2018 13th IEEE International Conference on Automatic Face & Gesture Recognition

Biomechanical-based Approach to Data Augmentation for One-Shot Gesture Recognition

Maria E. Cabrera and Juan P. Wachs
School of Industrial Engineering
Purdue University, West Lafayette, IN, USA

Abstract— **Most common approaches to one-shot gesture recognition have leveraged mainly conventional machine learning solutions and image based data augmentation techniques, ignoring the mechanisms that are used by humans to perceive and execute gestures, a key contextual component in this process. The novelty of this work consists on modeling the process that leads to the creation of gestures, rather than observing the gesture alone. In this approach, the context considered involves the way in which humans produce the gestures the kinematic and biomechanical characteristics associated with gesture production and execution. By understanding the main "modes" of variation we can replicate the single observation many times. Consequently, the main strategy proposed in this paper includes generating a data set of human-like examples based on "naturalistic" features extracted from a single gesture sample while preserving fundamentally human characteristics like visual saliency, smooth transitions and economy of motion. The availability of a large data set of realistic samples allows the use state-of-the-art classifiers for further recognition. Several classifiers were trained and their recognition accuracies were assessed and compared to previous one-shot learning approaches. An average recognition accuracy of 95% among all classifiers highlights the relevance of keeping the human "in the loop" to effectively achieve one-shot gesture recognition.**

I. INTRODUCTION

Gestures have been a subject for multidisciplinary research including linguistics, cognitive sciences, computer science and engineering given their relevance to human-human and human-machine interaction. Researchers have studied how gestures are produced, perceived and mimicked, as well as how computer systems can detect and recognize them. However, the overlap between these areas of research has been rather narrow. For example, in gesture recognition, a vast number of approaches leverage machine learning and vision-based techniques [1], but there is little consideration to the cognitive process, perception, or how gesture production aspects play a role in that recognition.

Some of the major challenges regarding gesture recognition lie on representation and robust learning. Recognizing gestures has the intrinsic difficulty of grouping common traits within a gesture class considering their high variability due to human nature, and conversely the ability to perceive key changes that make two gesture classes different [2].

Most common models used to discern between gesture classes, include empirical parameters and high-level descriptors to encompass the complexity and diversity among gestures. These parameters can be fine-tuned with relative ease when multiple examples of the same gesture class are available; this is known as N-shot learning [3]. As the number of samples dramatically decreases to one, entering the one-shot learning domain, traditional approaches do not perform well. This difficulty stems not only from lack of training data but also because the bulk of machine learning algorithms are focused on N-shot problems, where N is often large [4]. The focus of this paper is to propose an approach leveraging state-of-the-art machine learning classification methods [5]–[7], typically used for N-shot learning, by virtually generating an entire set of naturalistic training data from a single example. Naturalistic in the sense that the data produced shares human like characteristics such as kinematic or biomechanical aspects.

Extracting a relevant representation from a single example brings forth challenges. As fewer gesture examples become available, a biomechanical characterization of such gestures could be leveraged for context. We propose employing a model based on physiological constraints and quantifications of human variability to suggest how real humans might replicate a previously seen gesture.

Using this model repetitively leads to the creation of an artificial gesture data set. This data generation process involves the solution of the inverse kinematic (IK) problem for the human arm [8]. The set of IK solutions are concatenated sequentially while constraining the overall hand trajectory to minimum jerk, smooth changes at joint level, and minimum energy expenditure. Such combination of constraints reflects realistically gestures that are comfortable and have low muscular strain. This technique is referred throughout the paper as the Backward approach, as opposed to an existing technique called the Forward approach [9].

The contribution of this work is the development of the artificial data process used for one-shot gesture learning. This approach includes physical aspects of gesture generation, from kinematics and biomechanical constraints.

II. BACKGROUND

Most research published and surveyed in the area of gesture recognition, are in the field of human-computer interaction [10], human-robot interaction [11] and assistive technologies [12], [13]. The methods discussed on the general field of gesture interaction, can be placed in a continuum scale according to the conceptual approach used to tackle the problem of gesture recognition (see Fig. 1).

On one end of the continuum, there are methods that are outcome-driven towards gesture classification (focused only

978-1-5386-2335-0/18/$31.00 ©2018 IEEE
DOI 10.1109/FG.2018.00016

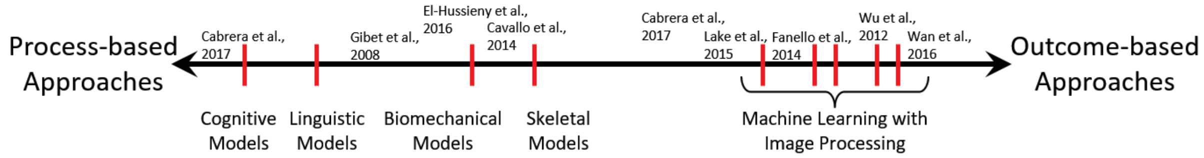


Fig. 1. Continuum for Gesture Production and Recognition

on the data); and on the other, the process-driven methods that rely in the perception and execution processes leading to the gestures. Within this continuum most works are on the outcome-driven side.

Similarly, previous one-shot gesture recognition approaches are mostly situated towards the outcome-driven end of the continuum in Fig. 1. The majority of those methods have relied on computer vision techniques to use a variety of descriptors (e.g. motion and orientation descriptors) in order to train and further classify gestures using a single training instance [14]–[16]. Wan et al. [17] report data augmentation by synthesizing artificial versions of the acquired sample using various temporal scales, and adapting their models to cope with gestures performed at varying speeds.

The limitation of these techniques is that they are not generalizable to new gestures and the overall recognition accuracies are below minimal performance requirements to be used in practice. This calls for new methods that should focus on the process of gesture generation, rather than only the outcome. This process shares most likely components of cognition, physiology and biomechanics. Cabrera and Wachs [9] reported a human-centered approach for data augmentation involving anthropometric features to enlarge the gesture data set artificially. This Forward approach lacked consideration of biomechanical components in their gesture generation process.

The work presented by Lake et al. [18] shows similarities to our proposed method by incorporating human-centered learning to their approach towards model-learning; given a single example of a drawn character, they use generative models using learned primitives to classify, parse and generate similar drawings for the same character. Their work is based on a Bayesian program learning framework for human-level learning that allows previous experience with related concepts to ease learning of new concept in terms of compositionality, causality and learning. Rezende et al. [19] proposed deep generative models capable of one-shot generalization, performing experiments on the MNIST [20] and PIE [21] data set. However, their approaches require large amounts of data to avoid overfitting. Some efforts have been reported to fine-tune fully trained deep learning models [22], [23] using an external memory representation and adapt them to an unseen class based on a single example. These approaches were applied to image data sets.

Jang et al. [24]discuss a visual analytics system supporting identification and characterization of gesture patterns from motion tracking data. Kinematic based techniques often rely on motion capture systems. Such approaches rely on wearable markers and accelerometers which are then detected and further tracked through special hardware and software. Ruiz et al. used such a setting towards the study and design of biomechanical models to describe five different kick gestures used in Taekwondo [25]. Performance metrics related to the execution of the motions were obtained by systematic analysis of the gathered data. A different approach to motion tracking is done using infrared and RGB-D sensors, like Microsoft Kinect. A feasibility study to detect upper limb behavior using Microsoft Kinect 2 [26] showed the collected data to be of sufficient quality to perform objective motor behavior in individuals to classify different levels of upper limb impairment. Cavallo et al. [27] used a biomechanical model to assess surgical performance in Minimal Access Surgery using acceleration, jerk, and energy expenditure.

Jerk minimization and energy expenditure are methods commonly used to model and plan trajectories performed by human arm and robotic manipulators [28], [29]. For example, Yazdani [30] predicted human motion based on minimum jerk. However, their work was limited to a planar motion at shoulder level. Similarly, Zhou et al. [31] restricted the shoulder movements to achieve planar arm motion and effectively determined arm trajectories. They used low metabolic costs with biomechanical models for the arm trajectories. However, such designs reduce the complexity of the inverse kinematics and musculoskeletal representation of the human arm.

Gibet et al. [32] used biomechanical models for gesture production. They analyzed gestures from a motor control perspective finding some commonalities in gesture execution related to invariance in velocity profiles and minimum jerk. This supports the idea that the arms and hands follow smooth trajectories during motion.

In the context of cognition-based approaches, Cabrera et al. [33] found a relationship between the timing of mu oscillations and kinematic inflection points. The fact that positive correlations have been observed between gesture performance and spikes in electroencephalographic (EEG) signals above the motor cortex, supports the hypothesis that inflection points in gesture performance are associated with distinct neural responses.

There is still a lack of research merging the cognitive and physiological aspects of gesture perception and production with the computational aspects of recognition that can be applied to one-shot gesture recognition effectively. The aim of this work is to close the gap in the continuum presented in Fig. 1 by leveraging biomechanical features as input for more conventional outcome-driven approaches.

III. METHODOLOGY

This section presents the concepts and implementation details adopted to achieve one-shot gesture recognition using the biomechanical principles of human gesture production. Our approach consists of the following main steps: (1) collecting a single gesture observation, expressed in terms of skeleton data; (2) artificially augmenting the single example per class given, to produce a large set of human-like gesture instances; (3) training state-of-the-art classifiers to recognize the class of a new gesture instance, and (4) assessing performance in terms of accuracy. The gestures considered are motions performed by peoples upper limbs and acquired using a Microsoft Kinect 2 sensor. Implementation details describe the training and testing of multiple classifiers on a publicly available data set, to compare their recognition accuracy with the state-of-the-art.

A. Overview of One-Shot Learning Method

The one-shot learning method presented here is an alternative approach than the one explained in [9] referred as the Forward approach. The Forward approach is based on a reduced anthropometry model (users shoulder and hand position) applied to one gesture sample to generate new artificial gesture instances. The instances are extracted from Gaussian Mixture Models around inflection points. Conversely, the approach presented in this paper (referred as the Backward approach) is mainly based on inverse kinematics solutions on the gesture trajectorys inflection points.

Let $\mathscr{L}$ describe a lexicon formed by N gesture classes $\mathscr{G}_i$, $\mathscr{L} = \{\mathscr{G}_1, ..., \mathscr{G}_i, ..., \mathscr{G}_N\}$. Each gesture class contains gesture instances g^i_j, with j number of instances. Each gesture instance is a concatenation of trajectory points in three dimensions (3D) extracted from the centroid of the hand, $g^i_k = \{(x_1, y_1, z_1), ..., (x_H, y_H, z_H)\}$, where H is the total number of points within that gesture instance. Using one instance per class g^i_1, a set of inflection points $\mathbf{x_q} = (x_q, y_q, z_q)$ are extracted, where $q = 1, ..., Q$ and $Q < H$.

The overview of the proposed approach is summarized in Fig. 2. Using a model to describe the kinematics and biomechanics of the human arm, inverse kinematics solutions are found at each inflection point $\mathbf{x_q}$. These sets of IK solutions at each inflection point are combined, and smooth trajectories with low strain are generated on the joint space. These trajectories are concatenated and stored as new gesture instances $\hat{g}^i_j$. Once an entire data set of artificially generated instances is created, state-of-the-art classifiers are trained. This work proposes a methodology for one-shot gesture learning that can be used with any classification technique and for this reason multiple classifiers are considered. Next, we will describe in detail each of the steps in Fig. 2

B. Backward Approach: Artificial Trajectory Generation leveraging biomechanics of the human arm

Given one gesture example g^i_1 from class i, where $i = 1, ..., N$ key points ($\mathbf{x_q}$) in the gesture trajectory are extracted, with $q = 1, ..., Q$. These inflection points can be found by computing the derivative in the gesture trajectory. Inflection points are associated with abrupt changes in speed and orientation of a gesture, signaling the concatenation of different gesture phases [34]. In this work, the different gesture phases are considered the trajectories between each two inflection points.

First, IK solutions are calculated so the human arm can reach the positions $\mathbf{x_q}$. This modeling follows the fundamentals of serial-link robot kinematics and Denavit-Hartemberg (D-H) notation [35], [36]. The MATLAB Robotic Toolbox [37] and notation for the D-H parameters [38] were used to determine the kinematic model of the human arm (Fig. 3 and Table I). The last column in Table I represents the absolute joint constraint (AJC_i) per degree of freedom.

TABLE I

DENAVIT-HARTEMBERG (D-H) PARAMETERS FOR HUMAN ARM MODEL

k	θ (rad)	d	a	α (rad)	AJC_i (degrees)
1	q_1	0	0	$-\pi/2$	-45 to 180
2	$q_2+\pi/2$	0	0	$-\pi/2$	-45 to 130
3	$q_3+\pi/2$	L_1	0	$\pi/2$	-60 to 180
4	q_4	0	0	$-\pi/2$	0 to 150
5	$q_5+\pi$	L_2	0	$\pi/2$	-70 to 85
6	$q_6+\pi/2$	0	0	$-\pi/2$	-20 to 40
7	$q_7+\pi$	0	0	$\pi/2$	-90 to 90

Let $\mathcal{S}^i_q$ be the set of possible IK solutions s^v_q for each inflection point, where $v = 1, ...V_q$. The different s^v_q were obtained by varying the configuration parameters for the toolbox function, such as number of iterations, error tolerance, variable step size and step size gain. Once all the sets $\mathcal{S}^i_q$ were determined, a recursive function (Algorithm 1) combined all solutions in the sets $\mathcal{S}^i_q$ exhaustively and generated multiple trajectories for each gesture phase.

The function $\mathcal{F}$, used to generate trajectory phases between inflection points, is explained in the following subsection. Two different strategies were considered, namely minimum jerk and energy expenditure. Each of these two strategies are used between two inflection points ($\mathbf{x_q}$ and $\mathbf{x_{q+1}}$) given IK solutions for both (s^v_q and s^u_{q+1}). The output of Algorithm 1 is a list of artificially generated gesture instances $\mathbf{\hat{g}^i_j}$.

The complexity of this algorithm is $\mathcal{O}(Q, V)$ based on asymptotic order analysis, where $V = \max(|\mathcal{S}^i_q|)$ for all i and q. One of the advantages of this algorithm is that it has tail recursion which makes it easily converted to a distributed algorithm. The maximum number of generated trajectories for class i is determined by $\prod_1^Q |\mathcal{S}^i_q|$.

C. Strategies for gesture phase generation

Using the IK solutions for two consecutive inflection points in the gesture example g^i_1, two different strategies for trajectory generation are considered: minimum jerk (J) and minimum energy expenditure (EE) as described earlier. The minimum jerk strategy allows to generate smooth trajectories with small progressive changes in the joints, consistent with motion observed in humans. The minimum energy expenditure strategy is associated with trajectories which are comfortable and have low muscular strain.

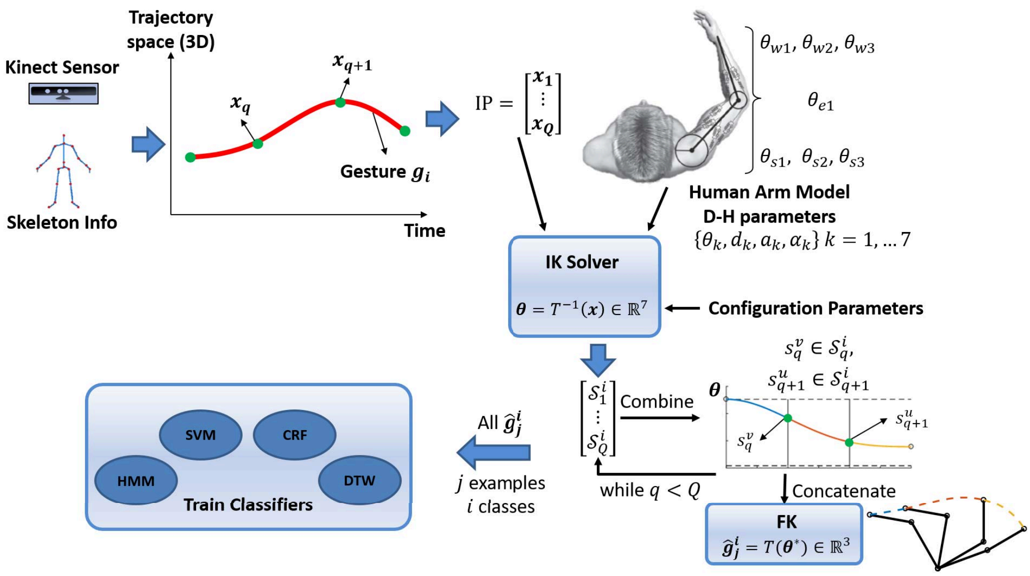


Fig. 2. Overview of the proposed methodology

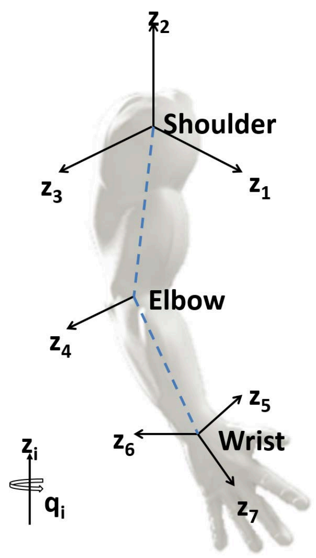


Fig. 3. Kinematic Model of the human arm with reference frames associated to degrees of freedom for each joint

Algorithm 1 Recursive artificial example generation

Input: index of inflection point q, IK solution s_q^v, and array of concatenated gesture phases (3D trajectories) $GestPhase$

$all\_Gestures = [\ ]$ ▷ All artificial gestures are stored here
gen_GestPhase (1 , [ ] , [ ])
function GEN_GESTPHASE(q , $IKsol$, $GestPhase$)
  if $q < Q$ **then**
    for all s_q^v with $v \leftarrow 1, \ldots, V_q$ **do**
      if $not\_empty(IKsol)$ **then**
        gen_GestPhase ($q+1$, s_q^v , $\mathcal{F}(s_q^v, IKsol)$)
      else ▷ This goes through all s_1^v
        gen_GestPhase ($q+1$, s_q^v , [ ])
      end if
    end for
  else ▷ $GestPhase$ is a complete gesture
    $all\_Gestures.append$ ($GestPhase$)
  end if
end function

Output: list of all artificially generated gestures $\hat{\mathbf{g}}_\mathbf{j}^\mathbf{i}$ of a given class in $all\_Gestures$

Minimum jerk (Eq.1) was applied on the joint space, using s_q^v and s_{q+1}^v as the initial and final condition, and considering that initial and final velocity was zero. Additional kinematic constraints were included to maintain the angles for each joint within the human operating range (AJC_k). A fifth-order polynomial function was used for each joint to determine the coefficients achieving minimum jerk. Once all the fitted polynomials were determined for each joint, forward kinematics (FK) were used to determine the 3D trajectories for the hand.

$$
\begin{gathered}
\min J \rightarrow min \sum_{k=1}^{K} \int \dddot{\theta}_k \\
\text{s.t. } \{\theta_1, \ldots, \theta_K\}_{init} = s_q^v, \{\theta_1, \ldots, \theta_K\}_{final} = s_{q+1}^v \\
\frac{d(\{\theta_1, \ldots, \theta_K\}_{init})}{dt} = \frac{d(\{\theta_1, \ldots, \theta_K\}_{final})}{dt} = 0 \\
\theta_k \leq AJC_k
\end{gathered} \tag{1}
$$

Where K is the degrees of freedom. In the case of the human arm, $K = 7$.

The minimum energy expenditure strategy (Eq. 2) is based on torque calculations for each joint, using Lagrange-Euler (Eq. 3). Only inertial ($M(\theta)$) and gravitational forces ($G(\theta)$) were considered. Inertia constants were obtained from standardized anthropometric data [39].

$$\min EE \rightarrow min \sum_{k=1}^{K} \int |\tau_k \times \dot{\theta}_k| \tag{2}$$

$$\tau_k = M(\theta)\ddot{\theta} + G(\theta) \tag{3}$$

Analogous to the jerk strategy, third-order polynomial functions were used for each joint, initial and final conditions were given by the previously obtained IK solutions. Forward kinematics were used to determine the 3D trajectories for the arm from the estimated joint values.

D. Implementation Details

The proposed approach was implemented and tested on a publicly available data set from Microsoft Research Cambridge (MSRC-12) [40]. This data set consisted of sequences of human movements, representing 12 different iconic and metaphoric gestures related to gaming commands and interaction with a media player. The number of gesture classes in the lexicon was reduced to 8 by excluding gestures that involved leg motion or movement of the whole upper body (i.e. bow).

From this data set, 100 gesture instances from each class were used as the testing set, for a total of 800 gesture motions. Eight additional instances, one for each gesture class, were used to extract the inflection points representing the gesture class, and from those 200 artificial observations per class were created. This artificial data set was used as training data. Examples of the gestures in this lexicon are shown in Fig. 4.

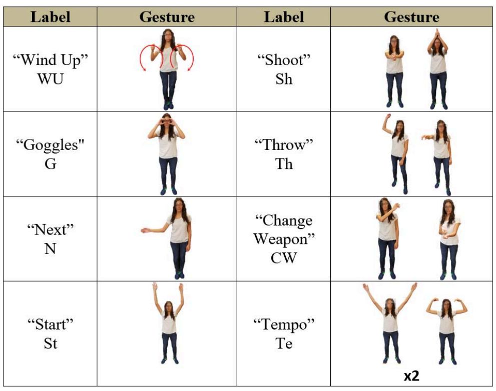


Fig. 4. Microsoft Research Data Set MSRC-12. Selected 8 gesture subset

Four different classification algorithms were trained using the artificially generated data set and their performances compared in terms of recognition accuracy ($Acc_{\%}$).

Recognition accuracy was determined through response operating curves (ROC). This was done by selecting a free parameter, in this case related to the confidence for the predicted class, and using different thresholds to obtain multiple pairs of true hit and false alarm rates.

The selected classification algorithms are commonly used in state-of-the-art gesture recognition approaches, namely: Hidden Markov Models (HMM), Support Vector Machines (SVM), Conditional Random Fields (CRF) and Dynamic Time Warping (DTW). In the case of HMM and SVM, a one-versus-all scheme was used, while CRF and DTW used a multi-class scheme assigning the label with the highest likelihood.

The feature vector considered as input for the classifiers contained the rate of change in position between two consecutive points, both in magnitude and orientation for both hands. Orientation was described as three different angles with respect to each axis.

IV. RESULTS

This section presents the results in terms of recognition accuracy for all classification methods mentioned previously. Using the implemented backward approach, a data set of human-like examples was generated for training, originating from a single example for each selected class of the MSRC-12 data set.

ROC for all classifiers are shown in Fig. 5. Five different thresholds were used for all classifiers. Overall accuracies were obtained by calculating the area under the curve (AUC). Overall accuracies are summarized in Table II.

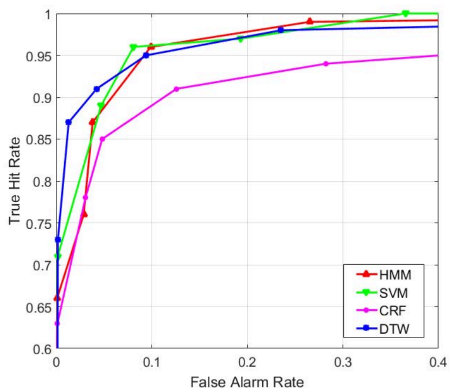


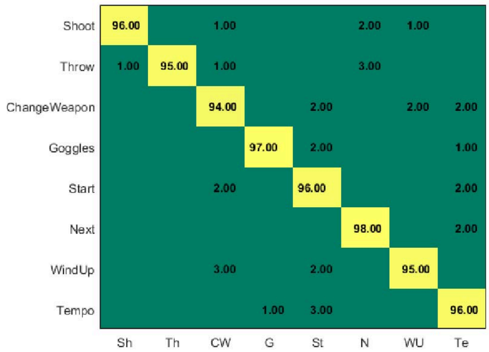
Fig. 5. ROC curves for all four classifiers: HMM, SVM, CRF, DTW

TABLE II

OVERALL ACCURACIES FOR ALL CLASSIFIERS

MSRC-12	HMM	SVM	CRF	DTW
AUC	96.6%	97.18%	91.51%	96.43%

The free parameter used to generate the ROCs was adjusted to the shortest Euclidian distance to the $(1,0)$ point for each classifier. Trained classifiers were tested on gesture examples from the MSRC-12 data set. The results are shown using confusion matrices for all classifiers in Fig. 6.



(a) HMM. Accuracy: 95.88%

(b) SVM. Accuracy: 96.13%

(c) CRF. Accuracy: 92.12%

(d) DTW. Accuracy: 96.25%

Fig. 6. Confusion matrices for all trained classifiers: a) HMM, b) SVM, c) CRF, d) DTW

Previous results reported by Ellis et al. [41] reached 88.7% accuracy, while [42] achieved 91.82%. The previously mentioned forward approach reported in [9] achieved an average of 89.2% among the same classifiers.

Artificial generation processes were compared using a k-fold cross-validation with $k = 10$. The results are shown in Fig. 7. A t-test between $ACC_{\%}$ for each classification method was conducted, with the approach used as the independent variable. All classifiers trained with the Backward approach had significantly higher recognition accuracy than the classifiers trained with the Forward approach. These results indicate a superior performance of the Backward over the Forward approach for every classifier used.

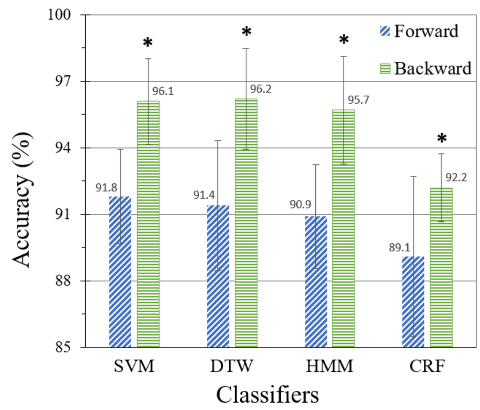


Fig. 7. Recognition accuracies for all implemented classifiers using two different approaches for artificial gesture generation. Backward approach was significantly higher in all classifiers

V. CONCLUSION

This paper presents a novel approach to achieve one-shot gesture recognition. This approach is based on kinematic and biomechanical characteristics associated with gesture production. Using a compact representation from a single example per class, an augmented data set of human-like samples is generated and used to train classifiers. Four classifiers commonly used in state-of-the-art gesture recognition were trained to evaluate the effect of the approach on the overall performance. Recognition accuracy was measured for each classifier using a subset of the publicly available data set of Microsoft Research MSRC-12. The recognition accuracy of all classifiers showed a significant improvement over previous approaches reported in literature, with an average recognition of 95%. These results highlight the relevance of including the biomechanical aspects of human motion as context to achieve one-shot gesture recognition effectively.

VI. ACKNOWLEDGMENTS

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award No. W81XWH-14-1-0042. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the funders.

REFERENCES

[1] K. Liu and N. Kehtarnavaz, "Real-time robust vision-based hand gesture recognition using stereo images," *Journal of Real-Time Image Processing*, vol. 11, no. 1, pp. 201–209, 2016.

[2] H. J. Caulfield and K. Heidary, "Exploring margin setting for good generalization in multiple class discrimination," *Pattern Recognition*, vol. 38, pp. 1225–1238, Aug. 2005.

[3] C. Jost, P. D. Loor, L. Ndlec, E. Bevacqua, and I. Stankovi, "Real-time gesture recognition based on motion quality analysis," in *2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*, pp. 47–56, June 2015.

[4] R. B. Mapari and G. Kharat, "Real time human pose recognition using leap motion sensor," in *2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, pp. 323–328, Nov. 2015.

[5] M. Elmezain, A. Al-Hamadi, S. Sadek, and B. Michaelis, "Robust methods for hand gesture spotting and recognition using hidden markov models and conditional random fields," in *The 10th IEEE ISSPIT*, pp. 131–136, IEEE, 2010.

[6] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, pp. 3592–3607, 2011.

[7] J. Y. Chang, "Nonparametric Feature Matching Based Conditional Random Fields for Gesture Recognition from Multi-Modal Video," *IEEE TPAMI*, vol. 38, pp. 1612–1625, Aug. 2016.

[8] Y. Wang and P. Artemiadis, *Closed-form inverse kinematic solution for anthropomorphic motion in redundant robot arms*. PhD thesis, Arizona State University, 2013.

[9] M. E. Cabrera and J. Wachs, "A Human-Centered approach to One-Shot Gesture Learning," *Frontiers in Robotics and AI*, vol. 4, 2017.

[10] A. Chaudhary, J. L. Raheja, K. Das, and S. Raheja, "Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey," *arXiv preprint arXiv:1303.2292*, 2013.

[11] D. Xu, X. Wu, Y.-L. Chen, and Y. Xu, "Online Dynamic Gesture Recognition for Human Robot Interaction," *Journal of Intelligent & Robotic Systems*, vol. 77, pp. 583–596, Apr. 2014.

[12] R.-H. Liang and M. Ouhyoung, "A real-time continuous gesture recognition system for sign language," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pp. 558–567, IEEE, 1998.

[13] H. Jiang, B. S. Duerstock, and J. P. Wachs, "User-Centered and Analytic-Based Approaches to Generate Usable Gestures for Individuals With Quadriplegia," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 3, pp. 460–466, 2016.

[14] D. Wu, F. Zhu, and L. Shao, "One shot learning gesture recognition from RGBD images," in *2012 IEEE Computer Society Conference on CVPRW*, pp. 7–12, June 2012.

[15] S. R. Fanello, I. Gori, G. Metta, and F. Odone, "Keep it simple and sparse: Real-time action recognition," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 2617–2640, 2013.

[16] J. Konen and M. Hagara, "One-shot-learning gesture recognition using hog-hof features," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 2513–2532, 2014.

[17] J. Wan, G. Guo, and S. Z. Li, "Explore Efficient Local Features from RGB-D Data for One-Shot Learning Gesture Recognition," *IEEE TPAMI*, vol. 38, pp. 1626–1639, Aug. 2016.

[18] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015.

[19] D. Rezende, I. Danihelka, K. Gregor, D. Wierstra, and others, "One-shot generalization in deep generative models," in *International Conference on Machine Learning*, pp. 1521–1529, 2016.

[20] L. Deng, "The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]," *IEEE Signal Processing Magazine*, vol. 29, pp. 141–142, Nov. 2012.

[21] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pp. 53–58, IEEE, 2002.

[22] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, and others, "Matching networks for one shot learning," in *Advances in Neural Information Processing Systems*, pp. 3630–3638, 2016.

[23] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "One-shot Learning with Memory-Augmented Neural Networks," *arXiv:1605.06065 [cs]*, May 2016. arXiv: 1605.06065.

[24] S. Jang, N. Elmqvist, and K. Ramani, "GestureAnalyzer: visual analytics for pattern analysis of mid-air hand gestures," in *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pp. 30–39, ACM, 2014.

[25] A. T. Ruiz, J. M. Fernandez, and S. A. Jimenez, "Biomechanical Characterization of Five Sporting Gestures of Taekwondo," in *VI Latin American Congress on Biomedical Engineering CLAIB 2014*, (Parana, Argentina), pp. 285–288, Springer, 2015.

[26] B. Dehbandi, A. Barachant, D. Harary, J. Long, K. Z. Tsagaris, S. Bumanlag, V. He, and D. Putrino, "Using data from the Microsoft Kinect 2 to quantify upper limb behavior: a feasibility study," *IEEE journal of biomedical and health informatics*, 2016.

[27] F. Cavallo, G. Megali, S. Sinigaglia, O. Tonet, and P. Dario, "A biomechanical analysis of surgeons gesture in a laparoscopic virtual scenario," *Stud Health Technol Inform MMVR*, vol. 14, pp. 79–84, 2006.

[28] C.-H. Chiu, J.-C. Wang, and C.-F. Chen, "Minimum energy expenditure of arm and leg motions," *International Journal of Sport and Exercise Science*, vol. 1, no. 2, pp. 33–40, 2009.

[29] D. A. Kistemaker, J. D. Wong, and P. L. Gribble, "The cost of moving optimally: kinematic path selection," *Journal of neurophysiology*, vol. 112, no. 8, pp. 1815–1824, 2014.

[30] M. Yazdani, G. Gamble, G. Henderson, and R. Hecht-Nielsen, "A simple control policy for achieving minimum jerk trajectories," *Neural Networks*, vol. 27, pp. 74–80, 2012.

[31] L. Zhou, S. Bai, M. R. Hansen, and J. Rasmussen, "Modeling of human arm energy expenditure for predicting energy optimal trajectories," *Modeling, Identification and Control*, vol. 32, no. 3, p. 91, 2011.

[32] S. Gibet and P.-F. Marteau, "Analysis of human motion, based on the reduction of multidimensional captured dataapplication to hand gesture compression, segmentation and synthesis," in *International Conference on Articulated Motion and Deformable Objects*, pp. 72–81, Springer, 2008.

[33] M. Cabrera, K. Novak, D. Foti, R. Voyles, and J. Wachs, "What makes a gesture a gesture? Neural signatures involved in gesture recognition," *12th IEEE International Conference on Automatic Face and Gesture Recognition*, May 2017.

[34] S. Kita, I. v. Gijn, and H. v. d. Hulst, "Movement phases in signs and co-speech gestures, and their transcription by human coders," in *Gesture and Sign Language in Human-Computer Interaction* (I. Wachsmuth and M. Frhlich, eds.), no. 1371, pp. 23–35, 1998.

[35] R. P. Paul, *Robot manipulators: mathematics, programming, and control: the computer control of robot manipulators*. Richard Paul, 1981.

[36] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot modeling and control*, vol. 3. Wiley New York, 2006.

[37] P. I. Corke, "A robotics toolbox for MATLAB," *IEEE Robotics & Automation Magazine*, vol. 3, no. 1, pp. 24–32, 1996.

[38] P. I. Corke, "A Simple and Systematic Approach to Assigning Denavit ndash;Hartenberg Parameters," *IEEE Transactions on Robotics*, vol. 23, pp. 590–594, June 2007.

[39] R. Drillis, R. Contini, and M. Bluestein, *Body segment parameters*. New York University, School of Engineering and Science Research Division, NY, 1966.

[40] S. Fothergill, H. Mentis, P. Kohli, and S. Nowozin, "Instructing people for training gestural interactive systems," in *Proceedings of the SIGCHI Conference*, pp. 1737–1746, ACM, 2012.

[41] C. Ellis, S. Z. Masood, M. F. Tappen, J. J. Laviola Jr, and R. Sukthankar, "Exploring the trade-off between accuracy and observational latency in action recognition," *International Journal of Computer Vision*, vol. 101, no. 3, pp. 420–436, 2013.

[42] M. Ramrez-Corona, M. Osorio-Ramos, and E. F. Morales, "A non-temporal approach for gesture recognition using microsoft kinect," in *Iberoamerican Congress on Pattern Recognition*, pp. 318–325, Springer, 2013.