Tina Memo No. 2016-013
To be presented at ICARCV 2016.

Stereo Vision Based Autonomous Navigation for 3-DOF Systems in Unstructured Environments.

Jingduo Tian, Neil Thacker and Alexandru Stancu.

Last updated
1 / 8 / 2016

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

Abstract

A stereo vision based autonomous navigation method for 3-DOF systems is presented in this paper. It is able to tackle the learning and recognition problem of generic scenes in an unstructured environment, providing motion-planning capability to control all the 3 DOFs of a robotic system. In this method, 3 spatial constraints are generated from a single visual recognition to estimate the robot pose. A feedback strategy is utilised for robot motion control, without the necessity of knowing any explicit distance information of the environment. The performance of the proposed method is evaluated in a novel wire-frame simulation environment, under the perturbation of multiple uncertainty sources. Autonomous navigation is achieved with good accuracy in the simulation environment, while preserving high robustness to all the uncertainty sources.

1 Introduction

Vision-based autonomous navigation in unstructured environments has attracted great research interest in the past 3 decades. The ultimate goal is to conduct visual perception, reasoning and associated path-planning without human intervention, given limited information of the environment structure. The key issue is to concurrently estimate the position of a robotic system from the visual recognitions, and use this to guide a path-planning. The path-planning requires all the robotic degree of freedoms (DOFs) to be specifically controlled over time. For holonomic systems, which are feasible design choices in vision-based robotic research field, there are 3 DOFs that need to be considered. All the 3 DOFs need to be controlled by at least 3 spatial constraints, which need to be extracted from the visual recognition results.

This paper presents a stereo-vision-based navigation method that explicitly constrains 3 DOFs of a robotic system, suitable for tasks in unstructured environments. This is achieved by matching the geometric co-occurrences of a sampled scene with previously-learnt 3D stereo edge models, thereby calculating the spatial relationship in between. Stereo-vision data, in comparison with monocular data or omni-directional data, provides the ability to predict the scene changes across a range of view angles. This property allows an additional spatial constraint, i.e. the out-of-image-plane orientation, to be extracted. An improved robust scaled shape recognition algorithm is proposed, which provides 3 spatial constraints to support the path-planning of holonomic systems. In addition, a feedback strategy is utilised for robot motion control, without the necessity of knowing any explicit distance information of the environment.

The proposed navigation method is evaluated in a novel wire-frame represented simulation environment. This environment is designed to represent real-world edge data, under the perturbation of multiple uncertainty sources. Our method demonstrates autonomous navigation in an unstructured environment, while maintaining robustness to all the simulated uncertainty sources.

In the rest of this paper, Section 2 provides a review on the related work. The proposed method is detailed in Section 3, where the environment learning and navigation are provided. Section 4 describes the simulation evaluation of the proposed method. Section 5 concludes the contribution of this work.

2 Related Work

Autonomous navigation has been widely studied in the research field of robotics, especially on applications in unstructured environments. It is an essential aspect in the realisation of fully autonomous robotic systems, which are expected to conduct environment learning, decision making and motion planning. As a priority, a map that can sufficiently represent a working environment needs to be generated. Regarding the map building methods, there are mainly two paradigms followed, i.e. metric maps [1, 2] and topological maps [3, 4, 5, 6, 7]. Although metric maps provide explicit distance information and are appropriate for obstacle avoidance, they are difficult to acquire and maintain due to the data association problem. Topological maps store visual information from distinct positions to form an abstract environment representation, therefore largely avoiding the problem in data association.

Given a topological map and an approximated robot location, autonomous navigation aims to guide the robot to reach a target position, using the stored visual information for path-planning. Path-planning requires the relative robot pose to be estimated from the visual information, so that the motion control can be achieved via visual servoing. [7] proposes a method to transfer omnidirectional images into 'bird-eye' views, and to extract the

corridor boundaries from these images. These boundaries are used as bounding boxes to constrain the mobile motion, thereby achieving path-planning with obstacle avoidance. SIFT features [8] are utilised in [3, 4] for distinctive point matching between different omnidirectional images. The positions of the matching points are used to calculate the relative robotic poses by solving a set of polynomials under the epi-polar constraints. Path-planning is thereby conducted based on the calculated robotic poses. [5] proposes an approach to categorise corridor types based on the Kinect point-cloud data. Associated motion control is determined by the corridor type that the robot is travelling through. In [6], a global descriptor based on *Fourier* transformation is used for image matchings. The relative robotic altitude is estimated by a coordinate reference system, to support the path-planning in autonomous navigation tasks.

In the related work, research on stereo-vision-based autonomous navigation approaches is still quite limited. Moreover, no approach has been proposed able to respectively constrain 3 DOFs on robotic motion, given a monocular image as the only visual input.

3 The Proposed Method

The proposed method aims to tackle the autonomous navigation task of 3-DOF systems in unstructured environments, which contains two aspects, i.e. the environment exploration and the environment navigation. The exploration obtains the visual representation of an environment in the format of a topological map, suitable to support the vision-based navigation across a spatial range. The navigation then uses this visual representation to concurrently estimate the system pose, thereby controlling all the 3 DOFs to converge towards the reference values via a feedback control strategy. The robotic system is therefore able to achieve a target location by conducting local navigations between consecutive topological nodes along an optimal path on the map. In the presented method, 3D stereo edge models are employed to build the visual representation of an unstructured environment.

3.1 Topological Map

Topological maps, in comparison with metric maps, employ an abstract graph to represent distinctive positions of an environment. They are more appropriate for applications in unstructured environments where learning is more of a priority than precision. In order to acquire a topological representation, the robot first applies an elementary motion planning strategy (e.g. follows the walls and avoids obstacles) to achieve environment exploration. The robot stops at different positions with the stereo vision system rotating a full round to obtain stereo images at specific view angles (e.g. 30 °respectively). The obtained images are processed by the *Canny* detection [9] and a 'stretch window correlation' stereo matching method [10] to reconstruct 3D stereo edge models. These 3D models are stored in a topological database, forming a topological map which represents the positions where the 3D models are acquired.

3.2 Local Navigation

Assuming the initial robot location has been approximated to its nearest topological node. The local navigation approach then guides the robot to the position of an adjacent topological node, constraining all the 3 DOFs using an improved scaled shape recognition algorithm. This algorithm is designed based on [11], where only 2 spatial constraints are provided in the original algorithm.

The proposed method uses Pairwise Geometric Histograms (PGHs) [11, 12] in the representation and recognition of geometric features of unstructured environments. PGHs are feature descriptors with the format of 2D histograms. They enter the probability of geometric co-occurrences between edge features to represent the physical structures visible in a scene [13]. Given a polygonised feature, a PGH first chooses an edge as the reference line. It then encodes all the other edges into a 2D histogram depending on their perpendicular distances and relative angles to the reference line. Multiple PGHs are calculated for representing a single polygonised feature, with each 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 [14]. There is another distinction of PGHs called the representational completeness [13]. It refers to the ability to inversely reconstruct a shape from the descriptor representation. This completeness property guarantees the PGH distinctiveness in the representation and recognition of arbitrary shapes. PGHs also provide a functional degree of invariance to environmental uncertainties, such as illumination, image noise, cluttering, edge fragmentation and occlusion [15]. 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 [16].

A holonomic mobile robot has 3 DOFs, i.e. X_r, Y_r and θ. The transformation matrix between the robot $[X_r, Y_r]$ and the world coordinate $[X_c, Y_c]$ can be expressed as,

$$\begin{bmatrix} X_r \\ Y_r \end{bmatrix} = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \end{bmatrix} - \begin{bmatrix} X \\ Y \end{bmatrix} \tag{2}$$

where θ represents the robot orientation, and $[X, Y]$ is the robot centroid position under the world coordinate.

As shown in Figure 1 (a), the proposed local navigation contains four steps, namely scene line labelling, model out-of-plane orientation estimation, model position estimation and model scale estimation. In Figure 1 (b), $[X, Y, \theta]$ represents the robot pose in the world coordinate and $[X_s, Y_s, \theta_s]$ is the pose of a reference node in the world coordinate. The ultimate goal is to respectively eliminate the residuals on ΔX, ΔY and $\Delta\theta$.

The proposed method uses the out-of-plane orientation to constrain ΔY, the position to constrain $\Delta\theta$ and the scale to constrain ΔX. All the constraints work coordinatively to navigate the robot towards a desired pose ($[X_s, Y_s, \theta_s]$).

1. Scene Line Labelling

The aim of scene line labelling is to label each model feature with a scene feature which shows the largest correspondence. This is achieved through PGH matchings between all the scene features and model features. A crucial problem that needs to be tackled is the possible recognition failure caused by linear scaling of the model features in a sampled scene, i.e. the scaling problem. Our solution is to scale the model along a scale range and enter different scaled models into a database, applying recognition to all the scaled models in order to find the best match. The database is generated in a way that, within the scale range of each two consecutive model entries, the similarity score between any two scaled models should always exceed a pre-defined threshold (e.g. a *Bhattacharyya* score of 0.9), as shown in Figure 2. The matching score of PGHs is expected to decline monotonically along the scaling [11]. By setting new database entries at the threshold border, the similarity score is guaranteed to exceed the threshold at any scale along the database.

2. Model Out-of-plane Orientation Estimation

Given a 3D stereo edge model of a scene, it is able to reliably predict the scene changes across a limited view-angles. The matching score of PGHs is proved to reduce monotonically across a range ($\pm 10^o$) of out-of-plane orientation [15]. By matching a sampled scene with the model rotated at multiple view-angles, the spatial relationship with respect to out-of-plane orientation can be estimated, which is then used to constrain the robot horizontal motion on Y_r axis.

In this method, a 3D stereo edge model is rotated to 3 view-angles a_-, a_0 and a_+, with the sampled scene matching to each of them, obtaining 3 match scores b_-, b_0 and b_+. By comparing the values of b_-, b_0 and b_+, the spatial relationship shown in Figure 3 can be obtained. The control signal U_Y can be generated for controlling the robot motion along Y_r axis, which is expressed in Equation (3). Where μ is an arbitrary positive number representing the proportional gain of a feedback control loop, determining the convergence speed of the control system.

It is important to choose appropriate values of view-angles a_-, a_0 and a_+, because they determine the steady state accuracy of ΔY. As an example, a large value leads to a fast system response, but reduces the steady state accuracy, and vice versa. A coarse-to-fine approach is therefore proposed to determine appropriate values for the view-angles. This approach starts by using a large view-angle $\pm a_{max}$ (e.g. $\pm 3^o$) for fast convergence. Once these values become too coarse to constrain ΔY, they are iteratively decreased until they reach a pre-defined minimum value $\pm a_{min}$. Afterwards, to further increase the steady state accuracy, a quadratic curve is fitted using b_-, b_0 and b_+ to find the convex optima that best estimates the current view-angle. The procedure is demonstrated in Figure 4. The combined control signal U_Y can be expressed as,

if $|a_\pm| > |a_{min}|$,

$$U_Y = \begin{cases} \mu \cdot |a_-| & b_- > b_0 > b_+ \\ 0 & b_0 > b_-\ ,\ b_0 > b_+ \\ -\mu \cdot |a_+| & b_+ > b_0 > b_- \end{cases} \tag{3}$$

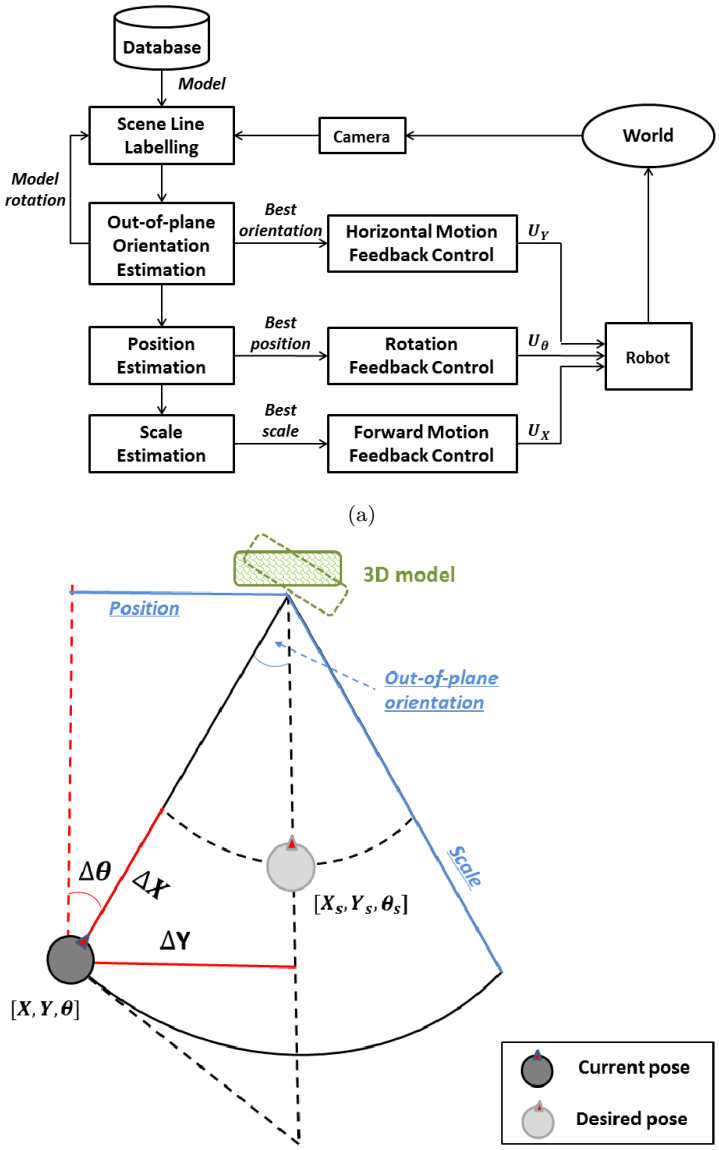

Figure 1: *(a) The flow-chart of the local navigation procedures. (b) The graphical description of the local navigation method.*

else,

$$U_Y = \begin{cases} \mu \cdot |-a_{min}| & b_- > b_0 > b_+ \\ \mu \cdot a_p & b_0 > b_- \,,\; b_0 > b_+ \\ -\mu \cdot |a_{min}| & b_+ > b_0 > b_- \end{cases} \tag{4}$$

where a_p is the value that gives the largest b_p on a quadratic function $a = f(b)$, which is fitted by 3 samples $[b_-, -a_{min}]$, $[b_0, a_0]$ and $[b_+, a_{min}]$. In addition, the rotated model which possesses the largest similarity (match score) with the sampled scene is used for the following model position estimation and model scale estimation procedures, to reduce the recognition ambiguity caused by out-of-plane orientation.

3. Model Position Estimation

This stage aims to estimate the position of a model in the image plane, calculating the distance between the image centre and the model position. In previous steps, each of the model lines has been matched to a corresponded scene line. In this step, the lines are used to determine the model position via a probabilistic scheme which calculates the most likely position estimation, considering the line fitting error.

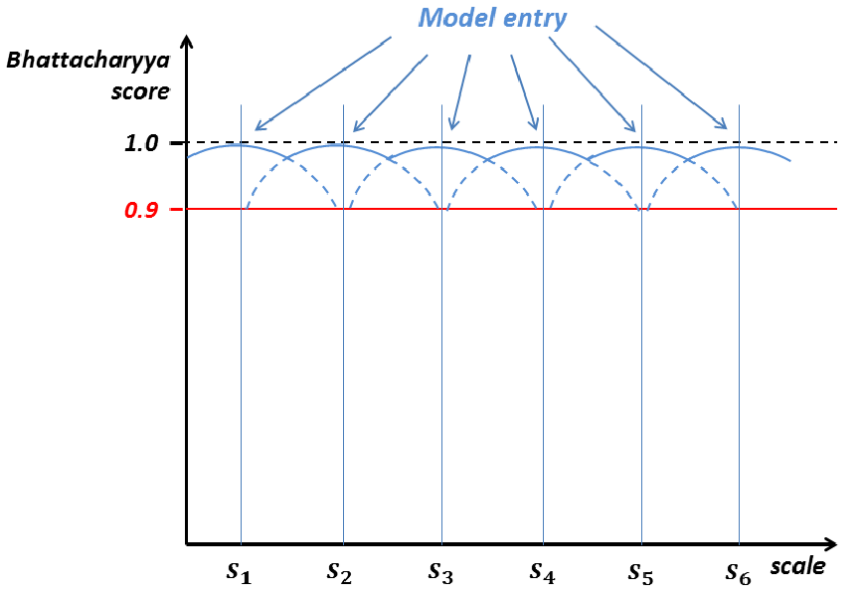

Figure 2: *Model entries in scene line labelling procedure.*

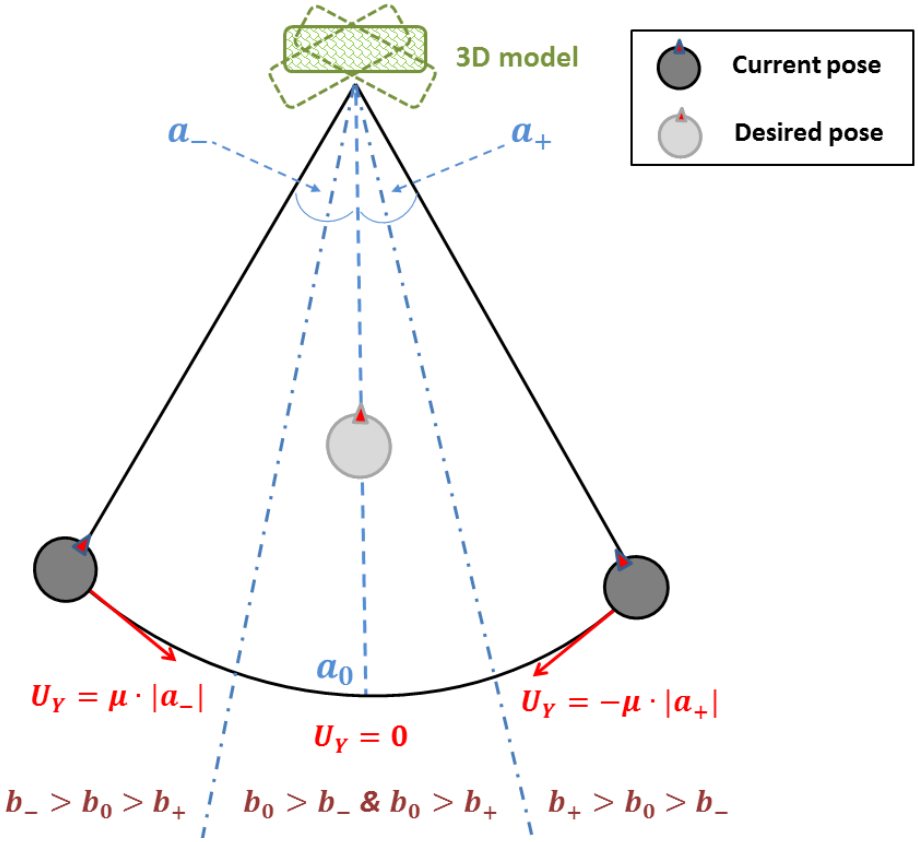

Figure 3: *The control signal with respect to out-of-plane orientation.*

A reference point is firstly extracted from the model lines to represent the position of the model. In order to eliminate the edge fragmentation effect caused by the edge detection, the centroid is defined as the reference point which is calculated based on line intersections. The perpendicular distance d_i from each model line to the centroid is calculated and stored for the following position and scale estimations.

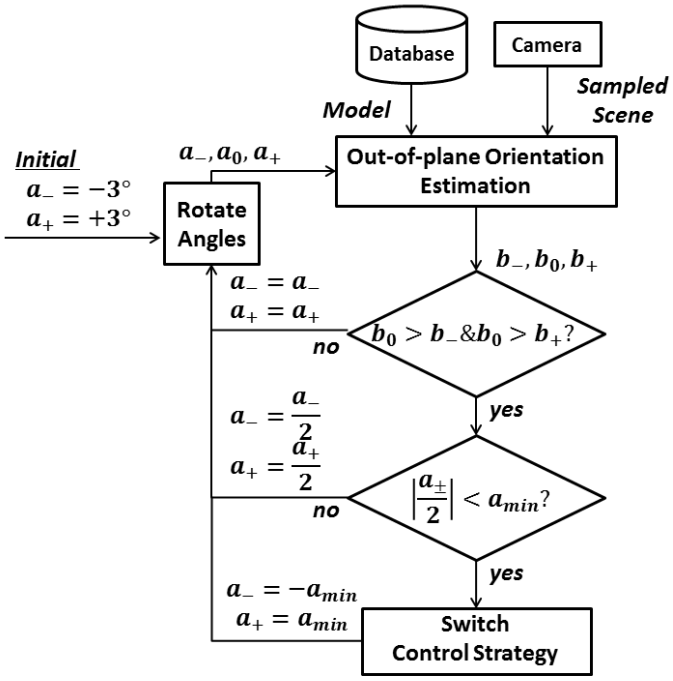

Figure 4: *The proposed coarse-to-fine approach for view-angle selection.*

The line fitting error can be approximated by circular *Gaussian* regions on the position of each line end points, as shown in Figure 5. This error model is proposed and validated in [11]. By shifting two non-parallel scene lines to corresponding perpendicular distances d_i, they are expected to have an intersection on the centroid. With error propagation, an ellipse region is obtained representing the probability distribution of the centroid position. The ellipses calculated by all the scene line intersections are entered into a 2D *Hough* space, obtaining the peak as the best position estimation. This peak gives the estimated model position in the current sampled scene. By comparing it with the model centroid position, a control signal $U_\theta = \nu \cdot |x_m - x_s|$ can be generated to constrain $\Delta\theta$. With x_m and x_s representing the position of the model and the position in the scene, ν is an arbitrary proportional gain of the control loop.

4. Model Scale Estimation

Model scale is an useful image evidence indicating the relative distance from the scene to the robot. In the scene line labelling procedure, an initial model scale estimation has been acquired. In this stage, the accuracy of scale estimation is largely increased by a probabilistic scheme. Given an estimated model position and a list of matched scene lines, the actual perpendicular distances p_i between the model position and lines are calculated. These distances are compared with d_i (from the previous step) giving $a_i = \frac{d_i}{p_i}$ indicating the scale estimations. By entering all the a_i into a 1D *Hough* space, a peak a_p is obtained as the best scale estimation of the model. Thus, the control signal is derived as $U_X = \lambda \cdot (1 - a_p)$, with λ an arbitrary number, being as the proportional control gain.

3.3 Global Navigation

Given a topological map, the global navigation aims to guide a robot to travel between the current node towards a target node. This is achieved by iteratively conducting local navigation using visual references from consecutive nodes, until the target is reached. A node is expected to be reached by the robot, if all the 3 spatial constraints are smaller than pre-defined threshold values. In addition, the robot orientation is assumed to be known from an on-board compass. Once the robot reaches a node, it then rotates to the direction towards the next node. The 3D edge model which is best aligned with this direction is chosen as the visual reference for the next local navigation. On the topological map, between the starting node and a target node, an optimal path is determined by the A^* algorithm providing a path with the least number of nodes.

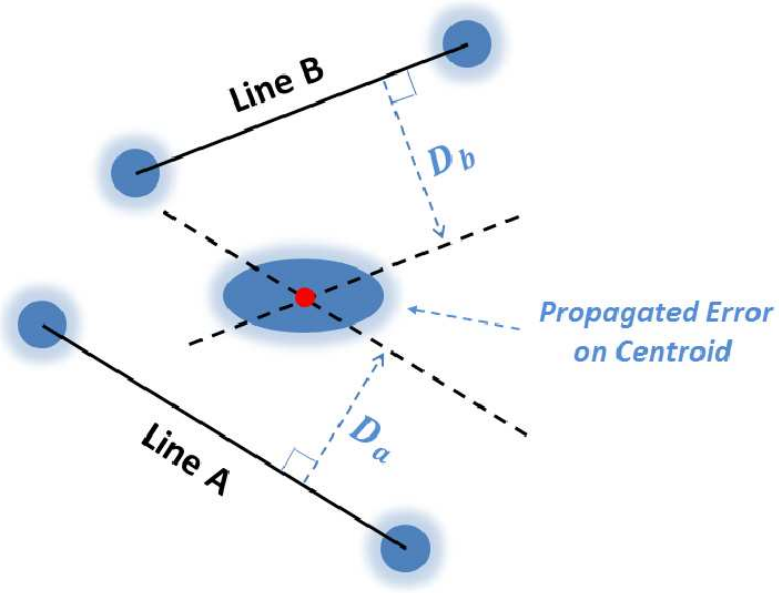

Figure 5: *Line fitting error propagating to centroid position estimation.*

4 Simulation Evaluation

A novel wire-frame presented simulation environment [17] is utilised for quantitative evaluation of the proposed navigation method. Multiple real-world uncertainty sources are modelled and implemented in the simulation system, including image noise, lateral edge shifting, edge detection loss, geometric approximation error, corner detection loss, corner match error, stereo match error and camera calibration error. The environment contains over 7000 features of a real-world work space with the volume of approximately 20m $\times$ 15m $\times$ 2.5m. 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. Such a simulation environment provides ground truth information with realistic visual data, suitable for quantitative evaluation and optimisation of vision-based robotic systems. Figure 6 shows multiple examples of the original environment and the uncertainty perturbations.

A topological map that has 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, as shown in Figure 7. At each node, 3D models are taken at 12 equally separated view-angles (i.e. 30^o respectively). These 3D models are perturbed by simulated uncertainty sources with specific magnitudes. In the simulation experiments, the robot starts at a known position and attempts to reach a target topological node through a calculated optimal path. Consecutive nodes are employed to iteratively navigate the robot towards the ultimate goal. Multiple navigation experiments are conducted by choosing various starting and targeting positions. Several visual recognition and localisation results are demonstrated in Figure 8, where clear overlaps between the models and sampled scenes can be observed. The resultant motion trajectories from the navigation experiments are shown in Figure 9. Moreover, the mean residual between the robot poses and the approached nodes is provided in Table 1. It shows that the target nodes are reached through optimal paths on the topological map, while all the 3 robot DOFs are constrained with good accuracy. In addition, small tracking residual is obtained between the robot motion and the topological paths, which demonstrates the robustness of the proposed navigation method under all the simulated uncertainty perturbations.

5 Conclusion

A stereo-vision-based autonomous navigation method is presented in this paper, which aims to undertake tasks in unstructured environments. The method utilises PGHs for the representation and recognition of environment appearances, generating a topological database across a spatial range. The visual recognitions are used in a feedback scheme to iteratively constraint the robot pose towards the desired values. The proposed method is capable of ex-

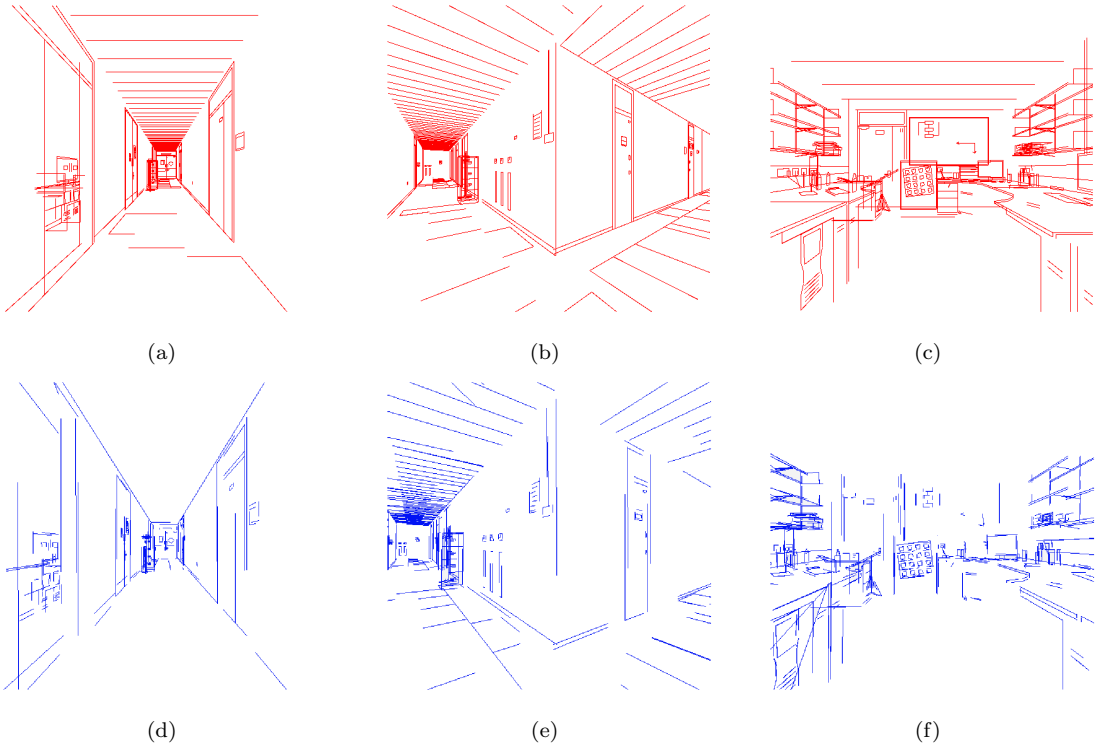

Figure 6: *(a - c) the ground truth and (d - f) the perturbed environment under all uncertainty sources.*

Table 1: Mean residual between robot poses and approached nodes

Experiments	ΔX (cm)	ΔY (cm)	$\Delta\theta$ (°)
1^{st} (black)	19.24 ±25.29	5.18 ±4.70	0.28 ±0.25
2^{nd} (pink)	17.06 ±17.37	7.54 ±5.71	0.56 ±0.84
3^{rd} (brown)	10.97 ±7.24	3.27 ±1.15	0.29 ±0.20
4^{th} (red)	6.25 ±6.66	3.07 ±2.42	0.27 ±0.32

tracting 3 spatial constraints from a single visual recognition, which satisfies the control requirement of holonomic robots with 3 DOFs. Given a starting position and a target location, the robot is able to conduct autonomous navigation through an optimal topological path. A novel wire-frame represented simulation environment is employed for quantitative evaluation of the proposed navigation system. The simulation system is designed to represent the real-world geometric characteristics and the associated possible uncertainty sources. A dense topological map with 72 nodes is utilised for simulation evaluation and performance analysis. The simulation results demonstrate a high navigation stability with good accuracy, while preserving robustness towards all the simulated uncertainty sources.

Future work include an optimisation of the existed system design via a large scale Monte-Carlo analysis. This requires a large number of simulations to be conducted using various control parameter settings, in order to acquire reliable statistic property for performance quantification. An optimal parameter setting can be obtained which provides good accuracy and high efficiency. The quantified system performance also allows direct comparison between algorithm structures, to guide further system improvements prior to physical implementation on real-world robotic systems.

References

[1] S. I. Roumeliotis and G. A. Bekey, "Collective localization: A distributed kalman filter approach to localization of groups of mobile robots," in *ICRA'00*, vol. 3, 2000, pp. 2958–2965.

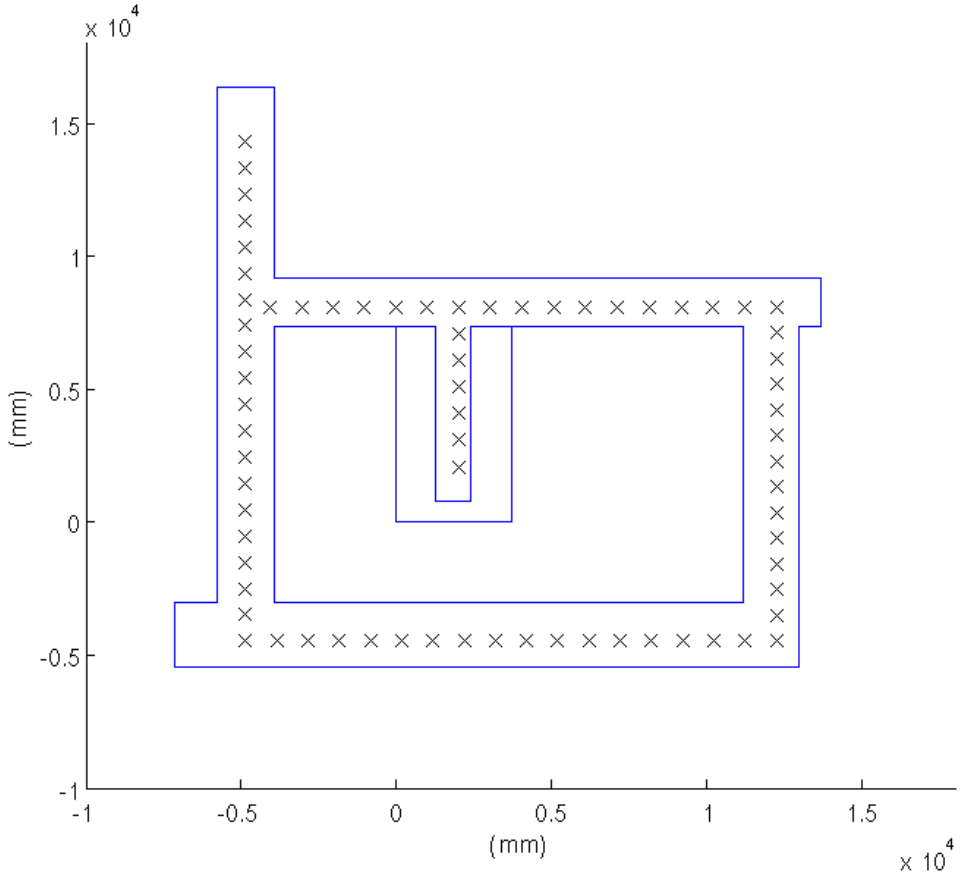

Figure 7: *A topological map with 72 nodes, covering the majority space.*

[2] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte carlo localization for mobile robots," in *ICRA'99*, vol. 2, 1999, pp. 1322–1328.

[3] T. Goedemé, M. Nuttin, T. Tuytelaars, and L. Van Gool, "Omnidirectional vision based topological navigation," *International Journal of Computer Vision*, vol. 74, no. 3, pp. 219–236, 2007.

[4] L. Maohai, W. Han, S. Lining, and C. Zesu, "Robust omnidirectional mobile robot topological navigation system using omnidirectional vision," *Engineering applications of artificial intelligence*, vol. 26, no. 8, pp. 1942–1952, 2013.

[5] H. Cheng, H. Chen, and Y. Liu, "Topological indoor localization and navigation for autonomous mobile robot," *Automation Science and Engineering, IEEE Transactions on*, vol. 12, no. 2, pp. 729–738, 2015.

[6] F. Amorós, L. Paya, O. Reinoso, D. Valiente, and L. Fernandez, "Towards relative altitude estimation in topological navigation tasks using the global appearance of visual information," in *VISAPP'14*, vol. 1, 2014, pp. 194–201.

[7] N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor, "Omni-directional vision for robot navigation," in *Omnidirectional Vision, 2000. Proceedings. IEEE Workshop on*, 2000, pp. 21–28.

[8] L. DG, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 8, no. 6, pp. 679–698, 1986.

[10] S. Crossley, "Robust temporal stereo computer vision," Ph.D. dissertation, University of Sheffield, 2000.

[11] A. Ashbrook, N. Thacker, P. Rockett, and C. Brown, "Robust recognition of scaled shapes using pairwise geometric histograms." in *BMVC'95*, 1995, pp. 503–512.

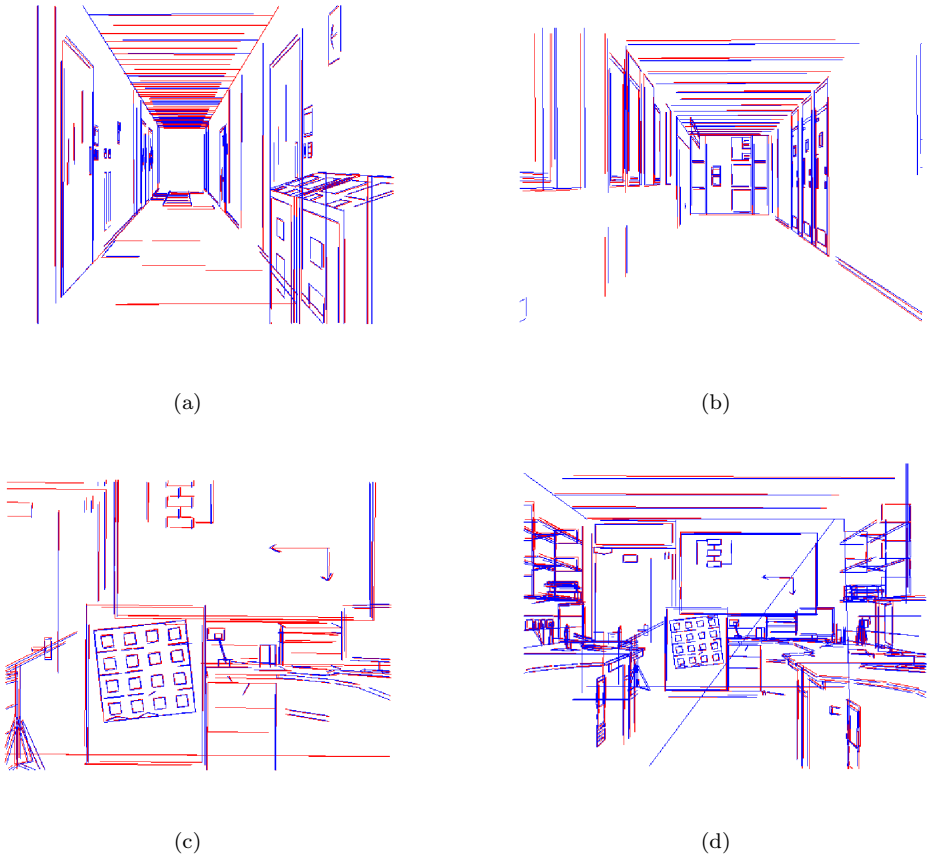

Figure 8: *Visual recognition and localisation between models (red) and sampled scenes (blue).*

[12] A. Evans, N. Thacker, and J. Mayhew, "The use of geometric histograms for model-based object recognition." in *BMVC'93*, 1993, pp. 429–438.

[13] N. Thacker, P. Riocreux, and R. Yates, "Assessing the completeness properties of pairwise geometric histograms," *Image and Vision Computing*, vol. 13, no. 5, pp. 423–429, 1995.

[14] J. Pinel, *Biopsychology*. Pearson Education, 1997.

[15] S. Coupe, "Machine learning of projected 3d shape," Ph.D. dissertation, University of Manchester, 2009.

[16] F. Aherne, N. Thacker, and P. Rockett, "The bhattacharyya metric as an absolute similarity measure for frequency coded data," *Kybernetika*, vol. 34, no. 4, pp. 363–368, 1998.

[17] J. Tian, "Quantitative optimisation of a vision-based robotic localisation and navigation algorithm," http://www.tina-vision.net, [Online]. Available: http://www.tina-vision.net/docs/memos.php. [Accessed 16-May-2016].

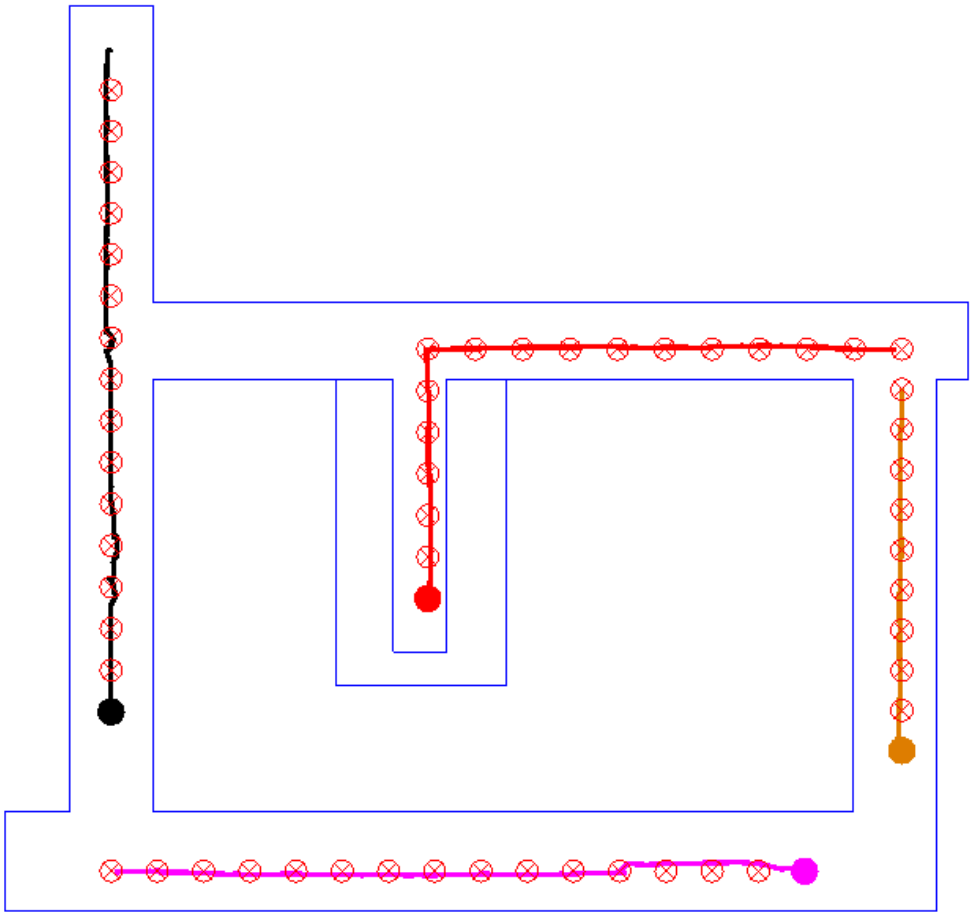

Figure 9: *The motion trajectories of multiple navigation experiments, where the filled dots represent the target nodes.*