Image Rejection and Match Verification to Improve Surface-Based Localization

James Mount, Michael Milford
Queensland University of Technology, Australia
j.mount, michael.milford@qut.edu.au

Abstract

The capability to localize is paramount for many mobile robots and autonomous vehicles. Key attributes of localization systems include reliability, accuracy, low-latency, minimal cost and robustness to variation in environmental conditions. Typical approaches by self-driving car systems incorporate some combination of LiDAR, camera, radar and GPS sensing technologies; which are all suboptimal with respect to one or more of the aforementioned key attributes. This paper presents new research that translates previous work on surface-based positioning systems, which have appealing latency and accuracy properties, to road networks in the context of autonomous car positioning. To achieve the required performance and robustness to appearance change caused by phenomena such as day-night cycles, we develop two new data-driven statistical and learning-based techniques. One performs self-evaluation with regards to the suitability of the current camera query image, while the other retrospectively determines the quality of the query-reference image comparison outcome. Multiple new road surface datasets spanning day and night cycles were utilized to evaluate the system. Results show that both contributions significantly improve place recognition performance, decreasing the median estimated image position error from $24m$ to $0.26m$. The strengths and limitations of our approach and how it could complement current positioning techniques to improve vehicle positioning capability are discussed.

1 INTRODUCTION

In recent years the focus on autonomous vehicles has rapidly increased in both the research community and the public sector [Levinson and Thrun, 2010]. Normal consumer cars are now entering the market with ever increasing assistive technologies, such as BMW's ConnectedDrive, and Tesla's Autopilot [Guizzo and Deyle, 2012; Luettel *et al.*, 2012]. One of the major hurdles to overcome before we see fully autonomous vehicles for commercial sale on public roads is the development of systems that are reliable and robust to the variety of conditions and environments faced by vehicles today.

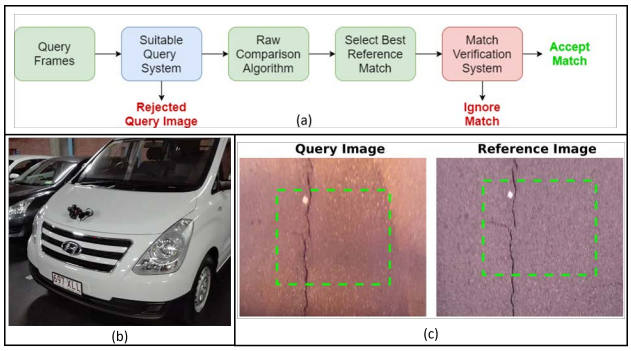

Figure 1: presents an overview of the contents of the paper. (a) presents the system architecture of the approach taken, (b) shows the Sony A7s camera mounted onto the Hyundai iLoad and (c) provides an example of a successful match between night and day road surface images.

An underlying requirement for many autonomous car systems, and indeed most autonomous mobile platforms, is vehicle localization. Typically the key requirements of such an approach are accuracy, low latency and reliability, with cost also being another driver for mass-produced vehicles.

This paper extends on the techniques developed for surface-based positioning systems for ship hull inspection and warehouse navigation. These systems have the key attributes of being accurate and having low latency but typically lack robustness to appearance variation. We extend this work to the domain of road networks and autonomous car positioning and improve the robustness

to condition variances, mainly day-night cycles, Figure 1.

To improve the robustness of the previous surface-based positioning systems with respect to appearance variation we develop two new data-driven statistical and learning-based approaches. The first method developed allows the rejection of query images that are likely to result in a false-positive match. The second system performs an evaluation on the query-reference match outcome of the comparison algorithm and determines its probability of being a true-positive result. These systems were developed using an exhaustive analysis of various statistics. These statistics were computed from the query image and the results of the comparison algorithm.

We show that these two new systems increase the robustness of the raw comparison technique to appearance changes caused by phenomena such as diurnal cycles. This resulted in a decrease in the median image position estimate error from $24m$ to $0.26m$. We have evaluated the two novel systems on several new road-surface datasets which span day and night cycles in a variety of built conditions, and are made available to the public.

The paper proceeds as follows. In Section 2 a literature review on visual surface recognition, vehicle localization and place recognition is provided. In Section 3 an overview of the methodology undertaken to develop the two new systems is discussed and the algorithms employed presented. Section 4 summarizes the equipment used and details the datasets gathered, while Section 5 presents the results. Finally Section 6 discusses the outcomes and the future work.

2 RELATED WORK

Current state-of-the-art autonomous vehicles, such as Carnegie Mellon University's (CMU) Junior or Stanford's Stanley combine traditional inertial navigation systems (INS) with range-based localization techniques to estimate the vehicle's pose in an environment [Levinson and Thrun, 2010; Thrun, 2010]. While these state-of-the-art robotic vehicles have optical cameras onboard, they do not use vision based techniques to augment their relative pose estimate, although this has been shown to improve localization especially in GPS denied environments and extended blackouts [Miller *et al.*, 2011].

There are a variety of vision based localization techniques including feature, pixel and object based methods. Feature based methods, such as SIFT, SURF, and ORB, have been shown to allow for vision based localization [Se *et al.*, 2001; Murillo *et al.*, 2007; Mur-Artal *et al.*, 2015]. However, object based localization is becoming an alternative to abstract feature based methods, and the work performed in [Vasudevan *et al.*, 2007] and [Salas-Moreno *et al.*, 2013] highlights this ability.

Pixel based techniques compare previously generated image maps with current sensor data and utilize a variety of algorithms such as sum of squared differences and cross-correlation normalization [Kelly, 2000; Kelly *et al.*, 2007; Milford *et al.*, 2004; Davison *et al.*, 2007]. Furthermore pixel based localization has been demonstrated to work in a variety of conditions [Mount and Milford, 2016; Kelly, 2000; Kelly *et al.*, 2007; Milford and Wyeth, 2012].

Kelly et al. thoroughly demonstrated that floor localization using pixel techniques for autonomous vehicles is feasible in warehouse environments with controlled lighting using a monocular camera [Kelly, 2000; Kelly *et al.*, 2007]. However, the work cannot be directly transferred to the localization of autonomous cars because,

- Outdoor environments suffer from highly dynamic weather and lighting conditions, and
- Road networks have a decreased persistent appearance compared to warehouse floors.

The autonomous vehicle localization system in [Miller *et al.*, 2011] provides additional evidence that a vision-based system can be utilized within autonomous cars to further improve current localization techniques. However, the system outlined by Miller et al. employed two cameras and was specifically aimed at urban environments, and hence cannot necessarily be extended to non-urban scenarios.

Other fields that investigate and utilize visual surface techniques in various applications include visual odometry, ship hull navigation and inspection, as well as power line and pole monitoring.

Visual odometry employs varying computer vision techniques, such as structure for motion and optical flow. There have been several papers discussing the performance and applicability of visual odometry in car-like vehicles and its ability to supplement localization [Campbell *et al.*, 2005; Johnson *et al.*, 2008; Parra *et al.*, 2010; Dille *et al.*, 2010; Nourani-Vatani and Borges, 2011; Gonzalez *et al.*, 2013].

The research presented on underwater visual ship hull inspection and navigation further demonstrates that vision based surface localization is feasible even in challenging conditions [Kim and Eustice, 2009; Hover *et al.*, 2012; Kim and Eustice, 2013; Ozog and Eustice, 2014]. However, the algorithms deployed were operating on slow moving robots, hence they are not directly exportable to fast moving autonomous ground vehicles. There is also ongoing work on the visual inspection of electrical and telecommunication infrastructure, such as the work presented in [Sa *et al.*, 2015; Whitworth *et al.*, 2001].

Robust place recognition or vision-based localization

systems often utilize a post comparison verification technique to identify if the outcome is likely to be a true-positive. These techniques include geometric verification in feature-based approaches, such as that employed in FAB-MAP 2.0 [Cummins and Newman, 2011], and patch verification techniques in pixel-based systems [Milford *et al.*, 2014]. While feature-based comparison systems implicitly reject images with low feature densities, hence increasing robustness to false-positive matches, there appears to be no pixel-based comparison technique, to the authors' knowledge, that rejects a query image unsuitable for matching. Limitations of rejecting a query image prior to performing a pixel-based comparison algorithm, will be discussed in Section 6.

3 APPROACH AND SYSTEM DEVELOPMENT

This section discusses the three main sub-systems developed and utilized within this paper, which are:

1. The algorithm to compare a query image to the reference dataset,
2. The development of a system to classify the suitability of query image to perform localization, and
3. The development of a match verification system to determine if a reference match is likely to be a true-positive.

Figure 2 shows an abstracted schematic of the approach taken and how the three main sub-systems connect. Our contributions are mainly within sub-system 2 and 3, however we believe the core algorithm utilized as part of the first sub-system, normalized cross-correlation, has not been demonstrated within a localization task where the query and reference images are drastically perceptually different.

In order to compare the performance of various systems and thresholds during development we were required to reduce precision-recall curves down to a single metric. We determined that the most suitable metric was to compute the area under the precision-recall curve up until a given recall. We consider this a suitable metric, as using the total area under the curve causes a bias towards systems with high-recall even if they have a lower precision at low-recall. Given one of the targeted domains of our system is self-driving cars, where a localization technique should estimate the position of a car every 1-5 metres, the area under the curve up until a recall of 20% was considered appropriate. At 20% recall, if 100% precision was achieved, the system would successfully localize 1 in every 5 frames on average. Therefore a system with a frame rate of $50Hz$ was traveling at $100km/hr$, a successful localization would occur every $3m$. We denote this metric as the *20% Recall Curve Area*.

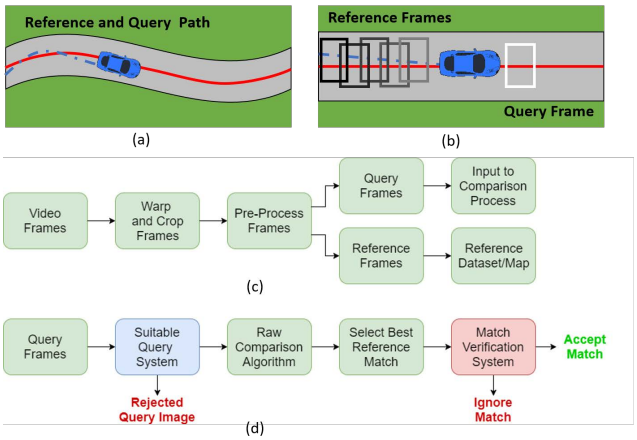

Figure 2: represents an abstracted schematic of the image data collection and the system architecture. (a) demonstrates the reference and query path taken, while (b) is a "zoomed" portion of (a) highlighting the reference and current query frames. (c) shows the pipeline for the pre-processing steps to generate the image datasets. (d) presents the flow diagram of the system architecture with the raw comparison, suitable query and match verification components in green, blue and red respectively.

3.1 Comparison Algorithm

The query frames were compared to all reference frames using normalized cross-correlation (NCC), which is given by,

$$R(u,v) = \frac{\sum_{x,y}(T(x,y)-\bar{T})(I(u+x,v+y)-\bar{I})}{\sqrt{\sum_{x,y}(T(x,y)-\bar{T})^2 \sum_{x,y}(I(u+x,v+y)-\bar{I})^2}} \quad (1)$$

where T and I are the query template region and reference image respectively, and R is the resultant correlation matrix for the current query-reference frame comparison. The maximum value within R is stored in a comparison matrix, C, for each query-reference comparison. A match is reported if the reference frame with the maximum comparison score for the current query image is above a threshold (K_t);

$$M(i) = \begin{cases} j, & \text{if } max(C_i) < K_t \\ nomatch, & \text{otherwise} \end{cases} \quad (2)$$

where M is a vector which contains the matched reference frames, C_i is all the reference scores for the current query and j denotes the index for the reference with the maximum comparison score. For our experiments we assumed that the orientation of images remained constant, as it has been demonstrated that rotation invariant template matching is plausible [Ullah and Kaneko, 2004] and

that through the use of additional sensors, such as a compass, the orientation of all images could be normalized.

3.2 Image Rejection System Approach and Development

The goal of this system was to determine if a query image was suitable for localization. Two classification methods were developed and compared, these being: linear support vector machines (SVM) and basic threshold classifiers. The processed query image (QI) and autocorrelation (AC) matrix were used to derive statistical data, including the mean, median of absolute deviation (MAD), entropy, and standard deviation. This statistical data was then employed by each of these classification methods and their performance evaluated by sweeping through a range of thresholds from 0 to 1, thereby producing a precision-recall curve for each threshold.

The linear SVM classifiers were trained using a supervised approach with the manually aligned ground truth query-reference frame alignment and the raw comparison results, $M(i)$, from *Comparison 2*, see Table 1. Each SVM classifier predicted the probability that a query image was suitable for localization purposes.

The initial analysis involved deploying each individual statistic for use in both the linear SVM and the basic threshold classifiers. The results of this analysis are presented in Figure 3a and b. The results indicate that all linear SVMs produced on-par or subliminal results to the baseline performance of the raw comparison algorithm. However, the standard deviation, entropy and MAD for the basic threshold classifier resulted in an increase in the 20% Recall Curve Area.

A second analysis was then conducted using the three statistics with an improved performance, standard deviation, entropy and MAD. A combination of these three statistics in both linear SVM and basic threshold classifiers were evaluated. Although the individual SVMs for these parameters did not have an increased performance relative to the baseline, it was postulated that a SVM trained on a combination of these statistics could produce an improved performance. The evaluation of this second analysis showed that no SVM or threshold classification system which employed a combination of these three statistics resulted in an improved performance relative to using a single statistic and a basic threshold classifier. Hence, it was determined that the entropy of the query image in conjunction with a basic threshold classifier would be used to establish the suitability of a query image for use in the raw comparison algorithm.

Entropy was selected due to having the highest overall performance increase, for the largest range of thresholds. Two example images that were rejected and passed by the image rejection system are presented in 3c and d respectively.

3.3 Match Verification System Approach

The goal of the match verification system was to determine if the selected reference match was likely to be a true-positive. We hypothesized that this could be done using the resultant correlation matrix (CM) between a query and its reference match. It was observed that the maximum comparison score was not always the true reference match and when inspecting the resultant correlation matrices of false-positive reference matches there was not always a clearly defined "hot-spot", Figure 4e. Typically the resultant correlation matrices for the ground truth reference image or for true-positive query-reference matches had a singular global maxima, Figure 4d. We developed and evaluated two classification methods: linear SVMs and basic threshold classifiers. The statistics of the resultant correlation matrix were used as parameters for each classification system and were compared against a range of thresholds between 0 and 1. The statistics evaluated included the mean, MAD, entropy, standard deviation and maximum.

The linear SVM classifiers were trained in the same way as the approach taken in the previous section. However, when the raw comparison algorithm reported a false-positive, the statistics of the resultant matrix for the ground truth reference match were appended to the the training data. Each trained SVM predicted the probability of how likely a reported match was to be a true-positive.

The initial analysis involved employing each individual statistic for use in both a linear SVM and a basic threshold classifier. The results of this analysis are presented in 4a and b. The results demonstrated that several linear SVMs, such as those for the standard deviation and entropy, had an increased performance over the baseline. However, no basic threshold classification system achieved a performance greater than the baseline.

A second analysis was conducted with the statistics which proved to have an increased performance in the initial linear SVM tests (mean, standard-deviation, entropy and MAD). This involved employing combinations of these statistics in linear SVMs. Figure 4c shows the results of this analysis. The two greatest performers are the standard deviation and MAD linear SVM as well as the mean, standard deviation and MAD linear SVM. It was decided to utilize the linear SVM with predictor variables of standard deviation and MAD as the match verification system. This system was chosen over the mean, standard deviation and MAD system as it had a more consistent performance and the decline in performance within the higher thresholds range was more gradual. Examples of successfully kept and rejected, as well as unsuccessfully kept, resultant matrices are presented in Figure 4d-f.

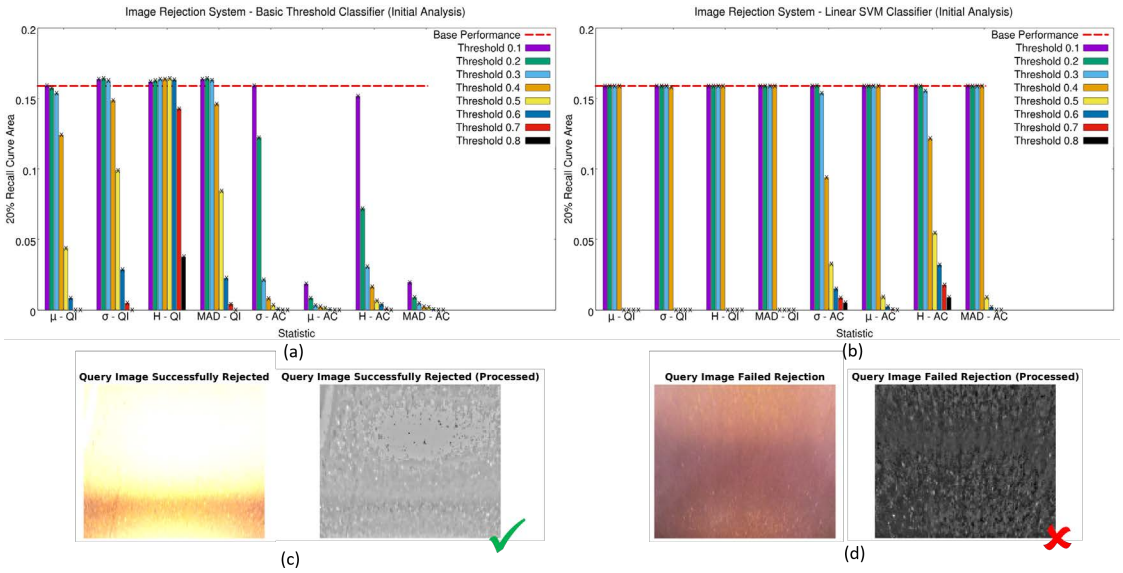

Figure 3: analysis of potential query image rejections systems utilizing (a) basic threshold classifiers, and (b) linear SVMs. The 20% Recall Curve Area metric across a range of thresholds was used to evaluate the performance of each system relative to the baseline of the raw comparison algorithm (red dashed line). The value used for the threshold is in the legend name. (c) and (d) show an example of a rejected and passed image by the suitable query system respectively. Symbols: μ - mean, σ - standard deviation, H - entropy, MAD - median of absolute deviation, QI - query image, AC - auto-correlation.

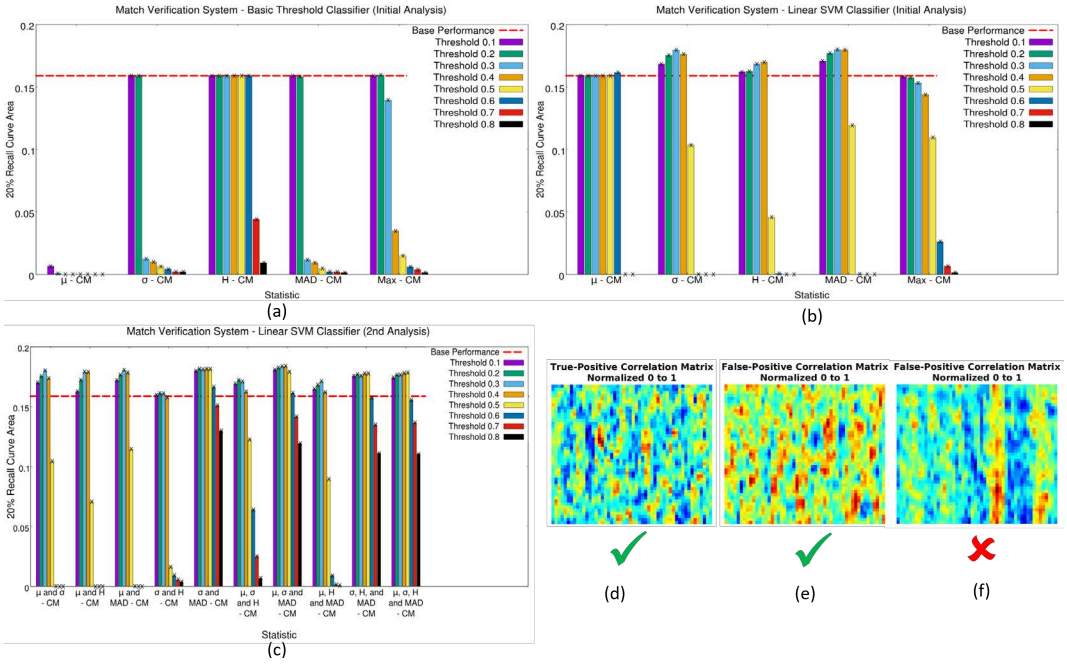

Figure 4: the initial analysis of the potential match verification systems utilizing (a) basic threshold classifiers, and (b) linear SVMs. The second analysis (c) utilized combinations of promising predictor statistics with linear SVMs. The 20% Recall Curve Area across a range of thresholds was used as the performance metric to evaluate the potential match verification systems. The red line in (a)-(c) show the baseline performance of the raw comparison algorithm. (d)-(e) show examples of query-reference resultant matrices for a true-positive and two false-positive matches respectively. (d) and (e) were successfully kept and rejected by the match verification system, while the second false-positive (f) was unsuccessfully kept. Symbols: CM - correlation matrix.

4 EXPERIMENTAL SETUP

The following section describes the experimental setup, details the dataset acquisition and generation process as well as the key parameter values. All the processes were performed on a standard Linux 64-bit desktop machine running MATLAB 2016b and custom OpenCV C++ software.

4.1 Hardware and Equipment

To acquire the image datasets a consumer grade camera, the Sony A7s, with a standard lens was mounted to the bonnet of a Hyundai iLoad van, Figure 1b. The Sony camera is a full-frame DSLR with superlative pixel wells, enabling images to be gathered in environments with extremely low lighting conditions.

4.2 Image Dataset Acquisition and Ground Truth

All image datasets were gathered by taking video using the Sony DSLR camera, with reference image sets been collected during the day and query sets been collected during the day and night. The video frames were then cropped so that they started and ended viewing the same area, all video frames were then extracted to create the image datasets.

To ground truth a query set to a reference set, approximately every 30th query image was manually aligned to a reference image, at the image level. Linear interpolation, with the assumption of linear speed, was then utilized to align those query frames that were not manually aligned. Additionally for one comparison, *Comparison 3*, we manually ground truth the query frames relative to the reference frames at the pixel level. To do this we,

1. Aligned the overlapping reference frames relative to each other (there was overlap in all frames),
2. Using the geometric pixel scale and the pixel shift between reference frames computed a ground truth position,
3. Aligned approximately every 30th query frame to a reference frame at the pixel level allowing the position to be computed via the geometric pixel scale, and
4. Using the assumption of linear speed, linear interpolation was applied to find the position of all other query frames.

4.3 Image Set Pre-Processing

Prior to performing any image comparisons, all images were warped to achieve a bird's eye view of the road surface. The images were then histogram equalized, cropped, resolution reduced and patch normalized, Figure 5.

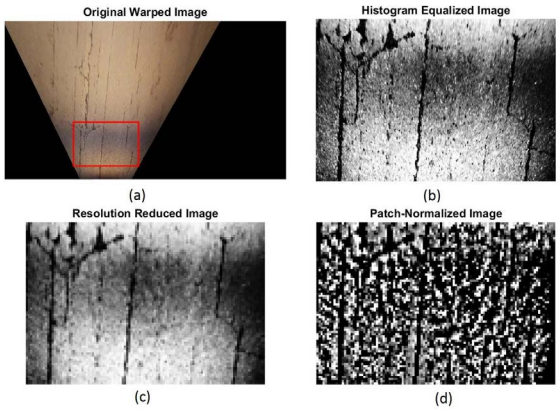

Figure 5: presents an example of the image pre-processing steps applied to a night time image (a). The pre-processing steps are, in order, converting to grayscale and histogram equalization (b), resolution reduction (c) and finally patch normalization (d). The red rectangle in (a) shows the cropped area.

4.4 Image Datasets

Six datasets of approximately 800m long from two different road surfaces were captured and evaluated throughout this experiment. Three datasets were collected during the day and the three night datasets were collected approximately 10 hours later. The paths captured contained a variety of elements, including pedestrian crossings and large surface cracks in a variety of lighting conditions. Figure 6 illustrates the area captured and Table 1 summarizes the datasets and the four comparisons made.

Table 1: The image datasets employed throughout the experiment and the comparisons they were used in. Comparison 2 and 3 are used as the training and validation datasets respectively.

Dataset Name	Dataset Area	No. of Frames	Used in Comparison
West Run 1 Day	Richmond Road West	2196	1,2,3 (Ref)
West Run 2 Day	Richmond Road West	2209	1 (Query)
West Run 1 Night	Richmond Road West	2192	2 (Query)
West Run 2 Night	Richmond Road West	2162	3 (Query)
East Run Day	Richmond Road East	2281	4 (Ref)
East Run Night	Richmond Road East	2283	4 (Query)

4.5 Key Parameters

The key parameters used in the pre-processing and raw comparison algorithm are given in Table 2. These parameters were determined empirically over a range of development datasets.

5 EVALUATION AND RESULTS

5.1 System Performance

Figure 7a-d show the precision-recall curve for each of the four systems for several datasets. The day-day com-

Figure 6: the two routes traversed along Richmond Road, Morningside, Brisbane for dataset generation. The orange and blue line indicates traverses in the east and west directions respectively.

Table 2: Key Parameter List for Pre-Processing and Raw Algorithm

Parameter	Value	Description
I_x, I_y	$112, 76$	Image Matching Resolution
P_{size}	5	Patch Normalization Radius
T_x, T_y	$51, 35$	Query Template Size for NCC
K_f	± 5	Frame Match Tolerance

parison performance of the raw comparison algorithm is outstanding, but day-night comparison result in a major performance reduction. The suitable query, match verification and the combined systems all have an improved performance relative to the raw comparison algorithm. However, they do slightly decrease the performance of day-day comparisons. This decrease in performance during day-day comparisons is of little concern as the system could be turned off during the day using temporal information. As shown by Figure 7d and e the systems are transferable to new data.

5.2 Metric Localization Results

Figure 8 and Table 3 demonstrates that if the true-positive match is reported by the system then accurate metric localization is still possible under appearance variation when using NCC. Figure 8a and the first two rows in Table 3 show the distance error between the ground truth and estimated positions for the query images where the system reported a match. The major observation is that the median distance error of the combined system is significantly reduced from 23.9m to 0.264m as our system greatly reduces the number of false-positives reported. The estimated query positions were determined by,

$$P_q = P_r + g * s \tag{3}$$

where P_q is the estimated position for the current query frame, P_r is the ground truth position for the matched reference frame, and g and s are the geometric scale of a single pixel and the row and column translational shifts respectively. The translational shifts are the distance from the centre to the peak of the resultant correlation matrix, $R(u, v)$ for the selected reference match.

Figure 8b shows the localization error if the estimated query image positions were utilized to estimate the position of an autonomous vehicle using a filter and the reasonable assumption that the system could be supported with motion information, such as a visual odometry system. While the medians of the two box plots are significantly high, it should be noted that only a simple weighted averaging filter was implemented and still the combined system showed a decrease in error. The estimated position is given by,

Table 3: Metric Localization Results for Comparison 3

System	Median	Minimum
Raw Algorithm Distance Error	23.9m	0.036m
Combined System Distance Error	0.264m	0.036m
Raw Algorithm Position Error	31.1m	0.915m
Combined System Position Error	15.0m	0.386m

$$P_e(i) = \begin{cases} w_1 * (P_e(i-1) + \delta) + w_2 * (P_r + g * s), \\ \quad \text{if ref match generated} \\ (P_e(i-1) + \delta), \\ \quad \text{otherwise} \end{cases} \tag{4}$$

where $P_e(i)$ is the i^{th} estimated position for the platform, δ is the amount the platform has moved since the last iteration, and w_1 and w_2 are the filter weights for the weighted averaging filter. To simulate a visual odometry system, in order to provide motion information (δ), applied noise to the difference in ground truth position between query frames, noise up to $\pm 1.5\%$ of the value of δ was added.

Additionally, as can be seen in Table 3 the minimum distance error achieved is less than 10cm, even by the raw comparison algorithm. This supports the idea that if a system reports a true-positive match, NCC can be used to accurately position a autonomous vehicle even if there is appearance variation. Furthermore Figure 8 and Table 3 both support the argument that the two systems developed improve the robustness of the base algorithm even.

6 DISCUSSION AND FUTURE WORK

In this paper we have extended the previous work on surface-based positioning systems developed for ship hull inspection and warehouse navigation for the context of autonomous vehicles on road networks. We developed two novel systems that improved the robustness of the raw comparison algorithm; through the rejection of unsuitable query images and a retrospective evaluation of the reported query-reference match. Using several datasets we demonstrated that these two systems improved the performance by increasing the precision at low recall rates.

The rejection of unsuitable query images prior to performing a comparison can be seen as wasting potential

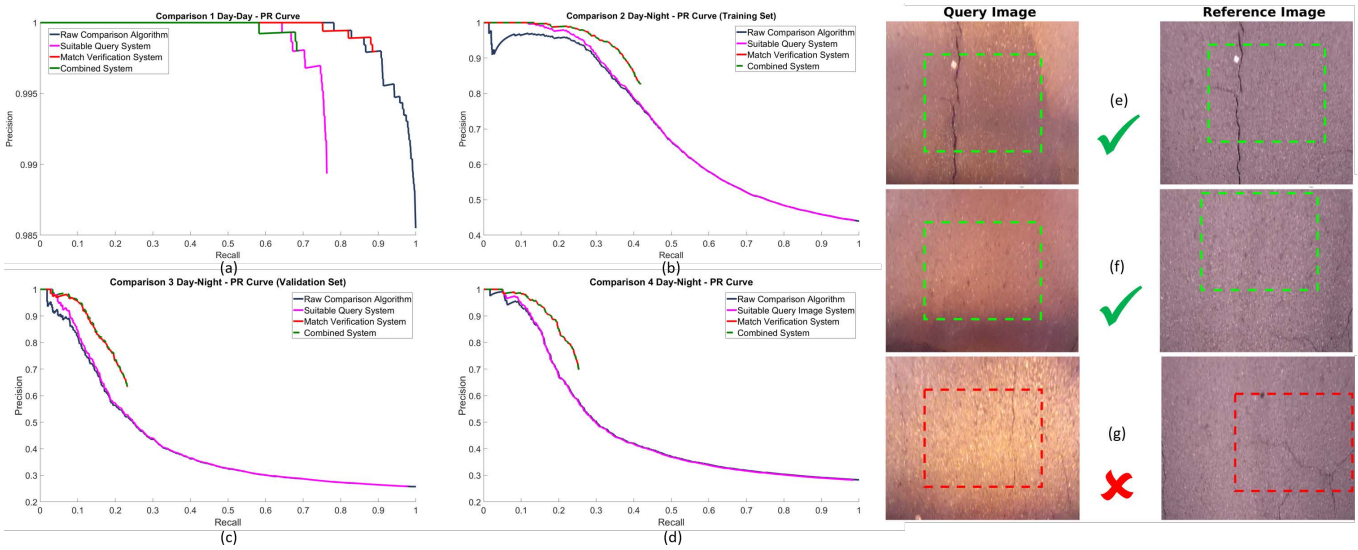

Figure 7: shows the precision-recall curves for the individual systems (raw comparison algorithm, suitable query system, correct match system) and when the systems are combined for several datasets. (a) is for *Comparison 1* a day-day comparison, (b) is for *Comparison 2* a day-night comparison as well as the training set, (c) is for *Comparison 3* a day-night comparison as well as the validation set and (d) is for *Comparison 4* a day-night comparison on a completely unseen road surface.

information, but this verification does not need to occur prior to the comparison. The query image entropy threshold could be used post comparison and additionally could only be used in the case where no strong match was generated. This is an avenue we are investigating for future work. However, if no image match will be generated when the query image entropy is below a certain threshold, irrelevant of how strong the match is, it should be done pre-comparison in order to save unnecessary computation, especially if a system can only compare every n^{th} incoming query image.

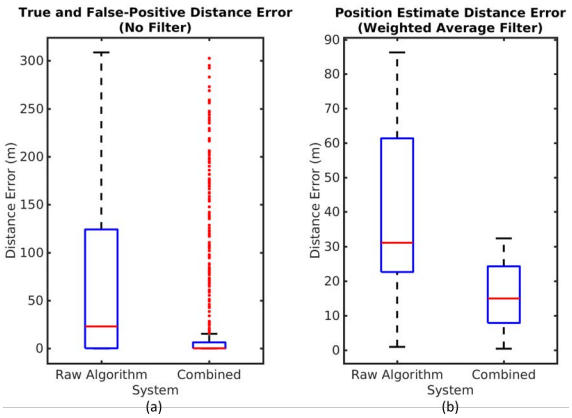

Figure 8: results for the metric query image position distance errors (a) and the platform position estimation (localization) errors (b).

We further demonstrated the capability that when the system reports the correct reference match it could accurately estimate the metric position of a query image.

The computational and storage requirements for the system is feasible using current computer hardware found in typical self-driving car prototypes. The major overhead of the system is the image comparison which grows linearly with the number of images searched. Current GPU hardware could search a 2 lane $1km$ stretch in less than a second, and current consumer SATA hard drives could store entire city road networks.

Compared to existing localization systems, the approach presented has different strengths and weakness, meaning it would serve as a useful complementary positioning capability. The first limitation is that the system requires a prior visual map of the environment; but pre-mapping is a requirement shared by many of the approaches to self-driving cars including those by Google and Uber. For these companies adding a downward facing camera, if not already available, would incur minimal cost and enable access to this complementary positioning capability. The second limitation is that road surfaces change, due to gradual wear and tear, damage and step changes caused by construction and resurfacing. Both could be addressed through the combination of crowd-sourced, ever updating surface maps, and local dead reckoning. Local dead reckoning is a fallback approach that has been proposed by major companies and startups operating in this space such as Oxbotica

[Nelson *et al.*, 2015; Linegar *et al.*, 2015]. In areas were paved roads are not present, such as dirt or gravel-based roads, other positioning technologies such as GPS would be required. The third limitation is the robustness to appearance changes caused by other phenomena, such as rain. These would be more challenging and we are currently investigating the resilience of the system to water on the road, with normal rain conditions appearing to be feasible, although deep pools of muddy water lead to positioning failures.

Future work will pursue a number of research avenues beyond the ones previously mentioned. One focus is to improve the accuracy and reliability, as well as implement techniques that allow for rotation variance and there is no technical reason position estimates of 20mm or less cannot be achieved. Another focus is incorporating motion information to allow for local rather than global searches.

Surface-based place recognition and localization offers a promising complementary positioning capability for autonomous vehicles that offer the advantages of high accuracy, low latency, computational feasibility and relatively inexpensive sensor requirements. The safety-critical nature and reliability expectations surrounding self-driving cars are so high that most critical systems will have multiple redundancies; we hope that surface-based positioning techniques will further improve the reliability of these systems as they continue to develop and be trialed at ever larger scales.

Acknowledgements

JM and MM are with the Australian Centre for Robotic Vision at the Queensland University of Technology. This work was supported by an ARC Centre of Excellence for Robotic Vision Grant CE140100016. MM is supported by an Australian Research Council Future Fellowship FT140101229. JM is supported by an Australia Postgraduate Award and a QUT Excellence Scholarship.

References

[Campbell *et al.*, 2005] Jason Campbell, Rahul Sukthankar, Illah Nourbakhsh, and Aroon Pahwa. A robust visual odometry and precipice detection system using consumer-grade monocular vision. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 3421–3427. IEEE, 2005.

[Cummins and Newman, 2011] Mark Cummins and Paul Newman. Appearance-only slam at large scale with fab-map 2.0. *The International Journal of Robotics Research*, 30(9):1100–1123, 2011.

[Davison *et al.*, 2007] Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. Monoslam: Real-time single camera slam. *IEEE transactions on pattern analysis and machine intelligence*, 29(6), 2007.

[Dille *et al.*, 2010] Michael Dille, Ben Grocholsky, and Sanjiv Singh. Outdoor downward-facing optical flow odometry with commodity sensors. In *Field and Service Robotics*, pages 183–193. Springer, 2010.

[Gonzalez *et al.*, 2013] Ramon Gonzalez, Francisco Rodriguez, Jose Luis Guzman, Cedric Pradalier, and Roland Siegwart. Control of off-road mobile robots using visual odometry and slip compensation. *Advanced Robotics*, 27(11):893–906, 2013.

[Guizzo and Deyle, 2012] Erico Guizzo and Travis Deyle. Robotics trends for 2012. *IEEE Robotics & Automation Magazine*, 19(1):119–123, 2012.

[Hover *et al.*, 2012] Franz S Hover, Ryan M Eustice, Ayoung Kim, Brendan Englot, Hordur Johannsson, Michael Kaess, and John J Leonard. Advanced perception, navigation and planning for autonomous in-water ship hull inspection. *The International Journal of Robotics Research*, 31(12):1445–1464, 2012.

[Johnson *et al.*, 2008] Andrew E Johnson, Steven B Goldberg, Yang Cheng, and Larry H Matthies. Robust and efficient stereo feature tracking for visual odometry. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 39–46. IEEE, 2008.

[Kelly *et al.*, 2007] A. Kelly, B. Nagy, D. Stager, and R. Unnikrishnan. Field and service applications - an infrastructure-free automated guided vehicle based on computer vision - an effort to make an industrial robot vehicle that can operate without supporting infrastructure. *IEEE Robotics Automation Magazine*, 14(3):24–34, Sept 2007.

[Kelly, 2000] Alonzo Kelly. Mobile robot localization from large-scale appearance mosaics. *The International Journal of Robotics Research*, 19(11):1104–1125, 2000.

[Kim and Eustice, 2009] Ayoung Kim and Ryan Eustice. Pose-graph visual slam with geometric model selection for autonomous underwater ship hull inspection. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 1559–1565. IEEE, 2009.

[Kim and Eustice, 2013] A. Kim and R. M. Eustice. Real-time visual slam for autonomous underwater hull inspection using visual saliency. *IEEE Transactions on Robotics*, 29(3):719–733, June 2013.

[Levinson and Thrun, 2010] J. Levinson and S. Thrun. Robust vehicle localization in urban environments using probabilistic maps. In *2010 IEEE International*

Conference on Robotics and Automation, pages 4372–4378, May 2010.

[Linegar *et al.*, 2015] Chris Linegar, Winston Churchill, and Paul Newman. Work smart, not hard: Recalling relevant experiences for vast-scale but time-constrained localisation. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 90–97. IEEE, 2015.

[Luettel *et al.*, 2012] Thorsten Luettel, Michael Himmelsbach, and Hans-Joachim Wuensche. Autonomous ground vehiclesconcepts and a path to the future. *Proceedings of the IEEE*, 100(Special Centennial Issue):1831–1839, 2012.

[Milford and Wyeth, 2012] M. J. Milford and G. F. Wyeth. Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights. In *2012 IEEE International Conference on Robotics and Automation*, pages 1643–1649, May 2012.

[Milford *et al.*, 2004] Michael J Milford, Gordon F Wyeth, and David Prasser. Ratslam: a hippocampal model for simultaneous localization and mapping. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 1, pages 403–408. IEEE, 2004.

[Milford *et al.*, 2014] Michael Milford, Eleonora Vig, Walter Scheirer, and David Cox. Vision-based simultaneous localization and mapping in changing outdoor environments. *Journal of Field Robotics*, 31(5):780–802, 2014.

[Miller *et al.*, 2011] Isaac Miller, Mark Campbell, and Dan Huttenlocher. Map-aided localization in sparse global positioning system environments using vision and particle filtering. *Journal of Field Robotics*, 28(5):619–643, 2011.

[Mount and Milford, 2016] J. Mount and M. Milford. 2d visual place recognition for domestic service robots at night. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4822–4829, May 2016.

[Mur-Artal *et al.*, 2015] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31(5):1147–1163, 2015.

[Murillo *et al.*, 2007] Ana Cris Murillo, José Jesús Guerrero, and C Sagues. Surf features for efficient robot localization with omnidirectional images. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 3901–3907. IEEE, 2007.

[Nelson *et al.*, 2015] Peter Nelson, Winston Churchill, Ingmar Posner, and Paul Newman. From dusk till dawn: Localisation at night using artificial light sources. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 5245–5252. IEEE, 2015.

[Nourani-Vatani and Borges, 2011] Navid Nourani-Vatani and Paulo Vinicius Koerich Borges. Correlation-based visual odometry for ground vehicles. *Journal of Field Robotics*, 28(5):742–768, 2011.

[Ozog and Eustice, 2014] P. Ozog and R. M. Eustice. Toward long-term, automated ship hull inspection with visual slam, explicit surface optimization, and generic graph-sparsification. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3832–3839, May 2014.

[Parra *et al.*, 2010] Ignacio Parra, MA Sotelo, David Fernández Llorca, and Manuel Ocaña. Robust visual odometry for vehicle localization in urban environments. *Robotica*, 28(03):441–452, 2010.

[Sa *et al.*, 2015] Inkyu Sa, Stefan Hrabar, and Peter Corke. Outdoor flight testing of a pole inspection uav incorporating high-speed vision. In *Field and Service Robotics*, pages 107–121. Springer, 2015.

[Salas-Moreno *et al.*, 2013] Renato F Salas-Moreno, Richard A Newcombe, Hauke Strasdat, Paul HJ Kelly, and Andrew J Davison. Slam++: Simultaneous localisation and mapping at the level of objects. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1352–1359, 2013.

[Se *et al.*, 2001] Stephen Se, David Lowe, and Jim Little. Vision-based mobile robot localization and mapping using scale-invariant features. In *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, volume 2, pages 2051–2058. IEEE, 2001.

[Thrun, 2010] Sebastian Thrun. Toward robotic cars. *Commun. ACM*, 53(4):99–106, April 2010.

[Ullah and Kaneko, 2004] Farhan Ullah and Shun'ichi Kaneko. Using orientation codes for rotation-invariant template matching. *Pattern Recognition*, 37(2):201 – 209, 2004.

[Vasudevan *et al.*, 2007] Shrihari Vasudevan, Stefan Gächter, Viet Nguyen, and Roland Siegwart. Cognitive maps for mobile robotsan object based approach. *Robotics and Autonomous Systems*, 55(5):359–371, 2007.

[Whitworth *et al.*, 2001] CC Whitworth, AWG Duller, DI Jones, and GK Earp. Aerial video inspection of overhead power lines. *Power Engineering Journal*, 15(1):25–32, 2001.