Real-Time Object Detection and Recognition System Using OpenCV via SURF Algorithm in Emgu CV for Robotic Handling in Libraries

Corina Monica Pop, Gheorghe-Leonte Mogan, and Răzvan Gabriel Boboc

Abstract—This paper describes an experimental system that has been designed, implemented and tested for object recognition and tracking in still, respectively dynamic images – successive video frames captured in real time (live) with a web camera – based on Intel's open source computer vision functions library, OpenCV (*Open Source Computer Vision*). We propose a real-time object recognition system in intelligent library environments. The system consists of two key modules: feature extraction and object recognition. Specific detectors such as SIFT (*Scale-Invariant Feature Transform*) and SURF (*Speeded Up Feature Robust*) are efficient methods that provide high quality features, yet are too computational for use in real-time applications.

This paper also proposes a low complexity, robust object recognition and tracking method using advanced real time feature matching. It combines Microsoft Visual Studio 2008 Express Edition C# with OpenCV Function library using SURF algorithm in Emgu CV to develop the software. The tests showed that the proposed system and method are more efficient and more robust than in most traditional applications.

Index Terms—Computer vision, object recognition, object tracking, OpenCV, SURF.

I. INTRODUCTION

Research in the field of library automation is characterized by a long history of robot assistance applications such as book cataloguing, retrieval and return [1], [2]. The studies focused on solving the technical problems at fixed locations in the library. Mobile robots [3], [4] which are both autonomous and capable to detect and react to environmental factors have recently attracted the attention of libraries. Some libraries have introduced mobile robots and use them to guide student users towards the adequate shelf [5], [6]. However, there are only a few cases in which educational research has addressed the field of service robot applications [7]–[9] designed for use in libraries. These studies have emphasized the general trend which regards robots as adequate agents to ensure user guidance and involvement in learning activities.

References [9]–[12] as well as others emphasize systematic user involvement in their studies on the improvement of robot structures in order to facilitate tasks such as resource location, detection, and recognition in a library (books, journals, CDs etc.) [13]–[17], devised for robot handling.

Object detection is one hot research topic in computer

vision due to its numerous applications.

Object recognition and tracking are major tasks in several computer vision applications, such as Augmented Reality (AR), interactive systems and robotic systems.

Some of the main advantages of vision systems, in robotic applications, are the simplicity of the algorithm, the low cost, and the reduced need for maintenance, while aspects such as fast and effective identification are still unsolved. Even though adequately efficient and accurate algorithms have been developed, the processing speed still fails to meet the modern manufacturing requirements [18].

Most of the objects tracking approaches based on feature matching are highly computational and less robust in various environments. To efficiently track an object in a video sequence, at first, feature points are extracted from the object of interest. The extracted objects then recognize the target object, and the detected object is continuously tracked on the input stream [19]-[21]. Object recognition consists of two main steps. The first step refers to extracting a feature of interest in a target object. The second one is matching a corresponding point on a target video sequence. In object recognition based on feature, the extraction of an accurate feature on the target has influenced the performance of object recognition. The performance of recognition can be improved by using a large number of feature points, which constitute the extracted area of interest. However, it is impossible to recognize the object in a real time environment given the increased computing complexity. On the other hand, accurate recognition cannot be achieved. Thus, the algorithm of object recognition requires accurate detection of the object and less computational complexity in real time surveillance systems [22]. One of the most efficient methods in real-time object recognition is the SURF algorithm.

II. PROBLEM STATEMENT AND ASSUMPTIONS

The method to be described in this paper was intended for:

- 1) finding the position of objects on image/frame
- 2) recognizing the object requested by the user for robotic handling

Following assumptions were taken into consideration:

- 1) environment lighting is controlled and constant
- 2) the space designed for system operation is static
- 3) the system will only recognize the object requested by the user.

Manuscript received September 22, 2017; revised October 30, 2017.

The authors are with the Faculty of Mechanical Engineering, Transilvania University of Brasov, Brasov, 500024 Romania (e-mail: popcorina@unitbv.ro, mogan@unitbv.ro, razvan.boboc@unitbv.ro).

III. OVERVIEW

This paper considers a mobile robot platform Pioneer 3 All - Terrain (P3-AT) featuring:

- a rigid structure which provides support to additional equipments
- a touch screen monitor- for simple and friendly interaction, designed to help users to obtain the required information about library book status which is of interest to them
- a monitor-mounted frontal view webcam for environment, objects, and visitor surveillance, (library users: students, master students, doctoral students, teachers etc.)
- 4) a Radio-Frequency Identification (RFID) reader to identify the books on a shelf
- 5) a robotic manipulator provided with a prehension system designed for gripping books (manipulating system).

The robot is designed to perform in indoor library spaces, more precisely, a reading room.

The block diagram showing the basic robot design is presented in Fig. 1.



Fig. 1. Block diagram showing the basic robot design.

The project consists of two parts:

- the hardware which includes the mobile robot platform (mobile), the additional equipments, their mount and the gripping mechanism
- 2) the software which includes robot controller, image processing, and the wireless data transmission systems.

This paper focuses mainly on the software aspects of the project.

IV. THE PROPOSED METHOD

Various object recognition methods have been investigated over the two last decades. More recently, SIFT [23] and its variants such as PCA-SIFT (*Principal Components Analysis*) [24] and GLOH (*Gradient Location-Orientation Histogram*) [25] have been successfully applied for many image matching applications.

The goal of this study is to develop a book detection, recognition and tracking system using OpenCV library via SURF in Emgu CV capable to capture the interest objects that appear in front of camera. One of OpenCV's goals is to provide a simple computer vision infrastructure. The system is designed as a tool for detecting and recognizing the requested books for robotic handling. This paper proposes a real-time object recognition system in intelligent library environments using the new SURF algorithm. The object recognition system is made up of two key modules: feature extraction and object recognition.

A. Feature Extraction using SURF

Image representation, image classification and retrieval, object recognition and matching, 3D scene reconstruction, motion tracking, texture classification, robot localization, and biometrics systems, all rely on the presence of stable and representative features on the image. Image feature is a symbolic representation of the contents of an image through the detection of interest points in the given image and the extraction of features from these points [26].

The SURF feature is a enplane rotation and scale invariant feature. It contains descriptor and detector in the point of interest. The detector identifies the point of interest in the image and the descriptor describes the features in the point of interest and develops the feature vectors of the interest. The Hessian matrix $H(x, \sigma)$ in a image *I* at a point x=(x,y) in *x* at σ scale is defined as follows

$$H(x,\sigma) = \begin{bmatrix} L_{x}(x,\sigma) & L_{y}(x,\sigma) \\ L_{y}(x,\sigma) & L_{y}(x,\sigma) \end{bmatrix}$$
(1)

)

where $L_{xx}(x, \sigma)$ is the convolution value of the Gaussian second order partial derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image *I* in point *x* and similarly for $L_{xy}(x, \sigma)$ and $L_{yy}(x, \sigma)$. The scale and the area of the point of interest are chosen by depending on the

B. Object Recognition

quality given by the Hessian matrix [27].

The SURF algorithm is used in object recognition due to its powerful attributes, including scale, translation, lighting, contrast, and rotation invariance [28].

- The algorithm consists of four main parts:
- 1) Integral image generation,
- 2) Fast-Hessian detector (interest point detection),
- 3) Descriptor orientation assignment (optional),
- 4) Descriptor generation.

The flowchart of the SURF algorithm is shown in Fig. 2.



Fig. 2. Flowchart of SURF algorithm.

Fig. 3 shows the detailed description of the algorithm. There are two images: one is captured from the camera (observed image) and the other is the image containing the cover of the requested book (object image). In the next step both images are converted into grayscale, in a preliminary phase. As mentioned before, the SURF algorithm contains an interest point detection sequence that uses a *Hessian corner detector*. After conversion, the *SURFDetector()* function from Emgu CV library can be applied. This function contains a threshold parameter, which determines how large the Hessian filter output should be so that a point could be used as an interest point. A larger value will result in fewer interest points. In our case, this parameter was set to 500. In the following step, features of both the object image and the observed image

are extracted using the *DetectFeatures()* function and then compared to determine a match, using *MatchFeature()*. Next, the Emgu CV functions *VoteForUniqueness*, and *VoteForSizeAndOrientation* are used, in order to check for and eliminate the matched features that are not unique or not rotated and scaled beyond the fixed boundaries. After satisfying this condition, a homography matrix is set up for combining the object image and the observed image by using the corresponding key points after applying the *GetHomographyMatrixFromMatchedFeatures()*function. Finally, both observed image and object image are merged into one single image to be displayed while the key-points and matching lines between them are drawn.



Fig. 3. Detailed description of the SURF algorithm.

V. EXPERIMENTS AND RESULTS

In order to simplify the book handling process several conditions must be met regarding size, weight, maximum width, and upright position of the books on the shelf with spaces between covers, or the type of book cover.

In accordance with the tests, the main stages that describe a fully automated (robotic) handling of a book in a real environment are presented in Fig. 4.

Book identification and positioning of the manipulation system in front of it are based on the RFID method. The robot obtains the information about book size and weight from the RFID tag attached to the book and uses it to program the required movements (gripping, rotation, and prehension force of the prehension device) in order to extract and carry the requested document.

At first, the book is detected by a webcam while a software module recognizes it using computer vision. The user interface displays two image boxes and the application works as follows:

 "Choose Image" mode (Fig. 5) is used to identify the similarities on two images. The image containing the requested object (the queried image, e.g. a table or a book shelf) is uploaded to the left-hand box while the on the right-hand image is uploaded the object to be found (the template matching, e.g. the book or its spine).

In Fig. 5, book retrieval is performed on an image containing multiple information materials (books, journals, etc). The book that must be detected in the left-hand box is used as input in the right-hand box.



Fig. 4. The steps describing robotic handling of a book in real environment. a), b) – identifying the book and positioning the manipulation device in front of it; c), d) – gripping the book; e), f) – lifting and extracting the book from the shelf; g) – moving to the user area; h) – releasing the book; i) – return of manipulation device to initial position.



Fig. 5. Points of interest between requested object and queried image

The left-hand box will now display two images by concatenation of the right-hand template matching with the queried image which contains the requested book.

After detecting the book, the points of interest are extracted (the areas marked by a red circle) by pressing the button PerformSURF in the left-hand box. A descriptor is computed for each area. Next, a distance is calculated between the list of points of interest in the template matching (the requested book) and the other frame (the retrieved book). The user can decide whether to display these features (Fig. 5) or not (Fig. 6).

2) In the "Webcam" mode (Fig. 6), in the right-hand box will be uploaded the image of the requested object (template matching, book spine) while the left-hand box will display the video (stream) captured from the real scene by the webcam.

In this study, system performance is evaluated by using a database containing cover and spine images of the books.



Fig. 6. Recognizing the requested book spine

VI. CONCLUSION

Applying the SURF algorithm for detecting and recognizing the books in a library proved to be a good decision, due to following advantages: the algorithm is invariant to lighting changes and observation point changes, the computing speed of the SURF algorithm is 3-4 times higher than that of SIFT, it can be adapted for parallel computing because every Hessian matrix can be estimated in parallel, and it provides good accuracy. This study illustrates the multiple possibilities to use vision in autonomous robots and service robots, respectively. By using only one webcam the path towards an object was mapped and localized efficiently. Further research will greatly enhance the possibilities of this type of sensors.

REFERENCES

- [1] B. K. Kim, T. Sugawara, K. Ohara, K. Kitagaki, and K. Ohba, "Design and control of the librarian robot system in the ubiquitous robot technology space," in Proc. *RO-MAN* 2008 – *The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 616-621, 2008.
- [2] A. Morales, M. Prats, P. J. Sanz, and A. P. Pobil, "An experiment in the use of manipulation primitives and tactile perception for reactive grasping," in *Proc. Robotics: Science and Systems - Manipulation Workshop - Sensing and Adapting to the Real World*, Atlanta, USA, 2007.
- [3] T. Bräunl, Embedded Robotics: Mobile Robot Design and Applications with Embedded Systems, Springer Science & Business Media, 2008.
- [4] G. Dudek and M. Jenkin, Computational Principles of Mobile Robotics, Cambridge University Press, 2000.
- [5] J. Hahn, M. Twidale, A. Gutierrez, and R Farivar, "Methods for applied mobile digital library research: A framework for extensible way-finding systems," *The Reference Librarian*, vol. 52, no. 1 -2, pp. 106-116, 2010.
- [6] D. Meere, I. Ganchev, M. O'Droma, M. O'hAodha, and S. Stojanov, "Evolution of modern library services: The progression into the mobile domain," *M-libraries 2: A virtual library in everyone's pocket*, London. England: Facet, pp. 6l-72, 2010.
- [7] C. W. Chang, J. H. Lee, P. Y. Chao, C. Y. Wang, and G. D. Chen, "Exploring the possibilities of using humanoid robots as instructional tools for teaching a second language in primary school," *Educational Technology and Society*, vol. 13, no.2, pp. 13-24, 2010.
- [8] L. C. Lin, "An integrated framework for the development of radio frequency identification technology in the logistics and supply chain anagement," *Computers & Industrial Engineering*, vol. 57, no. 3, pp. 832–842, 2009.

- [9] S. Woods, K. Dautenhahn, and J. Schulz, "The design space of robots: Investigating children's views," in *Proc. Robot and Human Interactive Communication*, pp. 47-52, 2004.
- [10] S. Kiesler and J. Goetz. "Mental Models and Cooperation with Robotic Assistants," in *Proc. Conference on Human Factors in Computing Systems* (CHI'02), ACM Press, pp. 578-579, New York, 2002.
- [11] M. Mikawa, Y. Morimoto, and K. Tanaka, "Guidance method using laser pointer and gestures for librarian robot," *RO-MAN*, 2010 IEEE, pp. 373-378, 2010.
- [12] A. P. Saygin, T. Chaminade, H. Ishiguro, J. Driver, and C. Frith, "The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid actions," *Social Cognitive and Affective Neuroscience*, vol. 7, no. 4, pp. 413-422, 2012.
- [13] T. Anderson, A. Druin, K. Fleischmann, E. Meyers. L. Nathan, and K. Unsworth, "Children, technology and social values: Enabling Children's voices in a pluralistic world," in *Proc. American Society Information Science and Technology*, vol. 46, no. 1, pp. 1-9, 2009.
- [14] E. T. Dresang, "Access: The information-seeking behavior of youth in the digital environment," *Library Trends*, vol. 54, no.2, pp. 178-196, 2006.
- [15] A. Druin, "The role of children in the design of new technology," *Behaviour and Information Technology*, vol. 21, no. 1, pp.1-25, 2002.
- [16] I. L. Guha, A. Druin, and J. A. Fails, "How children can design the future," *Human-Computer Interaction, Users and Applications*, Springer Berlin Heidelberg, pp. 559-569, 2011.
- [17] E. Mazzone, J. C. Read, and R. Beale, "Understanding children's contribution during informant design," in *Proc. the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 2*, British Computer Society, pp. 61-64, 2008.
- [18] J. Kodagali and S. Balaji, "Computer vision and image analysis based techniques for automatic characterization of fruits – A review," *International Journal of Computer Applications*, vol. 50, no. 6, pp. 6-12, 2012.
- [19] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," ACM Computing Surveys, vol.38, no. 4, pp.1–45, December 2006
- [20] K. Werner and M. Kampel, "Interest Point based Tracking," Proceedings of the 20th International Conference on Pattern Recognition (ICPR'10), pp. 3549–3552, Istanbul, Turkey, IEEE, August 2010.
- [21] M. Du, J. Wang, J. Li, H. Cao, G. Cui, J. Lv, and X. Chen, "Robot robust object recognition based on fast SURF feature matching," in *Proc. 2013 Chinese Automation Congress* (CAC'13), pp. 581–586. Changsha, China, IEEE, November 2013.
- [22] Y. H. Lee, H. Ahn, H. J. Cho, and J. H. Lee, "Object recognition and tracking based on object feature extracting," *Journal of Internet Services and Information Security*, vol. 5, no.3, pp. 48-57, 2015.
- [23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints, cascade filtering approach," *International Journal of Computer Vision* vol. 60, no.2, pp. 91–110, 2004.
- [24] Y. Kee and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Conference* on Computer Vision and Pattern Recognition, pp. 506–513, 2004
- [25] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005
- [26] M. Hassaballah, A. A. Abdelmgeid, and H. A. Alshazly, "Image features detection, description and matching," *Image Feature Detectors and Descriptors*, Springer International Publishing, pp. 11-45, 2016.
- [27] F. Guo, X. Luo, and Y. Liu, "Research on feature extraction and match method based on the surf algorithm for mobile augmented reality system," *International Industrial Informatics and Computer Engineering Conference* (IIICEC), pp. 615-619, 2015.
- [28] B. B. Swapnali, S. Kayastha Vijay, and K. H. Varsha, "Feature extraction using surf algorithm for object recognition," *International Journal of Technical Research and Applications*, vol.2, no.4, pp. 197-199, 2014.



Corina Monica Pop is a graduate of *Transilvania* University of Brasov, Faculty of Technological Engineering, Department of Industrial Robots. She works at *Transilvania* University Library of Brasov, the IT department. In 2014 she became PhD student at the Faculty of Mechanical Engineering, in the field of industrial engineering.

She has a professional experience as a member in 6 research projects and 2 Leonardo da Vinci Mobility

Projects.



Gheorghe-Leonte Mogan received his bachelor degree in industrial technology from *Transilvania* University of Brasov. In 1995 he obtained the PhD degree from *Transilvania* University of Brasov, Faculty of Mechanical Engineering. He is a professor and PhD supervisor at *Transilvania* University of Brasov, Department of Automotive and Transportation.

He has written more than 50 research papers in the fields of mechatronics, robotics, CAD / CAE, optimization of mechanical structures, finite element analysis. He has an extensive professional experience as a coordinator of 10 academic research projects, 2 of which European (FP5 and FP6).



Răzvan Gabriel Boboc was born in Brasov, in 1986. He received his bachelor degree in electrical engineering, telecommunications specialization from *Transilvania* University of Brasov, Romania, in 2009. In 2011 he graduated the master program with specialization in communication Networks. In 2015 he obtained the PhD degree from *Transilvania* University of Brasov, Faculty of Mechanical Engineering with the thesis "Natural human-robot

interaction for assistive robotics applications". He is currently scientific researcher at *Transilvania* University of Brasov, IIR (*Industrial Informatics and Robotics*) laboratory, Department of Automotive and Transportation Engineering, where he conducts education and research activities in the fields of robotics, human-robot interaction, virtual reality, augmented reality, 3D CAD modeling, computer aided design.