Improving 3D Keypoint Detection from Noisy Data using Growing Neural Gas

Jose Garcia-Rodriguez[1], Miguel Cazorla[2], Sergio Orts-Escolano[1], and Vicente Morell[2]

[1] Department of Computing Technology at University of Alicante. Spain
[2] Instituto de Investigación en Informática at University of Alicante. Spain
{jgarcia,sorts}@dtic.ua.es
{miguel,vmorell}@dccia.ua.es

Abstract. 3D sensors provides valuable information for mobile robotic tasks like scene classification or object recognition, but these sensors often produce noisy data that makes impossible applying classical keypoint detection and feature extraction techniques. Therefore, noise removal and downsampling have become essential steps in 3D data processing. In this work, we propose the use of a 3D filtering and down-sampling technique based on a Growing Neural Gas (GNG) network. GNG method is able to deal with outliers presents in the input data. These features allows to represent 3D spaces, obtaining an induced Delaunay Triangulation of the input space. Experiments show how the state-of-the-art keypoint detectors improve their performance using GNG output representation as input data. Descriptors extracted on improved keypoints perform better matching in robotics applications as 3D scene registration.

Keywords: GNG, Noisy Point Cloud, Visual Features, Keypoint Detection, Filtering, 3D Scene Registration

1 Introduction

Historically, humans have the ability to recognize an environment they had visit before based on the 3D model we unconsciously build in our heads based on the different perspectives of the scene. This 3D model is built with some extra information so that humans can extract relevant features [1] that will help in future experiences to recognize the environment and even possible objects presents there. This learning method has been transferred to mobile robotics field over the years. So, most current approaches in scene understanding and visual recognition are based on the same principle: keypoint detection and feature extraction on the perceived environment. Over the years most efforts in this area have been made towards feature extraction and keypoint detection on information obtained by traditional image sensors [2], existing a gap in feature-based approaches that use 3D sensors as input devices. However, in recent years, the number of jobs concerned with 3D data processing has increased considerably due to the emergence of cheap 3D sensors capable of providing a real time data stream and therefore enabling feature-based computation of three dimensional environment properties like curvature, getting closer to human learning processes.

The Kinect device[3], the time-of-flight camera SR4000[4] or the LMS-200 Sick laser[5] are examples of these devices. Besides, providing 3D information, some of these devices like the Kinect sensor can also provide color information of the observed scene. However, using 3D information in order to perform visual recognition and scene understanding is not an easy task. The data provided by these devices is often noisy and therefore classical approaches extended from 2D to 3D space do not work correctly. The same occurs to 3D methods applied historically on synthetic and noise-free data. Applying these methods to partial views that contains noisy data and outliers produces bad keypoint detection and hence computed features does not contain effective descriptions.

Classical filtering techniques like median or mean have been used widely to filter noisy point clouds [3] obtained from 3D sensors like the ones mentioned above. The median filter is one of the simplest and wide-spread filters that has been applied. It is simple to implement and efficient but can remove noise only if the noisy pixels occupy less than one half of the neighbourhood area. Moreover, it removes noise but at the expense of removing detail of the input data.

Another filtering technique frequently used in point cloud noise removal is the Voxel Grid method. The Voxel Grid filtering technique is based on the input space sampling using a grid of 3D voxels to reduce the number of points. This technique has been used traditionally in the area of computer graphics to subdivide the input space and reduce the number of points [4]. The Voxel Grid method presents some drawbacks: geometric information loss due to the reduction of the points inside a voxel and sensitivity to noisy input spaces.

Based on the Growing Neural Gas [5] network several authors proposed related approaches for surface reconstruction applications [6]. However, most of these contributions do not take in account noisy data obtained from RGB-D cameras using instead noise-free CAD models.

In this paper, we propose the use of a 3D filtering and down-sampling technique based on the GNG [5] network. By means of a competitive learning, it makes an adaptation of the reference vectors of the neurons as well as the interconnection network among them; obtaining a mapping that tries to preserve the topology of an input space. Besides, GNG method is able to deal with outliers presents in the input data. These features allows to represent 3D spaces, obtaining an induced Delaunay Triangulation of the input space very useful to easily obtain features like corners, edges and so on. Filtered point cloud produced by the GNG method is used as an input of many state-of-the-art 3D keypoint detectors in order to show how the filtered and down sampled point cloud improves keypoint detection and hence feature extraction and matching in 3D registration methods.

[3] Kinect for XBox 360: http://www.xbox.com/kinect Microsoft

[4] Time-of-Flight camera SR4000 http://www.mesa-imaging.ch/prodview4k.php

[5] LMS-200 Sick laser: http://robots.mobilerobots.com/wiki/SICK_LMS-200_Laser_Rangefinder

In this work we focus on the processing of 3D information provided by the Kinect sensor. Because the Kinect is essentially a stereo camera, the expected error on its depth measurements is proportional to the squared distance to the scene.

The rest of the paper is organized as follows: first, a section describing briefly the GNG algorithm is presented. In section 3 the state-of-the-art 3D keypoint detectors are explained. In section 4 we present some experiments and discuss results obtained using our novel approach. Finally, in section 5 we give our conclusions and directions for future work.

2 GNG Algorithm

With Growing Neural Gas (GNG) [5] method a growth process takes place from minimal network size and new units are inserted successively using a particular type of vector quantization. To determine where to insert new units, local error measures are gathered during the adaptation process and each new unit is inserted near the unit which has the highest accumulated error. At each adaptation step a connection between the winner and the second-nearest unit is created as dictated by the competitive Hebbian learning algorithm. This is continued until an ending condition is fulfilled, as for example evaluation of the optimal network topology or fixed number of neurons. The network is specified as:

- A set N of nodes (neurons). Each neuron $c \in N$ has its associated reference vector $w_c \in R^d$. The reference vectors can be regarded as positions in the input space of their corresponding neurons.
- A set of edges (connections) between pairs of neurons. These connections are not weighted and its purpose is to define the topological structure. An edge aging scheme is used to remove connections that are invalid due to the motion of the neuron during the adaptation process.

This method offers further benefits over simple noise removal and downsampling algorithms: due to the incremental adaptation of the GNG, input space denoising and filtering is performed in such a way that only concise properties of the point cloud are reflected in the output representation.

3 Applying Keypoint Detection Algorithms to Filtered Point Clouds

In this section, we present the state-of-the-art 3D keypoint detectors used to test and measure the improvement achieved using GNG method to filter and downsample the input data. In addition, we explain main 3D descriptors and feature correspondence matching methods that we use in our experiments.

First keypoint detector used is the widely known SIFT (Scale Invariant Feature Transform) [7] method. It performs a local pixel appearance analysis at different scales. SIFT features are designed to be invariant to image scale and rotation. SIFT detector has been traditionally used in 2D image but it has been extended to 3D space. 3D implementation of SIFT differs from original in the use of depth as the intensity value.

Another keypoint detector used is based on a classical HARRIS 2D keypoint detector. In [8] a refined HARRIS detector is presented in order to detect keypoints invariable to affine transformations. keypoints. 3D implementations of these HARRIS detectors use surface normals of 3D points instead of 2D gradient images. Harris detector and its variants (Tomasi3D and Noble3D) have been tested in Section 4.

Once keypoints have been detected, it is necessary to extract a description over these points. In the last few years some descriptors that take advantage of 3D information have been presented. In [9] a pure 3D descriptor is presented. It is called Fast Point Feature Histograms (FPFH) and is based on a histogram of the differences of angles between the normals of the neighbour points. This method is a fast refinement of the Point Feature Histogram (PFH) that computes its own normal directions and it represents all the pair point normal diferences instead of the subset of these pairs which includes the keypoint. Moreover, we used another descriptor called CSHOT [10] that is a histogram that represents the shape and the texture of the keypoint. It uses the EVD of the scattered matrix using neighborhood for each point and a spherical grid to encode spatial information.

Correspondence between features or feature matching methods are commonly based on the euclidean distances between feature descriptors. One of the most used method to find the transformation between pairs of matched correspondences is based on the RANSAC (RANdom SAmple Consensus) algorithm [11]. It is an iterative method that estimates the parameters of a mathematical model from a set of observed data which contains outliers. In our case, we have used this method to search a 3D transformation (our model) which best explain the data (matches between 3D features). At each iteration of the algorithm, a subset of data elements (matches) is randomly selected. These elements are considered as inliers; a model (3D transformation) is fitted to those elements, the rest of the data is then tested against the fitted model and included as inliers if its error is below a threshold; if the estimated model is reasonably good (its error is low enough and it has enough matches), it is considered as a good solution. This process is repeated a number of iterations and the best solution is returned.

4 Experimentation

We performed different experiments on real data to evaluate the effectiveness and robustness of the proposed method. First, a normal estimation method is computed in order to show how estimated normals are considerably affected by noisy data. Finally, the proposed method is applied to 3D scene registration to show how keypoint detection methods are improved obtaining more accurate transformations. To validate our approach, experiments were performed on a dataset comprised of 90 overlapped partial views of a room. Partial views are rotated 4 degrees in order to cover 360 degrees of the scene. Partial views were captured using the Kinect device mounted in a robotic arm with the aim of knowing the ground truth transformation. Experiments implementation, 3D data management (data structures) and their visualization have been done using the PCL[6] library.

4.1 Improving normal estimation

In the first experiment, we computed normals on raw and filtered point clouds using the proposed method. Since normal estimation methods based on the analysis of the eigenvectors and eigenvalues of a covariance matrix created from the nearest neighbours are very sensitive to noisy data. This experiment is performed in order to show how a simple 3D feature like normal or curvature estimation can be affected by the presence of noise. In Figure 1 it is visually explained the effect caused by normal estimation on noisy data.

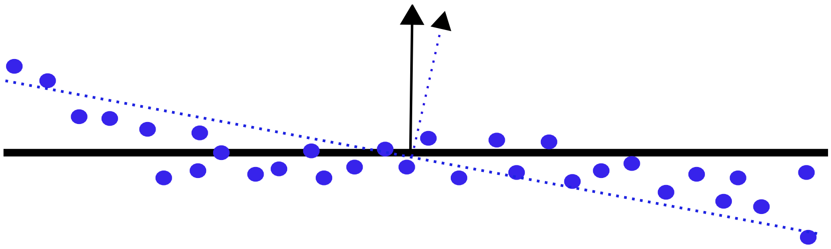

Fig. 1. Noise causes error in the estimated normal

Normal estimation is computed on the original and filtered point cloud using the same radius search: $r_s = 0.1$ (meters). In Figure 2 can be observed how more stable normals are estimated using filtered point cloud produced by the GNG method. $20,000$ neurons and $1,000$ patterns are used as configuration parameters for the GNG method in the filtered point cloud showed in Figure 2 (Right).

4.2 Improving 3D keypoint detectors performance

In the second experiment, we used some keypoint detectors introduced in section 3 in order to test noise reduction capabilities of the GNG method. RMS deviation measure calculates the average difference between two affine transformations: the ground truth and the estimated one. Furthermore, we used a fixed number of neurons and patterns to obtain a downsampled and filtered representation of the input space. Different configurations have been tested, ranging from 5,000 to 30,000 neurons and 250 to 2,000 patterns per epoch. Figure 3

[6] The Point Cloud Library (or PCL) is a large scale, open project [12] for 2D/3D image and point cloud processing.

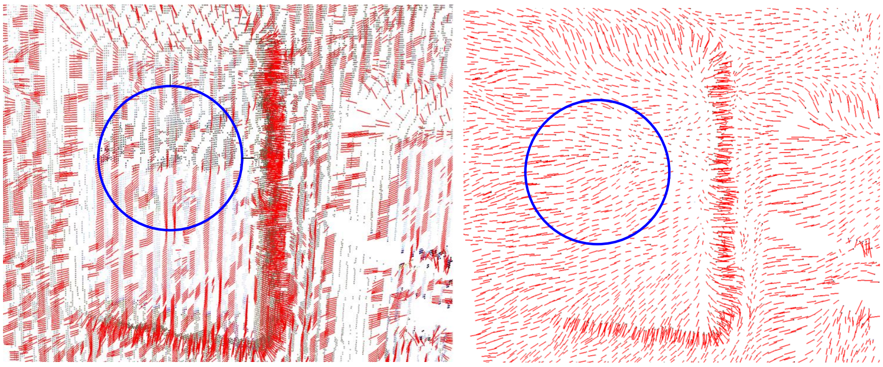

Fig. 2. Normal estimation comparison. Left: Normal estimation on raw point cloud. Right: Normal estimation on filtered point cloud produced by the GNG method

shows correspondences matching calculated over filtered point clouds using the proposed method.

In Table 1 we can see how using GNG output representation as input cloud for the registration step, lower RMS errors are obtained in most detector-descriptor combinations.

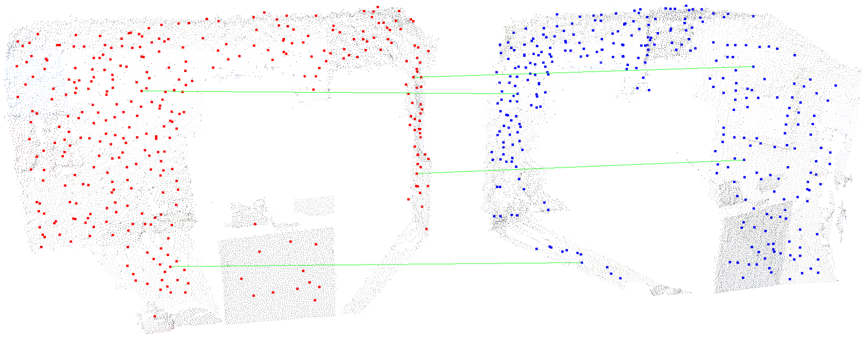

Fig. 3. Registration example done with the HARRIS3D detector and the FPFH descriptor using a GNG representation with 20000 neurons.

Experiments are performed using different search radius for keypoint detection and feature extraction methods. Search radius influences directly on the size of the extracted features, making methods more robust against occlusions. A balance between large and small values must be found depending on the size

Keypoint detector search radius = 0.1; Feature extractor search radius = 0.2

	SIFT3D/FPFH	**SIFT3D/SHOTRGB**	**HARRIS3D/FPFH**	**HARRIS3D/SHOTRGB**
Raw point cloud	0.1568	0.0359	0.3074	0.0490
GNG 20000n 1000p	0.1842	**0.0310**	**0.1158**	0.0687
GNG 10000n 500p	0.1553	0.0677	0.1526	**0.0466**
GNG 5000n 250p	**0.0435**	0.0957	0.1549	0.0566

	Tomasi3D/FPFH	**Tomasi3D/SHOTRGB**	**Noble3D/FPFH**	**Noble3D/SHOTRGB**
Raw point cloud	0.3604	0.0764	0.3074	0.0655
GNG 20000n 1000p	2.2816	0.0416	**0.1095**	**0.0653**
GNG 10000n 500p	**0.1783**	**0.0349**	0.1526	0.0790
GNG 5000n 250p	0.1808	0.0730	0.3438	0.0925

Keypoint detector search radius = 0.05; Feature extractor search radius = 0.2

	SIFT3D/FPFH	**SIFT3D/SHOTRGB**	**HARRIS3D/FPFH**	**HARRIS3D/SHOTRGB**
Raw point cloud	0.1568	0.0359	0.0329	**0.0362**
GNG 20000n 1000p	0.1842	**0.0310**	0.1170	0.0415
GNG 10000n 500p	0.1553	0.0677	**0.0769**	0.0626
GNG 5000n 250p	**0.0435**	0.0957	0.1549	0.0566

	Tomasi3D/FPFH	**Tomasi3D/SHOTRGB**	**Noble3D/FPFH**	**Noble3D/SHOTRGB**
Raw point cloud	0.1482	0.0532	0.1059	0.0991
GNG 20000n 1000p	**0.0702**	0.0666	0.2234	0.0518
GNG 10000n 500p	0.1027	**0.0257**	**0.0323**	**0.0196**
GNG 5000n 250p	0.1110	0.0446	0.0600	0.0327

Table 1. RMS deviation error in meters obtained using different detector-descriptor combinations. Different combinations are computed on the original point cloud (raw), and three different filtered point clouds using the proposed method. GNG output representation produces lower RMS errors in most detector-descriptor combinations.

of the presents objects in the scene and the size of the features we want to extract. For the used dataset, the best estimated transformations are found using keypoint detector search radius 0.1 and 0.05 and feature extractor search radius 0.2.

Experiments shown in Table 1 demonstrate how the proposed method achieves lower RMS deviation errors in the computed transformation between different 3D scene views. For example, the computed transformation is improved using the GNG representation as input for the Noble3D detector and SHOTRGB feature descriptor combination, obtaining a more accurate transformation. For the same combination, the proposed method obtains less than 2 centimetres error whereas original point cloud produces almost 9 centimetres error in the registration process.

5 Conclusions and future work

In this paper we have presented a method which is able to deal with noisy 3D data captured using low cost sensors like the Kinect Device. The proposed method calculates a GNG network over the raw point cloud, providing a 3D structure which has less information than the original 3D data, but keeping the 3D topology. It is shown how state-of-the-art keypoint detection algorithms perform better on filtered point clouds using the proposed method. Improved keypoint detectors are tested in a 3D scene registration process, obtaining lower RMS errors in most detector-descriptor combinations. Future work includes the integration of the proposed filtering method in a indoor mobile robot localization application.

References

1. Anne M. Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, January 1980.
2. M. Szummer and R.W. Picard. Indoor-outdoor image classification. In *Content-Based Access of Image and Video Database, 1998. Proceedings., 1998 IEEE International Workshop*, pages 42 –51, jan 1998.
3. Andreas Nuchter, Hartmut Surmann, Kai Lingemann, Joachim Hertzberg, and S. Thrun. 6d slam with an application in autonomous mine mapping. In *In Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1998–2003, 2004.
4. C. Connolly. Cumulative generation of octree models from range data. In *Robotics and Automation. Proceedings. 1984 IEEE International Conference on*, volume 1, pages 25 – 32, mar 1984.
5. B. Fritzke. *A Growing Neural Gas Network Learns Topologies*, volume 7, pages 625–632. MIT Press, 1995.
6. Y. Holdstein and A. Fischer. Three-dimensional surface reconstruction using meshing growing neural gas (mgng). *Vis. Comput.*, 24(4):295–302, March 2008.
7. David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
8. Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. In Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen, editors, *Computer Vision âĂŤ ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 128–142. Springer Berlin Heidelberg, 2002.
9. Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 3212 –3217, may 2009.
10. F. Tombari, S. Salti, and L. Di Stefano. A combined texture-shape descriptor for enhanced 3d feature matching. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 809 –812, sept. 2011.
11. Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.
12. Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.