Journal of Robotics, Networking and Artificial Life, Vol. 3, No. 3 (December 2016), 144-147

# Using Multi-Target Tracking and Identification TLD Algorithm for Intelligent Mobile Robot

**Jr-Hung Guo**
*Department of Electrical Engineering, National Chin-Yi University Of Technology, No.57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 41170, Taiwan*

**Kuo-Hsien Hsia***
*Department of Electrical Engineering, Far East University, No.49, Chung Hua Rd., Hsin-Shih. 74448 Tainan, Taiwan*

**Kuo-Lan Su**
*Department of Electrical Engineering, National Yunlin University of Science & Technology, 123 University Road, Section 3,Douliou, Yunlin 64002, Taiwan*
*E-mail: g9710801@yuntech.edu.tw, khhsia@cc.feu.edu.tw* ,sukl@yuntech.edu.tw*

**Abstract**

Image algorithms used herein are Tracking-Learning-Detection (TLD) and Speed UP Robust Features (SURF). TLD is used for target tracking and SURF is used for identifying targets. We use zoning identification, with the use of statistical probability to strengthen the efficiency of TLD and SURF. With such a method, the efficiency of image identification and target tracking can be enhanced so that the robot can simultaneously track and identify multiple targets.

*Keywords*: multi-target tracking, TLD, SURF, image identification, mobile robot.

## 1. Introduction

Image recognition and tracking technology are important issues on robotics research. Fabian et al. [1] used Microsoft Kinect and Simulink for instant object tracking. Marković et al. [2] used omnidirectional camera on a mobile robot to track and follow objects. Huang et al. [3] used the SURF algorithm to make the robot have the ability to follow the object. As can be seen from previous studies, image recognition and tracking technology usually require dedicated imaging equipment or complex algorithms. But this will cause an increase in the cost of mobile robot, or reduce the efficiency of the mobile robot. Therefore, how to enhance the efficiency of the mobile robot on image identification and tracking is the main research purpose of this paper.

The image tracking algorithm that we use is TLD [4] which was developed by Zedenek Kalal, University of Surrey, Czech. The TLD algorithm used for tracking single-target for a long time. In this paper we extend the TLD algorithm to multi-target tracking based on the original single-target tracking. But this algorithm cannot identify the tracked target. The TLD framework improves the tracking performance by combining a detector and an optical-flow tracker. Since this algorithm is based on optical flow, the accuracy of TLD cannot be high. So we use Speed UP Robust Features (SURF) to assist the targets identification and as an auxiliary for TLD algorithm to track the targets.

Published by Atlantis Press

SURF is based on the scale-invariant feature transform Scale-invariant feature transform (SIFT) to identify the object. Although SURF operation is relatively light to the SIFT, it will still affect the effectiveness of the mobile robot. Therefore, in this paper, the image is divided into some separate regions. Each region is detected in turn. Supplemented with the statistics and probability, it can assist TLD on tracking. The target tracking and identification process with SURF and TLD is shown in Fig. 1. The relevant algorithms will be explained in the next section.

## 2. Algorithms

### 2.1. *TLD*

TLD technology is divided into three parts. They are the tracker, the learning process and the detector. Tracker and detector operate in parallel in TLD. Both the tracking and detecting results will feed into the learning process. The new model after learning is feedback to the tracker and detector resulting in a real-time updating. The overall process ensures that even if the target appearance changes, the target can still be tracked. The architecture of TLD tracking is shown in Fig. 2.

TLD's tracker estimates the motion of the target using the Lucas-Kanada [6] [7] optical flow, which is a frame to frame tracking method. However, the precision of the optical flow tracking is not high, thus it often results in a tracking failure. The detector records the history of target locations and surface information, called learning, and will re-detect while tracking failure.

P-N learning [9] is a very important part of tracking in TLD. Here P is Positive Constraint, also known as P-expert or growing event, and N is Negative Constraint, also known as N-expert or pruning event. In practice, P-experts as well as N-experts will have a certain bias, but the study of Zdenek Kalal found that, although there is an error, under certain conditions, the error is acceptable, and performance testing module will therefore be improved. The P-N learning consists of four parts:

(i) A classifier for learning;
(ii) A training sample set;
(iii) Supervised learning;
(iv) P-N experts: to generate positive and negative samples in the learning process;

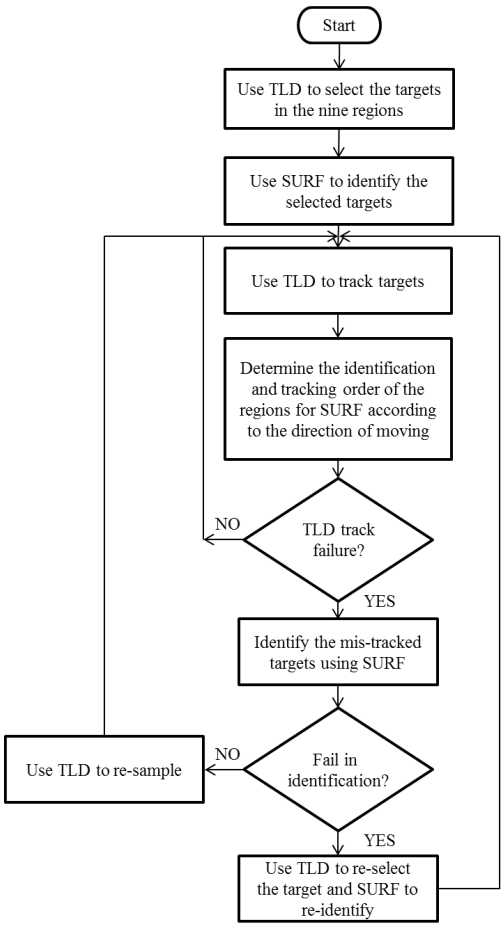


Fig. 1. Target tracking and identification process with SURF and TLD.


Fig. 2. TLD tracking architecture [4].

The relationship between these parts is shown in Fig. 3. The "random forest" classifier, which can instantly update and forecast, is used in the detector [4, 6]. The information for the feature detector is generated by random forests 2bitBP (2 bit binary pattern). The

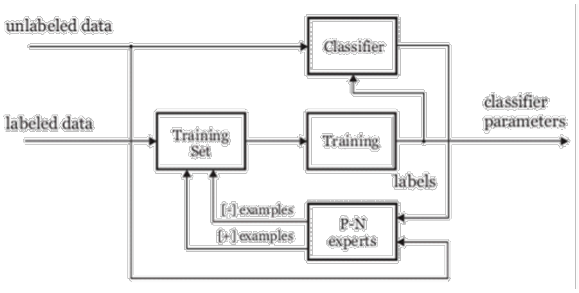


Fig. 3. TLD learning architecture [8].

gradient orientation information of a specific area will be converted into encoded output. 2bitBP is similar to harr-like feature which includes the feature type and the corresponding characteristic values. Therefore, when using nFeat type to express the object, each Fern is a Quad tree because of the 2bitBP characteristics.

TLD tracking module is based on the Median Flow tracker method. This method is performed on a frame. Select some pixels as the feature points in the previous frame and then find the corresponding positions in the current frame by the feature values. Then sort the displacements of these features pixels and find the median of these displacements. Finally let the pixels less than 50% of the median as the next feature pixels. Since this approach assumes that the tracked object is in the frame, it will fail once the object is out of the frame or obscured. The SURF will be used for confirmation at that time. Once it is confirmed, the new tracking target will be selected by the TLD.

### 2.2. *SURF*

SURF is a technology for image recognition and tracking based on SIFT. The main difference is that SIFT uses DoG (Difference-of-Gaussians) image, and SURF uses the determinant of a Hessian matrix to approximate the image. The Hessian matrix of a pixel in an image can be defined as [5]:

$$H(f(x,y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (1)$$

The determinant of H is:

$$\det(H) = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - (\frac{\partial^2 f}{\partial x \partial y})^2 \quad (2)$$

The sign of the determinant value can be used to classify the point and to determine the point is an extreme point or not. In the SURF algorithm, the function value $f(x, y)$ is replaced by the image pixel $I(x, y)$, and a second-order standard Gaussian function is chosen as a filter. Then elements of H matrix can be obtained by convolution and second-order partial derivatives as following:

$$H(x,y,\sigma) = \begin{bmatrix} L_{xx}(x,y,\sigma) & L_{xy}(x,y,\sigma) \\ L_{xy}(x,y,\sigma) & L_{yy}(x,y,\sigma) \end{bmatrix} \quad (3)$$

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (4)$$

where

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)} / 2\sigma^2 \quad (5)$$

and $\sigma$ is the scale. The determinant of the $H$ matrix of each pixel of the image can be obtained. Hence each point can be classified. For the ease of application, Herbert Bay proposed to replace the $L$ by an approximation [9]. The discriminant of $H$ matrix can be expressed as:

$$G(t) = \frac{\partial^2 g(t)}{\partial x^2} \quad (6)$$

Finally, what we want is a transformed image of the original image, because we have to look for the feature points on the transformed image and then reflect it onto the original image. And this is constituted by the approximate determinant of the Hessian matrix of each pixel with the approximate determinant defined by:

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (7)$$

Since SURF process the pixels at the image for Hessian matrix processing simultaneously on both X and Y directions, its efficiency is better than SIFT. But it still requires a large amount of computation. So we take the solution as follows:

(i) Identify region by region to reduce the amount of pixel processing. Because the entire image is divided into nine regions for TLD tracking, SURF will therefore use the same area for identification and tracking.

(ii) Determine the identification sequence according to the orientation. Since the orientation of the robot

| | | |
|---|---|---|
| 3 | 5 | 8 |
| 1 | 2 | 7 |
| 4 | 6 | 9 |

(a) turn left

| | | |
|---|---|---|
| 6 | 2 | 7 |
| 4 | 1 | 5 |
| 8 | 3 | 9 |

(b) forward

| | | |
|---|---|---|
| 8 | 5 | 3 |
| 7 | 2 | 1 |
| 9 | 6 | 4 |

(c) turn right

Fig. 4. Robot orientation relationship with SURF scanning order.

Fig. 5. Tracking and identification result by TLD and SURF.

can be known, we design the scanning sequence following the orientation of the robot, as shown in Fig. 4. The central region is of higher priority, then the upper and lower regions. With this strategy, the target tracking for the mobile robot can be of higher efficiency.

## 3. Experimental Results

A Microsoft's LifeCam Cinema was used in our experiments. Because of its large aperture and high-resolution, images with better quality can be obtained. For the image tracking and identification system, the entire image was cut into nine regions, and TLD was applied to have the tracked object, and then SURF was used for identification. The CPU of the computer for the experiment is INTEL I5-3470 with 8G memory, and the system was developed by EMGUCV 2.4 and Microsoft Visual Studio VB2010. The experimental results are shown in Fig. 5. The red box marks the tracking target of TLD, and the blue box marks the SURF tracking and identification results with the identification result of SURF marked at the upper right corner.

## 4. Conclusion

The TLD and SURF algorithms are successfully combined in this paper so that the robot's vision system can track and identify the targets. The entire image is divided into nine regions, in order to reduce the loading on image processing, and to enhance the efficiency of tracking and identification. You used a mid-level webcam in this paper resulting a lower image quality and higher failure rate in tracking and identification. High-level camera and CUDA (Compute Unified Device Architecture) can increase the quality of image and efficiency of image processing. Thus this algorithm can be used in a more complex environment. We believe that the algorithm can be widely used in robot systems in the near future.

## Acknowledgements

This work was partially supported by National Science Council of Taiwan. (NSC 104-2221-E-224 -015 -).

## References

1. J. Fabian, T. Young, J.C. Peyton Jones, and G.M. Clayton, Integrating the Microsoft Kinect with Simulink: Real-time object tracking example, *IEEE/ASME Transactions on Mechatronics*, 19(1) (2014) 249-257.
2. I. Marković, F. Chaumette and I. Petrović, Moving object detection, tracking and following using an omnidirectional camera on a mobile robot, *2014 IEEE Int. Conf. Robotics and Automation* (*ICRA*) (Hong Kong, China, 2014), pp. 5630-5635.
3. P.T. Huang, C.Y. Li, C.C. Hsu, and C.M. Hong, Object following based on SURF for mobile robots, *2012 IEEE 1st Global Conf. Consumer Electronics* (*GCCE*) (Tokyo, Japan, 2012), pp. 382-386.
4. Z. Kalal, K. Mikolajczyk, and J. Matas, Tracking-learning-detection, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(7) (2012) 1409-1422.
5. C. Evans, *Notes on the OpenSURF library*, (University of Bristol, Tech. Rep. CSTR-09-001, January, 2009).
6. B.D. Lucas, *Generalized image matching by the method of differences*, (doctoral dissertation, tech. report, Robotics Institute, Carnegie Mellon University, 1984).
7. B.D. Lucas, and T. Kanade, An iterative image registration technique with an application to stereo vision, in *Proceedings of the 7th Int. Joint Conf. Artificial Intelligence* (*IJCAI*) (1981) 674-679.
8. Z. Kalal, J. Matas, and K. Mikolajczyk, P-N learning: Bootstrapping binary classifiers by structural constraints, *2010 IEEE Conf. Computer Vision and Pattern Recognition* (*CVPR*) (San Francisco, CA, 2010) 49-56.
9. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, SURF: Speeded Up Robust Features, *Computer Vision and Image Understanding* (*CVIU*), 110(3) (2008) 346-359.