Using SIFT and Sliding Window to Detect and Invent Literature for Library Robot

Nguyen Phuong Nam, Nguyen Cong Nam, and Nguyen Truong Thinh
Department of Mechatronics, HCMC University of Technology and Education, Ho Chi Minh City, Viet Nam
Email: namlibra1998@gmail.com, ng.cong.nam98@gmail.com, thinhnt@hcmute.edu.vn

***Abstract*—In recent years, workload of librarian has increased even more in academic or university libraries. In each library, the librarian need to do enormous thing such as guiding, searching information, sorting, arranging, inventing the books or journals, magazines… and that is the daily workload for professional librarians. In this paper, a method of identifying books placed on bookshelves in the library is described like as integrating the identification systems into the robot to carry out service works as well as to reduce service manpower. Book identification's implementation for inventing the book or journals, magazines… focused on SIFT model in combination with Sliding Window method to recognize the spine of books on bookshelf. The experimental results were greatly potential. The accuracy of inventing with the method of SIFT is up to approximately 95% or more in the case of non-duplicated books.**

***Index Terms*—SIFT, sliding window, image processing, library robot, service robot**

I. Introduction

The everyday task of book statistics is still a boring and time-consuming task. To do this task requires the librarian to identify exact position and order of the book on the shelf. As usual, at the back of each book, there is a tag like barcode or RFID and these things can be read with the specialized reader. Each tag has information about the book and its location. However, we want to develop a different method of scanning books and directing robots with the camera mounted on the robot's gripper. Our robot has 4 degrees of freedom which including three degrees of freedom to help the robot can move freely in the library environment. The mission of the last degree of freedom in our robot is to help robot's gripper move up and down between each floor of a bookcase. The main contributions of our robot are as follows:

- Navigate the robot to exactly to the location of bookshelf. Then extract the positon based on camera install from above to help identify the directional vector of robot. Since this vector alignment must always be parallel to the vector indicating the direction of the bookshelf. This allows the camera's angle which is mounted on the robot's arm can be approximately 90 degrees with the bookshelf. It also means that the robot facing with the bookshelf.
- Detect book spine through SIFT model combines with a sliding window method. This is the mainly thing that we present in this paper. Based on the database of the cropped image of the back of the book, the algorithm can scan on the input image respectively and show the positon and the order of the book along with the pixel position on the photo frame.

Many researcher has shown the great contribute to develop an efficient method recognizing book on bookcase [1]-[7]. The previous work on book identification focused on the spine of the book, mainly by separating the book spine and applying OCR (Optical Character Recognition) module to analyze the writing on it. Then the results will be compared with the existing words in database to find whether if it matches or not. Nguyen Huu Quoc and Won Ho Choi et al [1] use a high-frequency filtering and thresholding in each frame of the camera. They look for the vertical lines of the book spine which means the boundary the book, then they will extract the text data on the cover and use handwriting recognition to read the title name of the book. Recently, Xiao Yang et al [2] has also introduced the method of using deep neural network to recognize the character on book spine. Combined with the processing techniques such as Hough Transform, their method promises to give better results than the OCR module, which was applied for a very long time. In our cases, in the past, we have tried to use the Hough Transform to find the edge the book. However, the negative side of this method is that they were typically affected by noise and lighting condition so the identification processing was very unstable. Therefore, we choose to find a new solution to fix problems. We apply the SIFT model to identify the book on the bookshelf. Through experiments, the SIFT descriptor has been proven to be very useful in practice for image matching and object recognition under real-world conditions. In addition, we have also used the method which called "Sliding window". Utilizing both of them, we are able to detect books on bookshelf in a various location. The performance results show the great potential in statistics and book searching.

II. SIFT Algorithm

The scale invariant feature transform (SIFT) algorithm, developed by Lowe [8]-[10], is an algorithm for image

Manuscript received September 3, 2020; revised May 20, 2021.

doi: 10.18178/ijmerr.10.7.386-391

features generation which are invariant to image translation, scaling, rotation and partially invariant to illumination changes and affine projection. SIFT algorithm can be used to detect distinct features in a picture. Once features are detected for two different images, one can use these features to answer questions like "are the two images taken of the same object?" and " Is the object present exactly in the second image if it appear in the first image ?" [11]. When computed of SIFT image features, there are four consecutive phases briefly described in the following:

A. Scale-Space Local Extrema Detection

This stage of the filtering attempts to spot those locations and scales that are identifiable from different views of the same object. This can be efficiently achieved employing a "scale space" function. Further it's been shown under reasonable assumptions it must be supported by the Gaussian function. The scale space is defined by the function:

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{1}$$

Where G(x, y, σ) is a variable-scale Gaussian, * is the convolution operator and I(x, y) is the input image. In the scale-space, there are various methods can be used to detect stable keypoint locations. Difference of Gaussians is one such method, locating scale-space extrema, D(x, y, σ) by computing the difference between two images, one with scale k times the other. D(x, y, σ) is then given by:

$$D(x,y,\sigma)=L(x,y,k\sigma)-L(x,y,k\sigma) \tag{2}$$

In order detect the local maxima and minima of D(x, y, σ), 8 neighbors around each point is compared with it at the same scale, and its 9 neighbors which are up and down one scale as presented in Fig. 1.

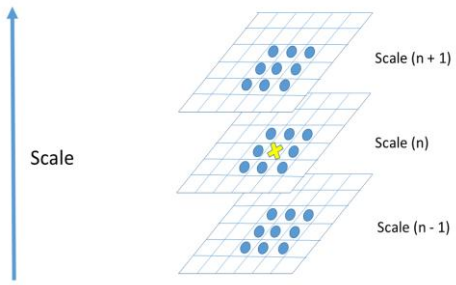

Figure 1. An extrema is defined as any value in the DoG greater than all its neighbors in scale-space.

If the minimum or maximum is the value of all these points then this point is an extrema. The search for extrema which excludes the first and the last image in each octave since they do not have a scale above and a scale below respectively. To increase the amount of extracted features, the input image will be treated by SIFT algorithm after it is doubled, which however increases the computational time significantly.

B. Keypoint Localization

As for the keypoints, the detected local extrema are good candidates. However, to make systems run more accuracy, they need to be exactly localized by fitting a 3D quadratic function to the scale-space local sample point. The quadratic function is computed using a second order Taylor expansion which has the origin at the sample point. After that, local extrema with low contrast and such that correspond to edges are not able to use because they're sensitive to noise.

C. Orientation Assignment

Once the SIFT-feature location is determined, a main orientation will be assigned to each feature based on local image gradients as shown in Fig. 2. For each pixel of the region which is around the feature location, the gradient magnitude and orientation are computed respectively as the equations following:

$$m(x,y)=\sqrt{\left(L(x+1,y)-L(x-1,y)\right)^2+\left(L(x,y+1)-L(x,y-1)\right)^2} \tag{3}$$

$$\theta(x,y)=\tan^{-1}\left(\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x+1,y)}\right) \tag{4}$$

The Gaussian window whose size depends on the feature octave weight the gradient magnitudes. In order to detect the local maxima and minima of D(x, y, σ), each point is compared with its 8 neighbors at the same scale, and its 9 neighbors up and down one scale. After having the result, if this value is the minimum or maximum of all these points then this point is an extrema.

D. Keypoint Descriptor

At the selected scale in the region around each keypoint where the local image gradients are measured. These are convert into a representation that allows for significant levels of local shape distortion and change in illumination.

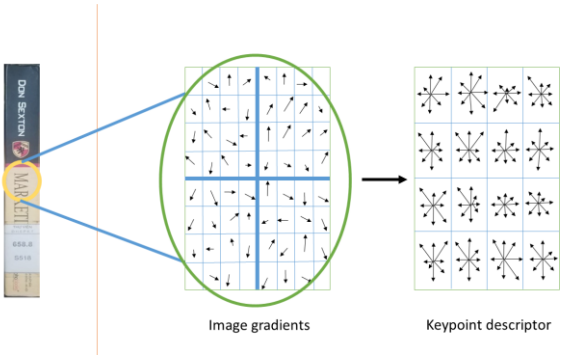

Figure 2. Orientation assignment.

III. Book Recognition

A. Preparation Step

Before starting to apply the SIFT model of identifying the book, we collect the database. At first, we take the pictures of the book individually, then in the next step, we will crop the image to make the border of the photo close to the edge of the book spine (Fig. 3). The main reason is that we want to remove all the excess scene elements around the book.

Figure 3. Crop the boundary of the book.

B. Book Recognition

Once the database of the books in the library has been synthesized as the cropped image. At this time, when implementing, the book recognition system only focuses on the image of the spine in the frame. It will extract keypoints and compute its descriptor then comparing it with the descriptor in the image which extracted from the database to see whether if it matches or not. It is obvious that in the processing, many of initial matches maybe incorrect due to a complicated background which have some ambiguous features. Therefore, we will take the priority on the cluster which have some features first than the individual feature matches because these cluster have much higher probability of detecting correct object. In addition, we will adjust the threshold of the number of the key descriptor. When a certain threshold is reached, the boundary of the book's spine will be formed with the detailed pixel of its location on the frame. To make robot navigate the book on the bookshelf, we transform those pixels which were extracted from above to real coordinates in outer world. Because of the invariance of SIFT to translations, rotations and scaling transformation in the image domain, moreover it is also robust to moderate perspective transformations and illumination variations so when attaching it to our systems, robot can detect book in different position such as inclined and horizontal position.

C. Combining with Sliding Window Method

The main duty of SIFT is to extract the features of the query image to compare with the feature in the image of database. Because of this, it is quite hard for it to identify multiple duplicates object in the same frame of the camera. As a matter of fact, bookshelf in a real life always have a number of duplicated books placed next to each other so in order to solve this problem, we decide to approach a method "Sliding Window". In fact, Sliding Window has been widely used in object detection like cars, faces or pedestrian. We want to use it in our book detection systems. The sliding window is simply a rectangle with a predefined length and width which length side equal to the height of the image and one more parameter name step, which can be adjusted according to situation. To identify the entire bookshelf, we choose a sliding window of size (x) and then combined with the SIFT to make the identification processing in that (x) area, this process will repeat until it reaches the end of the image LIKE AS Fig. 4. Moreover, between each process is a step size which is define the distance between the new window and the previous window. In each window, the output return whether the book has been detected or not. Usually, the sliding window scan multiple times from small to large sizes to optimize the object recognition but in the case of the library bookshelf, the robot's navigation is always at a fixed focal length with the bookshelf. It will ensure that the bookshelf will be accurately captured within in the frame of camera. Therefore, we would reduce the time to search for the books and have more time to optimize the systems. Specifically, the process of combining SIFT and sliding window is as follows: 1) Initializing window on image. 2) Analyzing features of window. 3) Classify and identify if there is enough threshold reliability. 4) Clean up data and move on to the next window.

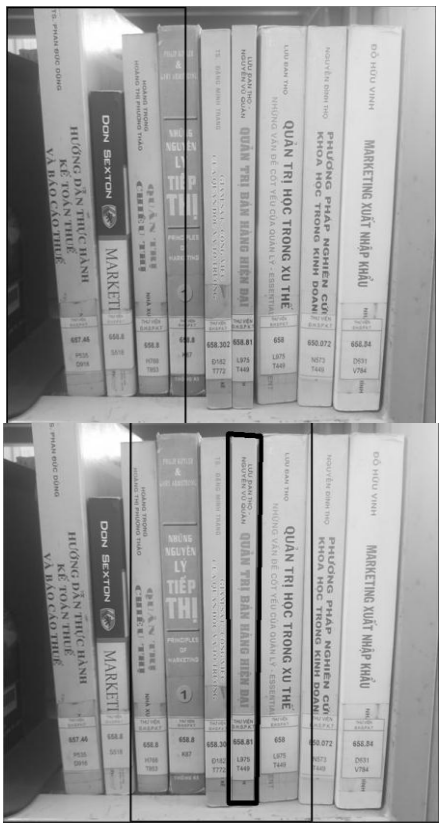

Figure 4. Combining SIFT with the sliding window to increase the performance of the algorithm.

IV. Experimental Results and Discussions

Our system able to identify books on shelves in a library environment. The result of the book recognition is shown in Figure. During scanning, even when the filter measures are applied to help minimize the noise in the image but when we doing some experiments, noise still exist. It makes the system have a wrong detection. Instead of detect a right boundary which close to the book spine, it shows a weird shape on the image such as distortion, or irregularly shaped borders. These problems can lead to unexpected results. When facing with them, we can adjust the threshold of features extraction which suits with the situation but in addition, we add a method to control pixels of the 4 corners vertices of the book and constraint conditions on the angle and length of the edges. Once all the conditions have been met, our systems will ensure that area is definitely the book spine, which will improve the efficiency of the algorithm.

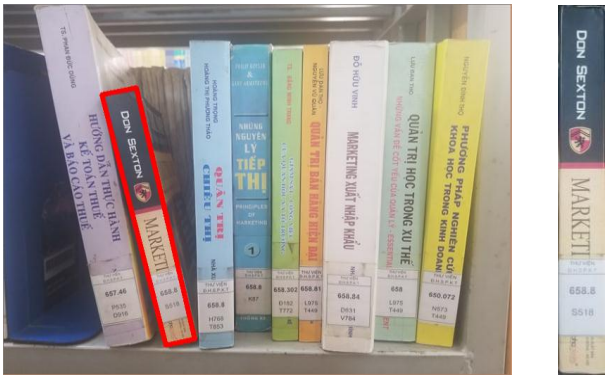

Figure 5. Book spine detection using SIFT: (a) Result image of bookshelf, (b) Database image.

The Robot will go to the shelf address to search for the book. For example, the book that Robot want to detect is in Fig. 5(b) and robot will go to the shelf where it located and take a photo. Then the system will operate to compare the captured image with the book in the database. It detects at which pixel coordinates of the input image match with the key descriptor of the trained image. The result of the process is shown in Fig. 5(a).

In case there are duplicated books on the shelf, when taking an experiment, based on applying SIFT and Sliding Window, we found that the program has successfully implemented the scan books even when two identical books are lined up next to each other (Fig. 6).

In terms of book statistics, the robot will have to process the identification for each book according to the available database. This is the mandatory thing in our system during the recognition process. Fig. 7 shows the result of demo book statistics. There is a period of time for the system to scan each book spine, which is shown in the Fig. 7. The interactive time to scan for a book is approximately 400ms for step size of 35 on a 720x540 pixels photo frame. The results are experimented shown in Table I. These is only 3 failures out of 10 trials, which could give 70% success rate. This is the approximately percentage that we have tested on nearly 70 bookshelves. Cause the limitation of robot's moving and space to do experiment so we could only test 10 bookshelves each time doing a research. Sometime the percentage of successful book recognition could be better due to the objective condition but we conclude the percentage which our robot performs most stable. During our experiments, our team had faced a number of problems, but the identification process still produced outstanding results. These problems are:

- Book maybe skewed and the ratio of books in images maybe larger or smaller than the rates in images in the database.
- The saturation of the book in the image sometimes is not the same as the original color due to the light in the library environment is unstable.
- The quality of camera is not good: This can cause a decrease in the image quality. So that when the system analyzes images, the results maybe weird or deviated from the database

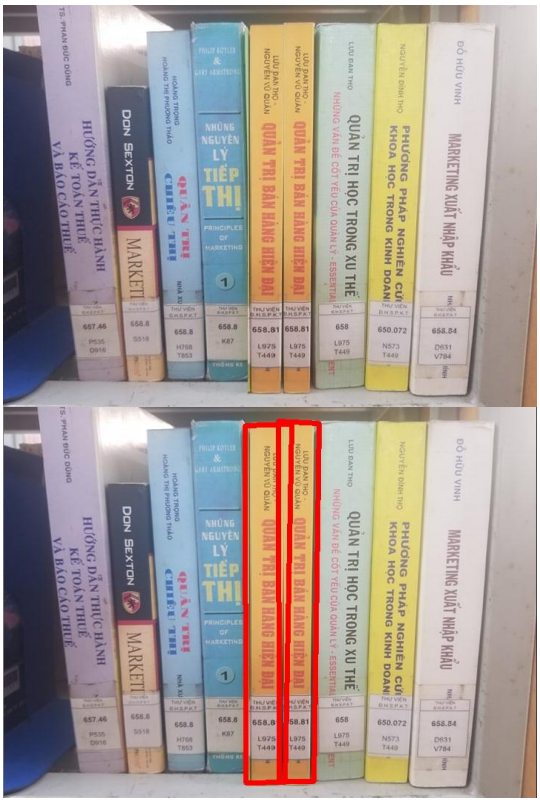

Figure 6. Book shelf which has a duplicate searching book, then result when combining SIFT and Sliding Window

Figure 7. The result image of bookshelf when applying fully process.

TABLE I. BOOK RECOGNITION EXPERIMENTAL RESULT

Bookshelf number	Real Book	Book recognized
1	9	9
2	8	9
3	11	10
4	8	8
5	9	9
6	10	10
7	11	9
8	9	9
9	10	10
10	10	10

● The shooting area of a camera is not bright enough due to the shadow of the robot.

To solve these problem, we intend to install LED lighting systems on robot so that it no longer has to depend on natural light conditions.

Figure 8. Experiment based library robot.

V. CONCLUSION

Sift has been widely applied by many object recognition systems because its invariant to rotation and scale. Thanks to the assistance of SIFT algorithm in combination with the sliding window method, we have implemented a book identification method to help the library management process. When applied to robots, it can help librarians to statistic the number of books in and out after a day. Based on the captured image, the identification system will extract the features of each book to compare with the database. The system has been tested on real images and gave great satisfactory results with the correct identification, approximately 70%. Due to the experiment, our system has shown potential in applying to library management and can be improved even more when combined with robots to not only identify books but also help robots interact do many things in library as the request of the users like retrieve and return the book where it is located.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Nguyen Phuong Nam, Nguyen Truong Thinh contributed to the analysis and implementation of the research, to the analysis of the results and to the writing of the manuscript. All authors discussed the results and contributed to the final manuscript. Besides, Nguyen Truong Thinh conceived the study and were in charge of overall direction and planning. Nguyen Truong Thinh is a corresponding author.

ACKNOWLEDGMENT

The authors wish to thank Ho Chi Minh City University of Technology and Education, Vietnam. This study was supported financially by HCMUTE Open Lab and Ho Chi Minh City University of Technology and Education, Vietnam.

REFERENCES

[1] N. Quoc and W. Choi. A framework for recognition books on bookshelves. In *Proc. International Conference on Intelligent Computing (ICIC'09)*, Ulsan, Korea, September 2009.
[2] X. Yang et al., "Smart library: Identifying books on library shelves using supervised deep learning for scene text reading," in *Proc. 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, Toronto, ON, 2017, pp. 1-4.
[3] D. Chen, S. Tsai, K. H. Kim, C. H. Hsu, J. P. Singh, and B. Girod, "Low-cost asset tracking using location-aware camera phones," Number 1, San Diego, California, USA, 2010.
[4] D. M. Chen, S. S. Tsai, B. Girod, C. H. Hsu, K. H. Kim, and J. P. Singh, "Building book inventories using smartphones," in *Proc. ACM Multimedia (MM'10')*, MM '10, Firenze, Italy, 2010. ACM.
[5] D. Crasto, A. Kale, and C. Jaynes, "The smart bookshelf: A study of camera projector scene augmentation of an everyday environment," in *Proc. IEEE Workshop on Applications of Computer Vision (WACV'05)*, Breckenridge, CO, January 2005.
[6] M. Loechtefeld, S. Gehring, J. Schoening, and A. Krueger. "Shelftorchlight: Augmenting a shelf using a camera projector unit," *UBIProjection 2010 - Workshop on Personal Projection*, 2010.
[7] K. Matsushita, D. Iwai, and K. Sato, ,Interactive bookshelf surface for in situ book searching and storing support," in *Proc. the 2nd Augmented Human International Conference*, New York, NY, USA, 2011.
[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
[9] D. G. Lowe, "Object recognition from local scale-invariant features," *International Conference on Computer Vision, Corfu, Greece*, pp. 1150-1157, September 1999
[10] D. G. Lowe, "Local feature view clustering for 3D object recognition," *IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii*, pp. 682-688, December 2001.
[11] T. Bakken, "R&I Research Note" Telenor ASA, 2007.

Nguyen Phuong Nam is Graduate student of Ho Chi Minh City University of Technology and Education (Vietnam). Major is mechatronics. He has three years of experience researching in mechatronics, intelligent control. In addition, He also got many scientific research awards.

Cong-Nam Nguyen is Graduate student of Ho Chi Minh City University of Technology and Education (Vietnam). Major is mechatronics. He has three years of experience researching in mechatronics, intelligent control. In addition, He also got many scientific research awards.

Nguyen Truong Thinh is Associate Professor of Mechatronics at Ho Chi Minh City University of Technology and Education (HCMUTE). He received his Ph.D in Mechanical Engineering at Chonnnam National University (Korea) in 2010 and obtained a positive evaluation as Associate Professor in 2012. His main research interests are Industrial Robotics, Service robotics, Mechatronics, Industrial Automation.