Developing a robot with computer vision for automating business processes of the industrial complex

N V Gavrilovskaya¹, V P Kuvaldin², I S Zlobina³, D E Lomakin⁴,⁵ and E E Suchkova⁵

¹ Department of Information Technology in the Russian State Agrarian University—Moscow Timiryazev Agricultural Academy, Moscow, Russia
² Department of Criminal Law, Criminal Procedure Law and National Security of Vyatka State University, Kirov, Russia
³ Department of Foreign Languages, Vyatka State University, Kirov, Russia
⁴ Department of Algebra and Mathematical Methods in Economy of Orel State University named after I.S. Turgenev, Orel, Russia
⁵ Orel branch of Emperor Alexander I St. Peters burg State Transport University, Orel, Russia

E-mail: zlobina2021@yandex.ru

Abstract. Robotic Process Automation technology is one of the solutions implemented in mechanical engineering and industrial production. A business process is automated by observing how a person is doing a task in a graphical user interface. Then the cyberphysical device repeats the same task. A significant problem of the implementation of such automated systems is that if the work is to be carried out through the interface of a virtual desktop, then it is impossible to obtain data on the elements of the graphical interface without computer vision. The object of the research is the automation of business processes of the industrial complex using computer vision. The purpose of the study is to apply computer vision methods to automate business processes and develop an appropriate software robot. The paper analyzes computer vision methods for recognizing user interface elements and text within these elements. Moreover, the authors have developed and tested a cross-platform application that reproduces a business process using computer vision. The test results have shown that selected computer vision methods will help automate industrial production business processes on any platform with a virtual desktop interface.

1. Introduction

Nowadays manual business processes are not efficient in terms of time and cost. There have been many attempts to automate them. The vast majority of work that a person does can and should be done by computer. RPA (Robotic Process Automation) technology is one of the automation solutions which is actively used in the industry. The manual business processes are automated by observing how a person is doing a task in a graphical user interface (GUI). Further, the software robot (Bot) repeats the same.

One of the main problems of RPA is that if the work is to be done through a virtual desktop interface, then it is impossible to obtain data about the elements of the graphical interface without computer vision [1]. Computer vision is an aspect of artificial intelligence that allows all RPA robots to see every element on a computer screen [1]. Thanks to this knowledge area in RPA, users are able to automate tasks in...
most of environment with virtual desktop interface (VDI) environments regardless of platform or operating system [2]. Robotic process automation has many applications, including healthcare and pharmaceuticals, financial services, outsourcing, retail, telecommunications, energy, utilities, etc. [3]. Applications of RPA are in demand all over the world. However, software for RPA is often limited to only one operating system and without computer vision [4].

The purpose of the study is to apply computer vision methods to automate the business processes of the industrial complex and to develop an appropriate software robot. The research has the following tasks: to review and analyze scientific and technical information; choose algorithms for searching for user's interface elements on different operating systems and searching for figures in images; design a robot with computer vision to automate business processes; choose application's architecture; create UML-models of the area; select software and development tools; implement algorithms; test programs.

Having analyzed the work of the Doronichi industrial complex, the authors have identified the following business processes with the same type of operations: filling out forms and documents, comparing tables, purchasing goods, organizing delivery, etc. All these processes require a lot of time and human resources. RPA technology can automate these processes. Having studied the literature, the authors have found significant shortcomings of analogues of RPA: inability to work through VDI and on MacOS; need for basic knowledge of programming.

The proposed solution is to get rid of these shortcomings and automate routine human activities in order to make the business process of the industrial complex be performed correctly and efficiently in time. The user needs to record this process. The bot collects all the necessary information about the elements of the graphical interface: data on the pressed keys of the keyboard and mouse, coordinates of clicks, etc. This is done when the user is doing it with the recorder running. Later the bot should be able to perform this process on its own.

2. Materials and methods

RPA is one of the business process automation technologies based on a software robot. The main feature of RPA is the use of graphical user's interfaces to manage and collect information. Control is a user interface element. A program robot (bot) is a type of software that performs tasks assigned to it using a GUI. Borders (edges) are curves in the image along which there is a sharp change in brightness or other types of irregularities, CED – Canny Edge Detection [2]. The edge detector of Canny is a borders (edges) detection operator that uses a multi-stage algorithm to detect a wide range of edges in images [5].

Convolution is an operation that shows the "similarity" of one function with a reflected and shifted copy of another. When working with images, convolution is the operation of calculating a new value for a given pixel, which takes into account the values of the surrounding neighboring pixels. The main element of the convolution is the convolution kernel - a matrix of arbitrary size and aspect ratio; the most commonly used is a square matrix (3x3 by default). One of the most important convolutions is the computation of derivatives.

Images are made up of pixels that give the picture a grayscale brightness value. In other words, the image is a two-dimensional matrix of numbers.

The operator Sobel is a discrete differential operator that computes an approximation of the brightness gradient of an image. This operator is used in image processing and computer vision, especially in algorithms of edge detection [6]. Gradient is a vector showing the direction of the steepest increase of a certain parametric, the value of which changes from one point in space to another. For example, CED uses the gradient of image for detection of edges [7].

OCR is Optical Character Recognition, VDI – Virtual Desktop Interface, RNN – a recurrent neural network, LSTM – Long Short-Term Memory. The most popular RPA apps available on the market are Blue Prism, UiPath, and Automation Anywhere. Both Blue Prism and Automation Anywhere can automate everything, directly in the environment where they are installed. But they do not have the ability to automate anything in a virtual machine, for example, VirtualBox, VMware, Hyper-V [8]. At the moment, there is no free version of Blue Prism. Users working with Blue Prism need programming skills. This tool is highly specialized [9]. UiPath is the next leading provider of robotic automation tools.
of industrial business processes. UiPath applies its RPA-based tools in three different versions for its consumers [10]. However, UiPath is missing on the Linux platform. Users without basic knowledge of programming cannot use this tool. Moreover, there is no support for two or more screens.

3. Results

There are two ways to search for user interface elements. The first way is to search for an item in the control tree using the API of operating systems Windows and Linux. The API of these operating systems is equipped with search tools for the user's interface element with which the interaction occurs. They allow to receive information about the location of the control on the screen and all its characteristics. However, one of the target platforms of the solution of RPA is MacOS. This operating system does not allow to obtain the necessary control information in connection with the security policy. Another problem of obtaining information about user's interface element through the API of operating systems is working through the virtual desktop interface. At the moment, it is not possible to get information about functional element via VDI.

The second way to find an item is to use computer vision. Prospective areas of computer vision applications that will be used in RPA are: image preprocessing (noise removal, contrast enhancement), image detail extraction (lines, borders, edges, and contours), text recognition in an image, detection of the occurrence in an image. In this way, it is supposed to allow the solution of RPA to work on MacOS and through VDI.

One method of removing noise from an image is a Gauss filter. It is often used fully or as part of other image processing algorithms. By highlighting the borders in the image, you can get a lot of information about the elements of the user interface in the computer's field. The information obtained after the selection of boundaries can indicate the location of the control and its area. Moreover, most methods for extracting borders in an image use image preprocessing algorithms. They allow to embed preprocessing in the border extracting method. One of the most efficient algorithms for finding boundaries in an image is Kenny's algorithm. Parameters of algorithm allows to recognize edges with different characteristics depending on specific implementation requirements. Text recognition (OCR) in the solution of RPA allows to search for the desired user's interface element not only by the image, but also by the name. Text recognition can be used to find text for interaction, for example, filling out forms.

Detecting the image entry is capable of detecting a user interface element when the process is played. One of the easiest ways to find the input of an image is through pattern matching. RPA suggests to use method of comparison of the pixels in the template and in the source image. At the time of search, the desktop is the source image, and a specific user's interface element is the template. The control is in the place where there is less difference between the template and the place of the original image. The above methods provide the solution of RPA with computer vision while recording and replaying the process. In addition, they allow RPA to work not only on operating systems of Windows and Linux, but also on MacOS and all platforms with VDI.

The application should provide the following features: the ability to record the process; the ability to reproduce the process with the specified parameters (for example, when buying goods on the site the parameters are: city of departure, route, employee name, passport data, etc.); cross-platform availability (Windows, Linux, MacOS); availability of computer vision to work with VDI. The requirements for work are: .Net Runtime (for servers: .Net Hosting Bundle); at least - Windows 7, Ubuntu 19.0 and MacOS Siera.

Development on .Net Core has the main advantage: the presence of cross-platform. This unique software platform already runs on Windows, Mac OS X and Linux. An application developed on the basis of .NET Core can run unchanged on all operating systems that have the specified platform. This allows .Net Core to promote its applications as software-defined rather than platform-defined. Visual Studio Community 2019 was chosen as the development environment. This IDE is free and supports development on the .Net Core platform. VS allows to create modular applications with different architectures. Windows Presentation Foundation (WPF) is similar to WinForms, but does not support
cross-platform support. The EtoForms library is used as a replacement to create the GUI, it is free for commercial use.

WinApi is used to search for user's interface elements on the Windows. All user's interface elements in Windows are represented as a tree, where the desktop is the root of the tree and the application windows are its branches. Each of these can contain elements such as menus, buttons, toolbars, and lists. With the help of this library, it is possible to get the location of the control, its Id and all the necessary information for the application of RPA to work.

OpenCvSharp is a cross-platform which has OpenCV wrapper for the .NET Framework. The Open Source Computer Vision Library is an open source software library for computer vision and machine learning. OpenCV was created to provide a common infrastructure of application. OpenCV was formed to supply computer vision and to accelerate the use of machine perception. OpenCV is free for commercial use.

Tesseract 4 is an open source text recognition (OCR) engine available under the Apache 2.0 license. It can be used with an API to extract typed text from images. Tesseract 4 is used by Google in Google Translate. Tesseract is free for commercial use. To create executable files, the Inno Setup utility was used. The Orca utility was used to create MSI distributors. Using these tools, you can create executable files for each available OS and place them in the application stores of AppStore and WindowsStore.

The solution is built on a modular architecture. To record the processes that need to be automated, the Recorder module was created. The Bot module was created to reproduce the processes. The NativeHelper and VisionParser modules retrieve information during recording and search for it during playback. In order to see what RPA recognizes during recording, the Marker module was created.

LSTM in Tesseract 4 supports Russian and English languages that are used in RPA. Here there is another problem: the required language is not available, or the text is written in a font that was not taught. In this case, it is possible for the neural network to learn this language or font. For this, you can apply "fine tuning". It is necessary to train the application on specific additional data. This might work with problems that are close to the existing training data, but differ in some subtle way, such as an unusual font. You can even work with a small amount of training data. First, you need to prepare the data that need to be transferred to Tesseract. You need one or more files that contain at least one (but preferably more) occurrences of each character in the font for training. Input files must be named in accordance with the agreement of the Tesseract: [language name].[font name].exp[number].[file extension].

For example, if there are 3 .png files with an English text in Arial font. Their names will be: eng.arial.exp0.png; eng.arial.exp1.png; eng.arial.exp2.png. Once the data is collected in one place and named correctly, you need to generate box files for them. These files tell Tesseract where each character is located. This is done through the command line. After installing the Tesseract.makebox – tells Tesseract to (only) create box files. Batch.nochop command – tells Tesseract not to use its algorithms for image segmentation. Next, you need to open each image file with qt-box-editor and fix Tesseract if it has any mistakes (if there are no errors, training is probably not required). The last thing to do is generate a*.traineddata file that can then be loaded into Tesseract. So, the app can recognize the characters that it was just taught. For automation, you first need to record the process. In the future, the robot will know what elements of the graphical interface it will interact with. To do this, you need to open the recorder. The recorder uses the same computer vision modules or native parser. You can change the selection by clicking on the grid. When recording a process, the user needs to see which elements the RPA can successfully track. This is done by highlighting with a marking module. If a native parser is used when writing, then operating system gives the area in which the control is located. In another case, when using a module with computer vision, you have to use Kenny's algorithm. If there is an interaction with an element of the user's interface, then the picture of the area in which the element is located is saved in such a way that during playback it will be possible to find this control using template matching. The user recording the process can observe which controls were interacted with, so that in case of an error, there would be a possibility to delete.
In order to get the text from the elements of the user interface by the method of Kenny, the boundaries of the control or the required text are found. After starting the bot, you need to select the process that the bot will execute. On the server, after authorization, the robot has an access to all processes loaded there as well as to all processes that are stored locally.

4. Conclusion
With the development of computer technology, more activities that were previously performed by humans can now be performed by various types of robots. This leads to the fact that workers can be free from performing monotonous tasks. In the industrial complex, there are many routine tasks (collecting information; weighing grain; receiving goods; paperwork; accounting for agricultural products; sorting; preparing a product for sale, etc.). RPA is a technology that allows to shift routine work from a person to a robot. Such routine operations are currently performed by employees of the industrial complex using a keyboard, screen and mouse.

A robot doing routine work instead of a human is a program that requires a keyboard, screen, mouse, and some kind of “vision”. It uses them virtually. The robot copes with the routine faster than a person, without mistakes associated with the "human factor", no vacations, days off and stops, with a predictable and easily controlled result. If there is computer vision in the implementation, it becomes possible to automate business processes on all platforms with the presence of VDI and on MacOS.

Thus, the main methods and algorithms that provide RPA of robots with computer vision were considered. Finding boundaries during process playback is useful in order to reduce the time for finding GUI elements. The image is presented as a set of related boundaries, which in turn makes it easier to obtain information and analyze the image. Kenny's boundary-finding method is considered one of the classic algorithms in computer vision and is one of the most effective. At the moment, one of the best tools for extracting text from an image is Tesseract OCR, which uses a recurrent neural network based on the LSTM architecture. This technology allows the robot to receive information from the elements of the user interface and use it in the future. Pattern matching allows the robot to find the right control in any operating system and when working through VDI. The requirements for the application were created.

The platform of .NetCore and object-oriented programming language C# were chosen for the implementation. The main advantage of this platform is the presence of cross-platform. Visual Studio Community 2019 was chosen as the IDE, since this development environment is free and has a wide range of development and debugging tools.

The software implementation consists of several modules. They have functions for working with computer vision, native search, recording and playback of processes. Kenny's algorithms and pattern matching were used to implement computer vision. Tesseract OCR version 4 was selected for text recognition in user interface elements. Version has an already trained neural network based on the LSTM architecture. OCR is free for commercial use.

Thus, the application was implemented using technology of RPA. Computer vision enabled automation of business processes on Windows, MacOS, Linux, and all platforms that support VDI. This implementation of RPA allows users of the industrial complex to automate business processes independently without knowledge of programming.

References
[1] Madakam S, Holmukhe R and Jaiswal D 2019 The Future Digital Work Force: Robotic Process Automation (RPA) J. of Inform. Syst. and Techn. Management. 16 1-18
[2] Canny J 1986 A Computational Approach To Edge Detection IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8. 6 679-98
[3] Peng X and Wang C 2020 Building Super-Resolution Image Generator for OCR Accuracy Improvement DAS 2020. Lecture Notes in Computer Science. 12116 145-160
[4] Zhou P, Ye W and Wang Q 2011 Canny Algorithm for Edge Detection J. of Computational Inform. Syst. 7(5) 1516-23
[5] Jozefowicz R, Zaremba W and Sutskever I 2015 An empirical exploration of recurrent network architectures *In International conference on machine learning*. 37 2342-50

[6] Graf M and Kvasnikov V 2018 The Construction of the Algorithm Study Based on the Mathematical Model of Motion *In ICTERI* 235-42

[7] Skvortsov E. A, Nabokov V I, Nekrasov K V, Skvortsova E G and Krotov M I 2019 Application of artificial intelligence technologies in agriculture *Agrarian Bulletin of the Urals*. 8 187

[8] Lyasnikov N V 2018 Digital agricultural sector of Russia: review of breakthrough technologies of the fourth technological order *J. Food Policy and Security*. 5(4) 169-82

[9] Wendt III R 2014 Principal Component Analysis of EBT2 Radiochromic Film for Multichannel Film Dosimetry *Intern. J. of Medical Physics, Clinical Engineering and Radiation Oncology*. 3 156-66

[10] Dong C, Loy C C, He K and Tang X 2016 Image super-resolution using deep convolutional networks *IEEE Trans. Pattern Anal. Mach. Intell*. 38(2) 295-307