Інформаційні технології та системи управління

UDC 004.89

N. Axak, O. Barkovskaya, H. Ivashchenko

Kharkiv National University of Radio Electronics, Ukraine

DEVELOPMENT OF THE HAND GESTURE RECOGNITION SYSTEM ON THE BASIS OF CLONAL SELECTION MODEL

Image processing method which allows to satisfy the requirements of the users with disabilities in a real-time mode was developed to minimize a delay between the serve of command and reaction of the system on it. The hand's gestures recognition system for automatic recognition of the given commands by the system was proposed. The principle of images comparison using clonal selection model was put in the basis for the proposed system.

Keywords: *gestures recognition, alternative means of communication, artificial immune system, CLONALG, CUDA.*

Introduction

Gestures recognition is the method of information input in a computer. Gestures recognition systems (GRS) provide an interface for the input of control commands in a computer. Also GRS allows reducing of the distance between human and a computer unlike peripherals input devices, such as a mouse or keyboard. Such systems can be a communication alternative for users with disabilities. Gesture recognition problem is a particular case of object recognition problem. Objects are chosen from a sequence of video frames. Gesture recognition is done in three main phases: detection, tracking and recognition. GRS which analyze the trajectory of different manipulators' motion (flashlight-pointer that radiates bright light) in the basis of the system is the most popular. Doing such flashlight-pointer gestures, one can be sure in high-precision recognition of commands.

Also one can use special markers whereby it is possible to distinguish object of interest in focus of digital camera or special bright color gloves which are unique for area of interest.

However, in spite of relative ease of implementation and low cost, their performance depends on many factors, the violation of which would lead to poor quality or failure of performance.

Various sensors and cameras can be used as additional devices for building a map of depth of objects' location in observed area; using of 3D-gloves or 3D-accelerometer allows to track the position of the manipulator in three-dimensional space, thereby ensuring a high recognition accuracy.

Literature review and problem statement

Human gestures are perceived through vision, it is a subject of great interest for computer vision researchers. A large number of works are devoted to pattern recognition problem nowadays.

For example, in [1] for pattern recognition problem the histogram analysis method is used, which consists in constructing two histograms with its subsequent analysis. The background image is separated from the foreground image. Image is converted to halftone images in shades of gray. Then two histograms for selected image of hands on the X and Y axes are constructed. The rule of histogram's constructing is following: the image is divided into N parts horizontally or vertically, for every part the average value of grayscale is calculated. Then a histogram analysis is carried out after the normalization of all image parts. This method depends on lighting and filtering techniques. Thus, there is a need to adapt to environmental changes. In some cases, it's quite difficult to trace the logic of the histogram changes when changing examined parts of the image (hand position relative to the camera). Therefore, the histogram analysis is the method only applicable to static gestures and it is a very time-consuming task.

Another study [2] presented gestural recognition model for management of video-player usage. A manipulator positioned as a pencil or handle of bright color is used for this purpose. This method perfectly fits the selection of objects on the static homogeneous background. In case of dynamically changing background there can be a coincidence of color matching of manipulator for management and object in videocamera.

Authors of work [3] were using histograms' analysis methodology. The main feature is that the offered model supposes the use of method of transformation of web cam or mobile phone camera in the sensor of depth that is analogical on functionality to Kinect sensor. A hand is distinguished by setting a points' threshold range of the object, which histogram is built on the next

steps. Disadvantage of this method is the need of purchasing additional devices.

In [4] the manual alphabet gesture recognition method ASL (American Sign Language) was presented. Gesture recognition is performed by extracting, preliminary image processing, and subsequent comparison of the removed normalized geometric skeletons of arms generated by the sensor. Comparison of the skeletons is based on Dynamic Time Warping algorithm (DTW), having polynomial complexity. Just as in [3; 5], it is supposed that additional devices are used like the Kinect for receiving map of depth of camera space.

The author of [5] focuses on building a robust gesture recognition system using the stereo camera systems with depth of the image detection, because usual camera is not able to provide all the necessary information for all objects in the field of view. Additionally, this method needs controlled lighting.

Contour image analysis technique is used in [6]. The gesture, which was referred to the system, is determined by the collected time statistics of palm contours movement. However, further image and contours analysis in a real-time mode is a time-consuming task.

In [7] a palm image analysis obtained by the WEB-cameras is performed. The process of recognition has such steps like palm image binarization to select its silhouette, as well as the construction of the skeleton. Skeleton representation simplifies the analysis of the shape of the palm, and can effectively determine the number of visible fingers, fingertips and coordinates of the center of the palm. Approaches based on the artificial intelligence methods, such as artificial neural networks and artificial immune systems (AIS) are currently actively developing [8]. There are various models based on the principles of immune system: clonal selection model, immune network model and others that can be used to solve the problems of pattern recognition.

Based on this analysis, one can conclude that at present time there is no system with opportunity to recognize hand gestures, excluding lighting quality and color of body, without additional devices to focus the camera at high speed.

The aim and tasks of the study

The aim is to develop a hand's gesture recognition system with receiving of binary images with recognized result from the WEB-cameras in a short time without using of additional manipulators. To achieve the aim it is necessary to solve the following tasks:

– develop the method of processing of the obtained data from the camera video stream;

– develop methods for real time image processing, which will minimize the delay between commands and system response to the given command;

– develop a model of gesture recognition based on artificial immune system for the automatic recognition of gestural command.

Morphological operations for images processing

To select the background we use the algorithm MOG2 Background Subtractor [9; 10] based on Gaussian Mixture Model. The proposed algorithm identifies specific colors and foreground objects or the part of background to be deleted.

Calculating of the mixture weights for the image (fig. 1) you can select colorful components which belong to image background.

The obtained image is a binary pixel mask (fig. 2), which describes objects in the image (walking down the street people or moving cars).

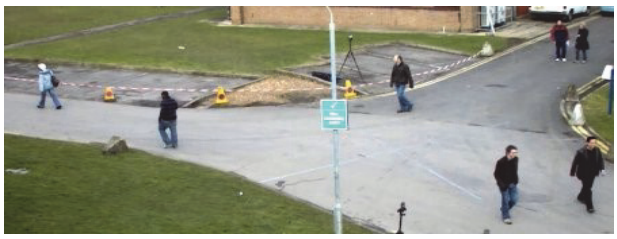


Fig. 1. Example of input image

Also, this algorithm provides information about the objects' shadow. This information is useful for modeling three-dimensional model of the scene, or to build the path of motion of an object. Gray dots represent the shadow mask objects.

Fig. 2. Foreground mask obtained
using MOG2 Background Subtractor

Denoising and images filtration is the next step of graphical image processing. Noise pixels are pixels that do not carry useful information about the object and make recognition process more difficult.

Morphological narrowing operation (erosion) and morphological extension (dilatation) are used for noise reduction in the computer vision tasks for clearing image from individual pixels, as well as from small groups of pixels. An image without noise is the result of these operations.

The loss of information about the object (shape of the object or its size) can be the disadvantage of these methods. Morphological extension is used to recover information about shape and size of the object. As a result, the objects of interests are clearly visible and can be used in further calculations.

Dilatation is the convolution of the image (fig. 3, a) or of the selected area of the image with some core

(fig. 3, b). The core may be of any shape and size. In this case, only the leading pixel is distinguished in the core for superposing with the current pixel while calculation of convolution. In many cases, a square or a circle with a leading pixel in the center is chosen as the core.

The core can be regarded as a template or mask. In general, dilatation is a sliding of the template across the image and applying of local maximum search operator for the intensity of the pixels in image.

This operation causes the growth of light areas on the image. Pixels, which are a result of the dilatation and became white, in the fig. 3, c are marked grey.

Bounding rectangle is constructed taking into account the most extreme white image pixels on each side of the image. It helps to determine the size and location of the desired object on the image.

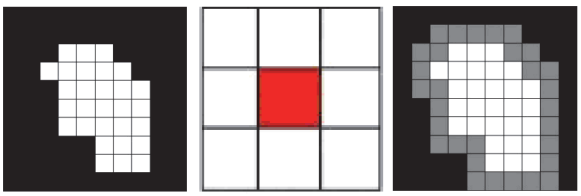


Fig. 3. Principles of the morphological extension: a – input image; b – core-mask (the leading element is in the center); c – processed image

Image recognition model based on clonal selection model

Clonal selection algorithm searches the most similar sample for unknown image, according to the affinity between the unknown one (antigens) and the most similar sample (antibodies). The hand gesture recognition model which based on clonal selection is presented in fig. 4.

Using the Hamming distance between the selected objects from the bounding rectangle and a set of templates from table 1 we will recognize the main gestures of the hand. It uses the limited number of possible control commands. Every command can be regarded as two-dimensional image. All gesture commands correspond to control command of the system (for example, volume adjustment, or file changing).

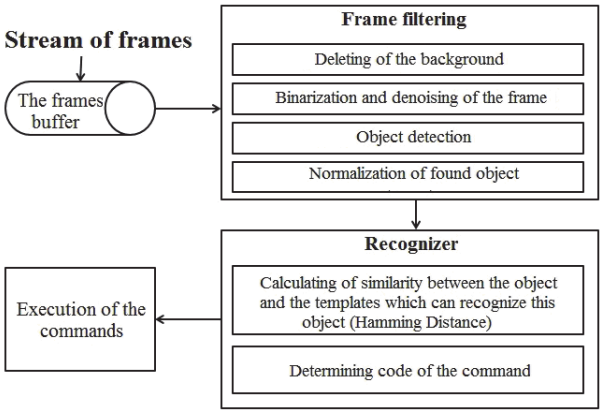
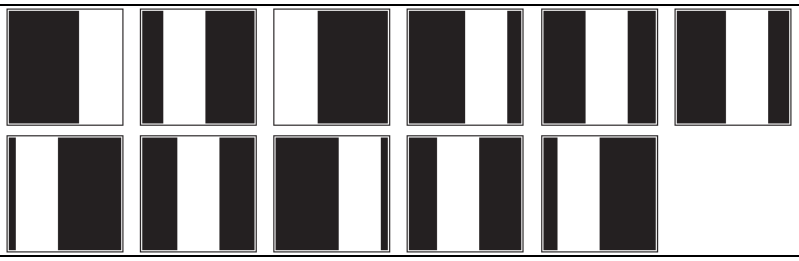
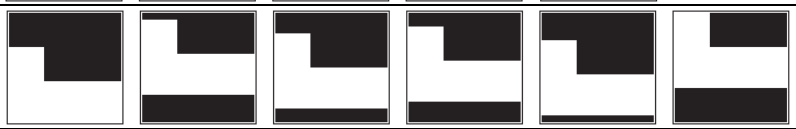
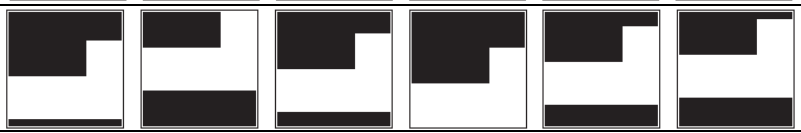


Fig. 4. The hand gesture recognition model

Table 1

Control commands and related set of test patterns

Control command	Code of the command	Set of test patterns
Stop/start	VERTICAL	
Forward	HORIZON-TAL_ NEXT	
Back	HORIZON-TAL_ PREV	
Enhance	DIAGONAL_R	
Reduce	DIAGONAL_L	

Hamming distance is widely used in the theory of artificial immune systems [10] to determine the affinity (measure of difference between the antigens and antibodies). For hand gesture recognition used clonal selection algorithm CLONALG. The general CLONALG model involves the selection of antibodies (candidate solutions) based on affinity either by matching against an antigen pattern. The following steps of the proposed method are separation of the object and its scaling.

For objects separation we should perform the following steps:

– analyze the aspect ratio of the bounding rectangle (size comparison of the rectangle sides to correct it and make it square);

– resize the bounding rectangle;

– copy the image information from the bounding rectangle.

The width of the rectangle should be extended according to formula 1 if the width of the bounding rectangle is less than the height:

$$p = \frac{h - w}{2}, \qquad (1)$$

where h – height of the rectangle;

w – width of the rectangle.

The value is required to calculate the affinity and indicates change of the rectangle size, while converting it into a square. The selection of the object should not be broken. Calculation of similarity values of two binary images occurs by the formula:

$$\mathrm{Aff}(Ag_i - Ab_i) = \sum_{n=1}^{N} \delta_n, \qquad (2)$$

where Ag – antigen, object for recognition;

Ab – antibody, template from the templates' population;

N – number of all pixels of the image;

$$\delta_n = \begin{cases} 1, & Ab_{in} = Ag_{in}, \\ 0. \end{cases} \qquad (3)$$

Each point of the selected object is cyclically compared with the point of one of the templates in the same coordinates. If the values of compared points are equal – a variable that stores the coefficient of similarity of images is incremented. This operation is repeated for each template image of controls command.

After comparison of the selected object with all templates, the coefficient of similarity of images transforms to the percentage of the number of equal points to the entire size of the image. Control command image represents by a set of binary pixels, the number of which depends on this command. The highest value of the coefficient of similarity of images determines the most appropriate template.

Software implementation of the hand gesture recognition model

Calculation of similarity of images is time-consuming operation. All calculations are performed by the CPU. Template images of control commands are selected in such a way that the calculation made in a short period of time. Increasing the size of template images can increase the accuracy of recognition. But the increasing of recognition accuracy decreases performance because of speed of calculations, which depends on the images size and on their quantity.

Using of graphics processors and the technologies of general-purpose computing for graphics processing units (GPGPU), such as CUDA on Nvidia, or CTM from AMD will increase the speed of the system and can be the solution of this problem, because the main requirement for GPU programming (data parallelism with a large number of the same computing operations of small computational complexity) is satisfied.

For modeling of the hand gesture trajectory recognition system used the computer Intel Core 2 Quad CPU Q8200 @ 2.33GHz with the NVidia graphics card GeForce GTX 460. Parallelization performed using C++ language and CUDA technology in Microsoft Windows 10. The GPU memory patterns (i.e. templates) store in in the graphics processing unit memory in the form of a one-dimensional array, describing two-dimensional matrixes of the templates. OpenCV library implements separation of the object and its scaling.

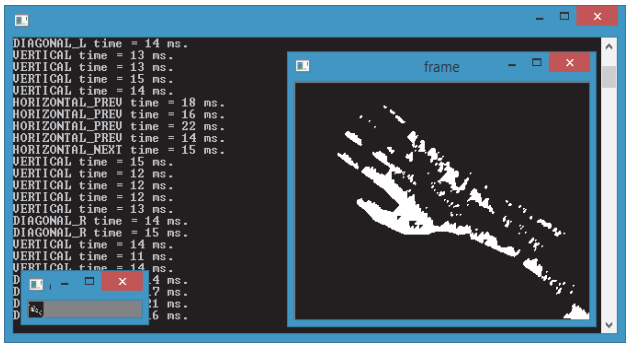


Fig. 5. The result of the program

Table 2

Recognition accuracy of input commands and time of execution of these commands

Code of the command	N	Recognition accuracy, %	The average time of recognition, ms.
VERTICAL	100	100	15,9
HORIZ_NEXT	100	10	16,9
HORIZ_PREV	100	17	17,0
DIAGONAL_L	100	89	17,7
DIAGONAL_R	100	88	17,2

Fig. 5 shows the result of hand gestural commands recognition algorithm. In the table 2 it can be seen recognition commands for HORIZ_NEXT and HORIZ_PREV codes showed the worst results of accuracy of recognition. This result is that the templates of these commands are almost identical, the difference between them is only in a small region of the template points.

Conclusion

The developed model of videostream data processing was proposed in the article. The difference between the existing models and described model is that it does not need additional manipulators for creating some specific trajectory for movement recognition except a camera which is located close to the user. It allows to make the system cheaper and increase it competitive ability.

The method for real-time image recognition which bases on the principles of images comparing using artificial immune system allows to minimize time delay between getting of videostream and making the decision about the gesture of the patient was proposed.

Automatic gesture recognition system which can be used like alternative source of input information for computer (for example, for users with disabilities) was developed and tested.

References

1. Kataev M.Y. The methodology of determining the hand gestures the observed with the camcorder / M.Y. Kataev, L.V. Shirokov // Proceedings of TUSUR University. – 2013. №27. – P. 45-49.

2. Fang Y.K. A Real-Time Hand Gesture Recognition Method / Y.K. Fang, K.Q. Wang, J. Cheng, H.Q. Lu // IEEE International Conference on Multimedia and Expo. – Beijing, 2007. – P. 995-998.

3. Ren J. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera / Z. Ren, J. Yuan, and Z. Zhang // ACM Int. Conf. on Multimedia. – 2011. – P. 1093-1096.

4. Nagapetyan V.E. ASL Fingerspelling Recognition / V.E. Nagapetyan // RUDN Journal of Mathematics, Information Sciences and Physics. – 2013. – . № 2. – P. 105-113.

5. Yong W. Using human body gestures as inputs for gaming via depth analysis / W. Yong, Y. Tianli, L. Shi, L. Zhu // IEEE Multimedia and Expo Int. Conf. – 2008. – P 993-996.

6. Domunian AA. Software tools for gesture recognition / A.A. Domunian // Proceedings of MIPT. – 2014. – № 4. – P. 107-110.

7. Mestetskiy LM. Real-time hand gesture recognition by planar and spatial skeletal models / L.M. Mestetskiy, A.V. Kurakin // Informatics and its Applications. – 2012. – № 1. – P. 114-121.

8. Castro L.N.D. Learning and optimization using the clonal selection principle / L.N.De Castro // Evolutionary computation. – 2002. – № 3. – P. 239-251.

9. Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction / Z. Zivkovic // Proc. 17th Robotics and Automation Int. Conf. – Cambridge, England, 2004. – P. 28-31.

10. Zivkovic Z. Efficient adaptive density estimation per image pixel for the task of background subtraction / Z. Zivkovic, F. van der Heijden // Pattern Recognition Letters. – 2006. – № 7. – P. 773-780.

Надійшла до редколегії 2.03.2017

Рецензент: д-р техн. наук, проф. М.М. Корабльов, Харківський національний університет радіоелектроніки, Харків.

РОЗРОБКА СИСТЕМИ РОЗПІЗНАВАННЯ ЖЕСТІВ НА ОСНОВІ МОДЕЛІ КЛОНАЛЬНОГО ВІДБОРУ

Н.Г. Аксак, О.Ю. Барковська, Г.С. Іващенко

У статті запропоновано метод обробки зображень в режимі реального часу, що може застосовуватися для обслуговування користувачів з обмеженими можливостями. Підхід дозволяє мінімізувати затримку між подачею команди і реакцією системи на неї. Запропоновано систему розпізнавання жестів руки, в основу якої покладено принцип зіставлення образів за допомогою моделі клонального відбору.

Ключові слова: *розпізнавання жестів, альтернативні засоби комунікації, штучні імунні системи, CLONALG, CUDA.*

РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ ЖЕСТОВ НА ОСНОВЕ МОДЕЛИ КЛОНАЛЬНОГО ОТБОРА

Н.Г. Аксак, О.Ю. Барковская, Г.С. Иващенко

В статье предложен метод обработки изображений в режиме реального времени, применяемый для обслуживания пользователей с ограниченными возможностями, позволяющий минимизировать задержку между подачей команды и реакцией системы на нее. Предложена система распознавания жестов руки, основанная на сопоставлении образов посредством модели клонального отбора.

Ключевые слова: *распознавание жестов, альтернативные средства коммуникации, искусственные иммунные системы, CLONALG, CUDA.*