A Method of Landmark Visual Tracking for Mobile Robot

Lijun Zhao, Ruifeng Li, Tianying Zang, Lining Sun, and Xufeng Fan

State Key Laboratory of Advanced Robotic Technology and System,
Harbin, China
{zhaolj,lrf100,ztying,lnsun,xffan}@hit.edu.cn

Abstract. Landmark tracking is key factor for mobile robots localization and navigation. This paper proposes a combined approach automatically to detect and track landmark. Firstly, a landmark is initially located in the image coordinates by features recognition- SIFT (Scale Invariant Feature Transform) and matching technology-RANSAC(Random Sample Consensus). Then based on similarity distance, tracking algorithm is called, which depends on adaptive particle filter. Furthermore, re-position strategy based SIFT is also presented to catch the landmark which was lost. Finally, the experimental results show that the proposed method achieves robust and real-time tracking of a landmark and has a practical value for robot visual.

Keywords: Visual target location; Robot vision; Visual target tracking.

1 Introduction

In indoor environments, targets such as walls, corners, doorways and even cartons are used as landmarks studied in mobile robot navigation, mapping and exploration. However, landmarks must have stable features and be easy to identify. Nowadays, methods based on features descriptors have been hot topics in target recognition because of their promising performance and invariant to many kinds of geometric and photometric transformations. Many scholars focus on various application researches: robot localization [1], target recognition [2,3] and object categorization [4,5]. Feature detectors could be traced back to the Moravec's corner detector [6], which searched the local maximum of minimum intensity changes. However, the drawback of such detector was anisotropic, noisy, and sensitive to edges. The Harris corner detector [7] was developed for overcoming those drawbacks, but it failed to deal with scale changes which always occur in robot vision. Lowe [5] introduced SIFT (Scale Invariant Feature Transform) technology to deal with this scaling problem. The SIFT's detector solved the local maximums of a series of DoG (Difference of Gaussian) images. Mikolajczyk and Schmid[8] proposed the GLOH (Significance of the Gradient Location and Orientation Histogram) algorithm, which was an extension of the SIFT descriptor. Ke and Sukthankar [9] proposed the PCA-SIFT (PCA based SIFT), which performed more efficiently than the SIFT descriptor. Similar to the PCA-SIFT, GLOH also applied PCA to reduce the dimension of the descriptor. Previous descriptor evaluations [8] revealed that under either scale changes or image blur, SIFT and

C. Xiong et al. (Eds.): ICIRA 2008, Part I, LNAI 5314, pp. 901–910, 2008.

GLOH obtained the best results. But SIFT did not require high-dimensional matrix, and was more suitable for application in real-time recognition than GLOH. Thus based on robust recognition, in this paper, the position and shape of target was also caught by initial matching those key features.

Object tracking is a classical issue in the field of robot vision. The main challenges come from robustness to variation of the target in the scene and occlusion problem. Mean shift (MS) based on kernel density gradient has recently gained a great deal of attention[10, 11, 12] and shown to be a successful approach in the pursuit of robust tracking , but the algorithm which is local optimal can not solve the global optimum. Particle filter [13] as a non-parameter global optimal algorithm is also widely used in tracking. Although the former executes rapidly, it is hard to deal with occlusion problem. The latter can better solve the problem, for visual tracking, but it possesses low efficiency when a large number of particles are used for tracking. In view of the tradeoff, some scholars had put forth a hybrid algorithm based on mean-shift and particle filter (PF) [14], but the method needed to designate target area in the visual image artificially, and failed to study algorithm efficiency issue. Furthermore, those methods could not adaptively adjust the number of particles according to requirements of real-time tracking. Focusing on those issues, this paper presents the hybrid algorithm to solve the drawback, and adaptively to adjust the number of particles according to the Bhattacharyya distance.

In this paper, the novel hybrid algorithm for landmark tracking includes the following two steps: target localization and tracking in image frame. The first step uses SIFT detector to detect key-point of objective imagine and its template. And target is located by RANSAC matching [15] SIFT descriptors between objective imagine and the template. The second calls the tracking algorithm or re-position algorithm according similarity distance.

The rest of the paper is organized as follows: Section 2 describes the target recognition and localization. In Section 3, the improved tracking strategy is introduced based on particle filters. Experiment results and data on mobile robot are provided in Section 4, and conclusions are drawn in Section 5.

2 A Target Recognition and Localization

SIFT descriptors are invariant to image translation, scaling, rotation, partially invariant to illumination changes and affine, those characteristic being suitable for recognition and tracking landmarks, even solving mobile robots being kidnapped.

2.1 A Landmark Recognition Based SIFT

The SIFT descriptor for each key-point $\vec{z}_{kp}$ (with scale σ_{kp} and orientation θ_{kp}) is a 128 dimensional vector which is created by first computing the gradient magnitude and orientation in the neighborhood of the key-point. It contains 16 orientation sub histograms, and each consists of 8 bins. In detail, SIFT algorithm have the following four steps:

(1) Building a scale space. The first stage searches over scale space using a DoG (Difference of Gaussian) function to identify potential interest points.

(2) Localization the features (key-point). Location and scale of each candidate point are calculated, and key points are selected on the stability.
(3) Assignment of key point orientation. One or more orientations are assigned to each key point based on local image gradients.
(4) Generation of key point descriptors. A descriptor is generated for each key point from local image.

BBF(Best Bin First)[16] matching algorithm often leads to some invalid matched points. To improve recognition accuracy rate, RANSAC algorithm is introduced to calculate the basis of matrix [17] to remove the invalid matched point pairs.

Among matched SIFT key-point pairs, we define vector U as candidate target vector and vector V as template vector, and $U, V \subset R^{3\times n}$. That is

$$U = \{u_i \mid i = 1, \cdots, n\} . \tag{1}$$

$$V = \{v_i \mid i = 1, \cdots, n\} . \tag{2}$$

$$u_i = (x_i^{(1)}, y_i^{(1)}, \sigma_i^{(1)})^T . \tag{3}$$

$$v_i = (x_i^{(2)}, y_i^{(2)}, \sigma_i^{(2)})^T . \tag{4}$$

where $x_i^{(j)}$, $y_i^{(j)}$ and $\sigma_i^{(j)}$ are respectively x-coordinate, y-coordinate, and scale of i-index key points, j=1, 2.

Epipolar geometry constraint matrix can be fully derived geometric relationships between the vector set *U* and *V*. Geometric constraint matrix is described by basement matrix *F*, as shown in the following type.

$$V^T FU = 0 . \tag{5}$$

Here $F \subset R^{3\times3}$, belief probability p=0.95, then the RANSAC algorithm removed the invalid matched key points is designed as follows:

```
RANSAC algorithm

  (1) Initialization: max_sample=1000,S0,p=0.95;
  (2) for 1: max_sample;
     (a)Random selection of 7-point from the U, V,
        and alculating the basis matrix F;
     (b)Calculating matched-point in U and V to meet
        constraint matrix F, then creating set S
     (c) if (#(S)>#(S0));
           S0=S;
```

$$max\_sample = \frac{\ln(1 - p)}{\ln(1 - ((\#(U) - \#(S))/\#(S))^7)} ;$$

```
      end for;
  (3) random re-selecting elements from S0 and
      calculation the basis matrix F;
  (4) removing invalid matched-point according to F;
```

2.2 Localization of a Landmark

Assume that RANSAC algorithm gains valid k matched-point pairs in set S,

$$S = \{ s_i(u_i, v_i) \mid i = 1, \cdots k \} . \tag{6}$$

and key points of target are expressed by $U' = \{ (x_j^{(1)}, y_j^{(1)}, \sigma_j^{(1)}) \mid j = 1, \cdots, k \}$. Then target position is

$$\begin{cases} x_{tp} = \sum_{j=1}^{k} x_j^{(1)} \Big/ k \\ y_{tp} = \sum_{j=1}^{k} y_j^{(1)} \Big/ k \end{cases} . \tag{7}$$

3 A Landmark Tracking Strategy

3.1 Similarity Distance

It is important that the concept of Bhattacharyya distance is introduced, because Bhattacharyya distance is a key factor of the tracking strategy.

Firstly, template of the target is converted to HSV color space, where H-component histogram is calculated and divided into m-region. The scope of H component is mapped to 0 - 255, so the color range for each region becomes 0-255/m. Matrix q is defined as the distribution of template H-component histogram,

$$\begin{cases} q = \{ q_u \}_{u=1\cdots m} \\ \sum_{u=1}^{m} q_u = 1 \end{cases} . \tag{9}$$

Similarly, the target size (location z in image frame) of the H component histogram:

$$\begin{cases} p(z) = \{ p_u(z) \}_{u=1\cdots m} \\ \sum_{u=1}^{m} p_u = 1 \end{cases} . \tag{11}$$

.Therefore, Bhattacharyya parameter is defined as:

$$\rho(z) = \rho[q, p(z)] = \sum_{u=1}^{m} \sqrt{q_u p_u(z)} . \tag{13}$$

Bhattacharyya distance which describes the extent of similarity between target and template is achieved from equation (14). Here, *d* (*z*) denotes similarity distance.

$$d(z) = \sqrt{1 - \rho[p(z), q]} . \tag{14}$$

3.2 Particle Filter Tracking

Particle filter describes the posterior probability with weighted particle set. Then particle set is denoted by $\{x_k^{(i)}\}_{i=1}^{N_s}$ and weight set is presented by $\{\omega_k^{(i)}\}_{i=1}^{N}$, and the number of particle is N_s:

$$p(\mathrm{x}_k \mid \mathrm{z}_\mathrm{k}) \approx \sum_{i=1}^{N_s} \omega_k^{(i)} \delta(\mathrm{x}_k - \mathrm{x}_k^{(i)}) . \tag{15}$$

And $\sum_{i=1}^{N_s} \omega^{(i)} = 1$, $\delta(\cdot)$ is *Kronecker* function. Method based on sampling importance re-sampling (SIR) can be formed with a choice of importance function

$$\mathrm{x}_k^{(i)} \sim q(\mathrm{x}_k^{(i)} \mid \mathrm{x}_{k-1}^{(i)}, \mathrm{z}_k^{(i)}) . \tag{16}$$

The value of weights is updated

$$\omega_k^{(i)} \propto \omega_{k-1}^{(i)} \frac{p(\mathrm{z}_k \mid \mathrm{x}_k^{(i)}) p(\mathrm{x}_k^{(i)} \mid \mathrm{x}_{k-1}^{(i)})}{q(\mathrm{x}_k^{(i)} \mid \mathrm{x}_{k-1}^{(i)}, \mathrm{z}_k)} . \tag{17}$$

Let $q(\mathrm{x}_k^{(i)} \mid \mathrm{x}_{k-1}^{(i)}, \mathrm{z}_k) = p(\mathrm{x}_k^{(i)} \mid \mathrm{x}_{k-1}^{(i)})$ then

$$\omega_k^{(i)} \propto \omega_{k-1}^{(i)} p(\mathrm{z}_k \mid \mathrm{x}_k^{(i)}) \propto \omega_{k-1}^{(i)} \cdot e^{-\lambda D^2[d(k)]} . \tag{18}$$

The shape of tracking window is denoted by width: w_{obj} and height: h_{obj}.

$$\begin{cases} w_{obj} = 2 \times \min\left(\left|x_{\max} - x_{obj}\right|, \left|x_{obj} - x_{\min}\right|\right) \\ h_{obj} = 2 \times \min\left(\left|y_{\max} - y_{obj}\right|, \left|y_{obj} - y_{\min}\right|\right) \end{cases} . \tag{19}$$

Updated particle position and weight are denoted by $\left\{p_{s(i)}, w_k^{(i)}\right\}_{i=1}^{N}$, so the center of target is:

$$\hat{p}_k = \sum_{i=1}^{N} \frac{s_{(i)} \times w_k^{(i)}}{w_k^{(i)}} . \tag{20}$$

and $\hat{p}_k$ is $\{\hat{x}_c, \hat{y}_c\}_k$.

3.3 Realization of Tracking Algorithm

The tracking algorithm should be characterized by real-time and stability. Only SIFT descriptor matching localization can not be directly used for target tracking because of its high cost and low efficiency of calculation. As far as carton tracking experiments concerned, only particle filter is more efficient than only mean shift in average iteration when tracking particle number is less than 200. Therefore, the proposed tracking strategies employ adaptive particle filter. Furthermore, considered temporarily lost target or occlusion problems, re-position strategy is also proposed to ensure tracking algorithm catching target again.

Bhattacharyya distance's two thresholds: t1 and t2 (t1<t2) need to be set, which decide the tracking method and adjust the number of particles. Thus we define such particle filter with auto-adaptive particle number as adaptive particle filter.

The tracking algorithm is described as:

```
Decision-making strategy
    for 1:sum_frame;
      (1)if(d(z)<t1)
         tracking with the number of particles: Ns=50;
      (2)if(t1<d(z)<t2)
         tracking with the number of particles:
         Ns=400*d(z);
      (3)if(t2>d(z))
         re-SIFT location strategy;
         if(target missing())
         exit();
      (4)Calculate Bhattacharyya distance;
         delay(20);
    end for
```

In the decision-making strategy , (1) denotes the result of current tracking step is the best, and a small number of particles are enough to track the target. (2) means that current target is close to its template, but the extent to the lower level, and requires of a big number of particles in the next step tracking. (3) deals with conditions such as target lost shortly, or serious occlusion of landmark.

```
Particle filter tracking algorithm
  (1)p_k=2p_{k-1}-p_{k-2} ; //p_k is target center at time k.
  (2)New particle random generated as p_k-center
      and weighted 1/N_s•
  (3)Calculating similarity distance•//Equation•14•
  (4)Calculating posterior probability•//Equation•18•
  (5)Updating weight value•
  (6)Getting target center p̂_k //Equation•20•
  (7)if( ‖p̂_k − p_k‖ ≤ ε_0 ) // ε_0 =2 pixels
       The shape of tracking window;// Equation•19•
       Break•
     else
       Goto (2)//re-sampling

Re-position strategy
   for i=1:m_times;
      (a)SIFT algorithm;
      (b)if(RANSAC==TRUE);
             break;
         else
             robot random move in direction of target
```

```
lost;
      (c)if(i==times)
            target missing();
   end for;
```

If function target missing() occurred, it means that target is already completely lost in the view field in limited time. In sum, the strategy based on Bhattacharyya distance possesses more efficient and robust target tracking than mean shift or general particle filter.

4 Experiments

In this paper, the target tracking algorithm is tested on real robot equipped with low-resolution camera in the head of a mobile robot as shown in Fig.1. The image is 320×240 pixels and video rate at 25 frames per second.

Fig. 1. The mobile robot for algorithm experiments

4.1 Target Localization

Indoor environment, carton as a natural landmark is tested for target localization in objective image frame as shown in Fig.2. Middle and right images in Fig. 2 are the original result of BBF and RANSAC matching respectively. The results show that reliability and accuracy of algorithm RANSAC are better than that of only BBF. The carton center location: x_{tp}=229.3 pixels, y_{tp}=115.6 pixels, and error is Δ_{max} = 10.7 pixels compared to actual center (233.5, 125). Localization algorithm based SIFT spends 0.6s in matching a frame image.

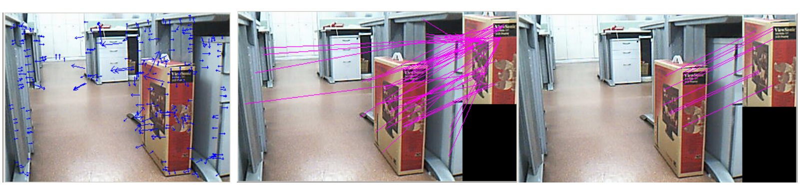

Fig. 2. SIFT and matched experiments

4.2 A Landmark Tracking Experiments

The thresholds of similarity distance are set: t1=0.15 and t2=0.35 in the proposed algorithm. Firstly, only mean shift algorithms experiment is shown in Fig.3. Fig.3 (a) denotes initial carton position is updated by SIFT+RANSAC method. However, from 200 frames, only mean shift tracking test is failed, which is caused by background noise.

(a) Frame 8 (b) Frame 73 (c) Frame 82 (d) Frame 204

Fig. 3. Only mean shift algorithm for target tracking

Fig.4 is a process of our strategy to track the target. At first, the target is initially positioned by SIFT, the strategy will be executed for the target tacking including multi-invariant test, as shown in image (a) and (b). To verify robustness of the tracking strategy, a case of the target lost is also tested. The goal lost is shown in image (c), and image (d) denotes that robot successfully catch the target again with re-position strategy. We can see that the entire tracking stage is smooth and similarity distance is almost between 0.15 and 0.35, in addition to 26 frames from 788 to 813.

(a) Frame 290 (b) Frame 529 (c) Frame 798 (d) Frame 812

Fig. 4. A comprehensive algorithm of a carton tracking

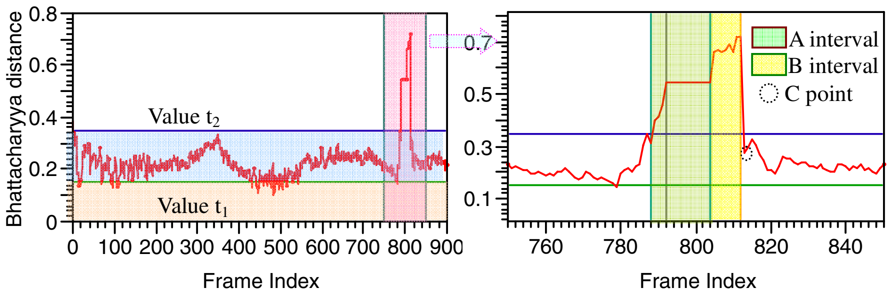

Fig. 5. Similarity distance description of process of Fig. 4 carton tracking

Data of tracking process are shown in Fig.5, here, t1=0.15, t2=0.35 in decision-making strategy. 900 frames tracking shows the proposed algorithm can successfully control carton tracing, and be effective. Re-position algorithm is firstly called from frame 788 according t2, but re-position algorithm is invalid for target completely lost between frame 792 and frame 804, and tries to catching carton when carton appears in the view from frame 805. At last the target is located at frame 812(C point in image).

5 Conclusions

The paper has presented a comprehensive landmark tracking method for mobile robot on based monocular camera. Experimental results demonstrate the effectiveness and robustness of the approach. The proposed tracking methods based on Bhattacharyya distance have multi-invariant features, and re-catching algorithm can effectively solve the similar case of occlusion. To improve the real time tracking, the number of relevant particles is linearly adjusted according to Bhattacharyya distance thresholds. Experiments of the real robot tracking carton show the tracking strategy achieves more robust tracking than mean shift algorithm and general particle filter, and has some practicable value in robot vision.

Future work involves vision based on multi-landmark tracking in environment and further improves real time of targets re-location and tracking.

Acknowledgments. The authors would like to thank the State Hi-tech Research Development Plan (program No. 2007AA04Z221) and Program for Changjiang Scholars and Innovative Research Team in University for supporting this research.

References

1. Se, S., Lowe, D., Little, J.: Vision-based mobile robot localization and mapping using scale-invariant features. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2051–2058 (2001)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transaction on Pattern Analysis and Machine Intelligence 24, 509–522 (2002)
3. Berg, A.C., Berg, T.L., Malik, J.: Shape matching and object recognition using low distortion correspondence. In: Proceedings of the IEEE International Conference Computer Vision Pattern Recognition, vol. 1(1), pp. 26–33 (2005)
4. Rothganger, F., Lazebnik, S., Schmid, C., Ponce, J.: 3D object modeling and recognition using affine-invariant patches and multi-view spatial constraints. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2(2), pp. 272–277 (2003)
5. Lowe, D.G.: Distinctive image features from scale-invariant key points. International Journal of Computer Vision 60(2), 91–110 (2004)
6. Moravec, H.: Towards automatic visual obstacle avoidance. In: Proceedings of the International Joint Conference on Artificial Intelligence, pp. 584–596 (1977)
7. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference, pp. 147–151 (1988)

8. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transaction on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
9. Ke, Y., Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors. In: Proceedings of IEEE Computer Vision and Pattern Recognition (2004)
10. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence 25(5), 564–577 (2003)
11. Collins, R.T.: Mean-shift blob tracking through scale space. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 234–240 (2003)
12. Cheng, Y.: Mean shift, mode seeking, and clustering. IEEE Trans. on Pattern Analysis and Machine Intelligence 17, 790–799 (1995)
13. Maskell, S., Gordon, N.: A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Processing 50(2), 174–188 (2002)
14. Zhaowen, W., Xiaokang, Y., Yi, X., et al.: Camshift Guided Particle Filter for Visual Tracking. In: IEEE Workshop on Signal Processing Systems, pp. 301–306 (2007)
15. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM 24(6), 381–395 (1981)
16. Beis, J.S., Lowe, D.G.: Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: Proceedings of computer vision and pattern recognition, pp. 1000–1006 (1997)
17. Hartley, R.I.: In Defense of the Eight-Point Algorithm. IEEE Trans. Pattern Analysis and Machine Intelligence 19(6), 580–593 (1997)