Single Object Recognition and Motion Tracking with Fast Multi-Frame Processing Algorithm

Zhen Gao* and Ishwar Singh
School of Engineering Technology
McMaster University,
Hamilton, Ontario, Canada
* Email: gaozhen [AT] mcmaster.ca

Abstract—**This work is focused on the experimental research of image processing for fast, accurate and efficient pattern recognition and motion tracking. To deal with the images and videos in real time as human beings is paramount about robot research if the robot wants to interact with the external environment in a humanoid way. In this research, object recognition and motion tracking are investigated simultaneously. An efficient and straightforward algorithm is developed to identify and track the motion trajectory of a single object. The experiments are conducted and the related results prove the effectiveness of the proposed method.**

Keywords- object recognition; motion tracking; image processing; motion trajectory

I. INTRODUCTION

Due to the rapid development of robotics and image processing technology, object recognition and motion tracking have been widely applied in different areas in the current industrialized and digitalized society [1-6]. Vision is one of the most important sensory functions if an autonomous mechatronic system wants to interact with the real-life environment and other objects or agents. That is why machine vision is researched by scholars from various areas including but not limited to computer science, electrical engineering, automation engineering, and mechatronics engineering [7-10]. Although motion tracking is an old topic, it is widely applied in many fields in decades. For robotics, if a robot possesses the functionality of real artificial intelligence, the equipped vision system is not only necessary but also should be advanced and perceptive. Currently, image processing is mainly based on mathematic calculation, i.e. matrix manipulation with n dimensions, typically, n = 2 or 3, which is very different with animal's natural vision processing mechanism.

For the application of object recognition and motion tracking, take an apple picking robot that autonomously picks apples in a farm as an example. First of all, it needs to identify what object is apple in a real-life environment based on the features of the object. Once it finds apples and their positions, its end-effector (gripper) may approach and perform the picking command. In case if the apple suddenly drops on the ground when implementing the picking procedure, the robot can detect the motion trajectory of the apple and track it on the ground. Finally, the robot can adjust the position information of the apple and pick it on the ground.

The main goal of this work is aiming at generating a doable computer vision module for robotics education, especially for the lab portion. A fast algorithm is developed to recognize the object and then record the motion trajectory in real-time based on multi-frame. Compared with some artificial intelligence approaches such as neural nets and support vector machines, the proposed method is much more straightforward and efficient. The proposed algorithm is also applicable in various scenarios including robot visual servo control, traffic monitoring, indoor precise positioning, and multi-agent collaboration.

II. METHODS

A. *Convert Video to Frames*

There are two main tools to investigate on pattern recognition and motion tracking, namely OpenCV and Matlab Image Processing Toolbox. This research is conducted based on the latter one. As a case study, the raw video is taken based on a mobile phone camera with 0.98 megabytes and dot MOV format. The function *VideoReader* is applied to acquire the raw video stream and save it as a file in the workspace. The number of frames is calculated based on the function of *NumberOfFrames*. The command to convert from video to frames is *read*. One issue needs to note is that order of the frame is alphabetic from the MSB to LSB (left to right). Therefore, if there is about thousands of frames, for example 2986, the first frame can be named as frame0001, and the last frame can be named as frame2986. The frames are written to the current folder in the format of dot PNG with *imwrite.*

B. *Background Reduction*

Due to the existing of noise in the raw frames, background subtraction is a paramount step for object identification. Background subtraction can be conducted in both color image and gray image. The main idea of background subtraction is to obtain the template image without the moving object. For example, if the camera wants to track a moving vehicle on a street, it needs to acquire the static street view without vehicle first. That image without moving traffic is called as a template background image. If the raw image with object is $F_{raw}(x, y)$, and the template image is $H_{tem}(x, y)$, it has,

$$T_{sub}(x, y) = F_{raw}(x, y) - H_{tem}(x, y) \quad (1)$$

where $T_{sub}(x, y)$ is the image with background reduction which is saved in the workspace.

III. 2-D Gaussian Filtering

The purpose of image filtering is to further reduce the noise when retaining the detailed characteristics of the raw image. As a necessary step for motion identification, the resulting performance of image filtering will directly affect the quality and correctness of the forthcoming operation. In this way, the unwanted noise signal, no matter with high or low frequency, can be somehow eliminated. In addition to reduce the noise, another main purpose of image filtering is to abstract the features of the object.

Following expression is used as the filter in this case [11].

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

The value of sigma will greatly affect the filtering effect. The larger sigma makes the image more smoothly; thus it also makes the image too vague. In this case, sigma = 2. The function can be used in the algorithm is Gaussian filter *imgaussfilt* with the coefficient sigma equals to 2.

Histogram of a frame represents the density distribution of hue, namely tone. Usually, horizontal axis of the histogram means color hue variations, and the vertical axis means the amount of pixels in the specific hue. Basically, for single object recognition, based on the observation of specific-hue threshold, the related object can be identified.

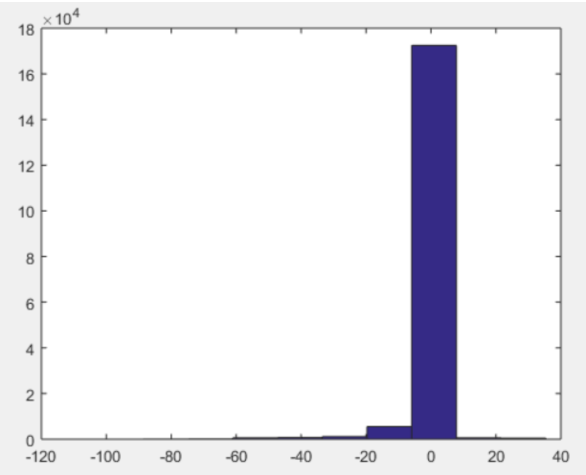


Figure 1. Histogram after Gaussian filter

Figure 1 shows the histogram after Gaussian filtering. Because the size of object is relative small compared with the background, the major hue distribution in the figure belongs to the background. Through observation, it can be found that the threshold of the object approximately below to -70 in x-axis.

IV. Object Recognition and Continuous Tracking

A. Object Recognition

After finding the threshold result of the raw image, a sequence of binary images which contain the object pixels information is generated. For a convex shaped object, a quick algorithm to find its center of mass is to calculate the average values of non-zero pixels in x axis and y axis. The average values represent the center pixel location of the identified object. If considering about the multi-object tracking, the issue become finding the multiple center points for multiple non-zero areas. That means the proposed single object recognition and motion tracking with fast multi-frame processing algorithm can be easily transplanted to multi-object recognition and motion tracking

B. Continuous Tracking

Continuous tracking is implemented based on the image addition of the sequence of raw image frames and the sequence of binary images which contain the identified center points. However, since the raw image is a color image with three dimensions; while the image which contains the center of mass of the object is a binary image. To achieve image overlay, following pseudo code should be conducted,

First channel = 1st dimension of raw image;
Second channel = 2nd dimension of raw image;
Third channel = 3rd dimension of raw image;

First channel of motion tracking image = First channel;
Second channel of motion tracking image = Second channel + binary image which contains the center of mass of the object × 255;
Third channel of motion tracking image = Third channel.

Which means the color of center of mass is green. It is based on the users' decision. If the recognized point is white, as shown in the following figure, the 2nd part of the pseudo code should be,

First channel of motion tracking image = First channel + binary image which contains the center of mass of the object × 255;
Second channel of motion tracking image = Second channel + binary image which contains the center of mass of the object × 255;

Third channel of motion tracking image = Third channel + binary image which contains the center of mass of the object × 255.

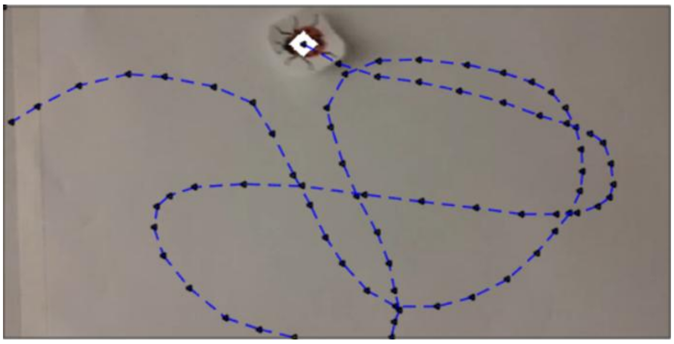


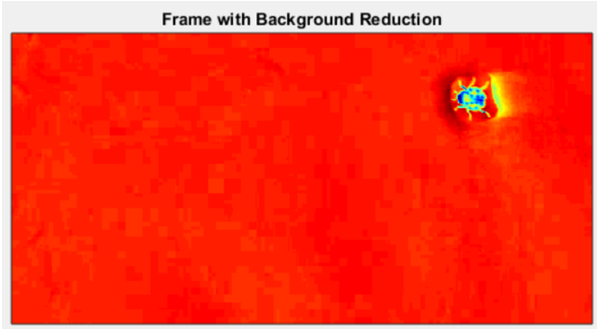
Figure 2. The identified object with white center point

V. RESULTS

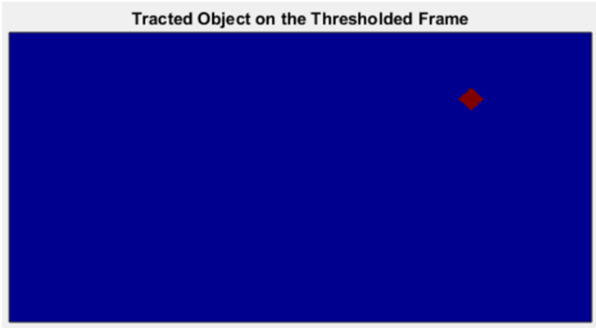
The proposed algorithm totally takes 2.707687 seconds including massive matrix operation and image reading/writing, since the motion tracking is not pre-trained and the algorithm itself finishes both object recognition and motion tracking. Compared with some pre-trained intelligent algorithm such as artificial neural network and support vector machine, the proposed method is relatively efficient and effective.

Three snapshots of single object recognition and tracking are given in figures 3-5, in which each graph shows (1) the original camera-taken image, (2) frame with background reduction, (3) Tracked object on the threshold frame, and (4) tracked and marked object with motion trajectory. Through the experiments, it can be found that the proposed algorithm for object identification and tracking under a relatively simple background is effective.

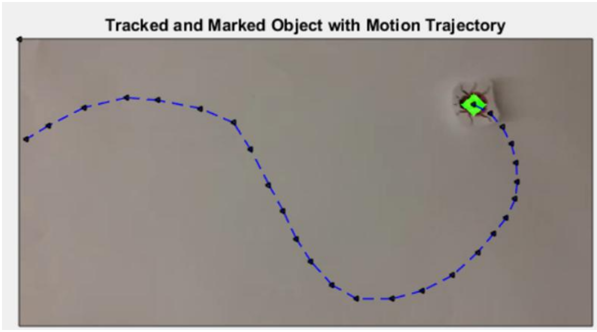
(a)



(b)



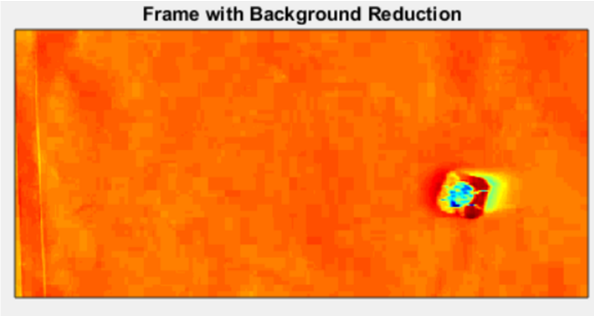
(c)



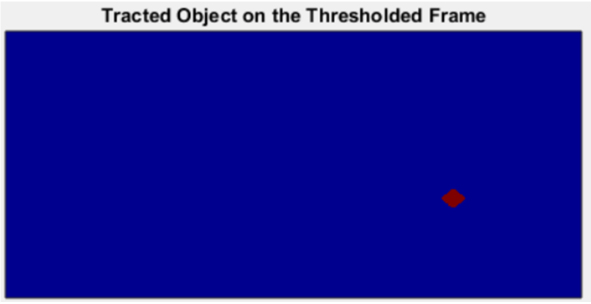
(d)

Figure 3. First snapshot of motion tracking; (a) raw image, (b) 2-D Gaussian filter, (c) center of mass of the object in a binary image, (d) tracked and marked object with motion trajectory

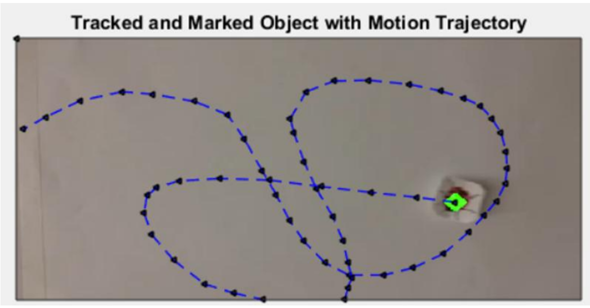
(a)



(b)



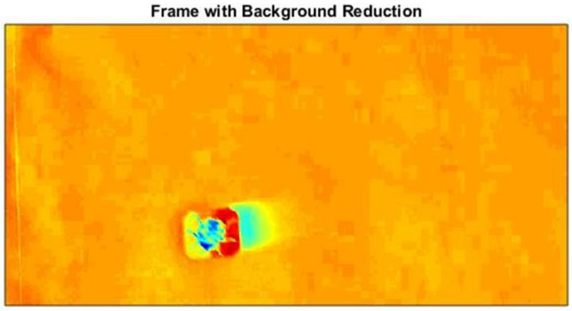
(c)



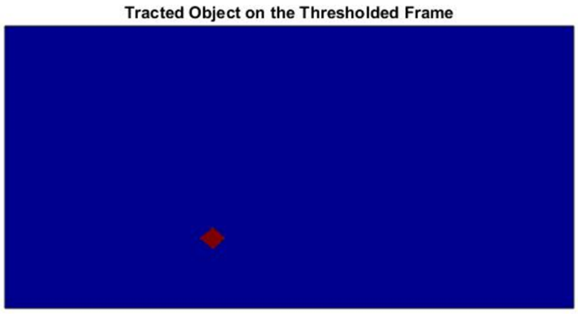
(d)

Figure 4. Second snapshot of motion tracking; (a) raw image, (b) 2-D Gaussian filter, (c) center of mass of the object in a binary image, (d) tracked and marked object with motion trajectory

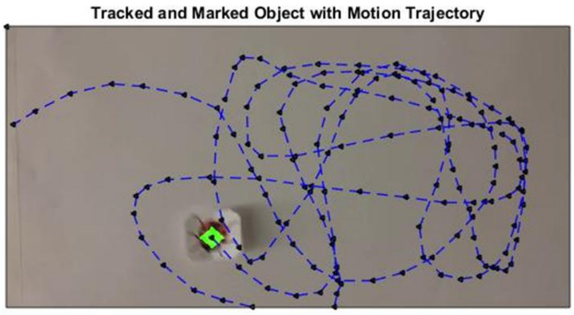
(a)



(b)



(c)



(d)

Figure 5. Third snapshot of motion tracking; (a) raw image, (b) 2-D Gaussian filter, (c) center of mass of the object in a binary image, (d) tracked and marked object with motion trajectory

VI. CONCLUSIONS

Nowadays, machine vision has become more and more important due to the popularity and increasing market of various applications of robotic, intelligent, and sensory systems. Object recognition and motion tracking are investigated together in this research since if the object needs to be found first, and then it can be tracked. A straightforward but efficient algorithm is developed to track the motion trajectory of a single object in a recorded video stream. Compared with some artificial intelligence approaches such as neural nets and support vector machines, the proposed method is relatively faster. The proposed method can be applied in many applications such as vision servo of robotic manipulation, traffic monitoring and traffic accident identification, indoor precise positioning, and so on.

REFERENCES

[1] A.P.C. Duarte, B.A. Silva, N. Silvestre, J. de Brito, E. Júlio, "Mechanical characterization of rubberized concrete using an image-processing/XFEM coupled procedure," Composites Part B: Engineering, vol. 78, 2015, pp. 214-226

[2] I. Lobato, D. Van Dyck, "MULTEM: A new multislice program to perform accurate and fast electron diffraction and imaging simulations using Graphics Processing Units with CUDA," Ultramicroscopy, vol. 156, 2015, pp. 9-17

[3] G. Lo Re, F. Lopresti, G. Petrucci, R. Scaffaro, "A facile method to determine pore size distribution in porous scaffold by using image processing," Micron, vol. 76, 2015, pp. 37-45

[4] M.V. Ananyev, D.I. Bronin, D.A. Osinkin, V.A. Eremin, R. Steinberger-Wilckens, L.G.J. de Haart, J. Mertens, "Characterization of Ni-cermet degradation phenomena I. Long term resistivity monitoring, image processing and X-ray fluorescence analysis," Journal of Power Sources, vol 286, 2015, pp. 414-426

[5] N. Arias, M. Méndez, JL. Arias, "The recognition of a novel-object in a novel context leads to hippocampal and parahippocampal c-Fos involvement," Behavioural Brain Research, vol. 292, 2015, pp. 44-49

[6] Y. Shibasaki, A. Hayata-Takano, K. Hazama, T. Nakazawa, N. Shintani, A. Kasai, K. Nagayasu, R. Hashimoto, M. Tanida, T. Katayama, S. Matsuzaki, K. Yamada, M. Taniike, Y. Onaka, Y. Ago, J.A. Waschek, K. Köves, D. Reglődi, A. Tamas, T. Matsuda, A. Baba, et al. "Atomoxetine reverses locomotor hyperactivity, impaired novel object recognition, and prepulse inhibition impairment in mice lacking pituitary adenylate cyclase-activating polypeptide," Neuroscience, vol. 297, 2015, pp. 95-104

[7] B.T.J. van Hagen, N.P. van Goethem, D.C. Lagatta, J. Prickaerts, "The object pattern separation (OPS) task: A behavioral paradigm derived from the object recognition task," Behavioural Brain Research, vol. 285, 2015, pp. 44-52

[8] B. Pyndt Jørgensen, L. Krych, T.B. Pedersen, N. Plath, J.P. Redrobe, A.K. Hansen, D.S. Nielsen, C.S. Pedersen, C. Larsen, D.B. Sørensen, "Investigating the long-term effect of subchronic phencyclidine-treatment on novel object recognition and the association between the gut microbiota and behavior in the animal model of schizophrenia," Physiology & Behavior, vol. 141, 2015, pp. 32-39

[9] E. Charles Leek, Lina I. Davitt, Filipe Cristino, "Implicit encoding of extrinsic object properties in stored representations mediating recognition: Evidence from shadow-specific repetition priming," Vision Research, vol.108, 2015, pp. 49-55

[10] M.A. As'ari, U.U. Sheikh, E. Supriyanto, "3D shape descriptor for object recognition based on Kinect-like depth image," Image and Vision Computing, vol. 32, Issue 4, 2014, pp. 260-269

[11] https://en.wikipedia.org/wiki/Gaussian_filter