AIP | Review of Scientific Instruments

Note: An object detection method for active camera

Yong Chen, Ronghua Zhang, Lei Shang, and Eric Hu

Citation: Review of Scientific Instruments **84**, 096108 (2013); doi: 10.1063/1.4821875
View online: http://dx.doi.org/10.1063/1.4821875
View Table of Contents: http://scitation.aip.org/content/aip/journal/rsi/84/9?ver=pdfcov
Published by the AIP Publishing

REVIEW OF SCIENTIFIC INSTRUMENTS **84**, 096108 (2013)

Note: An object detection method for active camera

Yong Chen,[1,2,a)] Ronghua Zhang,[1] Lei Shang,[1] and Eric Hu[2]
[1]*School of Energy Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2]*School of Mechanical Engineering, University of Adelaide, SA 5005, Australia*

(Received 13 May 2013; accepted 4 September 2013; published online 19 September 2013)

To solve the problems caused by a changing background during object detection in active camera, this paper proposes a new method based on SURF (speeded up robust features) and data clustering. The SURF feature points of each image are extracted, and each cluster center is calculated by processing the data clustering of k adjacent frames. Templates for each class are obtained by calculating the histograms within the regions around the center points of the clustering classes. The window of the moving object can be located by finding the region that satisfies the histogram matching result between adjacent frames. Experimental results demonstrate that the proposed method can improve the effectiveness of object detection. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4821875]

The movement of active camera (non-stationary camera) results in two difficulties. First, the background image always changes very quickly, and second, the scale of the object is continuously changing. Traditional methods of object detection consist of an optical-flow estimation algorithm and a motion compensation algorithm.[1] Optical-flow estimation algorithms are typically used to detect objects over a stationary background, whereas motion compensation algorithms are widely used in multiple areas.[2] Nevertheless, these methods cannot achieve satisfactory results with fast-moving objects or when the camera is being rotated. Detection and tracking information are complementary, with one transformed to the other by combining the motion vectors of feature points with a particle filter.[3] However, it is difficult to achieve accurate real-time tracking with these methods for an active camera.

To solve this problem, various algorithms based on speeded up robust features (SURF)[4] have been proposed. A new method based on super-pixels and SURF was presented for object detection.[5] In a further modification,[6] the target library was calculated and established via the object template, and then the process of matching between adjacent frames was implemented. However, these approaches cannot be successfully applied to object detection with a non-stationary camera.

Considering the rotation and scale changes that usually occur in active camera, it is necessary to calculate the motion vector in each frame such that the object template can be updated. To solve these difficulties, a new approach is proposed in this paper. First, the SURF feature points are extracted from each image. Each cluster center is then calculated by processing the data clustering of k adjacent frames. By calculating the histograms within the regions around the center points of the clustering classes, the templates of each class can be obtained. Then, by finding the region that satisfies the histogram matching result between adjacent frames, the window of the moving object can be located.

SURF can be used to extract feature vectors from video sequences.[6] To satisfy the requirements of real-time computation and accuracy, we employ the following methods in the SURF approach:

(1) A fast Hessian detector based on an approximation of the Hessian is used to obtain the key points. For any point (x, y) in the image, the Hessian matrix at scale σ is defined as

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}, \tag{1}$$

where $L_{xx}(x, \sigma)$, $L_{xy}(x, \sigma)$, $L_{yy}(x, \sigma)$ are the convolutions of the Gaussian template and the detection point.

(2) A box filter is used that is an approximation of the second-order Gaussian filter:

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

where D_{xx}, D_{xy}, D_{yy} are the filtered results of $L_{xx}(x, \sigma)$, $L_{xy}(x, \sigma)$, $L_{yy}(x, \sigma)$, respectively. For the same image, a different scale space is formed by modifying the size of the box. In addition, integration images are employed that significantly enhance the detection speed of the key point.

(3) To retain the invariance of the eigenvector, the SURF vector of each key point can be described by a 64-dimensional vector. First, the point of interest is detected by calculating the determinant of the Hessian matrix. If $\det(H_{approx}) > 0$, this point is a key point. Second, the key point is compared with the neighborhood value of a $3 \times 3 \times 3$ stereo region. If this is an extreme point, it will be selected as the candidate point. For each candidate point, the location and scale of the key point will be accurately determined by a function interpolation method. Third, the Haar-wavelet response values are calculated in the horizontal and vertical directions of the circle whose center is the key point. Therefore, the main direction of the vector at this point can be determined. Finally, the coordinate system is established according to the main direction. In addition, the key point is taken

[a)]Author to whom correspondence should be addressed. Electronic mail: ychencd@uestc.edu.cn.

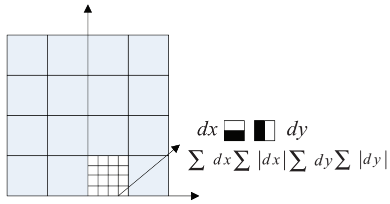

FIG. 1. Descriptor of the key point.

as the center, and a 20×20 region of pixels is chosen. The 20×20 region is then divided into four 4×4 sub-regions, as shown in Fig. 1. The Haar-wavelet responses in both the horizontal and vertical directions of each sub-region are added to form a four-dimensional vector $v = [\sum dx, \sum dy, \sum |dx|, \sum |dy|]$. Accordingly, the SURF vector of each key point can be described by a 64-dimensional eigenvector. Thus, the invariance under rotation of the key point eigenvector can be guaranteed.

Data clustering is the process that divides a data set into a number of different data sets. Among various data clustering algorithms, the k-means method[7] is simple and fast, k-means clustering is widely used in many fields. So we employ the k-means algorithm to implement the process of feature point data clustering. Equation (3) is selected as the optimized objective function. However, step 2 of the k-means is improved as: After the process of data clustering, the center point of each clustering set can be obtained. The calculation of the center point of each clustering set provides the original window. Therefore, the problem of object detection is to distinguish each clustering center point. These clustering center points may move under an active background. However, the data set that possesses the moving object will not change. Consequently, local-region histograms can be used to distinguish the clustering center points using the following equation:

$$E = \sum_{j=1}^{K} \sum_{x \in C_j} \|x - m_j\|^2. \tag{3}$$

A color histogram can be obtained by counting the occurrence frequency of the pixels in an image. Regional color histograms are an important feature in image registration.[8]

The color histogram of a rectangular region is obtained by calculating the histogram information of the region around the clustering center point. In this paper, the size of the region is 20×20 pixels. As the rectangular region is small, the number of bins is set to 4. The thresholds for different classes of intensity are defined as

$$f = \begin{cases} \lambda_1 & 0 \le f(x,y) \le 63 \\ \lambda_2 & 64 \le f(x,y) \le 127 \\ \lambda_3 & 128 \le f(x,y) \le 191 \\ \lambda_4 & 192 \le f(x,y) \le 255, \end{cases} \tag{4}$$

$$p_i = \frac{\sum \lambda_i}{N} \quad (i = 1, 2, 3, 4), \tag{5}$$

where $f(x, y)$ is the intensity of the pixel at location (x, y), λ_i is the threshold of the histogram series, and p_i is the probability of each series.

From the results of data clustering, it is clear that the feature clustering point changes under an active background. In this paper, it is assumed that the camera has the same direction of movement as the object. Thus, it is evident that the distance moved by the clustering point of the active object is less than that of the background points. Next, the threshold T is used to find the potential moving region, and the Euclidean distance is used to define the distance of movement D. The region in which D > T is the potential region of the moving object.

Following the above procedure, the regional histogram of the active region can be calculated, and the process of histogram matching can be implemented. Here, the Bhattacharyya distance[9] is employed to measure the similarity between two histograms. The Bhattacharyya distance between two continuous distribution functions is defined as

$$\rho[p, q] = \sqrt{\int p(u)q(u)du}. \tag{6}$$

We can now obtain the discretized result of the continuous integration in Eq. (6). Furthermore, substituting the two color histograms, $p = \{p^{(u)}\}u = 1, \ldots, m$, $q = \{q^{(u)}\}u = 1, \ldots, m$ into Eq. (6), we have

$$\rho[p, q] = \sum_{u=1}^{m} \sqrt{p^{(u)} q^{(u)}}, \tag{7}$$

where $\rho \in [0, 1]$. Hence, the Bhattacharyya distance can be expressed as

$$d = \sqrt{1 - \rho[p, q]}. \tag{8}$$

As the similarity between two histogram vectors increases, the distance d given by Eq. (8) decreases. A threshold T1 can be used to distinguish the moving region: *If (d < T1), this region is the area of the moving object, else it is background region.*

As the selected region is small, and the disturbance of each class is the background sky for our study object, it is necessary to distinguish the object window from the selected regions. After a series of tests, a value of T1 = 0.002 was found to be suitable in our experiments.

Object detection with a non-stationary background has frequently been the topic of object tracking and recognition. Instead of using a traditional motion model compensation method, in this paper, the correlation between k adjacent frames is used to implement the process of object detection. This is because of stark changes in the background between the k adjacent frames, while the circumstances of the moving object do not change notably. In this paper, the process of clustering SURF points is implemented via the k adjacent frames. Moreover, the regional histogram of each clustering center point is calculated. The histogram of the k subsequent frames and the k previous frames is then realized. After these procedures, the region of the moving object can be confirmed by finding the best matching window. Finally, the task of target detection can be accomplished using the active camera.

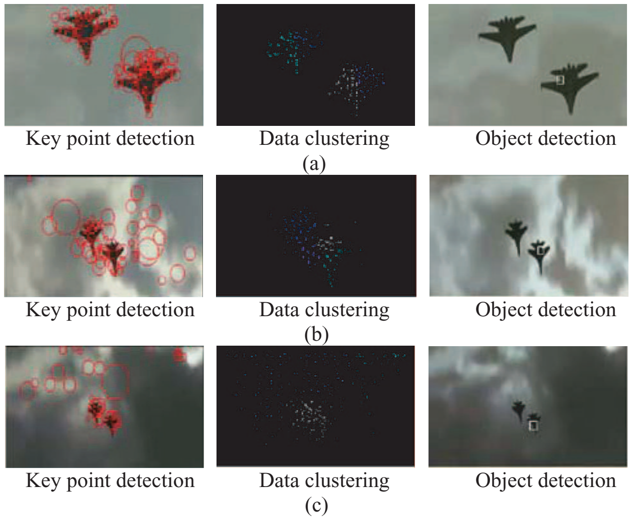

FIG. 2. Experimental results. (a) 38th frame. (b) 80th frame. (c) 90th frame.

The test video has a resolution of 320 × 166 pixels. In the test, the background has a significant effect on the detection accuracy of the object.

In the experiment, the target in the lower right corner is selected as the detected object. Figure 2(a) shows the result of detecting the SURF key points. Figure 2(b) shows the clustering result for the key points from five frames. Figure 2(c) shows the tracking performance of our improved approach. Here, three frames are chosen to implement the process of data clustering.

From the video, it can be seen that the scale parameter is constantly changing. The traditional motion model compensation method cannot obtain satisfactory results in circumstances with scale changes, whereas our proposed method can detect the object using a non-stationary camera by adjusting the data clustering center point.

Figure 3 shows the data clustering results of the subsequent and previous five frames. The red circle is the clustering center point of the matching region. The figure shows that the center point is always changing according to the motion of the object. Based on this characteristic, we can update the clustering location to accomplish the object detection.

From Fig. 4, we can see that only one pair of templates has a higher similarity histogram than the others in the subsequent and previous four frames. Meanwhile, the location of the matched histograms shows the motion window of the target. Therefore, the location of the moving object can easily be found.

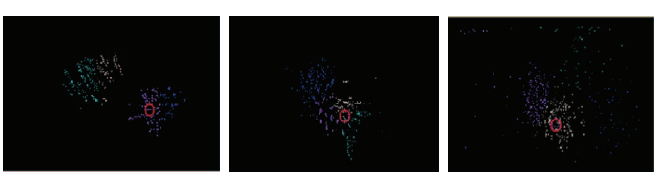

FIG. 3. Location of center points of the matched region.

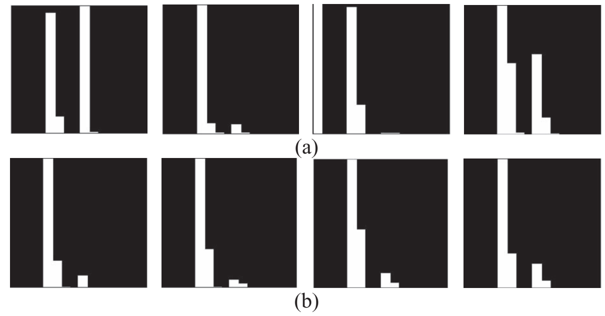

FIG. 4. Histograms of the subsequent and previous clustering center points. (a) Histogram of the clustering center point of the subsequent four frames. (b) Histogram of the clustering center point of the previous four frames.

To overcome the difficulty of object detection with a non-stationary camera, this paper proposed a new method based on SURF and a data clustering algorithm. The process of clustering SURF key points using a k-means algorithm was implemented. In addition, a regional histogram was used to remove disturbances. The location of the object can be determined by finding a suitable window. The proposed algorithm shows great robustness to rotation and scale changes of the objects.

This work was supported by the National Natural Science Foundation of China (61105030), the Fundamental Research Funds for the Central Universities (ZYGX2011J021), and the Scientific and Technical Supporting Programs of Sichuan Province (2013GZ0054).

[1] J. Heuer and A. Kaup, "Global motion estimation in image sequences using robust motion vector field segmentation," in *Proceeding of the 7th ACM International Conference on Multimedia* (ACM, 1999), pp. 261–264.
[2] Y.-W. He, B. Feng, and S. Yang, "Fast global motion estimation for global motion compensation coding," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE, 2001), pp. 233–236.
[3] Y. Chen, R.-H. Zhang, L. Shang, and E. Hu, "Object detection and tracking with active camera on motion vectors of feature points and particle filter," Rev. Sci. Instrum. **84**, 065109 (2013).
[4] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proceedings of the European Conference on Computer Vision* (Springer, Berlin/Heidelberg, 2006), pp. 404–417.
[5] Y. Liu, W. Zhou, H. Yin, and N. Yu, "Tracking based on SURF and superpixel," in *Proceedings of the Sixth International Conference on Image and Graphics* (IEEE, 2011), pp. 714–719.
[6] W. He, T. Yamashita, H. T. Lu, and S. H. Lao, "Surf tracking," in *Proceedings of the IEEE 12th International Conference on Computer Vision (ICCV)* (IEEE, 2009), pp. 1586–1592.
[7] H. Xiong, J.-J. Wu, and J. Chen, "K-means clustering versus validation measures: A data distribution perspective," IEEE Trans. Syst. Man Cybern. **39**(2), 318–331 (2009).
[8] M. Zeng and Y.-F. Li, "Region-based weighted histogram and its application to image contrast enhancement," in *Proceedings of the International Conference on Mechatronics and Automation* (IEEE, 2009), pp. 988–993.
[9] Z. P. Jiang, "Decentralized disturbance attenuating output feedback trackers for large-scale nonlinear systems," Automatica **38**(8), 1407–1415 (2002).