The Hand-gesture-based Control Interface with Wearable Glove System

Vladislav Berezhnoy, Dmitry Popov, Ilya Afanasyev and Nikolaos Mavridis
Institute of Robotics, Innopolis University, Universitetskaya Str. 1, Innopolis, 420500, Russia

Keywords: Gesture-based Control Interface, Glove System, Fuzzy C-means (FCM) Clustering, Arduino, V-REP Simulator.

Abstract: The paper presents an approach to building a gesture-based control interface with a wearable glove system and a real-time gesture recognition algorithm. The glove-based system is a wireless wearable device with hardware components, including Arduino Nano controller, IMU and flex sensors, and software for gesture recognition. Our gesture recognition methodology requires two stages: 1) Building a library of dynamic gesture models with the reference human gesture graphs; 2) Gesture capturing and evaluating with fuzzy c-means (FCM) clustering and constructing grammars of gestures by fuzzy membership functions. The system tests were provided with 6 different dynamic gestures to control position and orientation of a quadcopter in V-Rep simulator that has demonstrated encouraging results with a reasonable quality of real-time gesture-based quadcopter control.

1 INTRODUCTION

The gesture-based control interface has a great potential for more natural, intuitively understandable, customizable and convenient human-machine interaction, which can extend capabilities of widespread graphical and command line interfaces, which we use nowadays with mouse and keyboard. Therefore, development of advanced hardware and software approaches to hand-gesture recognition is important for many 3D applications such as control of computers and robots, interaction with the computer-generated environment (virtual or augmented reality), sign language understanding, gesture visualization, games control, enhancement of communication ability for disabled people, etc. Many recent studies discuss various methods to solve gesture recognition problem of certain gesture classes by computer vision based systems (Suryanarayan et al., 2010; Suarez and Murphy, 2012; Rautaray and Agrawal, 2015; Wachs et al., 2011; Zabulis et al., 2009), wearable sensor-based systems (Luzhnica et al., 2016; Park et al., 2015) or even integrated systems, which simultaneous use both vision-based devices and wearable sensors (Arkenbout et al., 2015; Mavridis et al., 2012). The classification of gesture classes depends on gesture difficulties, applications and level of recognition accuracy (e.g. surgical systems require higher accuracy than entertainment or communication applications) (Suarez and Murphy, 2012; Wachs et al., 2011). Computer vision-based techniques are one of the most frequently used approaches, which apply RGB camera and image processing algorithms (Manresa et al., 2005; Alfimtsev, 2008), Kinect sensors and depth maps (Suryanarayan et al., 2010; Suarez and Murphy, 2012; Ren et al., 2013; Dominio et al., 2014; Afanasyev and De Cecco, 2013) and Time-of-Flight (ToF) cameras (Gudhmundsson et al., 2010) for gesture tracking and hand motion detection. Although these solutions can be computationally expensive, they may suffer from lack of robustness in cluttered background or poor motion scenarios (e.g. by using just a single gesture). Moreover, they often demonstrate sensitivity to the environment, illumination conditions, scene, background details and camera parameters (resolution, frame rate, distortion, auto-shutter speed, etc.) that can affect recognition quality (Luzhnica et al., 2016; Wachs et al., 2011). Therefore, many investigations focus also on wearable gesture recognition systems with the ability to track dynamic gestures for the complicated work environment in real-time with reasonable computational cost and higher accuracies (Kumar et al., 2012; Luzhnica et al., 2016; Kenn et al., 2007; Battaglia et al., 2016). The overall goal of hand gesture recognition is to find similarity between an unknown performed gesture (called recognizable model) and a known class of gestures (patterns of gestures). Once the suitable hand gesture features have been extracted and a gesture set has been selected, gesture classification can be accomplished

Berezhnoy, V., Popov, D., Afanasyev, I. and Mavridis, N.
The Hand-gesture-based Control Interface with Wearable Glove System.
DOI: 10.5220/0006909304480455
In *Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2018) - Volume 2*, pages 448-455
ISBN: 978-989-758-321-6

by standard machine learning techniques or special-purpose classifiers (Suarez and Murphy, 2012), which are frequently based on Neural Networks (Gawande and Chopde, 2013), Bayesian networks (Suk et al., 2010), Hidden Markov Models (Bansal et al., 2011), etc. The main drawback of these methods is high computational complexity for forming gesture patterns and recognizing dynamic gestures that may limit their feasibility for real-time applications. Another approach with an attractive algorithm of dynamic gesture recognition based on fuzzy finite state automata (for human's wrist detection from video stream) was proposed in (Devyatkov and Alfimtsev, 2007; Alfimtsev, 2008), and inspired the authors of this paper to contribute and update this methodology for gesture recognition with inertial IMU and flex sensors built in a wearable glove.

In this paper, we focus on the development of a wearable gesture recognition system, which consists of sensor-integrated glove hardware and hand gesture recognition software for human control of computer-based objects and machines (see, Fig. 1). Our gesture recognition approach comes down to (1) tracking the trajectories of hand movements along the coordinate axes $x(t)$, $y(t)$ and $z(t)$; (2) building a recognizable model for the gesture $G[x(t), y(t), z(t)]$, using the tracked trajectories; (3) a comparison of the recognizable gesture model with the reference gesture patterns $E_i[x(t), y(t), z(t)]$ by computing the similarity function $C[G, E_i]$ to define a relation to the i-th gesture class. These wearable glove system and hand-gesture-based software were used for the creation of control interface to manipulate a quadcopter model in V-Rep simulator, demonstrating successful real-time gesture-based control of the quadcopter position and orientation with 6 different dynamic gestures.

The paper is organized as follows: Section 2 presents our wearable glove system for gesture recognition, Section 3 formalizes the dynamic pattern construction methodology, and Section 4 describes how we use the FCM clustering algorithm to recognize a gesture pattern. Finally, we test our approach to building a gesture-based control interface with the glove system and a real-time gesture recognition algorithm in Section 5 and conclude in Section 6.

2 DEVELOPMENT OF A WEARABLE GLOVE SYSTEM FOR GESTURE RECOGNITION

Numerous wearable glove-bases systems already exist (Dipietro et al., 2008). In our work, we have taken into account existing techniques but utilized a novel recognition approach.

In our system, to recognize an unknown gesture the following information is required:
1) Pitch, roll and yaw hand rotations relative to the surface;
2) Acceleration projections on each coordinate axis;
3) The numerical value of the bending for each finger.

The objectives 1 and 2 can be solved using the integrated sensor MinIMU-9 v2, which consists of an accelerometer, magnetometer, and gyroscope, and measures projections to calculate pitch, roll and yaw. The objective 3 is reached using the flex sensors. We organized gesture-based control interface with the wearable glove-based system, using sensors' connection to hardware with the Arduino Nano controller, Bluetooth (BT) wireless data transmission and Java application (see, a functional diagram in Fig. 1). The software executes gesture evaluation with fuzzy c-means (FCM) clustering algorithm (Bezdek et al., 1984) and computation of a similarity function between recognizable gesture models and dynamical reference gesture patterns from gestures' library. This glove-based system was tested by controlling an Unmanned Aerial Vehicle (UAV) in V-Rep simulator. The hardware components of the wearable glove system are shown in Fig. 2.

3 CONSTRUCTION OF DYNAMIC GESTURE PATTERNS

The method, which constructs dynamic gesture patterns performed by a hand motion, consists of two stages: 1) Capturing and tracking regions of interest $O_b[x(t), y(t)]$, containing $x(t)$ and $y(t)$ trajectory projections of hand motion on the coordinate axes over time; 2) Building a pattern $E_i[x(t), y(t)]$ and recognizable gesture model $G[x(t), y(t)]$, using the tracked trajectories. When performing a gesture for recognition, it is enough to execute it once. But for the construction of the reference gesture models (classes) the same gesture needs to be performed repeatedly. The motion trajectories of each repeated gesture, of course, are not the same. For example, each hand motion trajectory $(x(t), y(t))$, drawing the letter "N", will form a set of trajectories $(x_j(t), y_j(t))$ included in certain boundaries at the n-fold repetition of gesture $j = 1, 2, ..., n$, which are a gesture pattern E with own characteristic features, as shown in Fig. 3.

The Stage 2 of our approach is related to the methodology of dynamic gesture recognition based on fuzzy finite state automata proposed by (Devyatkov and Alfimtsev, 2007). As far as this methodol-

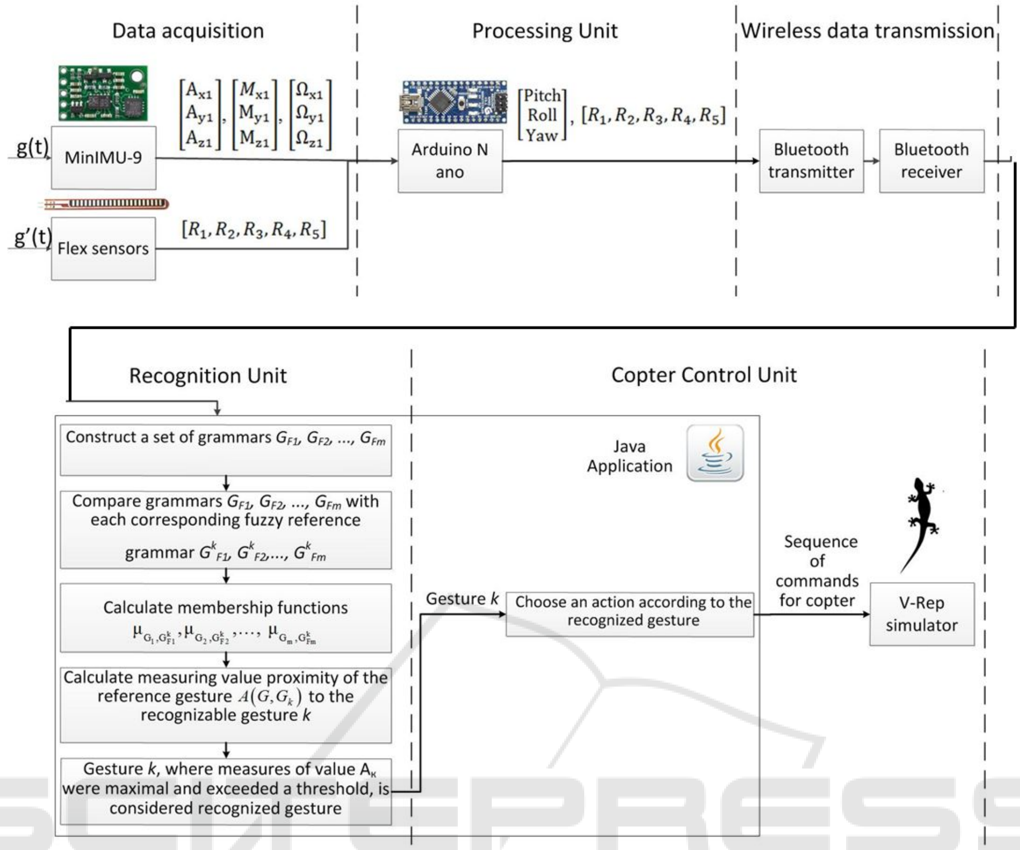


Figure 1: The functional diagram of wearable glove-system for gesture-based control of UAV.

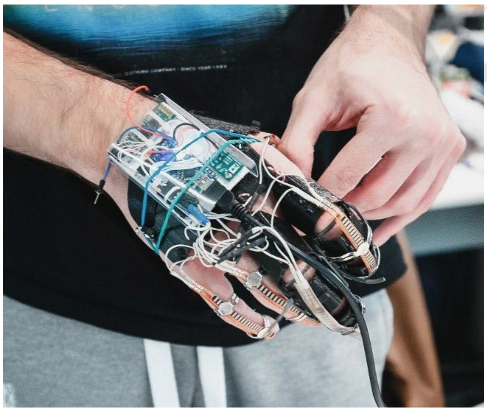


Figure 2: The hardware components of the glove system.

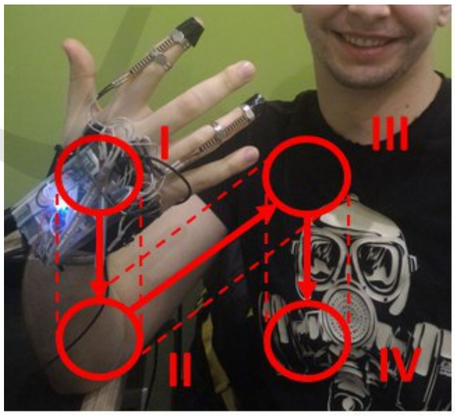


Figure 3: The wearable glove system with the dynamic gesture graph in the form of letter "N".

ogy was published in Russian, let us describe the main principles for the recognition of the gesture, drawing the letter "N". The generalization of gesture patterns E for different hand trajectories in the form of letter "N" can be represented as a graph shown in Fig. 3. According to this graph, vertex I includes a set of points (coordinates) that belongs to the beginning of the gesture, vertices II and III correspond to the points of the trajectory bending, vertex IV contains the end points of the trajectory. It serves as a basis for pattern gesture construction.

The first task of gesture pattern construction E_i according to trajectories $(x_j(t), y_j(t))$ is an algorithmic definition of movement trajectory points $(x_j(t_m), y_j(t_m))$ at time t_m, which correspond to the

m-th vertex of the graph. To solve this task, we used fuzzy c-means (FCM) clustering algorithm (Bezdek et al., 1984; Nayak et al., 2015).

Let us remind the main definitions for c-means clustering. The set of points for the motion trajectory relative to one vertex is called a cluster. The number of measurement points is denoted as N. Each c-th cluster includes a subset of the values of the characteristic features of the vectors $p_k = [p_{k1}, p_{k2}, ..., p_{km}]$, where $k = 1,...,N$ - is the total number of points, m - number of features. For the considered gesture, m = 4. The features of each k-th point are based on the trajectory $x(t)$ and $y(t)$ at the corresponding time t_k.

$$\begin{cases} p_{k1} = x(t_k), \\ p_{k2} = \frac{dx(t)}{dt}|_{t=t_k}, \\ p_{k3} = y(t_k), \\ p_{k4} = \frac{dy(t)}{dt}|_{t=t_k} \end{cases} \tag{1}$$

The clustering algorithm c-means is based on a method for minimizing an objective function, which should be created such way to: (1) Minimize a distance between a cluster's point and the cluster center; (2) Maximize a distance between clusters' centers.

Although the approach of applying FCM clustering algorithm to gesture classification is used long ago (Li, 2003; Wachs et al., 2002), we used the criterion known as the sum of squared errors within a class, which uses the Euclidean norm to describe the distance between vectors d_{ik} 3 presented in (Devyatkov and Alfimtsev, 2007). This criterion is denoted $J(u,v)$, where u is a partition of all points in clusters, and v is a vector of cluster centers, which corresponds to the partition u. The formula of the criterion (the objective function) would be the following:

$$J(u,v) = \sum_{k=1}^{N} \sum_{i=1}^{c} u_{ik} d_{ik}^2 \tag{2}$$

where d_{ik} - is a measure in Euclidean n-dimensional feature space between the k-th m-dimensional vector p_k and the i-th cluster center v_i, which is calculated by using the formula:

$$d_{ik} = |p_k - v_i| = \left[\sum_{k=1}^{N} (p_{kj} - v_{ij})^2 \right]^{1/2} \tag{3}$$

The coordinates of the cluster centers $v_i = \{v_{i1}, v_{i2}, ..., v_{im}\}$ are calculated by the formula:

$$v_{ij} = \frac{\sum_{k=1}^{N} u_{ik} p_{kj}}{\sum_{k=1}^{N} u_{ik}} \tag{4}$$

where u_{ik} - the characteristic function, A_i - i-th cluster, $i = 1, 2, ..., C$:

$$u_{ik} = \begin{cases} 1, \ if \ p_k \in A_i, \\ 0, \ if \ p_k \notin A_i. \end{cases} \tag{5}$$

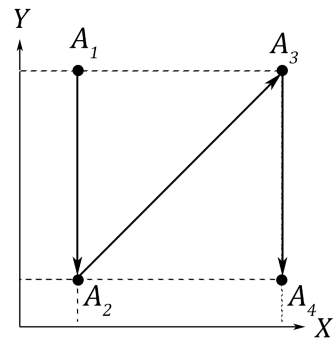


Figure 4: Graph of the gesture in the form of letter "N".

It is required to find the optimal partition u^* into clusters with centers v^*, for which the objective function value is minimal:

$$J(u^*, v^*) = \min_{u,v} J(u,v) \Big/_{u \in M_c} \tag{6}$$

where M_c - the set of all different partitions into C clusters.

We use a strategy of c-means clustering algorithm, which is known as iterative optimization that includes the following steps (Devyatkov and Alfimtsev, 2007):

1. Fix the number of clusters $C(2 < C < N)$ and select the primary partition for the set of trajectories of points on the A_i clusters. Then, perform the following steps for $r = 0, 1, 2, ...$
2. Compute the centers $v_i^{(r)}$ of all clusters defined by partition $u^{(r)}$.
3. Calculate the new features for all i and then k:

$$u_{ik}^{(r+1)} = \begin{cases} 1, \ if \arg\min_k d_{ik}^r \Big/_{i=\overline{1,N}}, \\ 0, \ otherwise. \end{cases} \tag{7}$$

4. Build a new partition of $u^{(r+1)}$.
5. If $u^{(r+1)}$ = $u^{(r)}$ then stop the process, considering the partition of $u^{(r+1)}$ as optimal. Otherwise, suppose $r = r + 1$ and go to the step 2.

This strategy (Devyatkov and Alfimtsev, 2007) allows to determine models of all reference gestures, which can be presented with graphs, where points of trajectories $(x_j(t), y_j(t))$ can be assigned to each vertex of the graph for dynamic gestures. It gives pattern gestures, where graph vertices correspond to clusters with their centers and edges relate to trajectory direction. The example of gesture graph, which has the shape of letter "N", is shown in Fig. 4, where vertices correspond to cluster A_1,A_2,A_3 and A_4. The coordinates of the cluster centers are shown on the axes.

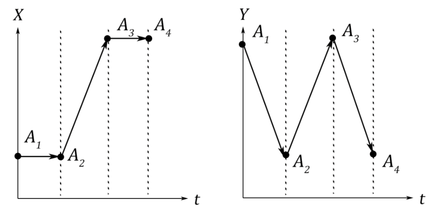


Figure 5: Gesture graph projections of the graph shown in the Fig. 4.

4 GESTURE RECOGNITION WITH FUZZY FINITE AUTOMATA AND GRAMMARS

Gesture graph in Fig. 4 does not contain information about the movement of the centers of clusters over time. To consider gesture over time, a model of dynamic gestures based on fuzzy finite automata should be build (Devyatkov and Alfimtsev, 2007). To do this, according to the gesture graph shown in Fig. 4, two graphs shown in Fig. 5 can be constructed. These graphs are obtained by moving the projection trajectories hands on time axis and the abscissa axis and also on the ordinate axis and the time axis.

The basic principles of automation grammar generation for gesture recognition is well presented in (Alfimtsev, 2008). Since it was published in Russian, let us describe the main statements. When considering one sample of a gesture projections $y_i(t)$, the sequence of $n+1$ samples $Y_i[t_0,t_n] = \{y_i(t_0), y_i(t_1), y_i(t_2), ..., y_i(t_n)\}$ of i-th projection of the same graph gesture for several consecutive time points $t_0, t_1, ..., t_n$ (for the time interval $[t_0,t_n]$) is called a signal. The set of samples $K(t) = \{y_1(t), y_2(t), .., y_m(t)\}$, where m - number of different projections of the same gesture graph at time t is called the reaction. The reaction sequence $K(t_0), K(t_1), ..., K(t_n)$, obtained by m projections of same gesture for several consecutive moments of time $t_0, t_1, t_2, ..., t_n$ (for a time interval $[t_0,t_n]$) is called a flow of reactions. Each sample $y_j(t_i)$ of the same signal corresponds the condition $b_j(t_i)$ of the finite automaton M_j. Then it can be introduced a function of outputs $\varphi(b_j(t_i)) = y_j(t_i)$ and the function of automaton transitions $f(b_j(t_i), t_{i+1}) = bj(t_{i+1})$ for a finite automaton M_j. Thus, each sample is a value of output function $y_j(t) = \varphi(b_j(t))$ of automaton M_j; each signal is a sequence of values of output functions $y_j(t) = \{y_j(t_0), y_j(t_1), ..., y_j(t_n)\}$ of the same automaton M_j; each reaction is the set $y(t) = \{y_1(t), ..., y_m(t)\}$ values of output functions of different automaton $M_1, M_2, ..., M_m$, and reaction flux is a sequence of $y(t_0), y(t_1), ..., y(t_n)$. Therefore, any automaton M_j corresponding to a projection of a gesture graph can be represented by its transition graph, where each graph vertex is marked with the symbol b_i, each pair of adjacent vertices (b_i, b_{i+1}), and edge directed from vertex i to vertex $i+1$ is marked with the symbol t_i in the alphabet $T = \{t_0, t_1, t_2, ..., t_{m-1}\}$. If write down all edges, the result will conclude a sequence of characters $t_1 t_2 ... t_{m-1} \Lambda$ (where Λ is an empty symbol, which may be omitted). This sequence can be considered as a word or a sentence of the language $L = L(G)$, generated by an automaton grammar $G = \{V, T, P, S = b_0\}$, where $V = \{b_1 b_2 ... b_{m-1}\}$, $T = \{t_1 t_2 ... t_{m-1} \Lambda\}$, $P = \{b_0 \to t_1 b_1,\ b_1 \to t_2 b_2, ...\ b_{m-2} \to t_{m-1} b_{m-1},\ b_{m-1} \to t_m b_m,\ b_m \to \Lambda\}$.

In the ideal case for gesture recognition a set of automatons $M_1, M_2, ..., M_m$ can be constructed such way that each automaton corresponds to one of the distinct grammars. Then the language corresponding to this automation could be unambiguously detected by the automaton grammar. However, in reality such an ideal situation is unattainable, because the person can not perform every new gesture in absolutely same way. Therefore, to cope with the uncertainty that arises when performing gestures, we need to move from deterministic to fuzzy automata. To do this, we construct for each distinct grammar G the corresponding fuzzy grammar G_F, based on the principles described in (Alfimtsev, 2008).

Each edge of the finite deterministic automaton corresponds two incident vertices b_i and b_{i+1} with the vertices' coordinates $[t_i, \varphi(b_i(t_i)) = y_i(t_i)]$ and $[t_{i+1}, \varphi(b_{i+1}(t_{i+1})) = y_{i+1}(t_{i+1})]$ respectively. Following (Alfimtsev, 2008), we will assume that the samples $y_i(t_i)$ of the same cluster corresponding to l different trajectories of the same gesture may vary within the standard deviation of the projection from the cluster center $v_i(t_i)$:

$$s_i = \sqrt{\frac{\sum_{l=1}^{N} \left[y_i^l(t_i) - v_i(t_i) \right]^2}{N}}, \tag{8}$$

where N – number of samples belonging to the cluster, v_i – the coordinate of the center of the i-th cluster, $y_i^l(t_i)$ – sample, belonging to i-th cluster. For simplicity, assume that s_i is the same for all i and equal to s. For each set of samples $y_i^l(t_i)$ we set the triangular membership function $\mu_i(y)$, defined by points $y_i^- = v_i - s$, $y_i = v_i$, $y_i^+ = v_i + s$, where $\mu_i(y_i^-) = 0, \mu_i(y_i) = 1$ and $\mu_i(y_i^+) = 0$.

The vertex b_i with coordinates (t_i, y_i) is replaced by set of vertex $b_{ri} \in B(b_i)$ with coordinates, changing within the interval $(y_i^- = y_i - s,\ y_i^+ = y_i - s)$. Each vertex b_{ri} corresponds to a specific coordinate, and the set $B(b_i)$ is calculated as a set of vertices of all the coordinates for which $y_i^l(t_i) > 0$. Then, instead of vertex b_{i+1} with the coordinates (t_{i+1}, y_{i+1}) we will have a set of vertices $b_{r(i+1)} \in B(b_{i+1})$ with coordinates, changing within the interval (y_{i+1}^-, y_{i+1}^+), and instead of one edge t_{i+1} (from vertex b_i to vertex b_{i+1}) we will have a plurality of all edges $\{(b_{ri}, b_{r(i+1)}) | b_{ri} \in B(b_i), b_{r(i+l)} \in B(b_{i+1})\}$ joining each vertex of the set $B(b_i)$ to each vertex of the set $B(b_{i+1})$.

More detailed description about how to generate grammars for both finite deterministic automaton and fuzzy finite automata for hand gesture recognition is presented in (Alfimtsev, 2008). Thus, two triangular membership functions: $\mu_i(y)$ and $\mu_{i+1}(y)$ are defined by triplets at the points $\{y_i^-, y_i, y_i^+\} and \{y_{i+1}^-, y_{i+1}, y_{i+1}^+\}$ correspondingly. Each of these functions is determined by the following expression:

$$\mu(y) = \begin{cases} \frac{y - y_k^-}{y_k - y_k^-}, \; if \; y_k^- \leq y \leq y_k, \\ \frac{y_k^+ - y}{y_k^+ - y_k}, \; if \; y_k < y \leq y_k^+. \end{cases} \quad (9)$$

where, $k \in (i, i+1)$. These functions define the measure of closeness to vertex coordinates to the "ideal coordinates", which correspond to the value of the membership function, equal to 1. It is assumed that the membership function of each edge $(b_{ri}, b_{r(i+1)})$ of incident vertices $b_{ri} \in B(b_i)$ and $b_{r(i+l)} \in B(b_{i+1})$ is defined as:

$$\mu_{(b_{ri}, b_{r(i+1)})}(t_{i+1}) = \min\{\mu_i(y_{ri}), \mu_{i+1}(y_{r(i+1)})\}. \quad (10)$$

Fuzzy grammar G_F obtained from the regular grammar G (Alfimtsev, 2008). The set of the rules P_F for the fuzzy grammar G_F will be the following:

$$\begin{gathered} P_F = \{b_{ri} \to t_{i+1} b_{r(i+1)}, \\ \mu(t_{i+1} b_{r(i+1)}) = \mu_{(b_{ri}, b_{r(i+1)})}, \\ i = 0, \ldots, n-1\}. \end{gathered} \quad (11)$$

According to (Alfimtsev, 2008), the grammar G with rules $\{b_{ri} \to t_{i+1} b_{r(i+1)}, \; i = 0, \ldots, n-1\}$ is comparable to the fuzzy grammar G_F, if there is a sequence of fuzzy rules $\{\mu_{(b_{ri}, b_{r(i+1)})}, \; i = 0, \ldots, n-1\}$, which takes place $b_i = b_{ri}$ for all $i = 0, \ldots, n-1$.

Thus, the dynamic gesture recognition algorithm, which uses a model based on fuzzy finite automata and the corresponding set of reference fuzzy grammars $G_{F1}, G_{F2}, \ldots, G_{Fm}$ will contain the following steps (Alfimtsev, 2008):

1. A gesture is treated with the same sampling steps along the time axis as the reference gestures, and with construction of a set of grammar $G_{F1}, G_{F2}, \ldots, G_{Fm}$.

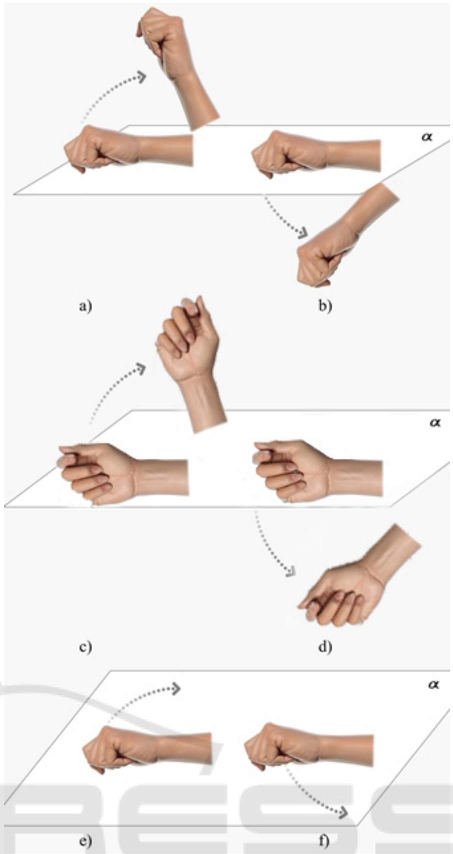
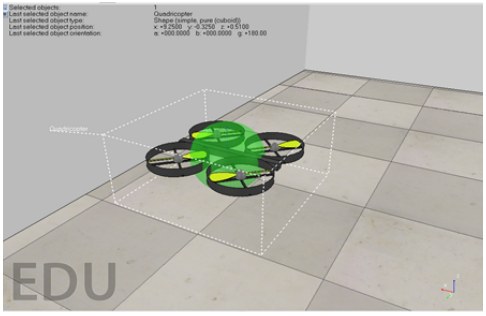
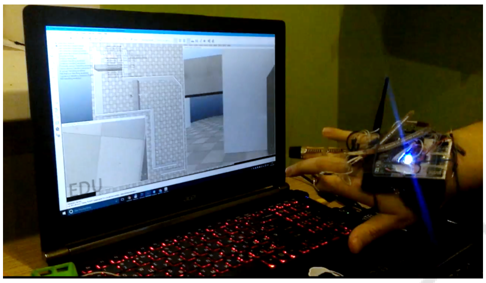


Figure 6: Gestures to control a quadcopter in V-Rep simulator: (a) move forward, (b) move backward, (c) move up, (d) move down, (e) turn right, (f) turn left.

2. The grammars $G_{F1}, G_{F2}, \ldots, G_{Fm}$ corresponding recognizable gesture with each corresponding fuzzy reference grammar $G_{F1}^k, G_{F2}^k, \ldots, G_{Fm}^k$, where $k \in \{1, \ldots, K\}$, and K is a number of recognizable gestures, are compared.
3. For the sets of fuzzy reference grammars $G_{F1}^k, G_{F2}^k, \ldots, G_{Fm}^k$, where comparison was successful, it is calculated the corresponding set of values for membership functions $\mu_{G_1, G_{F1}^k}, \mu_{G_2, G_{F2}^k}, \ldots, \mu_{G_m, G_{Fm}^k}$ according to the formula 10, and then the value of the measure A_k, which characterizes the similarity of the recognizable gesture to the reference gestures k, by the formula:
$$\begin{gathered} A(G, G_k) = A_k = \\ \max\{\mu_{G_1, G_{F1}^k}, \mu_{G_2, G_{F2}^k}, \ldots, \mu_{G_m, G_{Fm}^k}\}, \end{gathered} \quad (12)$$
4. The recognizable gesture is considered coincident with the pattern gesture k, for which the measure value A_k was maximal. If there is no successful comparison of grammars, then recognition of this gesture fails (i.e. the gesture was not recognized).



(a) The quadcopter in V-Rep environment.



(b) Control with a glove system (exp,).

Figure 7: Experiments with hand-gesture control of the quadcopter with the glove system in V-Rep simulator.

5 TESTS OF THE HAND-GESTURE CONTROL INTERFACE WITH THE GLOVE SYSTEM

After glove-based system hardware and software implementation, it was conducted experiments to understand the recognition ability. To test the system's recognition rate and demonstrate the advantages of gesture-based control, we decided to provide experiments to control UAV (in our case, the simulated quadcopter, Fig. 7a) with this glove system. To control position and orientation of the quadcopter in V-Rep simulator environment we selected 6 gestures (Fig. 6): move forward and backward, move up and down, turn left and right. The Fig. 7b shows an operator at the moment of the quadcopter control by the hand gestures during its flight in a V-Rep maze. The video with this experiment with the gesture control of the quadcopter with the wearable glove-based system in V-Rep simulator is achievable on YouTube (exp,). Before the UAV flight, the operator calibrates the glove system, recording in software the main gestures, which will be used for the control. By using only a set of predefined motions user could easily navigate UAV through the V-Rep maze. But it should be noticed that in this scenario the operator can only control the desired drone navigation points, whereas the built-in algorithms of V-Rep quadcopter control finds an optimal way to reach them. The tests show that the implemented wearable glove system for hand-gesture-based control is intuitive, easily adjustable and customizable through personal gesture library.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we present glove-based system hardware and software implementation, which organize the user-friendly hand-gesture-based control interface based on fuzzy finite state automata gesture recognition methods. These wearable glove system (with inertial IMU and flex sensors) and gesture recognition methodology were used to manipulate a quadcopter model in V-Rep simulator, demonstrating successful real-time gesture-based control of the quadcopter position and orientation with 6 different dynamic gestures. This methodology is based on fuzzy c-means (FCM) clustering algorithm with the sum of squared errors measure criterion (Devyatkov and Alfimtsev, 2007), which minimizes a distance between a cluster's point and the cluster center, and maximizes a distance between clusters' centers. These glove-based system' hardware and software can also be used in various applications to create human-computer interfaces. The software solution can be of interest in the situations, for instance in a teaching environment, to check the correctness of gestures, in case where gestures are many and the trajectory is important as much as the final position.

The methodology for gesture recognition based on fuzzy finite automata and grammars has many advantages and capabilities, for example: (1) Automatic creation of a pattern gesture model for each dynamic gesture. (2) Dynamic gesture model training with a small training set (only a few examples). (3) Reliable recognition of dynamic gesture trajectories in real-time, including occlusions and intersections. (4) Computational efficiency for such models. Fuzzy model has the computational complexity of $O(mn)$, where m - the number of fuzzy automata used for recognition, the n - the maximum number of states fuzzy finite automaton.

The future work may include extension of control interface capabilities to manipulation of larger number of robots, additional testing with a real quadcopter and comparison with alternative methods to clarify the learning time, control accuracy, etc.

ACKNOWLEDGEMENTS

This research has been supported by the grant of Russian Ministry of Education and Science, agreement: No14.606.21.0007, ID: RFMEFI60617X0007, "Android Technics" company and Innopolis University.

REFERENCES

The tests of the hand-gesture-based control interface with wearable glove system. Innopolis University: https://youtu.be/f9dfKCUuUvY.

Afanasyev, I. and De Cecco, M. (2013). 3d gesture recognition by superquadrics. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 2, pages 429–433. INSTICC.

Alfimtsev, A. (2008). *Research and development of methods for dynamic gesture capture, tracking and recognition*. PhD thesis, Bauman Moscow State Technical University.

Arkenbout, E. A., de Winter, J. C., and Breedveld, P. (2015). Robust hand motion tracking through data fusion of 5dt data glove and nimble vr kinect camera measurements. *Sensors*, 15(12):31644–31671.

Bansal, M., Saxena, S., Desale, D., and Jadhav, D. (2011). Dynamic gesture recognition using hidden markov model in static background. *International Journal of Computer Science Issues (IJCSI)*, 8(1).

Battaglia, E., Bianchi, M., Altobelli, A., Grioli, G., Catalano, M. G., Serio, A., Santello, M., and Bicchi, A. (2016). Thimblesense: a fingertip-wearable tactile sensor for grasp analysis. *IEEE Transactions on Haptics*, 9(1):121–133.

Bezdek, J. C., Ehrlich, R., and Full, W. (1984). Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203.

Devyatkov, V. and Alfimtsev, A. (2007). Recognition of manipulative gestures. *Bulletin of the Bauman Moscow State Technical University. Series of «Instrument making». In Russian.*, (3):56–74.

Dipietro, L., Sabatini, A. M., and Dario, P. (2008). A survey of glove-based systems and their applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(4):461–482.

Dominio, F., Donadeo, M., and Zanuttigh, P. (2014). Combining multiple depth-based descriptors for hand gesture recognition. *Pattern Recognition Letters*, 50:101–111.

Gawande, S. D. and Chopde, N. R. (2013). Neural network based hand gesture recognition. *International Journal of Emerging Research in Management and Technology*, 3:2278–9359.

Gudhmundsson, S. A., Sveinsson, J. R., Pardas, M., Aanaes, H., and Larsen, R. (2010). Model-based hand gesture tracking in tof image sequences. In *International Conference on Articulated Motion and Deformable Objects (AMDO)*, pages 118–127. Springer.

Kenn, H., Van Megen, F., and Sugar, R. (2007). A glove-based gesture interface for wearable computing applications. In *4th International Forum on Applied Wearable Computing (IFAWC)*, pages 1–10. VDE.

Kumar, P., Verma, J., and Prasad, S. (2012). Hand data glove: A wearable real-time device for human-computer interaction. *International Journal of Advanced Science and Technology*, 43:15–26.

Li, X. (2003). Gesture recognition based on fuzzy c-means clustering algorithm. *Department Of Computer Science The University Of Tennessee Knoxville*.

Luzhnica, G., Simon, J., Lex, E., and Pammer, V. (2016). A sliding window approach to natural hand gesture recognition using a custom data glove. In *IEEE Symposium on 3D User Interfaces (3DUI)*, pages 81–90.

Manresa, C., Varona, J., Mas, R., and Perales, F. J. (2005). Hand tracking and gesture recognition for human-computer interaction. *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, 5(3):96–104.

Mavridis, N., Giakoumidis, N., and Machado, E. L. (2012). A novel evaluation framework for teleoperation and a case study on natural human-arm-imitation through motion capture. *International Journal of Social Robotics*, 4(1):5–18.

Nayak, J., Naik, B., and Behera, H. (2015). Fuzzy c-means (fcm) clustering algorithm: a decade review from 2000 to 2014. In *Computational intelligence in data mining-volume 2*, pages 133–149. Springer.

Park, Y., Lee, J., and Bae, J. (2015). Development of a wearable sensing glove for measuring the motion of fingers using linear potentiometers and flexible wires. *IEEE Trans. on Industrial Informatics*, 11(1):198–206.

Rautaray, S. S. and Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54.

Ren, Z., Yuan, J., Meng, J., and Zhang, Z. (2013). Robust part-based hand gesture recognition using kinect sensor. *IEEE Trans. on Multimedia*, 15(5):1110–1120.

Suarez, J. and Murphy, R. R. (2012). Hand gesture recognition with depth images: A review. In *21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 411–417. IEEE.

Suk, H.-I., Sin, B.-K., and Lee, S.-W. (2010). Hand gesture recognition based on dynamic bayesian network framework. *Pattern recognition*, 43(9):3059–3072.

Suryanarayan, P., Subramanian, A., and Mandalapu, D. (2010). Dynamic hand pose recognition using depth data. In *Pattern Recognition (ICPR), 20th International Conference on*, pages 3105–3108. IEEE.

Wachs, J., Kartoun, U., Stern, H., and Edan, Y. (2002). Real-time hand gesture telerobotic system using fuzzy c-means clustering. In *Automation Congress, 2002 Proceedings of the 5th Biannual World*, volume 13, pages 403–409. IEEE.

Wachs, J. P., Kölsch, M., Stern, H., and Edan, Y. (2011). Vision-based hand-gesture applications. *Communications of the ACM*, 54(2):60–71.

Zabulis, X., Baltzakis, H., and Argyros, A. (2009). Vision-based hand gesture recognition for human-computer interaction. 34:30.