Visual Odometry based on Pulse-Coupled Neural Network

Luis Armando Cruz Prospero, Luis Armando Mejía-Lavalle
CENIDET
Mexico
lacp@cenidet.edu.mx

ABSTRACT: *Visual Odometry novel technique is presented. It work through the abstraction of the information contained in images in different times. Each pair of images is processed by a Pulse-Coupled Neural Network to perform the extraction signature describing the scene and thereby estimating the position of a mobile robot in a route. The results of proposed method is comparable with a traditional method.*

Keywords: Visual Odometry, Pulse-Coupled Neural Network, Binocular Vision, Signatures

Received: 10 October 2016, Revised 18 November 2016, Accepted 23 November 2016

1. Introduction

In recent years autonomous robotic navigation has become in a very important technique and is seen as a vital tool to tap the huge potential of robots in both military and civilian areas such as medicine, surveillance and rescue [1].

One of the main challenges of robotic navigation is to identify the location where the robot is on a path, this to make sure it follows the predetermined path to its destination. The process of estimating the robot location is known as odometry [2].

A wide range of devices has been used to calculate odometry including: Internal Navigation Systems or Inertial Measurement Units, Global Positioning Systems (GPS) among others [2]. However, it still has weaknesses that have prompted the search for new strategies.

Recently robotic navigation techniques have favored the use of vision sensors because of its size, weight, low power consumption and low price [3].

Visual odometry (VO) appear with implementation of vision sensors in robotic navigation techniques. VO is a potential answer

to face the needs and improved perception of mobile robots.

There are different techniques of Artificial Vision that have been proposed to date for the implementation of the VO. Within these techniques are those that have been developed with monocular vision systems and others that implemented stereo cameras [3].

This paper proposes calculating VO of stereo image taken from a video sequence acquired by binocular arrangement using Pulse-Coupled Neural Network (PCNN) to abstraction and description of the scenes given.

PCNN was born as suitable for digital image processing neural networks due to this paradigm is based on neuronal function of the visual cortex of mammals [4]. Therefore is submitted a new approach based on modified PCNN which simplifies the number of variables involved in the equations that define the network and allows to implement new ways of generating features that make it more robust to inherent changes in vision systems. This neural model is called Optimized Pulse-Coupled Neural Network (O-PCNN) [5].

The proposal is described in each section describing the characteristics reported in the literature and the results obtained in the experimental algorithm described.

In Section 2 a broader overview of the background and operating process of OV is presented. Section 3 presents the PCNN, describing the neuronal model and its operating mechanism. Then in Section 4 is describing the implementation of the proposed model while in Section 5 shows results of experimentation. Section 6 provides a discussion of the advantages and disadvantages of the proposed method and finally in Section 7 are presented the conclusions and future work of this research.

2. Visual Odometry

Odometry is the estimate of continuous change of the robot position in a series of steps over time. It is calculated from a collection of data collected by sensors during movement of the robot. In short, odometry it is used by mobile robots to estimate their relative position to a starting point in real time [6].

Odometry is widely used in systems of mobile robots. For its calculation have been implemented various sensors to collect the necessary data. The most representative sensor for this purpose is enconder rotation that is used to measure the distance of movement on wheels [7]. Other sensors such as gyroscope, accelerometer and GPS have been incorporated. Although these sensors may have direct measurements of the robot pose, they contain errors in the measurements that can drift over time [8].

VO is presented as a new alternative which implement visual sensors (cameras) for estimating the pose. The visual sensors represent an efficient alternative because they are more advanced technologies with low cost [9].

VO algorithms use sequential data over time, unlike traditional odometry, these extracted information from one or more cameras for tracking features [10].

Compared with traditional odometry, the VO has the following advantages [11]:

• Avoiding the inaccurate readings on encoders, error accumulation and other reasons when wheel robot slipping on smooth flat.

• It does not require the movement and the environmental priori information, only rely on the estimated camera motion information to determine the position and orientation of the robot, so this can ensure real-time and cost-effective.

• VO information is obtained through visual sensors, and contains the rich image characteristic information that is advantageous to the processing, convenient to target recognition, object detection and 3D scene reconstruction maps and other combined tasks to provide more adequate support for the synchronization of robot navigation and mapping.

3. Pulse-Coupled Neural Network

PCNN is a relatively new paradigm which in the field of Artificial Vision has shown potential in different aspects of digital image processing such as: segmentation, image denoising and feature extraction [12].

The first Pulse-Coupled Neural Model is known as Eckhorn Model, this model is biologically inspired by the observation of biological visual cortex; Eckhorn et al. discovered that the visual cortex creates binary images in an oscillatory manner in which different features are obtained in order to create the current image in the brain [13].

Eckhorn Model has an image as input which is grouped in pixels based on spatial proximity and similarity of brightness by the model. During the grouping, the model keeps small space gaps and minor local variations of intensity. This is a very desirable feature for image processing applications based on window. However the Eckhorn model has certain properties that decrease its utility in applications of digital image processing [14]:

• Mathematical analysis of the model is a difficult task.

• Grouping image pixels based on spatial proximity and similarity brightness is ambiguous.

• Determining the appropriate parameters for the model can be difficult.

The research above pointed new targets that search for relevant modifications would create a computer model that adapted the original, but at the same time mitigate the limitations that the Eckhorn Model had.

Johnson et al. propose modifications and variations to the Eckhorn Model with the aim of adapting its performance as algorithms of digital image processing. This new model is called PCNN which is an Artificial Neural Network (ANN) with a single two-dimensional layer [13].

PCNN is formed by pulse-coupled neurons virtually connected to the pixels of the input image with 1:1 correspondence (Figure 1). Because each image pixel is associated with a PCNN neuron, the network structure is constructed from the structure of the input image to be processed. [12].

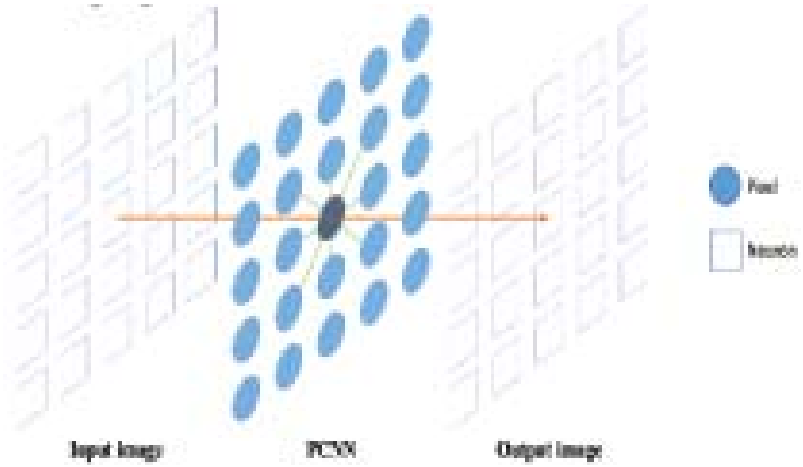

Figure 1. Neuron-Pixel correspondence

3.1 Optimized Pulse-Coupled Neural Network

PCNN has capacity to work on issues such as classification and image recognition, however the development of this paradigm has led to provide new modified models with the aim of improving the original model, at least in the following features [15]:

• Decreasing the computational demands of PCNN algorithm.

• The number of PCNN parameters reduction.

• Improving the quality of generated binary images that are used for object detection in image space and for image segmentation.

• Improving the invariance against geometrical transforms in images and noise invariance.

One of this modified PCNN is Optimized Pulse-Coupled Neural Network proposed by Forgáþ [15]. O-PCNN works reducing the

generating number of generated features to reach high image recognition performance.

In O-PCNN the feeding potential *Fij[n]* is defined by the intensity pixel *Sij* only while the linking potential *Lij[n]* is defined only by the convolution matrix *K[n]*. The internal activity of a O-PCNN neuron is shown in Figure 2 and is defined by the following equations [16]:

$$F_{IJ}[n] = S_{i,j} \tag{1}$$

$$K_{i,j}[n] = \Sigma_{kl}\, m_{ijkl} Y_{kl}[n\text{-}1]\, X_{kl}[n\text{-}1] \tag{2}$$

$$L_{i,j}[n] = K_{i,j}[n] \tag{3}$$

$$U_{i,j}[n] = F_{i,j}[n] * (1+ \beta L_{i,j}[n]) \tag{4}$$

$$X_{i,j}[n] = \frac{1}{1+e^{\,Ti,j\,[n\text{-}1]\,-\,Ui,j\,[n]}} \tag{5}$$

$$Y_{i,j}[n] = \begin{cases} 1 \text{ si } X_{i,j}[n] > 0.5 \\ 0 \quad \text{en otro caso} \end{cases} \tag{6}$$

$$T_{i,j}[n] = \begin{cases} V_T \text{sì} 1 \text{ si } Y_{i,j}[n] = 1 \\ \alpha_0 \quad \text{en otro caso} \end{cases} \tag{7}$$

where $F_{ij}[n]$ is input feeding potential and $L_{i,j}$ [n] input linking potential of neuron. S_{ij} is an intensity of pixel *i,j* in the input image which represents the intensity of given image element. Matrix elements *K[n]* are calculated by convolution of matrix *m*. $X_{i,j}$ is activation quantity of neuron that is defined by sigmoid transfer function, Y_{ij} is activation quantity of neuron that is defined by step function. $U_{i,j}$ is the activation quantity of neuron, T_{ij} is threshold potential of neuron, n is iteration step. Parameter ±o is threshold decay coefficient, β is linking coefficient and parameter V_T is coefficient of the threshold potential.

According to the advantages presented by the O-PCNN in this paper, the study of odometry will based on this neural paradigm as an alternative to using traditional methods of Artificial Vision.

4. Generating features using O-PCNN

The standard approach of feature generation by O-PCNN is based on series of virtual binary images generation. These binary images are produced by activity neurons for each iteration step. These series are representative features known as signature

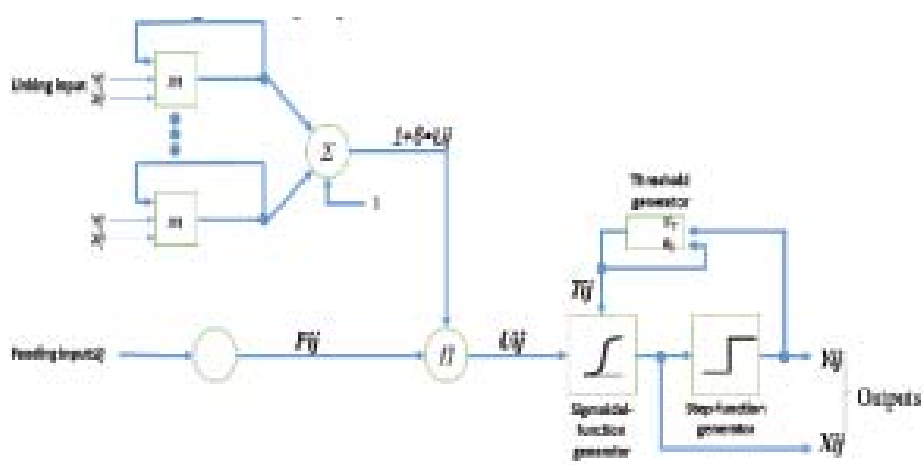

Figure 2. Diagram of O-PCNN neuron

which could be invariant to geometric changes due to the appearance of visualization, translation, scaling, rotation and brightness [17].

An example of an image signature is shown in Figure 3.

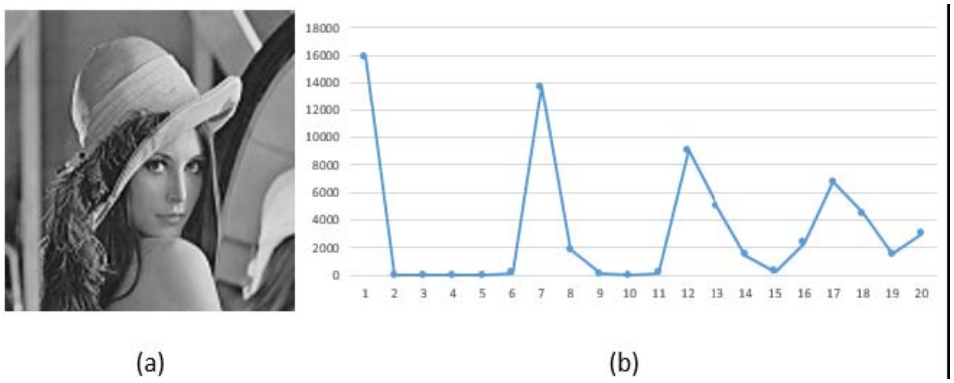

Figure 3. (a) Lenna image, (b) signature of (a)

The standard generating features *G[n]* in a defined number of iterations *i* is done by calculating the sum of activated neurons in each iteration network [18].

$$G[n] = \Sigma_{ij} Y_{i,j}[n] \tag{8}$$

However this method is affected by geometric transformations that hinder the process in the image recognition by signatures. When Forgáþ [15] introduced the O-PCNN provided new ways for generating signatures where not only the sum of activated neurons in different iterations is involved but also the acceptance of the X_{ij} from the signature generation process may introduce new methods.

Abul-Ela [18] proposed a new generating features method where the resulting signature is robust in large background or non-uniform light distribution as found in robot navigation process.

$$G[n] = \frac{\Sigma_{ij}(Y_{i,j}[n] * X_{i,j}[n] * FC}{\Sigma_{ij} S_{i,j}} \tag{9}$$

Where CF (Continuity Factor) is define by [18]:

First, we define an operator which is sensitive to the pixel intensity change, the operator happens to be the gradient. If the image is regarded as a function of two variables *A(x,y)*, then the gradient is defined as:

$$\nabla x2A = A(x+1, y) - A(x-1, y) \tag{10}$$

$$\nabla y2A = A(x, y+1) - A(x, y-1) \tag{11}$$

Calculate the continuity response (*Cr*), at any given pixel by the magnitude of the gradient vector witch is the length of the hypotenuse of the right triangle having sides (∇x and ∇y), and this reflects the strength of the continuity.

$$Cr = \sqrt{\nabla x2A^2 + \nabla x2A^2} \tag{12}$$

Cr range values are 0, 1 and 1.414. This value can be normalized and producing the *CF* by computing the following equation.

$$CF = Cr / \sqrt{2} \tag{13}$$

O-PCNN is applied to all the image to recognize the scene and thus knowing the position on the route. Also this concept is applied in singular *keypoints* in the image for tracking features.

5. Implementation

5.1 Experiment

To verify the qualities of the propose method in VO task is implemented methodology shown in Figure 4 where in part A there are 2 images in different time as input. Then in part B are applied SURF as feature detector and O-PCNN as general descriptor of the scenes and also works as singular descriptor of each *keypoint* obtained by SURF. The descriptions of the *keypoints* are given by each neuron activity obtained with Equation 14:

$$D_{ij}[n] = U_{ij}[n]\, X_{ij}[n]\, CF \tag{14}$$

Finally in part C the descriptions are matched by Brute Force algorithm and then the odometry is computed.

To do the experimentation they were used images of [19] database which were pre-processed to obtained images in grayscale of 128x128 pixels. The code was developed in C++.

5.2 Results

The analysis of results was done using 110 frames where each frame is a different time (*T*) in the experiments.

With the aim to be clear the results are divided to three parts. The first part point at the skill of the O-PCNN in image recognition where experimentation is made up of 34 frames classification in 5 key scenes of the path formed by the 110 frames. All of this using a Multi-Layer Perceptron as classifier. The result indicated a 97.1% of accuracy and it is shown in Confusion Matrix of Figure 5.

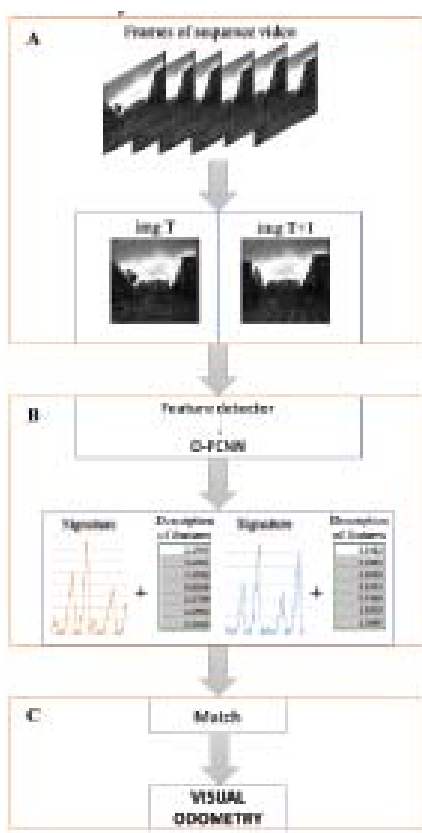

Figure 4. Implemented methodology

Second part is about the matching process where the accuracy was measured. The experimentation is made up of the evaluated of the 5 key scenes matching description with the next frames after. In Table 1 there are the results of matching using O-PCNN and SURF as descriptors.

Table 1 shown that the O-PCNN has a similar accuracy than the SURF algorithm. The different between their accuracy is 0.46% % where the SURF is better.

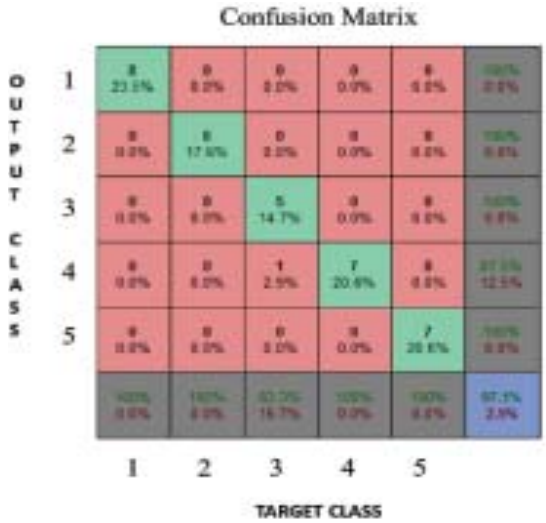

Figure 5. Confusion Matrix of classification

Algorithm	Accuracy O-PCNN	Accuracy SURF
KEY SCENE		
1 . . . ***T+21***	100%	100%
2 . . . ***T+21***	99.75%	99.93%
2 . . . ***T+21***	99.81%	99.98%
2 . . . ***T+21***	99.06%	99.5%
2 . . . ***T+21***	98.85%	99.3%
Total	**99.49%**	**99.74%**

Table 1. Algorithms performance

Finally the results of VO are shown is Figure 6. These results are proportional to the results obtained in the matching process because VO depends on it.

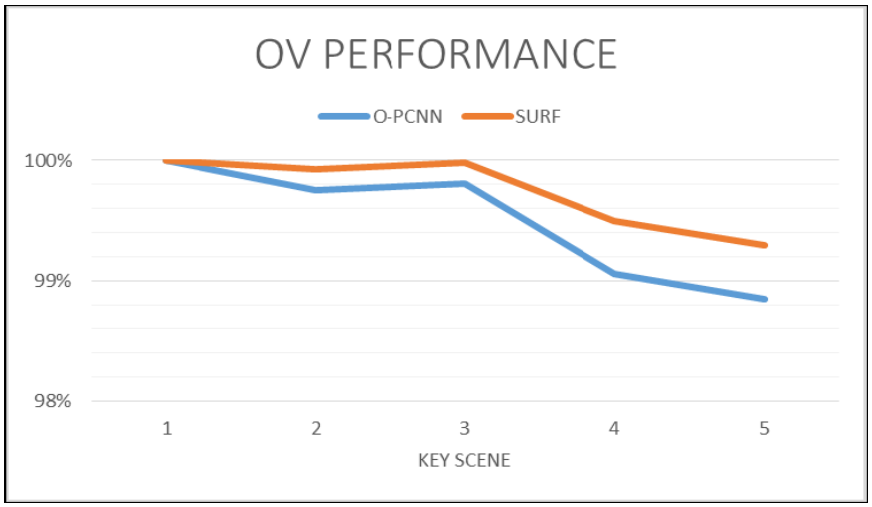

Figure 6. OV performance

The analysis of Figure 6 indicated that when the descriptions of an image in time *T* are matched with the descriptions of an image in time *T+n* where *n* is distant number the quality of matching decreases due to the objects that appear or disappear in image *T+n*. This affect both O-PCNN and SURF algorithms.

6. Discussion and conclusions

O-PCNN presents results that are comparable with those obtained by SURF. However SURF has 0.46% better performance than O-PCNN in OV task.

In another hand O-PCNN brings a recognition of the scene like is shown in part B of the methodology which with the description of the *keypoints* permit to know the place and the distance traveled by the robot at the same time.

Also worth mention that the O-PCNN belongs to the third generation of ANN which it is a technique in development where there are many study fields that can improve its performance.

7. Future work

Because this is a first contact with O-PCNN in VO task is necessary to continue with this research and experimentation with the aim of improve its performance which allow to find new methods of description.

In the same way is necessary work with some configuration of the O-PCNN that permit detect and describe *keypoints* by itself due to this could improve its accuracy and its general performance.

8. References

[1] Premaratne, P., Safaei, F. (2008). “Feature based Stereo Correspondence using Moment Invariant”, 4th International Conference on Information and Automation for Sustainability. Colombo.

[2] Mahmood, A., Baig, A., Ahsan, Q. (2012). “Real time Localization of Mobile Robotic Platform via fusion of Inertial and Visual Navigation System”, International Conference on Robotics and Artificial Intelligence (ICRAI), 2012.

[3] Atsushi, S., Tamura, Y., Kuroda, Y. (2010). “Visual Odometry Using Feature Point and Ground Plane for Urban Environments”, VDE VERLAG GMBH ’” Berlin ’” Offenbach.

[4] Eckhorn, R., Reitboeck, H., Arndt, M., Dicke, P. (1990). Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Comput*, vol. 2, p. 293–307.

[5] Forgáþ, R., Mokriš, I. (2008). Feature Generation Improving by Optimized PCNN, *In:* SAMI 2008 6th International Symposium on Applied Machine Intelligence and Informatics, 2008.

[6] Jha, A., Kumar, M. (2014). Two Wheels Differential Type Odometry for Mobile Robots, *In:* 3rd International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2014.

[7] Liang, Z., Xu, X. (2014). Improved Visual Odometry Method for Matching 3D Point Cloud Data, Control Conference (CCC), 33rd Chinese.

[8] Pei, Y., Kleeman, L. (2015). Online Robot Odometry Calibration over Multiple Regions Classified by Floor Colour, *In:* Control Conference (CCC), 2015 IEEE International Conference on Information and Automation.

[9] Choi, S., Hoon, J., Yu, W., Cho, J. (2011). What Does Ground Tell Us? Monocular Visual Odometry under Planar Motion Constraint, *In:* 2011 11th International Conference on Control, Automation and Systems (ICCAS).

[10] Ozawa, R., Takaika, Y., Kida, Y. et al., (2005). Using Visual Odometry to Create 3D Maps for Online Footstep Planning, *In:* 2005 IEEE International Conference on Systems, Man and Cybernetics.

[11] Liang, Z., Xu., X. (2014).Improved Visual Odometry Method for Matching 3D Point Cloud Data, *In:* 2014 33rd Chinese Control Conference (CCC).

[12] Wang, Z., Ma, Y., Cheng, F., Yang., L. (2010). Review of pulse-coupled neural network, Elsevier Imagen and vision computer, 28 (1) 5-13, Jan 2010.

[13] Eckhorn, R., Reitboeck, H., Arndt, J., Dicke., P. (1990). Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex, *Neural Comput.*, V. 2, p. 293–307.

[14] Kuntimad, G., Ranganath, H. (1999). Perfect Image Segmentation Using Pulse Coupled Neural Networks, *IEEE Transactions on Neural Networks*, 10 (3) May 1999.

[15] Forgáþ, R., Mokriš, I. (2008). Feature Generation Improving by Optimized PCNN, *In:*6th International Symposium on Applied Machine Intelligence and Informatics.

[16] Forgáþ, R., Mokriš, I. (2009). Algorithm for Pulse Coupled Neural Network Parameters Estimation, *In:* IEEE International Conference on Computational Cybernetics.

[17] Johnson, J. (1994). Time signatures of images, IEEE World Congress on Computational Intelligence, 1994.

[18] Abul-Ela, M., Samir, A., Tolba, M. (2011). Neutralizing Lighting non-homogeneity and Background Size in peNN Image Signature, *In*: International Conference on Computer Engineering & Systems (ICCES).

[19] Blanco, J. (2013). The Málaga Stereo and Laser Urban Data Set. MRPT Empowering C++ development in robotics.