

2012 AASRI Conference on Modelling, Identification and Control

Object Tracking Method Based on SURF

Hu Shuo^{a*}, Wu Na^b, Song Huajun^c

a Key Lab of Industrial Computer Control Engineering of Hebei Province, Institute of Electrical Engineering, Yanshan University, Qinhuangdao, 066004, China

b Key Lab of Measurement Technology & Instrumentation of Hebei Province, Institute of Electrical Engineering, Yanshan University, Qinhuangdao, 066004, China

c Institute of Information and Control Engineering, China University of Petroleum (East China), Dongying, 257061, China

Abstract

An object tracking algorithm based on SURF is presented in this paper. Interest points are detected by SURF detector in reference region which is located in the first frame manually. In the following sequences, the SURF feature points are extracted in a larger window which is selected as test region. In the matching stage, we calculate the Euclidean distance between the descriptor vectors of interest points in the test image and ones in the reference image. The proposed algorithm is implemented on an embedded platform with TI's DM6446 high performance processor. The experimental results show that our system implements a real-time tracking with robustness against appearance variations, scale change, and cluttered scenes.

© 2012 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).
Selection and/or peer review under responsibility of American Applied Science Research Institute

Keywords: Object track; SURF descriptor; feature detection; DM6446

* Corresponding author. Tel.: 0086-0335-8387556;
E-mail address: hushuown@gmail.com.

1. Introduce

Object tracking is a crucial work within the field of computer vision. The challenges of object tracking are partial occlusions, illumination change, object appearance variations and cluttered scenes. To cope with these problems, lots of algorithms and methods have been proposed^[1]. Aristidis Likas et al. propose a algorithm to track object by using the mean shift with Kalman filtering^[2], which combines accuracy and the robustness to occlusion. Recently, feature-based methods have been popular because of their promising results. Huiyu Zhou et al. proposed a algorithm combined mean shift with SIFT for the object tracking^[3]. In [3], the region of interests is corresponded in different frames by SIFT features. Meanwhile, mean shift is used to perform matching via colour histograms; In above cases, experimental results show the SIFT descriptors are invariant to illumination change and appearance variations, so they are able to overcome significant amounts of image variability. Unfortunately, in real-time application, the high dimensionality of the descriptor is a drawback of SIFT in matching stage^[4].

In 2006, SURF (speeded up robust features)^{[5][6]} was proposed by Herbert Bay, which is a rotation- and scale-invariant interest point descriptor and detector. It has the advantage of SIFT high precision while it is several times faster than SIFT^[7].

Ta, D.-N. et al. propose an SURF based method to track the object in video^[8], and SURF descriptor is used to optimize similarity search of interest points. The experimental result shows that the algorithm of SURF features extracting has smaller calculation and faster than SIFT.

In this paper, we propose an object track algorithm based on SURF and implement it on an embedded platform with TI's DM6446 high performance processor which can process data in a high speed for on-line applications.

The remainder of this paper is organized as follows. In Section 2, SURF features matching algorithm is introduced. We describe our hardware design for tracking system In Section 3. In Section 4, experiments are presented and we conclude the paper in Section 5.

2. SURF features matching algorithm

The steps of SURF algorithm contain three sections: interest points detecting, interest points describing and interest points matching. Interest points detecting uses a detector based on Hessian matrix, its stability and repeatability outperforms the existing state-of-the-art, e.g. a detector based on Harris. Interest points describing uses Haar wavelet^[9] to reduce the computational cost and enhance the robustness. The Block filtering to generate the second order Gaussian filtering approximately and the integral image to accelerate convolution, reduce the computational complexity and increases speed.

2.1. Integral Image.

Integral image is a feature representation method for original image. As it's shown in Fig.1, the dark region represents the integral value of the point(x, y), which is the sum of pixels in the dark region. Calculating the integration of original image $I(x, y)$, the integral image $I_{\Sigma}(x, y)$ will be obtained, where $I(x, y)$ is the pixel value of the point(x, y)^[10].

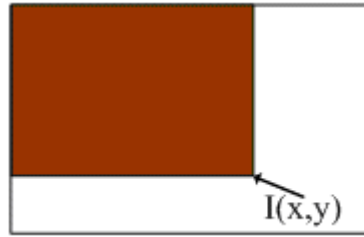


Fig.1 Integral Image

$$I_{\Sigma}(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(x, y) \quad (1)$$

2.2. Scale space representation.

According to SURF algorithm, interest points are detected at different scales, especially for the reason that similarity searching often depends on their comparison in images where they are detected at different scales^[6]. Generally, scale spaces are usually implemented as an image pyramid. In SRUF, block filters are used to approximate the Gaussian kernel, and the computational cost is independent of its size, which is critical for speed.

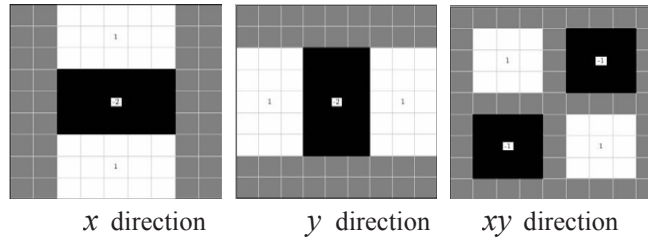


Fig.2 Box Filters

Weighting box filters' approximation^[11] for the second order Gaussian partial derivative in x- direction, y- direction and xy- direction, as shown in Fig.2, are denoted by D_{xx} , D_{yy} , D_{xy} . In Fig.3, white regions are positive, grey regions are zero and black regions are negative, respectively. To establish properly scale spaces for photometric change, box filters at different scales are required.

Fast Hessian features detection. According to SURF algorithm, Hessian matrix is used to detect interest points. Assuming $X = (x, y)$ is a point in an image, the Hessian matrix in x at scale σ is defined as follows:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (2)$$

Where $L_{xx}(x, \sigma)$ is the result of filters smoothing in point x . the xy means the first order derivative in x -direction and the second order derivative in y -direction.

2.3. SURF descriptor extraction.

The SURF descriptor is extracted by two steps: find a dominant orientation according to the circular region around the interest point and construct a square region along the dominant orientation for getting the descriptive information. The square regions are split up into 4×4 sub-regions. For each sub-region, we weight Haar wavelet responses with a Gaussian ($\sigma=3.3s$) in vertical and horizontal direction (denoted by d_y and d_x) at the interest point. Aiming at being invariant to rotation, $|d_y|$ and $|d_x|$ are summed up and denoted by $\sum |d_y|$, $\sum |d_x|$.

respectively. After normalizing the eigenvector, we can form a 4D vector in each sub-region:

$$v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \quad (3)$$

Then we calculate the eigenvectors in 16 sub-regions and construct a 64D (16*4) eigenvector. We define the eigenvector mentioned above as the SURF descriptor.

In match stage, we use the Euclidean distance of the eigenvector as the Similarity Measure of key points between two frame images for interest points matching.

3. Hardware Design for Real-Time Object Tracking System

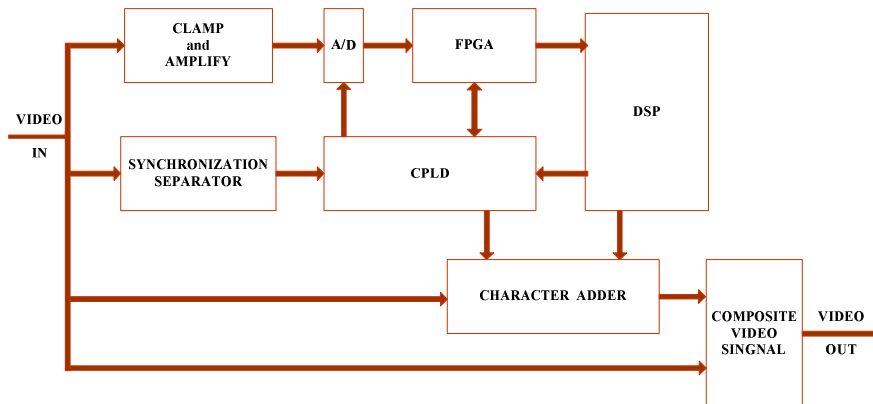


Figure 3. Hardware structure block diagram

In the proposed tracking hardware system is described as Fig.3. From the Fig.3, we can know that analogy video is preprocessed firstly by amplifier, filter and clamp operation. Then the video is converted to digital signal by A/D at 8-bit resolution. In our system, the chipset DM6446 is chosen as the digital signal processor (DSP), which includes two core chipset: DSP and ARM. DSP runs object tracking module and gets the object displacement information. Meanwhile, ARM runs the control and GUI (Graph User Interface) module which can receive command from other system and show all video. CPLD run the system logic control module and FPGA run image pre-processing module which has higher processing speed than DSP.

4. Experiment

The experiment is carried out as follows:

- (1) In the first frame, the reference (object) region is located manually, and interest points are extracted by the SURF detector in reference region.
- (2) In the following frames, the SURF feature points are extracted in a larger window which is selected as the test region.
- (3) In the matching stage, interest points in the test image are compared to ones in the reference image by calculating the Euclidean distance between their descriptor vectors. Because interest points are dispersive, we calculate the barycenter of the interest points as the candidate match displacement.
- (4) Update the reference image and prepare for next matching.

In our experiment, we use videos with difference scenes to test our system in term of robustness and speed. When the reference region is 80*60 while the test region is 160*120, the average computational time of matching between test region and reference region is 18ms.

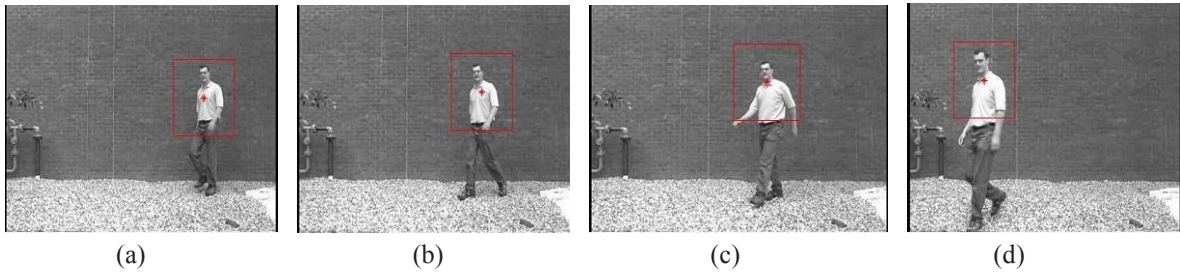


Fig.4 pedestrian tracking result

We first implement a pedestrian track and the experimental results as Fig.4 shown, where (a), (b), (c), (d) are the track results at frame 15, 30, 50, 78, respectively. As the experiment shows, our system can keep high detection accuracy when object appearance variation and scale change occur.



Fig.5 plane tracking result

Fig.5 shows the results of plane tracking, and (a), (b), (c), (d) are the track results at frame 215, 286, 323, 389, respectively. The background of scene is complicated compared to one in Fig.4. In the following sequences, the object has rotation, illumination changes and view variation, and our system implements a steady tracking due to robustness of the SURF descriptor. However, as the (d) shown, the mismatch occurs when the corresponding interest points disappear in the test image due to occlusion. Aim at influence of occlusion in term of robustness, we will improve the algorithm in the next work.

5. Conclusions

In this paper, we proposed a SURE based object tracking algorithm, in which interest points in the defined object are matched between consecutive frames by calculating the Euclidean distance between their descriptor vectors. We calculate the barycenter of the interest points as the candidate match displacement, which increase the detector's repeatability. We implement the proposed algorithm on an embedded platform with TI's DM6446 high performance processor. The experimental results show our system implements a real-time tracking with robustness against appearance variations, scale change, and cluttered scenes

Acknowledgements

This work is supported by the grants from the Science Foundation of Hebei province under Grant No. F2012203182.

References

- [1]. Hanxuan. Yang, Ling. Shao, Feng. Zheng, L. Wang, Z. Song. Recent advances and trends in visual

- tracking: A review. *Neurocomputing*, 74 (2011): 3823-3831
- [2]. Vasileios Karavasilis, Christophoros Nikou, and Aristidis Likas. Visual Tracking by Adaptive Kalman Filtering and Mean Shift. Springer-Verlag Berlin Heidelberg, 2010, 6040, p.153-162.
 - [3]. Huiyu Zhou, Yuan Yuan, Chunmei Shi. Object tracking using SIFT features and mean shift. *Computer Vision and Image Understanding*, 2009, 113, p.345-352.
 - [4]. D.Lowe. Distinctive image features from scale invariant keypoints. In *journal of Computer Vision*, 2004, 60(2), p.91-110.
 - [5]. Bay, H., Tuytelaars, T., Van Gool, L. SURF: Speed up robust features. In: *Proc. European Conf. of Computer Vision (ECCV)*, 2006:404–417.
 - [6]. Bay H, Tuytelars T, Van Gool L. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 2008(110): 346-359.
 - [7]. Q. Miao, G. Wang, C. Shi, X Lin, Z Ruan. A new framework for on-line object tracking based on SURF. *Pattern Recognition Letters*, 32 (2011): 1564-1571
 - [8]. Ta, D.-N., Chen, W.-C., Gelfand, N., Pulli, K., SURFTrac: Efficient Tracking and Continuous Object Recognition using Local Feature Descriptors, *Conference on Computer Vision and Pattern Recognition (CVPR'09)*, 2009:2937-2943.
 - [9]. Khashman Adnan, Dimililer Kamil. Image compression using neural networks and Haar wavelet. *WSEAS Transactions on Signal Processing*. 2008.5(4):330-339.
 - [10]. Freund Y. Boosting a weak learning algorithm by majority. *Information and Computation*, 1995(121)
 - [11]. Mikolajczyk K, Schmid C. An affine invariant interest point detector. *Computer Vision-*