Tina Memo No. 2016-012
Submitted to BMVC2016, Rejected.

A Monte-Carlo Simulation for Optimising a Stereo Vision Based Robotic System.

Jingduo Tian, Neil Thacker and Alexandru Stancu.

Last updated
14 / 5 / 2016

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

A Monte-Carlo Simulation for Optimising a Stereo Vision Based Robotic System

Abstract

The quantitative optimisation of vision-based robotic systems has remained a challenge for over 2 decades. The difficulties in this problem can be linked to the lack of ground truth and the inability to acquire sufficient statistical samples from real-world trials. Our solution is based upon a large scale Monte-Carlo approach for quantitative performance evaluation and optimization. A novel simulated environment is proposed and employed for optimising a stereo-vision-based global localisation approach via a large scale Monte-Carlo analysis. The simulation program is constructed in a way that models the major characteristic geometric uncertainties associated with edge-based stereo data. As a result, our approach shows robustness to multiple uncertainties and achieves a success rate of 92.82% with high efficiency under an optimised set of parameters.

1 Introduction

The visual capabilities of state of the art robotic vision systems is currently quite limited, e.g. autonomous navigation and SLAM [Lui and Jarvis(2012)], key issues in visual perception, reasoning and associated high level control yet need to be addressed in unstructured environments. A key problem to these tasks is how to continuously and precisely localise the robot given no global map or accurate internal sensory information. This problem can be categorized into two types, namely local localisation and global localisation. Local localisation aims to provide explicit robot poses in a local area, which requires the initial position to be approximately known. Global localisation assumes no initial position, i.e. the robot kidnap problem where the robot is moved without knowing the motion. Of the two, global localisation is more important, as it can tackle more serious problems such as loop-closure.

In this paper, a stereo-vision-based global localisation approach is proposed. This approach matches the geometric co-occurrence of a sampled scene across a previously learnt database, finding the most likely location estimation on a topological map. The database is generated at various positions in an environment, using 3D stereo edge models. 3D stereo edge models, in comparison with monocular edge models, provides the ability to predict the scene changes across a range of view angles. This property enhances the scene recognition robustness towards out-of-image-plane orientations and therefore improves the localisation accuracy. This global localisation approach contains a complex combination of functional modules, the instabilities of different modules inevitably propagate and accumulate throughout the system. Therefore, understanding of such processes is crucial for guiding the system design to deal with real-world problems, before physical implementation.

Systematic parameter optimisation is a challenge for improving robotic vision systems. As robots are goal-oriented designs, the criteria for evaluating the performance of a robotic parameter setting should be the reliability in accomplishing a specific task (i.e. success rate). Due to the precision of binomial evaluation processes, thousands of trails are required in order to statistically evaluate the performance. Whilst this is possible via a Monte-Carlo test, it would be impossible for real-world experiments. In this work we have therefore constructed a novel wire-frame representation of a real-world environment. Multiple dominant real-world uncertainty sources are modelled and implemented as plug-in functions. The key focus of the simulation construction is to simulate the data quality of 3D stereo edge models under real-world conditions. The concentration on edges minimises the computation burden by avoiding the requirement to render images. The considered uncertainty sources are: image noise, lateral edge shifting, edge detection loss, geometric approximation error, corner detection loss, corner match error, stereo match error and camera calibration error. Using the simulation program, a large scale Monte-Carlo analysis is applied to evaluate and optimise the proposed global localisation approach.

In the rest of this paper, Section 2 reviews the related work on vision-based global localisation approaches and the uncertainty sources in robotic simulations. The technical detail of a stereo-vision-based global localisation approach is provided in Section 3. Section 4 reports the design of a wire-frame simulated environment and the modelling of real-world uncertainty sources. Section 5 provides the result of a Monte-Carlo optimisation on the proposed localisation approach. Finally, Section 6 concludes the contribution of this work.

2 Related Work

Topological mapping and localisation, in comparison with approaches using metric maps, employs an abstract graph to represent distinctive positions of an environment. This is suitable for applications in unknown environments where learning is more important than accuracy. The use of local descriptors in topological localisation largely increases the processing speed in scene matching while enhancing the robustness towards occlusions, scale, rotation and illumination [Garcia-Fidalgo and Ortiz(2015)]. Many of the approaches [Zhang(2011), Johns and Yang(2011)] employ SIFT as a similar feature descriptor, and some of other approaches using SURF [F. Dayoub and Duckett(2011)], KLT [Rybski and Zacharias(2003)] and ASIFT [L. Maohai and Zesu(2013)]. One common aspect of these descriptors is that they use blobs in images as interesting features. In order to increase the computation speed, features with higher consistency and distinctiveness are more likely to be entered in to a topological database. Among the above-mentioned approaches, the utilised image acquisition devices follow two categorizes, omnidirectional cameras [Rybski and Zacharias(2003), F. Dayoub and Duckett(2011), L. Maohai and Zesu(2013)] and mono cameras [Zhang(2011), Johns and Yang(2011)]. The benefit of omnidirectional cameras is its invariant on camera orientations to represent a scene. However, omnidirectional cameras are more expensive than mono cameras, plus they suffer image distortions due to their mechanical structures.

The cornerstone of engineering is simulation based design. For sufficiently complex systems such as robotic systems, simulation provides the most efficient way to investigate performance and design choices. In related work, 3D virtual worlds are designed with a certain level of reality, aiming to assess the performance of various robotic algorithms. Several of the simulation environments [K. Okada and Kanehiro(2002), I. Ulusoy and Leblebicioglu(2004), Klaser and Wolf(2013), K. Asanuma and Arai(2004), Kučiš(2012), A. Hinkenjann and Millberg(2013), X. Zou and Lu(2012)] in robotic vision field use 3D rendered objects. Some of them are designed as noise-free environments [K. Okada and Kanehiro I. Ulusoy and Leblebicioglu(2004), Klaser and Wolf(2013)] which contradict the real-world conditions. Though [K. Asanuma and Arai(2004)] considers several elementary image variations, it is designed exclusively for testing on Robotic four legged robots. [Kučiš(2012)] simulated multiple camera uncertainties on rendered 3D objects, which provide a better-looking environment but also result in a huge computation burden. Synthetic rendered images disturbed by several image uncertainties are used in [A. Hinkenjann and Millberg(2013)] to test the performance of vision algorithms. Since it needs a pre-defined image sequence as the input data, it is unsuitable for simulations on robotic systems. [X. Zou and Lu(2012)] simulates camera calibration error and stereo matching error in a simulation to test a vision-based manipulator positioning system. However, additional errors such as image noise and illumination variation are not considered.

3 World Learning and Global Localisation

A world learning and global localisation approach is presented in this section. The aim of this approach is to explore an unstructured in-door area, employing stereo-vision data to construct a topological representation of the environment for later use in localisation.

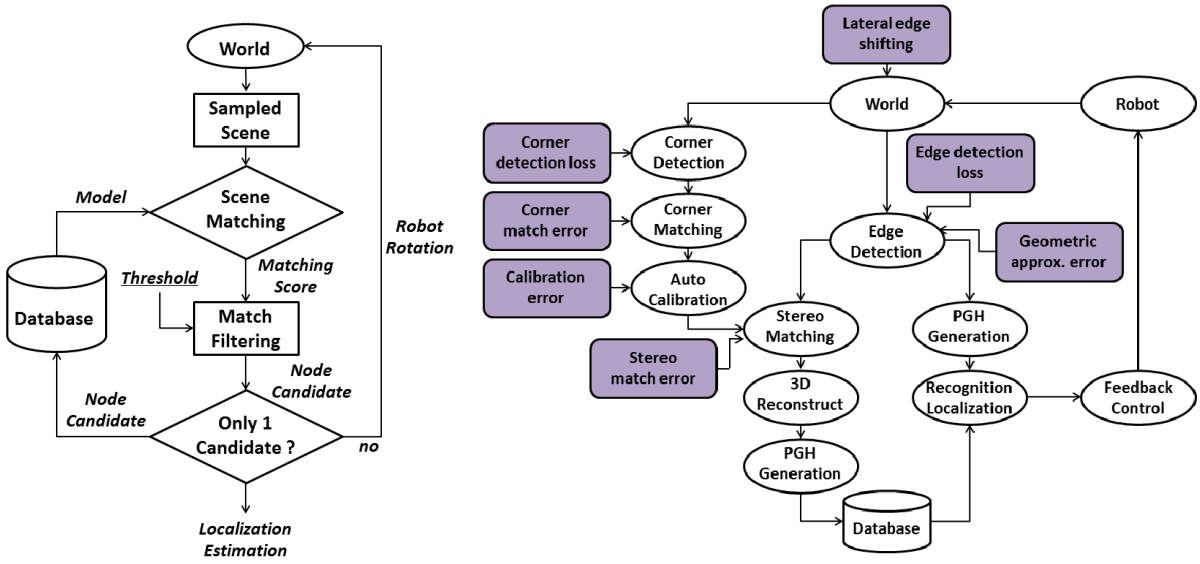
3.1 World Learning

The world learning procedure is to acquire visual information across a spatial range of an unstructured environment, sufficient to support the vision-based localisation. Simple path planning strategy (e.g. wall following with object avoidance) is applied to the robot for environment exploration. At various locations, the robot stops and the stereo vision system rotates (360 °) to acquire stereo image pairs with respect to multiple view angles (e.g. 30 °respectively). These image pairs are then post-processed by the *Canny* edge detection and a 'stretch window correlation' stereo matching algorithm [Crossley(2000)] to acquire 3D stereo edge models. These models are entered into a topological database, with each topological node representing the location at which a sequence of models is generated.

3.2 Global Localisation

Given a database generated from the world learning procedure, global localisation is equivalent to a shape recognition problem. This requires a scheme to systemically encode edge co-occurrences and also an appropriate calculation scheme for representation similarity between geometries. We encode the probability of relational geometry as probability densities, matching is conducted by assessing the probability density overlap. The practical algorithm is referred to Pairwise Geometric Histograms (PGH) [A. Ashbrook and Brown(1995), A. Evans and Mayhew(1993)].

PGH is an image feature descriptor in the format of a 2D histogram that encodes the probability of geometric co-occurrences between edge features which represent the physical structures visible in a scene [NA. Thacker and Yates(1995)]. Given a polygonised feature, a PGH selects an edge as the reference line, and then enters all the other edges into



(a) (b)

Figure 1: *(a) The flow-chart of global localisation approach. (b) The functional modules and their corresponding uncertainty sources (purple boxes).*

histogram bins depending on their perpendicular distances and relative angles to the reference line. Multiple PGHs are generated to represent a polygonised feature, with one PGH corresponding to one edge feature. The underlying principle of PGH is the argument that edge features are the most informative features defining the content and location of an object [Pinel(1997)]. Representational completeness is another distinction of PGH [NA. Thacker and Yates(1995)], which refers to the ability to inversely reconstruct a shape from the descriptor representation. The completeness property guarantees the PGH uniqueness in representing and later recognising arbitrary shapes. PGHs also provide a functional degree of invariance to environmental uncertainties, such as illumination, image noise, cluttering, edge fragmentation and occlusion [Coupe(2009)]. The matching similarity between PGHs is calculated using the *Bhattacharyya* score,

$$D_{Bhattacharyya} = \sum_{i=0}^{m} \sqrt{a_i}\sqrt{b_i} \tag{1}$$

where a and b are normalised histograms and m is the number of histogram bins. This match metric can be derived for independently distributed *Poisson* data, such as found in sample histograms and for PGH's [FJ. Aherne and Rockett(1998)].

Figure 1 shows the structure of the localisation approach. By processing a sampled image through *Canny* edge detection and thereby line fitting, a polygonised feature sample is acquired and encoded as PGHs. With an on-board compass, the current robot orientation is expected to be known. For each topological node, the learnt 3D stereo edge models which have angles adjacent to the robot orientation is matched to the sampled scene using the *Bhattacharyya* metric. The node which gives the largest match score B_{max} is used as a reference. Any other nodes that have match scores greater than $B_{max} - B_{th}$ is kept as competing candidates, where B_{th} is a pre-defined threshold. When no candidate is left, the reference node becomes the localisation result of this attempt. Otherwise the robot is rotated, acquiring a new sampled scene to match with the node candidates that are kept from the previous iteration. Between different iterations, the match scores are accumulated for the same nodes. A decay is applied on the threshold $B_i = B_{th}/i$, where i represents the i^{th} iteration.

4 Simulation Environment and Uncertainty Sources

In this section, the construction of a wire-frame represented 3D simulation environment is provided. This environment is further perturbed by multiple uncertainty sources, which are expected to have dominant effects towards real-world stereo-vision data.

4.1 A 3D Environment with Wire-frame Representation

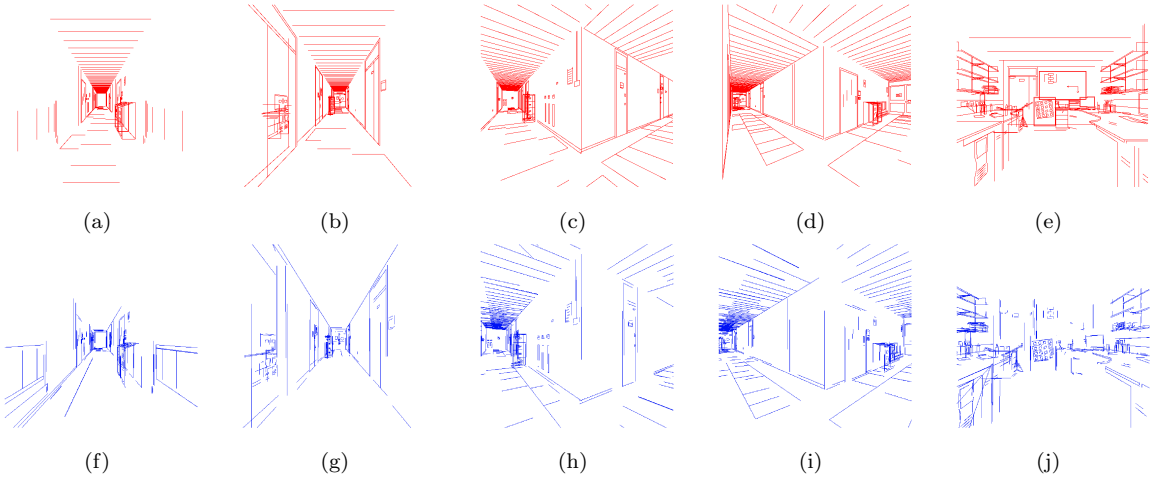


Figure 2: *The original environment (a - e) and the perturbed environment (f - j).*

A 3D environment is constructed which represents the characterisation of the real-world stereo data, suitable for evaluation of the proposed localisation approach. We firstly take 100 images of a potential work space with the volume of approximately 20m $\times$ 15m $\times$ 2.5m. Assuming the artificial lighting condition is static, by applying *Canny* edge detection the spatial distribution of genuine edges is acquired. The positions of these edges are manually measured and drawn in the simulation to an accuracy of around one centimetre. The simulation is defined as the ground truth, and this level of inaccuracy will not have significant effects on evaluations. As a result, about 7000 features are modelled. Multiple highly similar scene components (e.g. door frames) are also created as templates and used where possible, to test for robustness against worst case geometric ambiguities.

Due to the lack of surface information, fully automatic elimination of occluded edges is impossible. Instead, a view-dependency file is created in which a sequence of view-points are pre-defined as reference points. At each reference point, those edges that are expected to be visible are defined using a purpose built graphical interface. A reference point which is nearest to the current robot position is used to display all the visible edges. In this simulation, 236 uniformly distributed reference points are pre-defined. This mechanism allows us to adequately approximate the visible geometry at any location.

4.2 Uncertainty Sources

Due to data quality limitations, processing errors are inevitably caused by individual vision algorithms. Understanding these errors benefits system design and helps to avoid possible failures in real-world implementations. We investigate multiple dominant uncertainty sources in the proposed global localisation approach, as shown in Figure 1(b). Their joint effect is demonstrated in the simulated 3D environment in Figure 2.

The magnitudes of each uncertainty sources are not static, their values are determined by the specification of different environments (and algorithms). The default values are obtained by experimentations or algorithm analysis. These values are varied during the evaluation of a system design, in order to comprehensively test the system robustness under different environment and algorithm specifications.

Image Noise

Image noise is due to grey-scale measurements. They affect vision algorithms, including here the *Canny* edge detector and the subsequent geometrical approximation. These effects are approximated by the following lateral edge shifting and geometric approximation error.

Lateral Edge Shifting and Edge Detection Loss

Edges are usually represented by the discontinuities of brightness in an image, therefore they are highly susceptible to illumination. Due to blurring and lack of contrast with the background, genuine edges are not always detectable. A part of an extended edge feature can either be totally undetected or be detected with a biased orientation and location. When the detected edge locations are compared with the object model predictions, some systematic problems become apparent. Under distant illumination conditions (e.g. *Lambertian* reflectance models), a planar surface with homogeneous texture is assumed to have the same grey-level. Therefore, illumination changes apply systematically along an extended boundary of the surfaces resulting in lateral edge shifting [Coupe(2009)].

Given an image, edge detection loss is defined as the percentage of genuine edges that can not be repeatedly detected

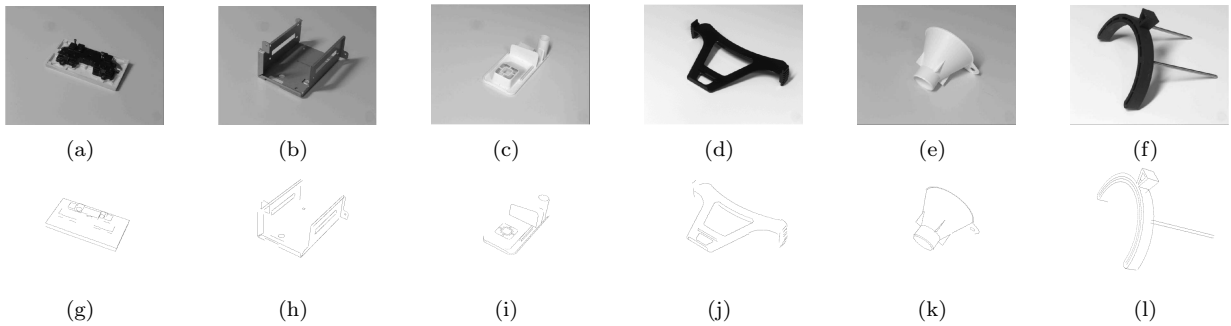


Figure 3: *A selection of test objects (a - f) and the re-projected features (g - l) [Coupe(2009)].*

by an edge detector. A quantitative verification using a power law model proposed in [Coupe and Thacker(2008)] is utilised to quantify the edge detection loss of the *Canny* edge detector. This approach manually creates 16 wire-frame models of a selection of referenced man-made objects constructed from a variety of materials. These wire-frame models are optimally projected onto their corresponding image edges, and the pixel alignments between model contours and image edges are quantitatively verified (examples shown in Figure 3).

The verification scores indicate the portion of artificially defined object contours that are detectable by *Canny*. Due to subjective definition of contour locations, the verification scores are inevitably biased from their true values. By doing a linear interpolation on the result reported in [Coupe(2009)], the bias is expected to be largely reduced. The result provide a quantitative approximation of the average *Canny* edge detection loss (11.25% with a range of 0% to 31.25%). In our simulation, the uncertainty of edge detection loss randomly removes ψ_e (by default 11.25%, range from 0% to 31.25%) of the length on each visible features. The uncertainty of lateral edge shifting then drifts each feature laterally within a range ψ_l (by default 3 $\pm$1.5 pixels), as suggested in [Coupe(2009)].

Geometric Approximation Error

Geometric approximation is used to fit distinctive image pixels into polygonised features, thereby allowing high-level feature analysis. The error on geometric approximation can be defined as the residual between a fitted geometric feature and its corresponding image pixels. A model that estimates the error distribution of line fitting on *Canny* edges has been proposed and validated in [A. Ashbrook and Brown(1995)]. Circular *Gaussian* uncertainty regions are defined at the end points of a fitted line feature, representing the error on geometric approximation. The STD of this *Gaussian* region is determined by line fitting threshold, which is used to define the break point of a line when fitting to the image pixels. In this work, the threshold is set to 0.15 pixels by default. Therefore, we simulate circular *Gaussian* noise regions with the STD of ψ_g (by default 0.15 pixels $\pm$0.05) on each of the line end points.

Corner Detection Loss and Corner Match Error

Corner detection loss is the proportion of corners that can not be re-detected in image 2 given they have been detected in image 1, due to the change on local image re-projection. Corner match error is the amount of incorrect corner matches that are accepted as correct matches, due to ambiguities on image patches around competing matching candidates.

In this work, the adopted corner matching algorithm [Thacker and Courtney(1992)] is based on the *Harris* and *Stephens* corner detector [Harris and Stephens(1988)]. As reported in [Thacker and Courtney(1992)], on average the typical repeatability of detecting the same corner in both images is about 85% (this varies for specific algorithm configurations and datasets). In the matching procedure, the camera geometry is used to determine a region in image 2 given the corner in image 1. As reported by [Thacker and Courtney(1992)], 1% of the accepted corner matches are incorrect. In the simulation, corner features are manually defined inside the 3D virtual environment. The uncertainty of corner detection loss randomly drops ψ_{cd} (15% $\pm$10%) corners in both images. The uncertainty of corner match error then randomly drifts ψ_{cm} (1% $\pm$1%) corners inside circular regions with a 10 pixel radius.

Stereo Match Error

For an image pair, stereo matching is performed along epi-polar constraints within a disparity range, matching geometric features with the largest correspondence. Due to huge ambiguities, however, those features that lie along epi-polar lines can not be properly matched. Some algorithms [Crossley(2000)] therefore inherently ignore those features before conducting the match. Stereo match error is the amount of incorrect matches that are accepted by a stereo matching process, and the elimination of those features parallel to epi-polar lines.

In our system, a 'stretch correlation' algorithm [Crossley(2000)] is used for stereo matching between geometric

features. According to a quantitative analysis in [Crossley(2000)], the typical mis-match rate is around 1% with a default disparity range of 20 pixels on each side. This algorithm automatically eliminate all the features which are $\pm 5^{o}$ parallel to their intersected epi-polar lines. For simplicity, the simulation system removes all the features that are $\pm 5^{o}$ parallel to a horizontal line. Then, ψ_s (1% varying from 0% to 10%) of the remaining feature are shifted randomly within the disparity range of 20 pixels in both images.

Camera Calibration Error

Camera calibration is used to estimate the intrinsic and extrinsic parameters of camera models from image evidence. We employ an automatic stereo calibration approach [Thacker and Mayhew(1991)] that uses matched stereo corner pairs from generic images, which is suitable for robotic applications in unstructured environments. Using real data, we derive a covariance matrix with respect to the calibrated parameters, and use it to specify the parameter error correlations.

A stereo model is defined as $S = F(f, a_x, a_y, o_x, o_y, k, R, T)$, with focal length f, aspect ratio (a_x, a_y), optical centre (o_x, o_y), radial distortion coefficient k, rotation and translation matrix R, T. A unified least-square cost function is then followed as,

$$\chi_t{}^2 = (a - a_t)^T C_a{}^{-1} (a - a_t) + \sum_i (y_i - \phi_i(a_t))^T W_i{}^{-1} (y_i - \phi_i(a_t)) \quad (2)$$

with a representing the true stereo parameters and a_t the parameter estimations at the t^{th} iteration. $\chi_t{}^2$ represents the cumulative residual between the true stereo parameters and the parameter estimations. y_i is defined as the image projections of a matched corner pair where $\phi_i(a_t)$ is the projection estimation given an epi-polar model ϕ_i with respect to an estimated parameter set a_t. The covariance matrix C_a is obtained by a sensitivity analysis. By conducting singular value decomposition (SVD) on the covariance C_a, the correlations and inverse variances of different parameter errors are derived. Using these, the predicted error distributions on individual parameters are calculated and simulated. The calibration uncertainty magnitude is defined as a variable of ψ_c (by default 1, from 0.1 to 10).

5 Monte-Carlo Analysis for Global Localisation

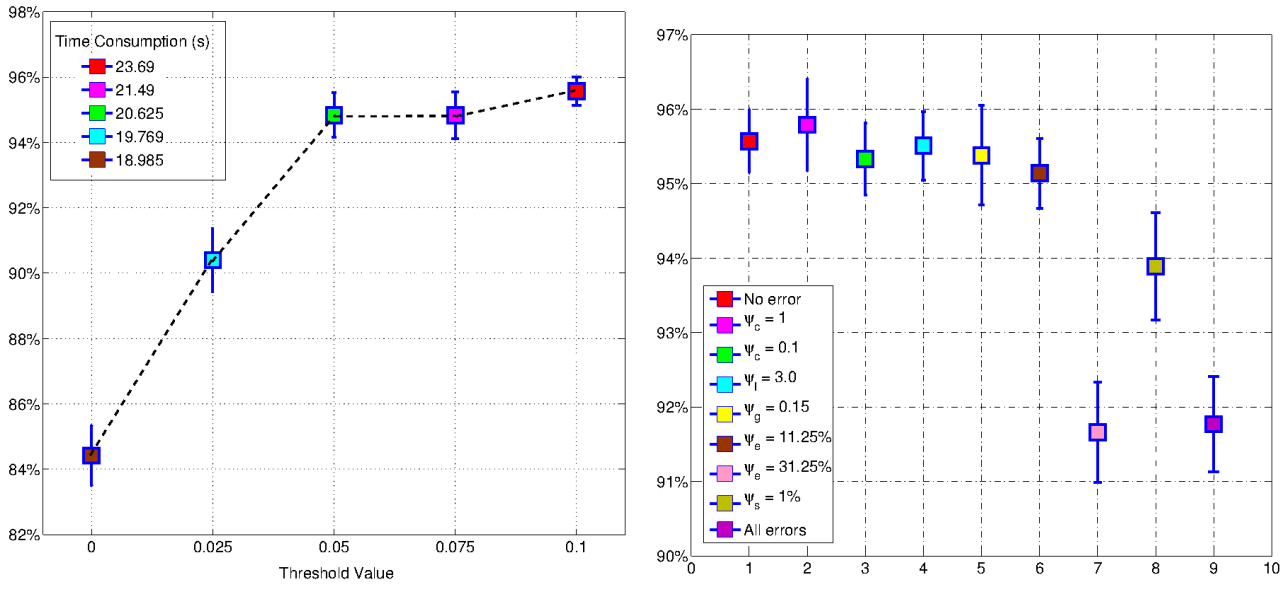
The aim of this section is to quantitatively evaluate and therefore optimise the localisation performance given different parameter settings and various uncertainty perturbations. An optimised parameter set is obtained providing high performance and efficiency. It also quantifies the effect of individual uncertainty source towards the overall performance, to guide the future design improvements.

A topological map of 72 nodes is generated by a virtual robot, with each node containing 12 3D stereo edge models. These nodes are uniformly distributed along a pre-defined path, covering the majority environment space. At each node, 3D models are taken at 12 equally separated view-angles (i.e. 30^{o} respectively). These 3D models are corrupted by simulated uncertainty sources with specific magnitudes. The localisation performance is defined as the success rate in localising the robot to its adjacent two nearest nodes, given random initial poses. For each parameter setting and each uncertainty study, over 900 simulations are repeated, which provides reliable statistics with a binomial observation error around 0.5%.

The threshold value B_{th} determines the number of candidate nodes that are kept by the candidate filtering. To investigate the correlation between the performance and efficiency with respect to different B_{th} values, an analysis is conducted using $B_{th} = 0, 0.025, 0.05, 0.075$ and 0.1, as shown in Figure 4 (a). As a result, $B_{th} = 0.01$ is determined as the optimised value, with which a success rate of 95.57% is acheived.

Next, the effect of individual uncertainty sources to the localisation performance is investigated. The uncertainty sources include edge detect loss ($\psi_e = 11.25\%$ and 31.25%), lateral edge shifting ($\psi_l = 3.0$), camera calibration error ($\psi_c = 0.1$ and 1.0), stereo matching error ($\psi_s = 1\%$) and geometric approximation error ($\psi_g = 0.15$). Simulations are also conducted using no uncertainties and all the uncertainties at their default values. Results shown in Figure 4 (b) proves that the proposed approach possesses high robustness towards lateral edge shifting, camera calibration error and geometric approximation error, while stereo matching error and intense edge detect loss both lead to performance declines.

PGHs are 2D histograms, the resolution is determined by the number of distance bins n_{dist} and angle bins n_{angle}. With greater resolution, the sensitivity of PGHs and the computation burden are increased. Simulations are conducted using multiple PGH resolutions, with $n_{dist} = 5, 15, 30$ and 60, plus $n_{angle} = 8, 16, 32$ and 64. All the uncertainty sources are applied at their default magnitudes. Figure 5 shows the optimal resolution is $n_{dist} = 15$



(a) (b)

Figure 4: *(a) Performance with different B_{th}. (b) Performance with uncertainty sources.*

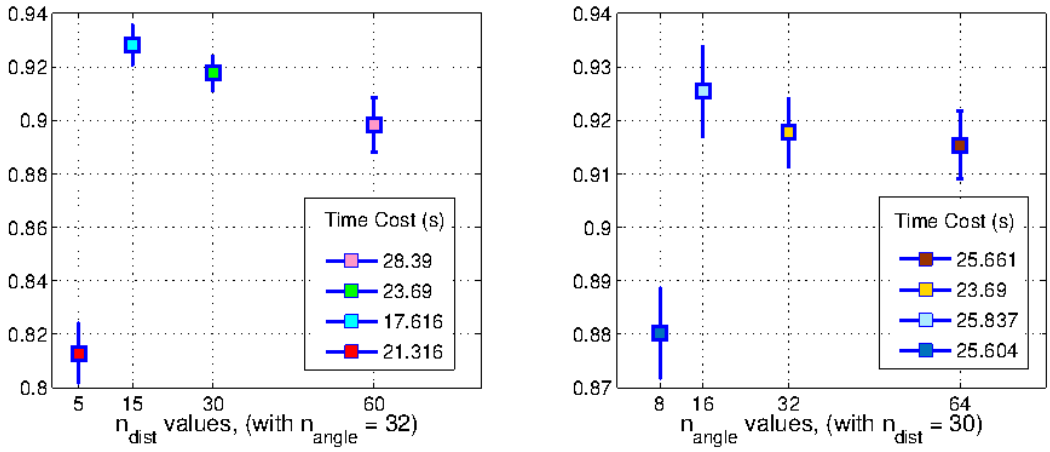


Figure 5: Performance with different n_{dist} (left) and different n_{angle} (right).

and $n_{angle} = 32$, with which the success rate reaches 92.82% at a time cost of 17.616s[1].

6 Conclusion

This paper presents a Monte-Carlo scheme for optimising a stereo-vision based robotic localisation system, under a wire-frame represented simulation environment. The simulation environment and its associated uncertainty sources are designed to represent real-world stereo edge data, while at the same time, minimising unnecessary computation burden caused by object rendering. The use of a realistic simulation environment provides the ground truth information for quantitative performance evaluations. Moreover, it allows hundreds of thousands experiments to be repeated in a reasonable time scale, which could be impossible for real-world trails. With the performance quantified, the system parameters are optimised, and the effects from specific uncertainty sources towards the ultimate output are acquired. All these provide essential evidence to evaluate the existing system design and to guide the future improvement.

[1] In previous work, 77 sets of totally 74110 simulations have been conducted in 4 months to guide and verify the design improvements. The initial success rate was 28.44% at a time cost of 2466.32s

The proposed robotic localisation system shows high robustness towards multiple uncertainty sources. Considering performance and efficiency, the optimised threshold is $B_{th} = 0.1$ and the PGH resolution is $n_{dist} = 15$ and $n_{angle} = 32$. With a dense topological map of 72 nodes (12 scenes per node) and all the uncertainties applied, the localisation performance reaches 92.82% at a time cost of 17.616s.

References

[A. Ashbrook and Brown(1995)] P. Rockett A. Ashbrook, NA. Thacker and CI. Brown. Robust recognition of scaled shapes using pairwise geometric histograms. In *BMVC'95*, pages 503–512, 1995.

[A. Evans and Mayhew(1993)] NA. Thacker A. Evans and JEW. Mayhew. The use of geometric histograms for model-based object recognition. In *BMVC'93*, pages 429–438, 1993.

[A. Hinkenjann and Millberg(2013)] T. Roth A. Hinkenjann and J. Millberg. Real-time simulation of camera errors and their effect on some basic robotic vision algorithms. In *CRV'13*, pages 218–225, 2013.

[Coupe(2009)] S. Coupe. *Machine Learning of Projected 3D Shape*. PhD thesis, University of Manchester, 2009.

[Coupe and Thacker(2008)] S. Coupe and NA. Thacker. Quantitative verification of projected views using a power law model of feature detection. In *CRV'08*, pages 352–358, 2008.

[Crossley(2000)] S. Crossley. *Robust Temporal Stereo Computer Vision*. PhD thesis, University of Sheffield, 2000.

[F. Dayoub and Duckett(2011)] G. Cielniak F. Dayoub and T. Duckett. A sparse hybrid map for vision-guided mobile robots. In *ECMR'11*, pages 213–218, 2011.

[FJ. Aherne and Rockett(1998)] NA. Thacker FJ. Aherne and PI. Rockett. The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 34(4):363–368, 1998.

[Garcia-Fidalgo and Ortiz(2015)] E. Garcia-Fidalgo and A. Ortiz. Vision-based topological mapping and localization methods: A survey. *Robotics and Autonomous Systems*, 64:1–20, 2015.

[Harris and Stephens(1988)] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50, 1988.

[I. Ulusoy and Leblebicioglu(2004)] U. Halici I. Ulusoy and K. Leblebicioglu. 3d cognitive map construction by active stereo vision in a virtual world. In *ISCIS'04*, pages 400–409. Springer, 2004.

[Johns and Yang(2011)] E. Johns and GZ. Yang. Global localization in a dense continuous topological map. In *ICRA'11*, pages 1032–1037, 2011.

[K. Asanuma and Arai(2004)] R. Ueda K. Asanuma, K. Umeda and T. Arai. Development of a simulator of environment and measurement for autonomous mobile robots considering camera characteristics. In *RoboCup 2003: Robot Soccer World Cup VII*, pages 446–457. Springer, 2004.

[K. Okada and Kanehiro(2002)] Y. Kino K. Okada and F. Kanehiro. Rapid development system for humanoid vision-based behaviors with real-virtual common interface. In *IROS'02*, pages 2515–2520, 2002.

[Klaser and Wolf(2013)] RL. Klaser and D. Wolf. Simulation of an autonomous vehicle with a vision-based navigation system in unstructured terrains using octomap. In *SBESC'13*, pages 177–178, 2013.

[Kučiš(2012)] M. Kučiš. Simulation of camera features. In *the 16th Central European Seminar on Computer*, pages 117–123, 2012.

[L. Maohai and Zesu(2013)] S. Lining L. Maohai, W. Han and C. Zesu. Robust omnidirectional mobile robot topological navigation system using omnidirectional vision. *Engineering applications of artificial intelligence*, 26(8):1942–1952, 2013.

[Lui and Jarvis(2012)] W. Dennis Lui and R. Jarvis. A pure vision-based topological slam system. *The International Journal of Robotics Research*, 2012.

[NA. Thacker and Yates(1995)] PA. Riocreux NA. Thacker and RB. Yates. Assessing the completeness properties of pairwise geometric histograms. *Image and Vision Computing*, 13(5):423–429, 1995.

[Pinel(1997)] JPJ. Pinel. *Biopsychology*. Pearson Education, 1997.

[Rybski and Zacharias(2003)] PE. Rybski and F. Zacharias. Using visual features to build topological maps of indoor environments. In *ICRA'03*, pages 850–855, 2003.

[Thacker and Courtney(1992)] N. Thacker and P. Courtney. Statistical analysis of a stereo matching algorithm. In *BMVC'92*, pages 316–326, 1992.

[Thacker and Mayhew(1991)] NA. Thacker and JEW. Mayhew. Optimal combination of stereo camera calibration from arbitrary stereo images. *Image and vision computing*, 9(1):27–32, 1991.

[X. Zou and Lu(2012)] H. Zou X. Zou and J. Lu. Virtual manipulator-based binocular stereo vision positioning system and errors modelling. *Machine Vision and Applications*, 23(1):43–63, 2012.

[Zhang(2011)] H. Zhang. Borf: Loop-closure detection with scale invariant visual features. In *ICRA'11*, pages 3125–3130, 2011.