraction technique.

The paper is organised as follows. Section 2 discusses previous work in this area and introduces Spatial Entropy as a metric to compare multi-modal sensors. Section 3 discusses the implementation for measuring the quality and consistency of image data from heterogeneous sensors. Section 4 describes the experiments and results and Section 5 proposes conclusions and directions for future work.

2 Background

Useful metrics can be applied to a range of sensor data and return a meaningful and comparable measurement that evaluates the quality of the data. Visual and infrared camera images are intrinsically different in that the former measures light intensity while the latter measures temperature and so cannot be compared in a straightforward way. Besides, perspective differences of the sensors make it difficult to compare sensor data directly. Quality metrics need to extract and evaluate commonalities in the data to provide some redundancy between heterogeneous sensors.

For robotic applications, as quality of perception is application-dependent, metrics should be analysed considering their relation with the performance of the application. Acquired images can be used in various crucial high level applications such as localisation, terrain modelling, motion detection, tracking or recognition/classification. Quality metrics should provide a prediction of how well an application will perform using the provided input data without necessarily having to perform the application.

The television and video industry has been developing quality metrics to attempt to quantify objectively how a human viewer would evaluate the quality of a video stream or image [17, 18]. While the metrics are generally developed to capture the errors caused by compression and transmission and are frequently tailored to the human vision system, there are many metrics that can still be relevant to the evaluation of UGV perception quality. For example, pictures that are colourful, well-lit, sharp with high contrasts are considered attractive to humans given the choice of dark, low contrast, blurry pictures. Most of these characteristics are also positive for perception applications on a UGV.

Previous work by the authors [2, 3] has identified Shannon Information and Spatial Information as useful metrics for evaluating the quality of sensor data that could be used across sensor modalities such as visual and infrared cameras. Shannon Information, which is a measure of entropy of the distribution of image intensities, can be used to identify images with very low amounts of information. However, CECs can have variable effects on the overall image luminosity distribution and this metric is very sensitive to variations in lighting conditions. Therefore, the evolution of Shannon Information could not reliably discriminate these situations. Spatial Information (SI) was introduced in the telecommunication industry as a metric to evaluate the quality of TV images [4]. To compute SI, edges in the image are first extracted using a Sobel filter. SI can then be defined as the first standard deviation of the distribution of intensities of the Sobel image. Assuming this distribution can be represented by a Gaussian distribution, SI can be interpreted as a representation of the amount of structure in an image. Since clouds of smoke or dust usually have no structure of their own and tend to obstruct the structure in the background of the scene, SI was shown to contribute significantly to the detection of such CECs [3]. Besides, these structure characteristics are known to be much less sensitive to lighting conditions. Furthermore, geometric structure is a common feature in both visual and infrared images, making SI a good image quality indicator, suitable to multimodal data comparison. However, SI failed to be interpretable as an overall image quality measurement whenever the Gaussian approximation was not valid, which was shown to happen quite often in the presence of CECs [2].

The authors recently introduced Spatial Entropy (SE) [2] as a quality metric to compensate for the identified flaws in SI. To compute SE, edges in the image are extracted using a Sobel filter with no threshold. SE is defined as a measure of the entropy of the distribution of intensities in the Sobel-filtered image. More specifically, the image I is composed of pixels with a range of intensity values. Similarly, the Sobel-filtered image is composed of pixels with a discrete set of possible intensity values ($i \in A_I$). If the probability of observing any particular intensity value, i, in the Sobel-filtered image is given by $P(i)$, SE is defined as [8]:

$$SE(I) = \sum_{i \in A_I} P(i) log_2 \frac{1}{P(i)} \tag{1}$$

and is expressed in average bits of information per observation. By using entropy, no assumption on the shape of the distribution is made for this metric. This method has been shown to provide a more reliable quality measurement for images than SI in particular in the presence of CECs [2].

3 Technical Approach

Quality metrics are dependent on the environment being perceived. In the absence of a reference, except in extreme situations when the overall level of quality is particularly poor (e.g. a black image), any particular level of quality metric does not immediately indicate a CEC. However, changes in the quality of the sensor data that are not explained by motion of the UGV or other dynamic features in the environment may be indicative of degrading quality due to a CEC. In this work, sensor data is evaluated prior to interpretation by applying lower bounds on the absolute values and lower bounds on the amount of change of quality metrics over time. The absolute values and variations in quality of data from multimodal sensors perceiving the same region of the environment can be used to help evaluate the quality of individual sensor data and to also check the consistency of sensor data prior to fusion. By using information about the quality of data from different sensors, poor quality data (potentially due to CECs) can be identified and the quality of one type of sensor data can be evaluated in comparison to another sensor.

Quality metrics provide three levels of evaluation as shown in Fig. 2. The first step considers the quality of the individual data by checking the absolute value of quality is within acceptable bounds. The second step evaluates the stability of the data quality over time. Sudden changes to the quality of data, in particular when quality suddenly drops, can indicate that further perception applications may fail. The final step involves ensuring that the quality of data between multiple sensors is consistent. Based on the results of this evaluation of sensor data, a decision of what data are appropriate for the RotE and what type of data, if any, should not be used in a regular fusion algorithm, nor for further interpretation.

Fig. 2 Filter sensor data prior to interpretation. Data from a visual and infrared camera are evaluated for quality, stability and then compared for consistency.

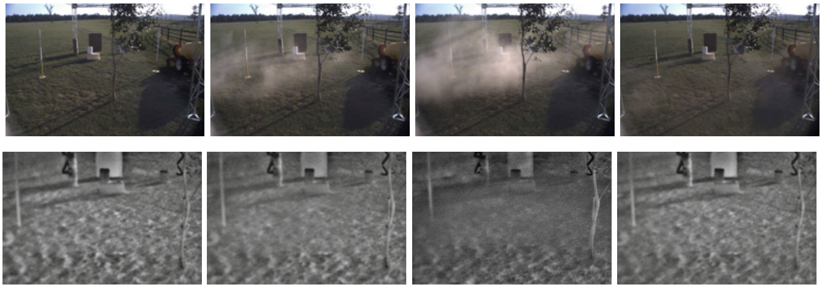
We set up a case study demonstration using Spatial Entropy to evaluate visual and infrared camera data that uses these three stages of interpretation and filtration. Further, this filter is used alongside a standard perception technique (SIFT) to demonstrate how filtering poor quality sensor data prior to application specific interpretation can improve perception quality and reliability.

4 Experiments and Results

This experimental study is based on public data sets presented in more details in [11]. These data were gathered by a UGV equipped with a wide variety of sensors, including laser range scanners, a radar, a colour camera and an infrared camera, in controlled and variable environmental conditions. These included known challenging environmental conditions such as the presence of airborne dust, smoke and rain. The study in this paper focuses on the effect of these conditions on visible and infrared camera images. Details of those cameras are the following:

- A Prosilica 1360 x 1024 resolution mono-CCD colour camera, acquiring images at a nominal frame rate of 10 images per second;
- A Raytheon 640 x 480 thermal infrared camera (images trimmed to 511 x 398 to remove black edges), with a spectral response range of 7-14μm and average frame rate of 12.5 images per second.

In order to assess the effect of challenging environmental conditions on sensor data, a reference environment was created and the stationary UGV was positioned so the sensors were perceiving this trial area. A variety of datasets were logged with dust, smoke and other conditions generated within the sensing area. Fig. 3 shows sample



(a) $t = 1s$ (b) $t = 10s$ (c) $t = 23s$ (d) $t = 35s$

Fig. 3 A few samples of reference images (top row: visual, bottom row: infrared). a) Clear conditions at $t = 1s$; b) Light dust cloud at $t = 10s$; c) Thick dust cloud at $t = 23s$; d) Very light dust covering most of image at $t = 35s$.

images from a dust dataset. In addition a number of different *dynamic data sets* were logged with the UGV moving in clear and challenging conditions. Specific information about the data sets used can be found in [11].

Overlapping areas of images from different cameras allow us to illustrate commonalities of sensor data from different sources of information about the same environment. Therefore, a smaller section of the visual camera images has been obtained by trimming the images to manually register with the field of view of the infrared camera. The resolution of the visual image has then been adjusted to match the resolution of the infrared image. A demonstration of the resulting pairs of images is shown in Fig. 4. Because the sensors are mounted in different physical positions on the vehicle, the positions of objects in the field of view do not necessarily match precisely. However, the information content of the two images is comparable. Thus, metrics computed on the trimmed images from both cameras are comparable.

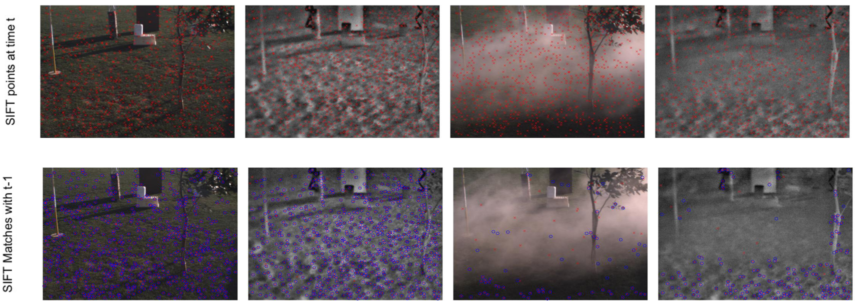
4.1 SIFT Feature Extraction and Matching for Visual and Infrared Cameras

In this paper, SIFT point [7] extraction and matching is used as an example of an interpretation technique illustrated in Fig. 1. In this system, the SIFT points themselves are considered as the representation of the environment (RotE). Visual odometry is an example of a higher level application that estimate motion by exploiting the SIFT points as features that are matched between images. If the number of features is small, if the points are concentrated in a small area of the image and/or if there are few correct matches between images, then the odometry solution will be poor [13]. A metric rating the quality of input images should match the level of quality of the application solution. In this case, quality metrics should predict when SIFT point extraction and/or matching will be sub-nominal and subsequently infer when visual odometry based on SIFT points is likely to be compromised.

In this work, software from VLFeat [16] was used to find SIFT points for the images in the data sets and match SIFT points between consecutive images. As a demonstration, the top row of Fig. 4 shows all the SIFT points (red crosses) that were found in clear conditions at $t = 1s$ and for thick dust at $t = 23s$. The bottom row of Fig. 4 shows the SIFT points that were found to match between the image (red crosses) and its preceding image (blue circles). A stationary UGV provides a ground truth for SIFT points because matched SIFT points should stay at the same location in images over time. In clear conditions (Fig. 4(a) and Fig. 4(b)), most SIFT points are observed to be paired on the image. However, in dusty conditions (Fig. 4(c) and Fig. 4(d)), there are less overall matched SIFT points and there are also many SIFT points that are found to match but are not in the same location in the image.

Fig. 5 shows the percentage of SIFT features that were matched in consecutive images and the average distance in pixels between the matched features (error) as dust and smoke were introduced to the environment. Note that in both of the data sets, the first few seconds (approximately 0-8 seconds for dust and 0-13 for smoke) have no dust or smoke and are therefore considered to show the reference for these static data sets.

In clear conditions, the visual camera data has more SIFT points, a higher percentage of features are matched and the overall error is lower than with the infrared camera which produces much noisier images. Therefore, it would be a preferable source of data for visual odometry. The appearance of dust in Fig. 5(a) significantly decreases the percentage of matches and additionally increases the error. Even if outliers are eliminated, the error of matching points increases, meaning that a final visual odometry solution will be less accurate than in clear conditions. However, in



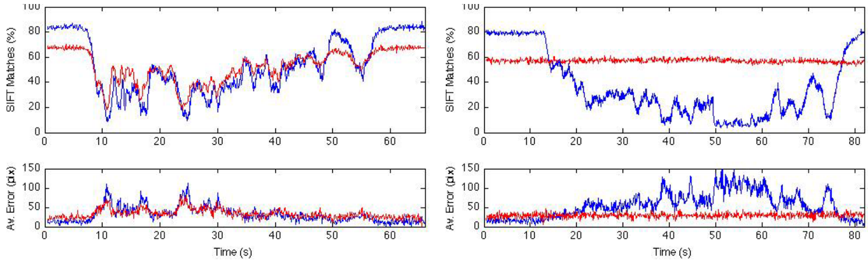
(a) Visual Image, Clear Conditions ($t = 1s$)

(b) Infrared Image, Clear Conditions ($t = 1s$)

(c) Visual Image, Thick Dust Conditions ($t = 23s$)

(d) Infrared Image, Thick Dust Conditions ($t = 23s$)

Fig. 4 Representative visual and IR images of the test environment for clear and dust conditions. The top row shows SIFT points (red crosses) identified in the image and the bottom row shows SIFT points that were found to match between the representative image (red crosses) and the preceding image (blue circles) in the data set.



(a) SIFT point evaluation in Dust Conditions (b) SIFT point evaluation in Smoke Conditions

Fig. 5 SIFT feature matching for visual (blue) and infrared (red) images perceiving the same test environment containing dust (left) and smoke (right). The overall percentage of matching points between consecutive images (top row) decreases while the error in matching (bottom row) increases in all cases except for smoke which does not affect the infrared images.

Fig. 5(b), smoke has a similar impact to dust in the visual camera images but does not affect infrared data. Therefore, by evaluating which type of sensor data is useful prior to interpretation, the best configuration of sensors to contribute to perception techniques such as fusion or visual odometry can be chosen leading to a better and more reliable perception solution.

It is not usually possible to evaluate the accuracy of SIFT matching points in normal operations of a moving UGV. By analysing the image data for quality prior to the SIFT algorithm (or other perception interpretation techniques), we can predict when the SIFT algorithm will fail by evaluating whether the data is suitable for further interpretation even before any interpration is done.

4.2 Spatial Entropy for a Stationary UGV

The acquired sensor data for the stationary UGV was evaluated using Spatial Entropy as a quality metric. The evolution of SE for both the dust and smoke datasets is shown in Fig. 6. The dashed green lines indicate the start and end of the presence of dust or smoke. Clear (nominal) conditions exist before and after dust and smoke were present in the environment. By gathering data for a variety of datasets in clear conditions, the nominal range of values and the typical variation of the SE metric for data obtained from a sensor can be observed. General thresholds can then be designed for the absolute value of SE and for the stability of SE for that sensor which can be applied in a larger range of environments.

The evolution of SE is again shown in (I) of Fig. 7 for dust (a) and smoke (b) datasets, this time with Higher ($SE_H = 3.3 Bits$) and Lower ($SE_L = 2.8 Bits$) thresholds (shown by the dotted lines). The stability of SE is shown in (II) for the visual camera (blue) and infrared camera (red). An image is rated between 0 and 1 for quality as shown in (III) of Fig. 7. A value of 1 acts as an alarm to indicate that the image is considered bad quality while a value of 0 indicates that the image is within

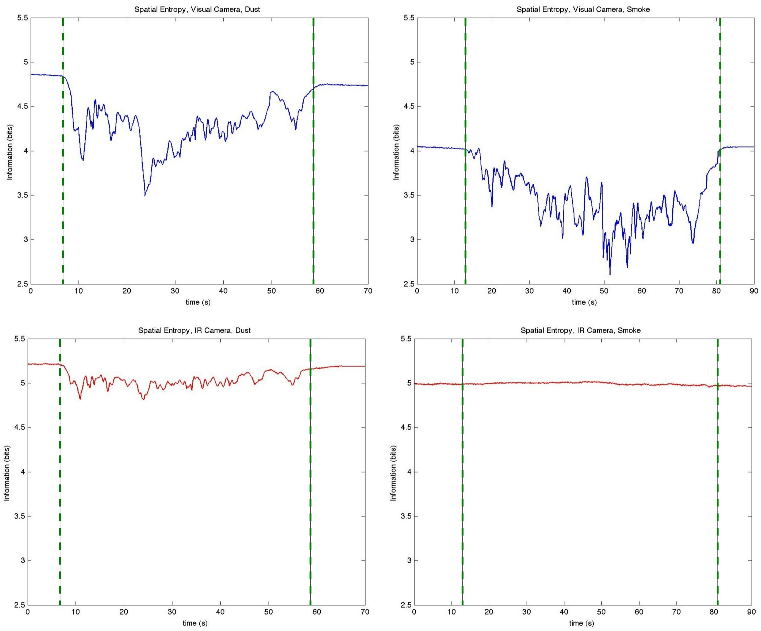
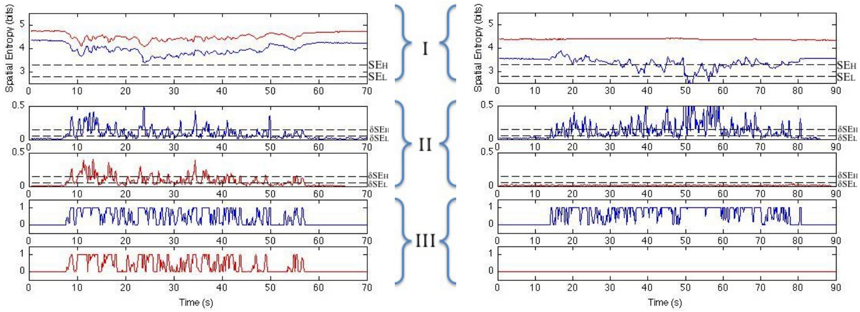


Fig. 6 Evolution of SE for *Dust* (left) and *Smoke* (right), Visual (top row) and IR Camera (bottom)

nominal limits and so is suitable for further interpretation. Although the alarm could potentially be used in a binary system, a value between 0 and 1 provides further discrimination of situations where the image is compromised but may still be useful.

As previously stated, any particular value for a quality metric, such as SE, does not indicate that the data is compromised except in extreme situations when the quality is very poor. A SE value greater than the threshold SE_H indicates that the image quality is within nominal limits and returns an output of 0 in the alarm indicator of quality. A value less than SE_H means that SE is outside this nominal state, while a value less than the SE_L threshold indicates an image is poor quality and should not be used for further perception. A value between SE_H and SE_L thresholds increases the alarm output linearly from 0 to 1. While the dust in this data set never causes the overall quality of the image to be definitively compromised, thick smoke conditions are found to significantly impact the quality of the image as the SE value drops below SE_H and then the SE_L threshold. The initial increase of SE in the presence of smoke is caused by a very thick, high contrast cloud on the periphery of the image which produces a strong edge from the Sobel filter and is therefore interpreted as more structure in the image. As the cloud moves across the image, the background is obscured by the generally unstructured smoke causing the eventual decrease in SE.

The stability of SE is calculated over a short time period (0.5 seconds - the reason for this time is explained in Section 4.3) by summing the difference of SE values



(a) SE evaluation in Dust Conditions (b) SE evaluation in Smoke Conditions

Fig. 7 Spatial Entropy evaluation for visual (blue) and infrared (red) cameras perceiving the same test environment containing dust (a) and smoke (b). A low value of SE or a high variation of SE over time indicates a challenging environmental condition for that sensor. (I) shows the evolution of SE, (II) show the measured stability of SE and (III) demonstrate an alarm that indicates a drop in quality of the data.

from five consecutive images from the mean SE value and is shown in (II) of Fig. 7. Thresholds δSE_H and δSE_L are based on nominal behaviour in clear conditions found in a variety of data sets. A stability value less than the δSE_L indicates that the image is behaving within nominal limits and returns an output of 0 in the alarm indicator of quality. A value greater than δSE_L means that SE is behaving outside this nominal state, while a value greater than the δSE_H threshold indicates an image is poor quality and should not be used for further perception. A value between SE_H and SE_L thresholds increases the alarm output linearly from 0 to 1. Both dust and smoke cause significant changes in the stability of the SE metric which are captured by the threshold.

By combining the output for both the absolute value and the stability of the metric, the quality of both the infrared and visual sensor data is found to be compromised during the dust dataset from approximately 8 to 57 seconds (Fig. 7(a)). However, there are no significant inconsistencies that are observed between the two sensors. i.e. the quality degrades similarly for both sensor data. For the case of smoke (Fig. 7(b)), the visual camera data is affected from approximately 15 to 80 seconds while smoke has no effect on the infrared data. Therefore, during this time period, infrared data should be preferred. This result aligns with the SIFT matching observed in Fig. 5(b) where the infrared error is less than the visual error during these times. Additionally, since the responses of infrared and visual cameras are not consistent in this situation, they should not be fused.

It is not always possible to evaluate the output of perception algorithms for accuracy, besides, perception algorithms can change between platforms. Fig. 5 demonstrated how SIFT point matching was compromised (fewer and poorer matches) when dust and smoke were in the environment. However, this reference evaluation required a stationary UGV to ensure that SIFT points from the environment matched over time. In normal operations, it is not as trivial to discern that the output of a perception technique is erroneous. We have shown in Fig. 7 that we can quickly and effectively predict whether the perception output will be reliable by monitoring the quality of the data that is an input to the perception module. The integrity and reliability of a system can be enhanced by using this analysis prior to any perception algorithm. SIFT is a computationally expensive operation which could be reduced by filtering out data that is considered compromised.

It should be noted that using this same method on three stationary datasets in clear conditions, each containing 900 images per type of camera, there was only one false positive. Additionally, the method was tested on a stationary data set with a dynamic component (a person walking around the environment) shown in Fig. 8. In this case, there were two false alarms in the infrared data caused by the sudden introduction of a very high contrast object which rapidly increased the measured structure in the image. This is similar to the case observed with the increase of SE seen in the smoke dataset, however, unlike smoke, once the initial sudden change in SE was captured, the continued movement of the object through the environment does not continue to cause an alarm in SE. This is because dust and smoke generally have little structure of their own so will have a negative impact on the amount of structure in an image. However, solid dynamic objects do not effect SE as abruptly because they add structure of their own while also obscuring the background environment.

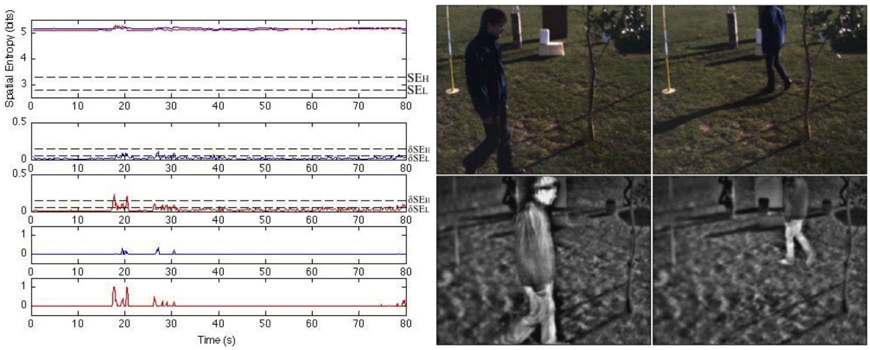


Fig. 8 Left: Spatial Entropy evaluation for visual (blue) and infrared (red) cameras perceiving the same test environment with a dynamic component (person walking) in the test environment. Right: Representative images from visual (top) and infrared (bottom) cameras of dynamic object in test environment at 20 (left) and 30 (right) seconds.

4.3 *Spatial Entropy for a Moving UGV*

Metrics measuring the quality of images will naturally be affected as the region of the environment that is perceived changes. Therefore it is important to maintain the background environment so that changes captured by the metrics reflect dynamic features in the environment (such as dust and smoke) and are not due to the motion of the UGV.

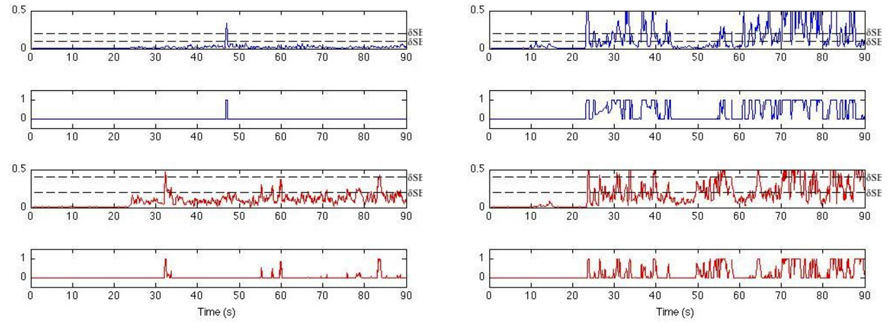
Common methods to compensate for the movement of the sensor consist of matching information contained in the data (e.g. optical flow [1] or registration [19]) to estimate an affine transformation that approximates the change from one image to the next, acquired after motion. However, in the context of identifying challenging conditions for perception, relying directly on sensor data is inappropriate as the purpose is precisely to evaluate the quality of that data. While any other sensor suite may be employed, the sensing data used to find the motion of the cameras must be fully independent of the cameras.

In this work, regions of images that overlap are found using a simple method to compensate for the motion of the vehicle. The motion of the camera sensor is calculated using navigation data (a cm-accuracy dGPS/INS system) available on the platform [11]. An affine transformation is then applied to approximate the motion between successive frame acquisitions from the camera such that a common area in the second image is projected into the first image. Once this transformation has been applied, the images can be cropped so that consecutive images approximately register.

The stability of SE is measured from a set of consecutive images in a dataset. However, if the region of the environment perceived in the first image is no longer perceived in the final image of the set, then there is no overlapping data to measure. The frame rate and field of view of the sensor and the dynamics of the vehicle will contribute to how many images can be used while guaranteeing an overlap of environmental data. For the dynamics of this particular system, five consecutive images (0.5 seconds of data) were found to ensure an overlap while maintaining a large enough set of images to measure the stability of SE. Thus, using the affine transformation described, at each timestep, the four previous images are transformed to overlap with the current image and then are trimmed to contain a common area.

The SE for datasets of the moving UGV in clear and challenging conditions are then evaluated using the same strategy outlined for a static vehicle but by evaluating the metric only on the common overlapping regions of the images. For the moving UGV the absolute value of SE indicates quality of the image. Therefore, the same thresholds used for the static case (SE_L and SE_H) are maintained for the overall SE measurement of the image. Since the affine transformation can not produce a perfectly overlapping image for moving sensor data, there are small changes in the SE from image to image despite the compensation for motion. Based on observations of datasets from a moving UGV in clear conditions, higher thresholds for stability (δSE_L and δSE_H) than those for the static case were found to identify the nominal behaviour of the stability.

Fig.9(a) shows the stability of SE for visual and infrared images on a moving UGV in clear conditions. In this data set, the UGV is initially stationary and then begins to move through the environment at 22 seconds. By compensating for the motion of the UGV, the variation of SE between consecutive images is maintained at a consistently low level as the UGV moves. In the visual data, there is only one false alarm produced at 47 seconds. It is seen that the stability of the infrared camera is affected more than the visual camera by the motion of the UGV. This is due



(a) Visual (blue) and infrared (red) image Spatial Entropy evaluation for a moving UGV in Clear Conditions

(b) Visual (blue) and infrared (red) image Spatial Entropy evaluation for a moving UGV in Dusty Conditions

Fig. 9 Spatial Entropy evaluation for visual (blue) and infrared (red) sensor data perceiving the same test environment from a moving UGV for clear (a) and dust (b) conditions. Rows 1 and 3 show the stability of SE for matching images. Rows 2 and 4 show an alarm indicating if the sensor data is judged to be poor quality for further perception applications.

to higher noise in infrared sensor data and a poor calibration leading to less ideal overlaps between some consecutive images.

Fig.9(b) shows the stability of SE for visual and infrared images on a moving UGV in dusty conditions. In this data set, the UGV is initially stationary. A very light dust cloud appears between 10 and 15 seconds. Although this dust cloud has an observable effect on the stability of the SE, it is not above the new thresholds and so is not identified in the alarm. A thicker dust cloud obscures the background from 22 until 43 seconds and the UGV begins to move through the environment at 23 seconds. There is no dust from 43-55s, then thick but variable clouds of dust are present until very thick dust is observed from 70 seconds and for the rest of the data set. The alarm identifies that the visual sensor data is compromised during the periods known to have dust affecting the environment. However, due to the higher level of noise due to motion in the infrared data, only high density dust clouds (70+ seconds) are identified as definitively causing the quality of both sensors to be compromised.

5 Conclusion

This paper has presented the application of a quality metric that is applied to data from heterogeneous imaging sensors to measure the appropriateness of the data prior to interpretation in a perceptual system. Spatial Entropy (SE) was introduced as a metric that provides comparable output for heterogeneous sensor input, in this case, visual and infrared images of the same test environment. Spatial Entropy is shown to decrease considerably in the presence of known challenging environmental conditions such as dust and smoke for infrared and visual sensor data, whereas dynamic features in the environment such as a person walking have a less pronounced effect. By setting thresholds for the nominal value and stability of SE, loss of quality due to challenging environmental conditions can be detected in infrared and visual images. The concept is illustrated specifically with dust and smoke conditions, and sensor data is evaluated prior to its use in a standard SIFT feature extraction technique. This paper has shown that this method can be used to identify poor image data. This can be used to filter inappropriate data (such as observed in dust) or to select the most appropriate sensor data. For example, to determine when smoke compromises visual image data but has no effect on the infrared images. The method was extended to the case of a moving UGV by using navigation data that is measured from other sensor sources. Overlapping images that perceive the same region of the environment were approximated by performing an affine transformation based on the motion of the vehicle.

Future work will involve studying how to take appropriate decisions in the perception system, based on the interpretation of the output alarm. For example, image data that is considered of insufficient quality may simply be removed to avoid jeopardising the interpretation that will be made by the perception system. Another option is to select the most appropriate sensor combination in response to the environmental conditions. More work is required to study the effect dynamic objects in the environment have on the SE of an image, in particular, the case when an object first appears at the edge of an image.

Acknowledgements. This work was supported in part by the Centre for Intelligent Mobile Systems (CIMS), funded by BAE Systems as part of an ongoing partnership with the University of Sydney, and the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

References

1. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. In: International Journal of Computer Vision (1994)
2. Brunner, C.J., Peynot, T.: Visual metrics for the evaluation of sensor data quality in outdoor perception. In: Performance Metrics for Intelligent Systems (PerMIS) Workshop (2010)
3. Brunner, C.J., Peynot, T., Underwood, J.: Towards discrimination of challenging conditions for UGVs with visual and infrared sensors. In: Australasian Conference on Robotics and Automation, ACRA (2009)
4. International Telecommunication Union (ITU-T). Subjective video quality assessment methods for multimedia applications (1999)
5. Kelly, et al.: Toward reliable off road autonomous vehicles operating in challenging environments. The International Journal of Robotics Research 25, 449–483 (2006)
6. Leonard, et al.: A perception driven autonomous urban vehicle. Journal of Field Robotics 25(10), 727–774 (2008)
7. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision (2004)
8. Mackay, D.: Information Theory, Inference & Learning Algorithms. Cambridge University Press (2007)

9. Narasimhan, S.G., Nayar, S.K.: Shedding light on the weather. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 665–672 (2003)
10. Nayar, S.K., Narasimhan, S.G.: Vision in bad weather. In: Proceedings of the Seventh International Conference on Computer Vision, vol. 2, pp. 820–827 (1999)
11. Peynot, T., Scheding, S., Terho, S.: The Marulan Data Sets: Multi-Sensor Perception in Natural Environment with Challenging Conditions. International Journal of Robotics Research 29(13), 1602–1607 (2010)
12. Peynot, T., Underwood, J., Scheding, S.: Towards reliable perception for unmanned ground vehicles in challenging conditions. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2009)
13. Scaramuzza, D., Fraundorfer, F., Siegwart, R.: Real-Time Monocular Visual Odometry for On-Road Vehicles with 1-Point RANSAC. In: IEEE International Conference on Robotics and Automation (2009)
14. Tan, R.T.: Visibility in bad weather from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
15. Urmson, et al.: Autonomous driving in urban environments: Boss and the urban challenge. Journal of Field Robotics 25(8), 425–466 (2008)
16. Vedaldi, A., Fulkerson, B.: VLFeat: An Open and Portable Library of Computer Vision Algorithms (2008), `http://www.vlfeat.org/`
17. Wang, Z., Bovik, A.C.: Modern Image Quality Assessment. Morgan & Claypool (2006)
18. Winkler, S.: Digital Video Quality: Vision Models and Metrics. John Wiley & Sons Ltd, New Dehli (2005)
19. Wyawahare, M.V., Patil, P.M., Abhyankar, H.K.: Image registration techniques: An overview. International Journal of Signal Processing, Image Processing and Pattern Recognition (2009)