2011 11th International Conference on Control, Automation and Systems
Oct. 26-29, 2011 in KINTEX, Gyeonggi-do, Korea

Vision based Autonomous Vehicle Navigation with Self-Organizing Map Feature Matching Technique

Kajal Sharma[1], Kwang-young Jeong[2] and Sung-Gaun Kim[3]

[1] Division of Mechanical and Automotive Engineering, Kongju National University, Cheonan, 331-717, Korea
(Tel : +82-41-521-9557; E-mail: kajal175@gmail.com)

[2] Division of Mechanical and Automotive Engineering, Kongju National University, Cheonan, 331-717, Korea
(Tel : +81-41-521-9250; E-mail: kyjeong@kongju.ac.kr)

[3] Division of Mechanical and Automotive Engineering, Kongju National University, Cheonan, 331-717, Korea
(Tel : +81-41-521-9253; E-mail: kimsg@kongju.ac.kr)

Abstract: Vision is becoming more and more common in applications such as localization, autonomous navigation, path finding and many other computer vision applications. This paper presents an improved technique for feature matching in the stereo images captured by the autonomous vehicle. The Scale Invariant Feature Transform (SIFT) algorithm is used to extract distinctive invariant features from images but this algorithm has a high complexity and a long computational time. In order to reduce the computation time, this paper proposes a SIFT improvement technique based on a Self-Organizing Map (SOM) to perform the matching procedure more efficiently for feature matching problems. Experimental results on real stereo images show that the proposed algorithm performs feature group matching with lower computation time than the original SIFT algorithm. The results showing improvement over the original SIFT are validated through matching examples between different pairs of stereo images. The proposed algorithm can be applied to stereo vision based autonomous vehicle navigation for obstacle avoidance, as well as many other feature matching and computer vision applications.

Keywords: Stereo Vision, Autonomous Vehicle, Feature Matching, SIFT, Self-Organizing Map.

1. INTRODUCTION

In recent years, due to the fast developments of computer vision techniques, studies on vision-based autonomous vehicle navigation have high prominence because of their potential in various applications. Vision based systems are employed in a wide range of robotic applications such as object recognition, obstacle avoidance, vehicle navigation and, more recently, in Simultaneous Localization and Mapping (SLAM) [1-4].

A variety of keypoint detectors have been developed in efforts to solve the problem of extracting points of interest in image sequences, Shi and Tomasi [5], SIFT [6], Speed Up Robust Features (SURF) [7] descriptor, affine covariant etc. These works mainly employ the same approach: extraction of points that represents regions with high intensity gradients. Schmid and Mohr [8] used Harris corners to show that invariant local feature matching could be extended to general image recognition problems wherein a feature is matched against a large database of images. Although these methods are capable to find qualitative correspondences, most of them are too slow to use in real-time autonomous vehicle navigation application. Lowe [6] overcomes such problems by detecting the points of interest over the image and scaling through the locations of the local extrema in a pyramidal Difference of Gaussians (DOG). The Lowe's features are invariant to image translation, scaling, rotation, and partially invariant to illumination changes and affine or 3D projection. Though having the aforementioned advantages, the SIFT algorithm has high complexity and requires long computational time, resulting in slow image matching speed. Some extensions of the SIFT descriptor have been proposed recently, for instance, Mikolajczyk and Schmid [9] experimentally compared the performances of several currently used local descriptors and found that the SIFT descriptors were the most effective, as they yielded the best matching results.

This paper improves the SIFT algorithm matching process, reduces the computational time, and solves the real-time problem. The work presented in this paper demonstrates increased matching process performance robustness with minimization of the computational time with the use of a Self-Organizing Map (SOM). The effectiveness of self-organizing map based feature matching has motivated the authors to use it for obstacle avoidance in autonomous vehicle navigation applications. It can be applied to a variety of computer vision problems based on feature matching including machine vision, object recognition, image retrieval and many others.

The most relevant contribution of this paper is the proposal of an improved time efficient SIFT feature matching technique based on Kohonen's Self-Organizing Map (SOM) neural network methodology in order to reduce the computation time while providing efficient feature matching. The interpretation of visual information and feature matching is done using self-organizing map introduced by Kohonen [10]. The unsupervised learning feature of the SOM is used to find out the winner neuron and matching is performed by associating similar winning pixels in the left and the right images of stereo pair images. Through experiments it was found that the proposed self-organizing map based feature matching yields highly distinctive feature matching which in comparison to other algorithms are faster to compute and provides better matching between different stereo images.

978-89-93215-03-8 98560/11/$15 ©ICROS

The remainder of this paper is organized as follows: Section 2 presents a brief introduction of SIFT algorithm. Section 3 presents a detailed view of the proposed Self-Organizing Map (SOM) based feature matching technique. Experimental results using stereo images and a discussion of the findings are presented in Section 4. Finally, we conclude and discuss future work in Section 5.

2. SCALE INVARIANT FEATURE TRANSFORM (SIFT)

Scale Invariant Feature Transform (SIFT) is a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene. The features are invariant to image scale and rotation, and are shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination [6]. To estimate the feature in stereo pair images, the image features that can be matched between the left and right images must be found. The following outlines the major stages of computation used to generate the set of image features.

2.1 Scale-space extrema detection

The first stage identifies key locations in scale space by looking for locations that are maxima or minima of a difference-of-Gaussian function.

2.2 Keypoint localization

The location and scale of each keypoint are estimated and unstable keypoints are eliminated.

2.3 Orientation assignment

One or more orientations are assigned to each keypoint location based on local image gradient directions.

2.4 Keypoint descriptor

The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

3. SELF-ORGANIZING MAP BASED FEATURE MATCHING

A Self-Organizing Map (SOM) is an unsupervised algorithm to map high-dimensional data to a lower dimensional space through a competitive and unsupervised learning process. In this paper, the feature vectors extracted from the Scale-Invariant Feature Transform (SIFT) are used to compose a topological map. The topological and metric relationship is obtained using a Self-Organizing Map (SOM) that performs vector quantization and simultaneously organizes the quantized vectors on a regular low-dimensional grid. In the proposed approach, a bumblebee stereo camera is used to provide a stereo view of the environment. Consecutive image pairs acquired by the camera are then matched to estimate the matching between the pair of images. Indeed, due to the high complexity and the long computational time of the original SIFT algorithm, a significant advance is the possibility of reducing computation time by improving the matching process using a Self-Organizing Map (SOM). Our methodology operates with scale invariant feature vectors instead of an image database for the input to the Self-Organizing Map (SOM). The vectors extracted from the SIFT feature are used to compose a topological map. The map is obtained using a self-organizing map based on the Kohonen neural network. We thereby obtain a 2D neuron grid. Each neuron is associated with a weight vector with 128 element descriptors. During the matching, feature vectors are presented as the inputs to the Self-Organizing Map (SOM). The learning algorithm is based on the concept of nearest neighbor learning. Once the network is trained, input data are distributed throughout the grid of neurons. The left image is considered as the reference image and the right image is considered as the matching image in the stereo pair, and they are expressed in terms of the winning neurons in the self-organizing map network. The next step is finding the winning neuron for each pixel of the right image. When the winning neuron is found, the pixel is associated to it. The matching is done between the pixels in the left and the right stereo pair images and feature matching is performed by associating the similar winning pixels. The same procedure is then performed on the pixels of the left image. In this step, after finding the winning neuron, it is also computed which pixel of the right image is the most similar to the pixels of the left image. The winning data is chosen by this procedure by matching the similar pixels in the left image and the right image. The matching between the pixels of the pair of stereo images is accomplished by iteratively following this procedure.

The Scale-Invariant Feature Transform (SIFT) is a well-known method to provide a set of keypoints detected in the scale-space that are characterized by a descriptor invariant to scale and orientation. For each pair of images, we detect the point of interest, compute SIFT descriptors, and perform stereo matching with the Self-Organizing Map (SOM). Both sets of features are the input to the self-organizing map, which computes stereo matching. Next, the feature association stage performs matching between sets of features that belong to the acquired stereo images.

Let us consider the set of input variables {xi} defined as the real vector X={x1, x2, x3…xk}∈ Rn. This input pattern vector is applied to the processing elements of the input layer. The processing elements of the input layer are connected with wi = [wi1, wi2….. win]T∈ Rn each element in the SOM grid. This grid contains the feedback layer region. We associate connection strength to every processing element of the feedback layer. The initial value of the WT is selected randomly. The input feature pattern vector X is then applied to the processing units of the input layer of the self-organizing map as shown in Fig. 1. The linear output of these processing units feeds the weighted input through feed forward connection to the SOM grid.

The activation of the jth process unit of the feedback layer can be represented by Eq. (1):

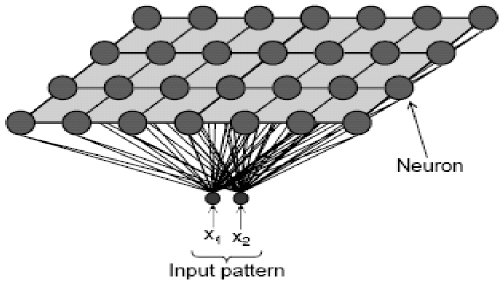


Fig. 1 Self-Organizing Map [10]

$$yi = \sum_{i=1}^{K} w_{ij} x_i \tag{1}$$

where j =1 to N (number of units in the feedback layer)

When a new input arrives, the topological map determines the neuron that best matches the input vector. A winning unit in Eq. (2), say P, corresponds to the minimal distance, and represents the pixel in the right image that could be a match in the left image selected among all the processing units of the feedback layer.

$$\sum_{i=1}^{K} (x_i - w_{P_i}) = \min \sum_{i=1}^{n} (x_i - w_{j_i}) \tag{2}$$

; for all j

Hence, during learning, the nodes that are topographically close to a certain geometric distance will activate each other to learn from the same input vector X and the weights associated with the winning unit P and its neighboring units r are updated by Eq. (3):

$$w_{iP}(t+1) = w_{iP}(t) + \lambda(P,r)[x_i(t) - w_{ij}(t)] \tag{3}$$

for i =1 to K and P =1 to N; here the λ (P, r) is the neighborhood function and it can be represented as:

$$\lambda(P,r) = \alpha(t).\exp[-\frac{\left\| R_P - R_r \right\|^2}{2\sigma^2(t)}]$$

where R_P refers to the position of the Pth unit in the grid, $\alpha(t)$ is the learning rate factor ($0< \alpha(t) <1$), and the parameter $\sigma(t)$ defines the width of the Gaussian function. $\sigma(t)$ gradually decreases so as to reduce the neighborhood region in successive iterations of the training process.

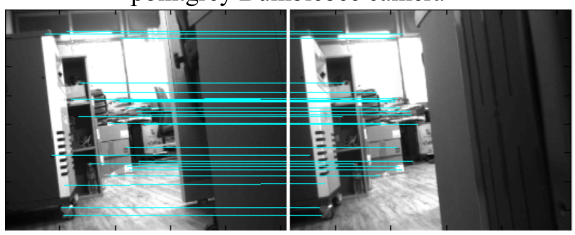
If the winning neuron is found, then we can perform matching between the pixels in the left and the right images. The SOM arranges feature vectors according to their internal similarity, creating a continuous topological map of the input space. After completing the steps above, feature matching between different image pairs acquired by the stereo vision camera can be evaluated. In order to compare the real feature matching performance, different tests are performed, and they are described in the next section.

4. EXPERIMENTAL RESULTS AND DISCUSSION

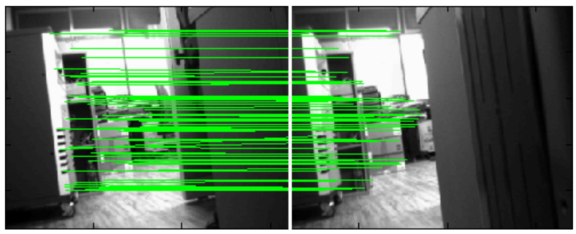
In our experiment, we use the Bumblebee 2 stereo vision camera for capturing the stereo images. The main advantage of using this camera system is that it comes mechanically pre-calibrated for lens distortion and camera misalignments. The Bumblebee stereo vision camera uses two Sony progressive scan CCDs, each with a HFOV up to 100 degrees, and communicates via an IEEE 1394 connection. It has a 12cm baseline and it can output 640x480 images at 48 FPS, or 1024x768 images at 20 FPS, via its IEEE-1394 (FireWire) interface.

The left and right stereo pair images are captured with the Bumblebee pointgrey camera and feature matching between left and right stereo pair images is performed with the original SIFT algorithm and the modified Self-Organizing Map (SOM) based feature matching algorithm. The left and right images captured by the stereo vision bumblebee pointgrey camera are shown in Fig. 2. Fig. 3 show the features that were obtained from one pair of stereo images acquired by our vision system with the original SIFT matching algorithm and also show the results obtained with the improved self-organizing map based feature matching algorithm.

Fig. 2 Left and right images captured by stereo vision pointgrey Bumblebee camera



(a) Original SIFT- 30 features matched in 0.07813 sec



(b) Self-Organizing Map based feature matching- 63 features matched in 0.03719 sec

Fig. 3 (a) and (b) shows the result obtained with the original SIFT matching algorithm and the result obtained with the improved self-organizing map based feature matching algorithm.

Experimental results on real stereo images show that the proposed algorithm performs feature group matching in a more time efficient manner than the original SIFT algorithm. The results of the SIFT improvement are validated through matching examples between different pairs of stereo images. Comparisons of the results obtained with the original SIFT matching algorithm and the improved Self-Organizing Map (SOM) based feature matching are shown in Table 1.

Table 1 Comparison of the results obtained with the original SIFT algorithm and Self-Organizing Map (SOM) based feature matching algorithm

Stereo Pair Images	Original SIFT		SOM based feature Matching	
	Time(s)	Matched Features	Time(s)	Matched Features
Pair 1	0.07813	30	0.03719	63
Pair 2	0.09130	19	0.03441	29
Pair 3	0.04816	14	0.03214	25
Pair 4	0.10483	68	0.07022	120
Pair 5	0.06730	39	0.03352	51
Pair 6	0.05773	16	0.03576	28
Pair 7	0.04773	16	0.04483	25
Pair 8	0.04374	8	0.04281	9
Pair 9	0.05611	51	0.05321	74

Two main experiments were conducted to identify the differences between the original SIFT and the proposed self-organizing map based feature matching. Fig. 4 shows the performance results using the original SIFT algorithm and the self-organizing map based feature matching on different sets of real time stereo images. The proposed approach results in reduced processing time for matching of the stereo images with efficient feature matching.

Fig. 4 Comparison graph showing the computation time for calculation of each stereo pair with the original SIFT algorithm and the modified self-organizing map based feature matching algorithm

5. CONCLUSION AND FUTURE WORK

This work proposed a new approach to feature matching in stereo images with visual odometry. The original SIFT algorithm developed by Lowe was modified by using the self-organizing map. The proposed approach was evaluated using a set of real scenarios, with different sets of stereo pair images. The results showed the advantages of the self-organizing map in relation to the original SIFT in terms of time performance. The experimental results show the effectiveness of the proposed approach. Considering time efficiency, our approach can be used as a foundation for the future development of stereo vision and path finding applications in the robotic field.

ACKNOWLEDGMENT

This work was supported by the research grant of Kongju National University and Brain Korea 21.

REFERENCES

[1] S. Belongie, J. Malik, and J. Puzicha, “Shape Matching and Object Recognition Using Shape Contexts,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, pp. 509-521, 2002.

[2] K. Hosoda, K. Sakamoto and M. Asada, “Trajectory generation for obstacle avoidance of uncalibrated stereo visual servoing without 3D reconstruction,” *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vol. 1, pp. 29-34, 1995.

[3] D. Murray and J. Little, “Using real-time stereo vision for mobile robot navigation,” *Journal of Autonomous Robots*, Vol. 8, No. 2, pp. 161-171, 2000.

[4] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, “MonoSLAM: Real-time single camera SLAM,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 6, pp. 1052-1067, 2007.

[5] J. Shi and C. Tomasi, “Good Features to Track,” *9th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600.

[6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, Vol. 60, No.2, pp. 91-110, 2004.

[7] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: Speeded Up Robust Features,” *Proc. of the European Conference on Computer Vision*, pp. 404-417, 2006.

[8] C. Schmid, and R. Mohr, “Local grayvalue invariants for image retrieval,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, 530-534, 1997.

[9] K. Mikolajczyk and C. Schmid, “Scale & affine invariant interest point detectors,” *International Journal of Computer Vision*, Vol. 60, No. 1, pp. 63-86, 2004.

[10] T. Kohonen, “The Self-Organizing Map,” *Proc. of the IEEE*, Vol. 78, No. 9, pp. 1464-1480, 1990.