Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej Nr 61

Politechnika Gdańska

Wydział Elektrotechniki i Automatyki

Gdańsk 2018

doi: 10.32016/1.61.01

THE COMPARISON OF EFFICACY THE GESTURES RECOGNITION ALGORITHMS BASED ON RGB AND RGB-D CAMERAS FOR INTUITIVE SMART HOME CONTROL

Arkadiusz AMBROZIAK, Michał RODAK, Adrian CHOJECKI, Piotr BORKOWSKI

Politechnika Łódzka, Wydział Elektrotechniki, Elektroniki, Informatyki i Automatyki
tel.: 42 631 26 63 e-mail: arkadiusz.ambroziak@p.lodz.pl

Abstract: This paper presents a comparison of the authors' presence detection and gesture interpretation algorithm for a RGB camera with a commercial algorithm for a RGB-D camera for smart home control. The author's presence detection algorithm is based on MOG2 algorithm employed for background learning and mathematical conversions for identification of position of an arm. The latter algorithm employs a depth camera to take pictures at which a human skeleton is overlaid and thus position of an arm in three-dimensional space is identified. The author's algorithm enables to achieve better efficacy of gesture recognition (statistically of around 20%) employing less hardware resources at the same time. This makes it perfect for smart home automation control applications especially for people with disabilities.

Key words: identification for control, smart home control, supervisory control, gesture control, RGB camera.

1. INTRODUCTION

Development of smart technologies enforces development of interaction methods between human and machines. Even though the machines are getting more sophisticated, there is a tendency to simplify the way in which human communicates with them. Most recent refinements in the field of human-machine interaction aim to create intuitive controls to enable users to use machines without special preparation and training. Nowadays, the most up-to-date method, which meets the above mentioned criteria, is gesture recognition. Gestures are used in emotion recognition (body language), robot control, entertainment (as a game controller), rising life comfort of the disabled.

This paper compares the efficacy of two algorithms which meet those criteria: commercial algorithm and an algorithm which was developed by the authors to enhance controls of a building automation system.

2. PROPOSED METHOD

The goal was to create an algorithm that would simplify smart home control also for people with disabilities. For this purpose the following assumptions were made: simplicity, easy implementation in existing and new buildings, intuitiveness, keeping hardware requirements at the lowest possible level.

The use of external devices such as gloves and special sensors, that a user would have to wear, is unintuitive and implies wearing personal sensors [1,2,3,4,5,6,7,8]. This control method could gain on popularity if smart clothes with built-in sensors were more common. The use of smartphones as the gesture input device is impractical too [9,10]. Most of home appliance manufacturers provide BMS applications that can be installed on a smartphone and allow user to control its devices but it is not very convenient solution for an average user and it may involve software platform limitations too. For these reasons, the use of cameras to detect gestures is much better solution, because the user will not need any external devices to make gesturesin home. The authors of this paper designed an algorithm for RGB cameras due to their popularity and availability. The use of RGB-D camera would enable to us to detect gestures in 3D, but at the same time it would increase the computing power demand required for image data analysis. Image analysis is performed by methods from OpenCV library. The use of multiple cameras to detect gestures is problematic for implementation in homes and has low efficiency [11,12,13]. The use of gestures database as HMM, SVM, neural networks limits the number of devices that can be controlled [14,15,16]. If the number of the controlled devices is increase or there is introduced any change in the gesture database, it requires teaching the system a new classifier what would make the system too complicated for other users to use. Instead of creating gestures database, a database of objects was created, with which user can interact. This simplifies controls and makes it easier for the new user to use. The gesture, which is recognized by the algorithm, has to be made so that it points at one of the objects from the database. It is the most intuitive way of indication for the human and does not require learning. Another assumption was simplicity of the algorithm, what results in lower hardware requirements. As a result, the algorithm can be run even on slower hardware, what facilitates its implementation and reduces the cost of maintenance and operation of the system.

2.1. The Authors' RGB algorithm

The authors' algorithm (Fig. 1) recognises gestures in two dimensional space.

The first stage of the algorithm is background training and it uses MOG2 algorithm from the OpenCV library [17]. One of its features is shade detection support what helps to eliminate errors caused by misinterpretation of a shadow as a human. Another feature of this algorithm is continuous learning ability what helps the system to adapt to changing conditions. For learning purposes the algorithm uses 100 subsequent frames of a film.

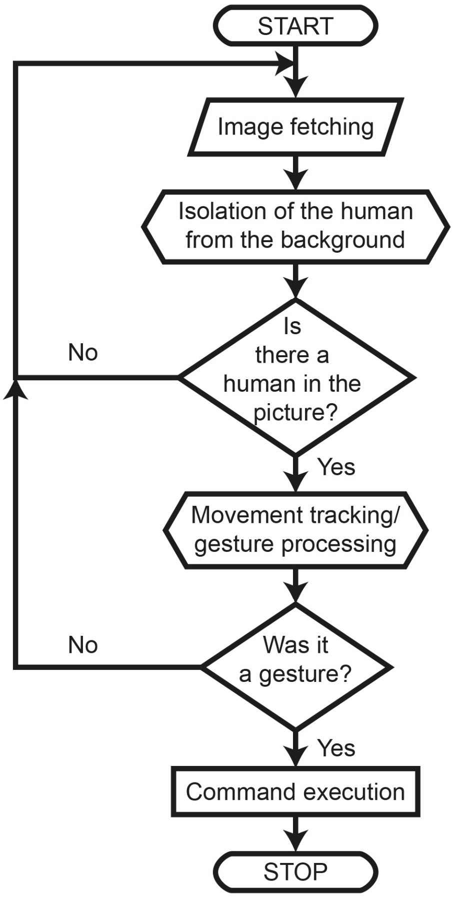

Fig. 1. Method of determining the pointing vector

When the background learning stage is finished the MOG2 algorithm produces an image consisting only of those elements which do not belong to the formerly learnt background. From this image shadow is removed and then the image is blurred in order to remove small inaccuracies.

The next stage covers identification of human silhouette. The image is divided into vertical strips of equal width. In each strip a number of white pixels is counted. Strips with the greatest number of white pixels are recognised as head with torso and legs, because of the fact that they have much greater space than arms and hands. Strips with the lower number of white pixels are interpreted as arms and hands.

Between lines on figure 2 is a region of interest, in which hands' contours are detected basing on their size. Basing on this, the contours are replaced with rectangles of the smallest possible area, thus creating vectors.

Subsequently, a vector created from position of both hands is formed. The algorithm detects motion of both hands at the same time.

The set of equations (1) allows to calculate the second point of the line at the end of which the vector is located.

$$\begin{aligned} P_{2,x} &= P_{c,x} + \cos(\alpha) \\ P_{2,y} &= P_{c,y} + \sin(\alpha) \end{aligned} \qquad (1)$$

where: $P_{2,x}$ is the x-coordinate of the P_2 point, $P_{2,y}$ is the y-coordinate of the P_2 point, $P_{c,x}$ is the x-coordinate of the P_c point, $P_{c,y}$ is the y-coordinate of the of the P_c point, α is the angle of inclination of the rectangle.

Basing on this, the algorithm calculates linear function's parameters (2).

$$y = ax + b \qquad (2)$$

where: $a = \dfrac{P_{2,y} - P_{c,y}}{P_{2,x} - P_{c,x}}$, $b = P_{c,y} - \alpha \cdot P_{c,x}$

Thus the vector value is averaged by values from 3 previous calculations what enables to keep the accuracy of the indication limiting sharp fluctuations of the vector at the same time.

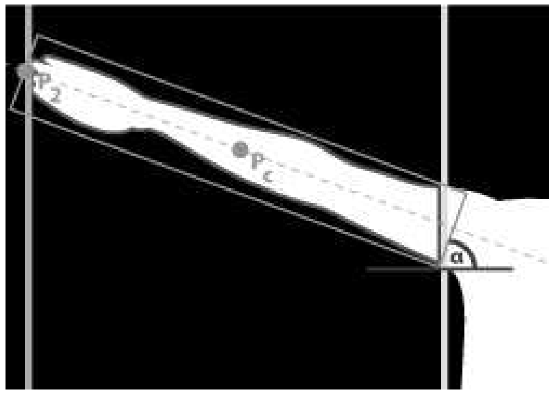

Fig. 2. Method of determining the pointing vector

An object with which a user wants to interact is described by a rectangle (Fig. 2). The algorithm translates the calculated vector accordingly to its direction through the line drawn by parameters from the formula (2). When it meets the rectangle on the object (3) it is recorded in a dedicated variable. To eliminate accidental commands, the vector has to indicate the object for the next 4 seconds. After that time the assigned command is executed (Fig. 3).

$$O_{z,x_1} \le x \le O_{z,x_2} \wedge O_{z,y_1} \le y \le O_{z,y_2} \qquad (3)$$

where: O_{z,x_1}, O_{z,x_2}, O_{z,y_1}, O_{z,y_2} are the x and y coordinates of the vertices of object z, *x* is the x-coordinate which is currently under review by the algorithm, *y* is the y-coordinate which is currently under review by the algorithm.

Fig. 3. Image obtained by the algorithm with additional information overlay

3. EXPERIMENTAL VERIFICATION

For the purpose of experimental verification a comparative research was carried for the following two algorithms: commercial algorithm, which is based on

RGB-D camera, and the authors' algorithm, which is based on RGB camera. The research was carried out at the Technical University of Lodz in the Department of Electrical Apparatus laboratory equipped with LCN building automation system [18]. It imitates a flat equipped with BMS system. Inside the laboratory there are designated pieces of smart home automation equipment (controlled objects), with which it is possible to interact by gestures, and an empty space (for the efficacy verification purposes). During the research standard behaviour of people in the laboratory was simulated, what created a custom data set and test setup. Then the research was carried out and its results were reworked statistically by statistical method based on confusion matrix for binary classification. The results were divided into four groups: true positive (the object was indicated and the algorithm identified it correctly), true negative (an empty space is indicated and the algorithm does not identify any objects), false positive (an empty space is indicated, but the algorithm identified a gesture as indication of an object), false negative (when an object was indicated but the object does not respond)

Number of samples in the research was 100 indications for each of arms including option with voice command if such option was available.

3.1. RGB-D algorithm

The commercial algorithm for RGB-D camera, chosen for the purpose of comparison and experimental verification, is in its final pre-release testing phase. The image was recorded by Asus Xtion Pro Live connected to a standard desktop PC, through which it is possible to control devices connected to the existing LCN building automation system [18].

The algorithm uses depth camera image onto which a human skeleton is overlaid. Basing on the formerly normalised hand and arm coordinates the algorithm creates a vector of their location. If the vector points at the formerly defined point, which represents a building device, the system awaits for a voice command (i.e.: turn on, turn off, dim) which will tell the building automation system what the building device should do. The algorithm allows to define location of only one hand and it is up to the user to decide which one.

It has a built-in voice command recognition tool, so for each of the commands it is required to create a sample voice command, which is mathematically described by the software tool delivered by the camera's manufacturer.

4. EXPERIMENTAL RESULTS

Basing on the researches results four parameters were calculated based on [19]: efficacy (accuracy - probability that a gesture is classified correctly), sensitivity (ability to sort out wrong indications), precision (probability that the algorithm responds appropriately on gesture), negative predictive value (NPV - probability that the algorithm responds appropriately to lack of gesture) .

Parameters are defined as follows [19]:
efficacy - is defined as a ratio between the correctly classified samples to the total number of samples,
sensitivity – is proportion between the positive correctly classified indications to the total number of positivie indiciations,
precision – is ratio between positive samples that were correctly classified as gesture to the total number of positive predicted samples,
NPV – is defined as proportion of negative samples that were correctly classified to the total number of negative predicted samples.

4.1. RGB-D algorithm

The algorithm was examined in the two variants of gesture recognition - without a voice command and with a voice command. The results are presented in the Fig. 4.

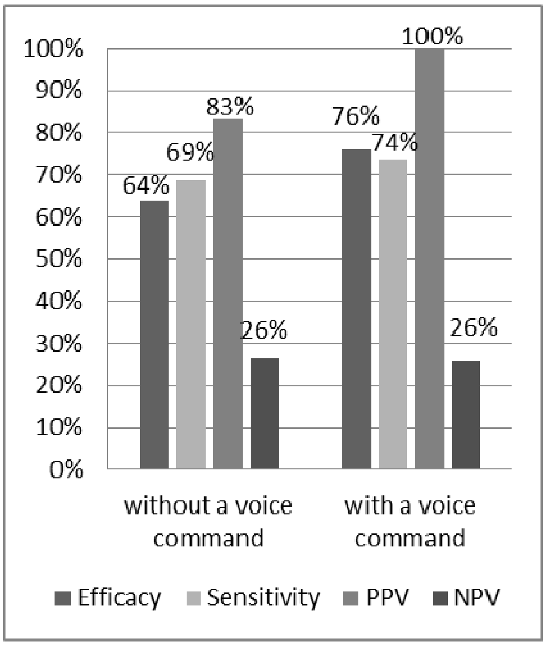

Fig. 4. Research results of RGB-D algorithm

The efficacy level of 64% (without a voice command) or 76% (with a voice command) means in practice that they may arise necessity to repeat the gesture so that it could be appropriately recognised. Despite high results of the efficacy, sensitivity and precision and low results of NPV for the variant with a voice command, the results for the variant without a voice command are average.

Application of voice commands efficiently eliminates wrong detection of a gesture and so voice commands act as an additional confirmation of an intention to use the indicated device. The maximum distance between a human and the camera is 3,5 m and this is caused by the hardware limitations of the RGB-D camera.

The gesture recognition was working flawless and the top CPU usage was 66,5% in case of RAM usage it is difficult to define because of recurring sharp rise from 300MB to 3GB over approximately 1,5 min what caused computer lagging. It also occurred several times that during normal talk a command was recognised even though the key word had not been spoken.

4.2. Authors' RGB algorithm

The algorithm was examined for gesture recognition made with both hands in two distances from the fixed camera. The research results are shown in the Fig. 5.

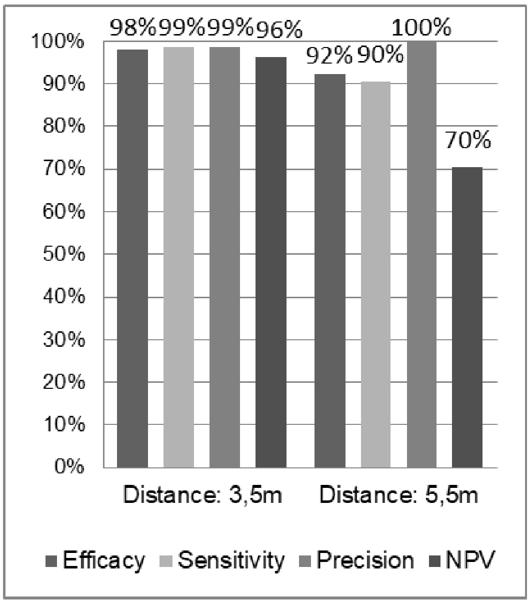

Fig. 5. Research results of RGB algorithm

The algorithm was tested on a personal computer and it was running flawlessly. CPU usage did not exceed 50% while RAM usage was at the level of 20MB. The efficacy of gesture recognition was at an acceptable level (92% in a distance of 5,5 m and 98% in a distance of 3,5 m) and it is expected that it could be applied in building automation systems easily. The precision and NPV performance indicators suggest that the risk of mistakes made by algorithm is smaller. Along with the increase in the distance from the camera efficacy and other parameters will decrease because it is more difficult to define arms location. Effective distance from which gestures are flawlessly recognised equals 5,5 m and above that distance pixels classified as arm are too small and therefore they are disqualified as a distortion.

In case of an instant change in lighting indoors occasionally occurred errors in gesture recognition. Temporary solution to this problem is background relearning but the authors plan to employ, apart from MOG2 algorithm, a human tracking algorithm.

5. CONCLUSIONS

In this paper the authors compared two algorithms dedicated for hardware solutions that enable intuitive control of smart home devices. The compared algorithms recognise gestures basing on RGB camera in two dimensional space or basing on RGB-D camera in three dimensional space.

The authors' algorithm for RGB camera consumes far less hardware resources and, at the same time, its efficacy is significantly higher than the algorithm for RGB- D camera even at the greater distance (5,5 m) and without any assistance of voice commands, and the probability of the results is far greater (basing on the precision and NPV test diagnostic indicators).

The achieved results are extremely important because they prove that the new algorithm, apart from the higher efficacy and operation over greater distance, offers also great potential for application without the necessity to upgrade existing hardware. The algorithm and this concept can be deployed across virtually any vendor of smart building automation system. Thus it creates a possibility to intuitively control home and building management systems what is extremely desirable nowadays. It has also great potential for application in facilities designed or upgraded for the people with motor impairments. Additionally, it is the first such a tool developed for controlling LCN-based building automation system. Relatively low hardware requirements enabled application of the authors' algorithm on Raspberry Pi3 microcomputer to reach sampling rates 20 frames per second in normal operation mode. That implies that it is possible to implement algorithm with low cost for end-user.

Acknowledgements

This work was supported in part by SILICON-TECH from Warsaw in Poland, which shared their commercial RGB-D–camera–based control algorithm and lent their hardware to carry out the research and comparison.

6. REFERENCES

1. Jaijongrak V., Chantasuban S., Thiemjarus S.: Towards a BSN-based gesture interface for intelligent home applications, ICCAS-SICE, Fukuoka, Japan, pp. 5613-5617, 2009.
2. Jingqiu W., Ting Z.: An ARM-based embedded gesture recognition system using a data glove, The 26th Chinese Control and Decision Conference, Changsha, China, pp. 1580–1584, 2014.
3. Kim D., Kim D.: An Intelligent Smart Home Control Using Body Gestures, International Conference on Hybrid Information Technology, Cheju Island, South Korea, pp. 439-446, 2006.
4. Wan Q., Li Y., Li C., Pal R.: Gesture recognition for smart home applications using portable radar sensors, 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, USA, pp. 6414-6417, 2014.
5. Kivimäki T., Vuorela T., Valtonen M., Vanhala J.: Gesture Control System for Smart Environments, 9th International Conference on Intelligent Environments, Athens, Greece, pp. 232—235, 2013.
6. Gonzalo P.J., Holgado-Terriza Juan A.: Control of home devices based on hand gestures, IEEE 5th International Conference on Consumer Electronics, Berlin, Germany, pp. 510-514, 2015.
7. Huang W., Jiang T., Liu Y., Liu W.: Applications of software radio for hand gesture recognition by using long training symbols, 9th International Conference on Signal Processing and Communication Systems, Cairns, Australia, pp. 1-5, 2015.
8. Liu T., Luo X. M., Liu J., Cui H.: Compressive Infrared Sensing for Arm Gesture Acquisition and Recognition, IEEE International Conference on Information and Automation, Lijiang, China pp. 1882-1886, 2015.
9. Batool A., Rauf S., Zia T., Siddiqui T., Shamsi J.A., Syed T.Q., Khan A.U.: Facilitating gesture-based actions for a Smart Home concept, International Conference on Open Source Systems and Technologies, Lahore, Pakistan, pp. 6-12, 2014.
10. Hung C.H., Bai Y.W., Wu H.Y.: Home outlet and LED array lamp controlled by a smartphone with a hand gesture recognition, IEEE International Conference on Consumer Electronics, Las Vegas, USA, pp. 5-6, 2016
11. Huang C.M., Chen Y.R., Fu L.C.: Visual Tracking of Human Head and Arms Using Adaptive Multiple Importance Sampling on a Single Camera in Cluttered

Environments, IEEE Sensors Journal, vol. 14, no. 7, pp. 2267-2275, 2014.
12. Juang C.F., Liang C.W., Lee C.L., Chung I.F.: Vision-based Human Body Posture Recognition Using Support Vector Machines, 4th International Conference on Awareness Science and Technology, Seoul, South Korea, pp. 150 – 155, 2012.
13. Chien C.Y., Huang C.L., Fu C.M.: A Vision-Based Real-Time Pointing Arm Gesture Tracking and Recognition System, IEEE International Conference on Multimedia and Expo, Beijing, China, pp. 983-986, 2007.
14. Ransalu S., Kumarawadu S.: A robust vision-based hand gesture recognition system for appliance control in smart homes, IEEE International Conference on Signal Processing, Communication and Computing, Hong Kong, China, pp. 760-763, 2012.
15. Wang R., Yu Z., Liu M., Wang Y., Chang Y.: Real-time visual static hand gesture recognition system and its FPGA-based hardware implementation, 12th International Conference on Signal Processing, Hangzhou, China, pp. 434-439, 2014.
16. Patras L., Giosan I., Nedevschi S.: Body gesture validation using multi-dimensional dynamic time warping on Kinect data, IEEE International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, pp. 301–307, 2015.
17. Zivkovic Z.: Improved adaptive Gausian mixture model for background subtraction, Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, vol. 2, pp. 28-31, 2004.
18. Borkowski P., Pawłowski M., Makowiecki T.: Economical Aspects of Building Management Systems Implementation, IEEE Trondheim PowerTech, Trondheim, Norway, pp. 1-6, 2011.
19. Tharwat A.: Classification assessment methods, 2018, https://doi.org/10.1016/j.aci.2018.08.003.

PORÓWNANIE SKUTECZNOŚCI ALGORYTMÓW ROZPOZNAWANIA GESTÓW OPARTYCH NA KAMERACH RGB I RGBD DO INTUICYJNEGO STEROWANIA INTELIGENTNYM BUDYNKIEM

Artykuł prezentuje porównanie autorskiego algorytmu wykrywania gestu dla kamery RGB z algorytmem komercyjnym dla kamery RGB–D, na potrzeby sterowania inteligentnym budynkiem.

Autorski algorytm oparty jest na algorytmie MOG2, wykorzystywanym do uczenia tła oraz w matematycznych przekształceniach w celu wykrycia pozycji ramienia. Na tej podstawie określany jest wektor wyznaczane przez rękę wskazującą na dany obiekt. Wskazywany element jest porównywany z bazą obiektów, następnie wykonywana jest związana z nim interakcja. Porównywany był on z komercyjnym algorytmem wykorzystującym kamerę głębi, która nakładała szkielet człowieka na obraz i pozycjonowała ramię w przestrzeni trójwymiarowej.

Opracowany algorytm pozwala na osiągnięcie wymiernie lepszych wyników w skuteczności rozpoznawania (o ok. 20% w ujęciu statystycznym) w stosunku do komercyjnego algorytmu, przy mniejszym wykorzystaniu zasobów sprzętowych. Pozwala to na zastosowanie algorytmu w istniejących oraz nowo powstałych budynkach mieszkalnych, wykorzystując do jego implementacji mikrokomputery.

Keywords: sterowanie inteligentnym budynkiem, sterowanie gestem, rozpoznawanie gestów, kamera RGB.