15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

A Visual-SLAM for First Person Vision and Mobile Robots

Takahiro TERASHIMA
Department of Computational Intelligence and Systems Science, Tokyo institute of technology
J3-13, Nagatsuta 4259, Midori-ku, Yokohama, 226-8503 Japan
terashima.t.aa@m.titech.ac.jp

Osamu HASEGAWA
School of Engineering, Tokyo institute of technology
J3-13, Nagatsuta 4259, Midori-ku, Yokohama, 226-8503 Japan
contact@soinn.com

Abstract

SLAM(Simultaneous Localization and Mapping) is one of the core subjects in computer vision and robotics. In order to avoid the effects of noise, SLAM systems need devises to remove the moving object such as human beings and cars in real-world environment. In this paper, we propose a method which excludes dynamic features and generate a map in crowded environment, called ICGM2.5. Experiments were conducted in indoor and outdoor crowded real environments. Experimental results show that our approach has superior performace compared to conventional approaches in terms of accuracy.

1 Introduction

SLAM (Simultaneous Localization and Mapping) is indispensable for mobile robots moving in unknown indoor and/or underground environments without a map. And Visual-SLAM is considered as SLAM with a simple system configuration. The method uses only cameras as external sensors and executes the SLAM only from the image information. Conventionally, Visual-SLAM systems are executed in environments where dynamic objects such as humans and cars do not exist. This is because accuracy descends by confusingly register dynamic objects unrelated to the environmental map as landmarks. [6] However, because the real-world environments are dynamic, mobile robots need to have a Visual-SLAM which can maintain accuracy in those environments. In this paper, we propose a Visual-SLAM system which is robust in dynamic real-world environments.

2 Related research

Studies that use only hand-held cameras to maintain the accuracy of Visual-SLAM in dynamic environments are few due to the restrictions. In order to suppress the influence from dynamic objects and accurately match the local feature, some methods of dividing local features obtained from an image into dynamic features and static features have been proposed. Kawewong et al have proposed PIRF (Position Invariant Robust Features) [3], and that showed better accuracy in dynamic environments[6].

Hua et al have proposed another approach named ICGM (Incremental Center of Gravity Matching) [2]. In ICGM, the centroids of static local features are first calculated, and vectors from the centroid to each local features are obtained. These vectors are compared with those of the previous frame, and if the vectors to each local features differ between frames, this local features are identified as dynamic local features.

For example, as in figure 1, features A, B, C, D, E exist in image I_t, and corresponding features A', B', C', D', E' exist in image I_{t-1}. Also, features A, B and C are static features, and it is unknown whether D and E are dynamic or static feature. The same applies to corresponding features. In this case, ICGM first obtains the center of gravity O of the static features A, B, C and similarly obtains the centroid O' of A', B', C'. Next, vectors from the centers of gravity O, O' to the features D, D', E, E' are calculated. Here, the vectors of $\overrightarrow{OD}$ and $\overrightarrow{O'D'}$ are different between frames, and the vectors of $\overrightarrow{OE}$ and $\overrightarrow{O'E'}$ are equal between frames. Therefore, feature D is identified as a dynamic feature and feature E as a static feature.

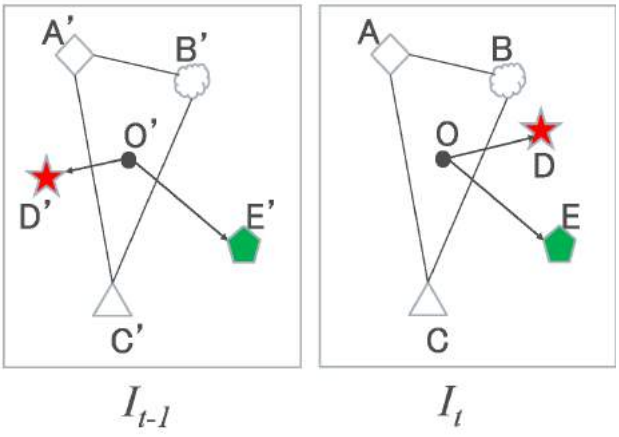


Figure 1. Concept of ICGM. Find the center of gravity O of the known static features A, B, C and calculate the vector from O to the unknown feature. If this vector differs between preceding and succeeding frames, that feature is identified as a dynamic feature.

Moreover, Kayanuma et al proposed ICGM 2.0 [1]. This is a method for improving the accuracy of ICGM by performing PIRF as preprocessing for ICGM, calculating the centroid after reducing the dynamic feature in advance. As a result, SLAM with higher accuracy than PIRF and ICGM is realized.

3 Proposed method

The purpose of this research is to divide dynamic features and static features in images and to perform highly accurate Visual-SLAM even in dynamic environments such as crowds. In consideration of ease of application to the system, the external sensor to be used is only a hand-held monocular camera. Therefore, in this research, we focus on ICGM 2.0 that can perform highly accurate Visual-SLAM even under these constraints, and propose ICGM 2.5 with improved accuracy.

3.1 ICGM2.5

ICGM 2.5 has the same concept as ICGM 2.0 and improves the problem on the algorithm. The outline is as follows.

First, as a problem on the algorithm of ICGM 2.0, there is a possibility that static features can not be selected as features to be used when obtaining the position of the center of gravity. Under the premise that known static features exist, ICGM calculates the centroid of randomly chosen features from them. However, in the real world it is unknown whether or not specific features are static features, so there is a possibility that system can not divide it well by choosing dynamic features at the time of center of gravity calculation. Therefore, in the proposed method, dynamic features are reduced in advance before centroid calculation of static local features. This approach reduces the possibility of including dynamic features among the local features used to determine the position of the center of gravity. The proposed method selects a feature with a shorter matching distance between the images I_t and I_{t-1} as a local feature for calculating the center of gravity to further improve the accuracy. The matching distance is an index indicating the degree of similarity between two matched features and is obtained from the descriptor of each feature. We sort the features in ascending order of matching distances and select k features from short ones to calculate the centroid of ICGM. With these improvements, the proposed method realizes accuracy improvement from the conventional method.

Next, we show the calculation of the centroid position and the deletion of the dynamic features after the static features are properly chosen. Given that CG is the position of the center of gravity and p_i is the coordinate of the local features determined to be static features, the position of the center of gravity can be obtained from the equation 1.

$$CG = \begin{bmatrix} X \\ Y \end{bmatrix} = \frac{1}{k}\sum_{i=0}^{k} p_i \tag{1}$$

Here, k is the number of static features used for center of gravity calculation, and in this study k = 5. Also, the elements of the vectors to the coordinates P of the arbitrary local features from the centroid position CG are defined as the equation 2.

$$CGV = CG - P = \frac{1}{k}\sum_{i=0}^{k} p_i - P \tag{2}$$

Algorithm 1 Algorithm of ICGM2.5

Require:
N : Number of sequential image
n_i : Number of local features of i^{th} image
$P_i = (p_{1,i}, p_{2,i}, ..., p_{n_{i-1},i}, p_{n_i,i})$: Set of local features in i^{th} image
$Dist_i = (dist_{1,i}, dist_{2,i}, ..., dist_{n_{i-1},i}, dist_{n_i,i})$: Set of matching distances in i^{th} image
for $i = 1$ **to** N **do**
  $GoodCenterOfGravity \leftarrow false$
  while $GoodCenterOfGravity = false$ **do**
    $s_1, s_2..., s_k \leftarrow$ Subscript of min($Dist_i$), second min($Dist_i$),..., k^{th} min($Dist_i$)
    $CG_i \leftarrow (p_{s_1,i} + ... + p_{s_k,i})/k$
    $CG_{i-1} \leftarrow (p_{s_1,i-1} + ... + p_{s_k,i-1})/k$
    $CGV_i \leftarrow CG_i - (p_{s_1,i}, ..., p_{s_k,i})$
    $CGV_{i-1} \leftarrow CG_{i-1} - (p_{s_1,i-1}, ..., p_{s_k,i-1})$
    Delete min($Dist_i$), second min($Dist_i$),..., k^{th} min($Dist_i$) from $Dist_i$
    if $RoD \leq Thr_{CG}$ **then**
      $GoodCenterOfGravity \leftarrow true$
      Delete $p_{s_1,i}, ..., p_{s_k,i}$ from P_i
      Delete $p_{s_1,i-1}, ..., p_{s_k,i-1}$ from P_{i-1}
    end if
  end while
  $CGV_i \leftarrow CG_i - P_i$
  $CGV_{i-1} \leftarrow CG_{i-1} - P_{i-1}$
  for $j = 1$ **to** n_i **do**
    if $RoD \leq Thr_{ICGM}$ **then**
      $P_i^{ICGM2.5} \leftarrow p_{j,i}$
      $P_{i-1}^{ICGM2.5} \leftarrow p_{j,i-1}$
    end if
  end for
end for

From this equation, in consecutive frames I_{t-1} and I_t, Vectors CGV_{T-1} and CGV_T representing the relationship between the centroid position CG and each features p are obtained. Using these vectors, calculate the ratio of difference RoD by the following equation 3.

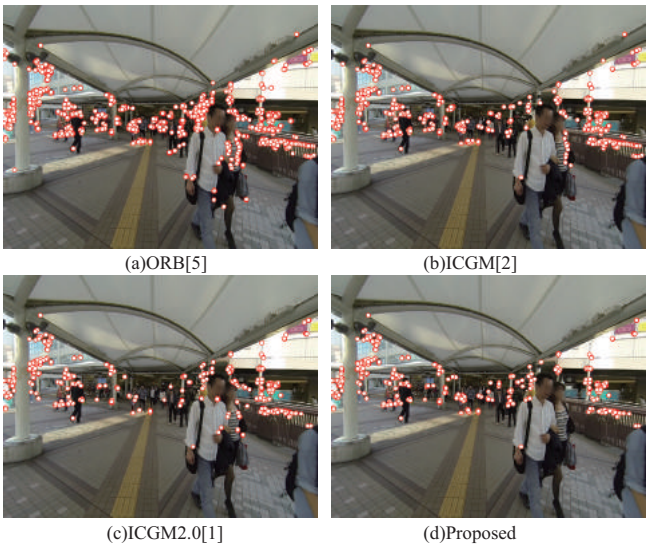
$$RoD = \frac{||CGV_T - CGV_{T-1}||}{||CGV_T|| + ||CGV_{T-1}||} \tag{3}$$

By comparing this RoD with the threshold Thr_{ICGM}, it is distinguished whether each feature is static features or dynamic features. $RoD \leq Thr_{ICGM}$ is identified as static features. On the other hand, the feature $RoD > Thr_{ICGM}$ is identified as dynamic features and deleted. Details of the algorithm of ICGM 2.5 are shown in Algorithm 1.

3.2 Deletion of dynamic feature

In this research, ORB [5] is used as the local feature. Figure 2 shows the comparison of the result of dynamic feature reduction using the proposed method and conventional method from this local feature. (A) shows a state in which no dynamic feature deletion processing is performed, and (b) shows a state in which dynamic feature deletion processing is performed by ICGM. (C) shows the state by ICGM 2.0, and (d) shows the state by ICGM 2.5 which is the proposed method. (a) extracts a lot of features from human beings which are dynamic elements in the image, (b), (c) and (d) reduce

dynamic features from (a). Among them, the proposed method reduces the dynamic features compared to (b) and (c), and the effectiveness of the proposed method can be confirmed.



(a)ORB[5] (b)ICGM[2]

(c)ICGM2.0[1] (d)Proposed

Figure 2. Comparison of feature points. (a) extracts feature points as they are. (b), (c) and (d) show how the feature points extracted from the human being, which is dynamic elements in the image, is reduced by each method. It can be confirmed that the proposed method shown in (d) most effectively removes dynamic features.

4 Experiment

In order to verify the superiority of the proposed method in Visual Odometry, we conduct comparative experiments with conventional methods. As an evaluation method, with reference to the evaluation method by [1], evaluations are performed using the error rate obtained by dividing the difference between the start and end points by the length of the whole Visual Odometry. Since the difference between the start and end points is the error accumulated when creating Visual Odometry, this error rate can be regarded as the average error appearing per unit distance. We selected two crowded environments indoor and outdoor as experimental environments and conducted experiments.

The threshold values and parameters used for experiments are shown. The threshold to determine the static / dynamic of the feature based on ICGM is $Thr_{ICGM} = 0.7$, and the threshold to measure the validity of the calculated center of gravity is $Thr_{CG} = 0.7$. The resolution of the camera is 1280 × 960, and the frame rate of the continuous image is 12fps.

4.1 Indoor evaluation experiment

The indoor experimental environment and its route are Indicated on Figure 3. The experiment was conducted near the Tokaido Line Shibuya station underground ticket gate. The experimental environment is about 100m × 50m, and the length of one round of the route is about 200m. We create a visual Odometry based on the continuous image shot according to the experimental route indicated by the red dashed line.

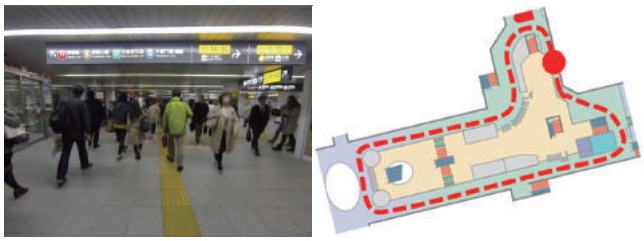


Figure 3. Indoor experiments environment

4.2 Result of indoor evaluation experiment

Indoor experiment results are shown in Table 1 and Figure 4. As shown in the Table 1, the proposed method significantly reduces the error rate compared with the conventional method. As can be seen from Figure 4, it can be seen that the proposed method generates Visual Odometry with smaller error between the start and end points. From these results, it is shown that the proposed method can more effectively invalidate the influence of dynamic features such as pedestrians.

Table 1. Indoor experimental results

Method	Error rate[%]	Time[s]
Proposed	0.46	1.22
ICGM2.0[1]	2.13	1.19
ICGM[2]	2.75	1.20
PIRF[3]	2.43	1.25
Libviso2[4]	5.76	1.16

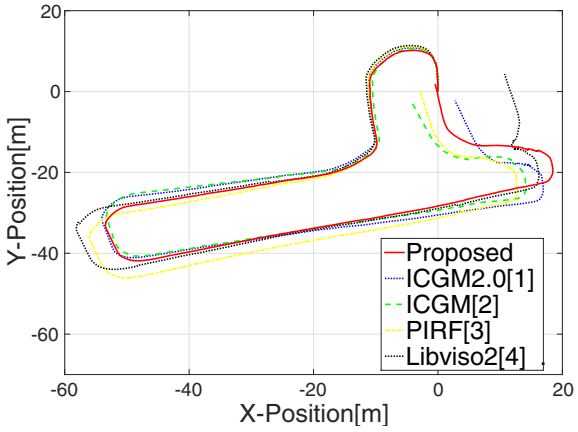


Figure 4. Indoor experimental results : Visual Odometry. The proposed method is closer to the start and end points of the loop than the conventional method. The system is drawing an accurate Visual Odometry.

4.3 Outdoor evaluation experiment

The outdoor experiment environment and its route are indicated on Figure 5. The experiment was con-

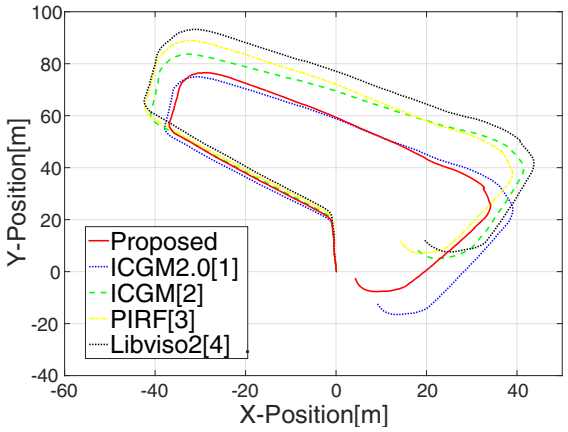


Figure 6. Outdoor experimental results : Visual Odometry. As in the indoor experiments, the proposed method is closer to the start and end points of the loop than the conventional method and draws accurate Visual Odometry.

ducted in the outdoor environment in front of Machida station. The experimental environment is about 80m × 100m, and the length of one round of the route is about 230m. We create a visual Odometry based on the continuous image shot according to the experimental route indicated by the red dashed line.

Figure 5. Outdoor experiments environment [7]

4.4 Results of outdoor evaluation experiment

Outdoor experiment results are shown in Table 2 and Figure 6. As shown in the Table 2, the proposed method significantly reduces the error rate compared with the conventional method. As can be seen from Figure 6, it can be seen that the proposed method generates a visual Odometry with a smaller error between the start and end points. From these results, It has been shown that the influence of dynamic features such as pedestrians can be more effectively invalidated in indoor and outdoor environments. Also, despite the greatly improved accuracy, the increase in processing time is kept to a very small extent.

Table 2. Outdoor experimental results

Method	Error rate[%]	Time[s]
Proposed	2.13	1.64
ICGM2.0[1]	6.22	1.41
ICGM[2]	8.22	1.59
PIRF[3]	7.52	1.54
Libviso2[4]	8.92	1.48

5 Conclusion

In this paper, we proposed a novel Visual-SLAM approach which is robust to the effects from dynamic objects in SLAM process.

In order to verify the superiority of the proposed method, we conducted experiments of Visual Odometry under indoor and outdoor crowded environments. Experimental results showed that the proposed method is superior to conventional methods in terms of accuracy.

References

[1] Toru Kayanuma, Osamu Hasegawa: "Simultaneous Localization and Mapping by Hand-Held Monocular Camera in a Crowd", Computer Vision and Image Media, 2015-CVIM-195, Vol.47, pp1-6, Jan, 2015.(in Japanese)

[2] G. Hua, O. Hasegawa, "A Robust Visual Feature Extraction Method for Simultaneous Localization and Mapping in Public Outdoor Environment", *Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII)*, 2015

[3] A. Kawewong, S. Tangruamsub, and O. Hasegawa, "Position invariant robust features for long-term recognition of dynamic outdoor scenes," *IEICE Trans. on Information and Systems*, Vol.E93- D, No.9, pp. 2587-2601, 2010.

[4] A. Geiger, J. Ziegler, and C. Stiller. "StereoScan:Dense 3D reconstruction in real-time." *IEEE Int. Veh. Symp.*, pp963-968, Baden-Baden, Germany, June, 2011.

[5] E.Rublee, V.Rabaud, K.Konolige, G.Bradski, "ORB: an efficient alternative to SIFT or SURF", *International Conference on Computer Vision (ICCV)*, No v , 2011.

[6] N. Tongprasit, A. Kawewong, and O. Hasegawa, "Pirf-nav 2: speeded-up online and incremental appearance-based slam in highly dynamic environment," *IEEE Workshop on Applications of Computer Vision (WACV)*, Jan, 2011.

[7] Geospatial Information Authority of Japan: "Aerial Photographs' Photography Record", http://mapps.gsi.go.jp/maplibSearch.do, Aug, 2016.