NRC Publications Archive Archives des publications du CNRC

Feature selection in Haptic-based handwritten signatures using rough sets

Sakr, Nizar; Alsulaiman, Fawaz A.; Valdes, Julio J.; El Saddik, Abdulmotaieb; Georganas, Nicolas D.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:

<https://doi.org/10.1109/FUZZY.2010.5584258>

2010 IEEE International Conference on Fuzzy Systems (FUZZ), 2010-07-23

NRC Publications Record / Notice d'Archives des publications de CNRC:

<https://nrc-publications.canada.ca/eng/view/object/?id=1f02918c-429c-4863-96f4-c4faeaf7ab30>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=1f02918c-429c-4863-96f4-c4faeaf7ab30>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

Feature Selection in Haptic-based Handwritten Signatures Using Rough Sets

Nizar Sakr, Fawaz A. Alsulaiman, Julio J. Valdés, Abdulmotaleb El Saddik, Nicolas D. Georganas

Abstract— This paper explores the use of rough set theory for feature selection in high dimensional haptic-based handwritten signatures (exploited for user identification). Two rough set-based methods for feature selection are analyzed, the first is a greedy approach while the second relies on genetic algorithms to find minimal subsets of attributes. Also, to further reduce the haptic feature space while maximizing user identification accuracy, a method is proposed where feature vectors are subsampled prior to the feature selection procedure. Rough set-generated minimal subsets are initially exploited to determine the importance of different haptic data types (e.g. force, position, torque and orientation) in discriminating between different users. In addition, a comparison between rough set-based methods and classical machine learning techniques in the selection of minimal information-preserving subsets of features in high dimensional haptic datasets, is provided. The criteria for comparison are the length of the selected subsets of features and their corresponding discrimination power. Support Vector Machine classifiers are used to evaluate the accuracy of the selected minimal feature vectors. The results demonstrated that the combination of rough set and genetic algorithm techniques can outperform well-established machine learning methods in the selection of minimal subsets of features present in haptic-based handwritten signatures.

I. INTRODUCTION

Haptics is a fast emerging technology that permits touch-enabled interaction with virtual environments. Its applications are wide, and range from surgical simulation to gaming and entertainment. Many current and future applications of haptic technologies in rehabilitation, security, training and simulation, medicine and entertainment, involve the analysis and interpretation of recorded haptic information in order to possibly reveal patterns and relationships in the data. However, regardless of the intended haptic application, the number of resulting features are significantly large (in the thousands range) as the recorded information typically consists of multidimensional time-varying data, e.g. force, torque, position, velocity, orientation, etc. Consequently, when dealing with such high-dimensional datasets, reduction of the number of acquired features through *feature selection* is highly desirable.

Feature selection refers to the process of identifying an optimum subset of features from a much larger set of potentially useful features [1]. Reducing the size of a large dataset offers numerous well-known advantages. These

benefits are especially apparent in a knowledge discovery system in which a classification model is built to distinguish between possible patterns in the data. With this in mind, probably the most important advantage is the fact that feature reduction can help alleviate the “curse of dimensionality” problem. Furthermore, excluding noisy, redundant and irrelevant features could potentially improve the accuracy and generalization properties of the classification model. Also, a smaller subset of features reduces the space complexity of the classification problem and consequently decreasing the computational overhead, and speeding up the learning process.

The dataset in this study consists of a collection of haptic-based handwritten signatures acquired using a haptic-enabled biometric application developed at DISCOVER laboratory of the University of Ottawa. Biometric systems provide a solution to ensure that protected services are solely accessible by a legitimate user. This is achieved while relying on users’ behavioral and/or physiological characteristics. In recent years, the possible use of haptic devices in biometric systems has been suggested to enable improved user identification/verification performance over more traditional techniques, such as those based on handwritten signatures.

Feature selection in high dimensional haptic data has been little explored in the literature. In [1], [2], Orozco et al. relied on the same dataset exploited in this paper in order to demonstrate the feasibility of a haptic based user authentication system. Feature selection is performed while relying on a relative entropy measure, which the author interpret as a pseudo-metric that defines the distance between an individual’s features and those of the entire population of users. The authors however, distribute the high dimensional attributes of each signature across different instances, i.e. position, velocity, orientation, torque and force data acquired at time t_1 are assigned to instance $Inst_1$, data acquired at time t_2 are assigned to instance $Inst_2$, etc., yielding a number of instances per user signature with only few attributes per instance. It is evident however that a more logical and adequate approach would be to assign all the generated haptic data attributes per signature to a single instance, i.e. each instance contains the entire (haptic-based) signature for a single user. This approach can lead to better analysis and interpretation of the haptic dataset, and improve discrimination between users. This however comes at the expense of having to deal with instances that possibly consist of thousands of attributes. In one of the authors’ recent papers [4] haptic feature vectors were redefined using the approach suggested above (all haptic data attributes associated with a single signature are

N. Sakr, F. A. Alsulaiman, A. El Saddik and N.D. Georganas are with the School of Information Technology and Engineering, University of Ottawa, Canada (E-mail: nsakr@site.uottawa.ca, fawaz@mcrilab.uottawa.ca).

J. J. Valdés is with the National Research Council Canada, Institute for Information Technology, Ottawa, Canada (E-mail: Julio.Valdes@nrc-nrc.gc.ca).

assigned to a single instance). The paper explored the use of genetic programming in finding analytic functions that can be exploited in discovering minimal knowledge-preserving subsets of features from high dimensional haptic datasets. In [5], a first effort was made to analyze the use of rough set-based feature selection techniques in the analysis of haptic datasets. Feature selection was however performed strictly on full datasets (rather than on training datasets) to investigate its potential use as a preprocessing step for feature generation techniques in visual data mining of haptic signatures.

This paper builds on the work presented in [5] and explores the use of rough set theory for feature selection in haptic-based signature identification systems (i.e. feature selection is strictly performed on training datasets). Two rough set-based methods for feature selection are analyzed, the first is a greedy approach while the second relies on genetic algorithms to find minimal subsets of features. Other contributions of this paper include: (1) to further reduce the haptic feature space while maximizing discrimination between users, a method is proposed where feature vectors are subsampled prior to the feature selection procedure. (2) A comparison between rough set-based methods and classical machine learning techniques in the selection of minimal information-preserving subsets of features in large haptic datasets, is provided. The criteria for comparison are the length of the selected subsets of features and their corresponding discrimination power. (3) minimal subsets generated using rough set-based approaches are also exploited to determine the importance of different haptic data types (e.g. force, position, torque and orientation) in discriminating between different users.

The rest of the paper is organized as follows. In Section II we introduce the basic concepts of the rough set theory and support vector machine classifiers. In Section III the experimental settings are described. In Section IV the results are provided. Finally, conclusive remarks are outlined in Section V.

II. CONCEPTS AND METHODS

In this section basic concepts of the rough set theory are introduced, followed with a concise description of two rough set-based feature selection algorithms. Finally, a brief introduction to Support Vector Machine (SVM) is provided.

A. Concepts of the Rough Set Theory

The rough set theory originally introduced by Pawlak [7] is a mathematical approach to deal with vagueness and uncertainty in data analysis. The theory is based on the assumption that some information is associated with every object of the universe of discourse. In contrast to crisp sets, rough sets cannot be uniquely defined in terms of their elements. Objects characterized by the same information are considered indiscernible. The *indiscernibility* relation constructed in this manner is the mathematical foundation of rough set theory. The purpose is to approximate a *rough* concept using a pair of crisp sets, determined through the *indiscernibility* relation. The crisp sets are called the lower and upper approximations

of the *rough* concept. The lower approximation consists of all objects that certainly belong to the set and the upper approximation contains all objects that possibly belong to the set. The *boundary region* is defined as the difference between the upper and the lower approximation, and consists of objects that cannot decisively be labeled as members or non-members of the set. A formal definition of the above notations and others required for a better understanding of this paper are given below.

a) *Information System*: an information system is a pair $\mathbf{A} = (U, A)$ where U is a non-empty finite set of objects and A is a non-empty finite set of attributes such that $a : U \rightarrow V_a$ for every $a \in A$, where V_a corresponds to the value set of a . An important subclass of an information system is called a *decision table*. It consists of any information system of the form $\mathbf{A} = (U, A \cup \{d\})$, where $d \in A$ is the decision attribute and the elements of A are the condition attributes.

b) *Discernibility Matrices*: A *discernibility matrix* M_A is defined as an $|U| \times |U|$ matrix of the information system $\mathbf{A}$. Each entry $M_A(x, y) \subseteq A$ consists of the set of attributes that can be exploited to discern between objects $x, y \in U$ as in Eq-1 and 2 .

$$M_A(x, y) = \{a \in A : \text{discerns}(a, x, y)\} \quad (1)$$

$$\text{discerns}(a, x, y) \Leftrightarrow a(x) \neq a(y) \quad (2)$$

where, $\text{discerns}(a, x, y)$ may be tailored to the application of interest.

c) *Indiscernibility Relations*: A discernibility matrix M_A defines a binary relation $R_A \subseteq U^2$. The relation R_A is referred to as an *indiscernibility relation* [8] (See Eq-3) with respect to A , and expresses which pairs of objects cannot be distinguished from each other.

$$xR_A y \Leftrightarrow M_A(x, y) = \emptyset \quad (3)$$

d) *Discernibility Functions*: A discernibility function [8] defines how an object (or a set of objects) can be discerned from a subset of the full universe of objects. Specifically, discernibility functions can be defined relative to a particular object, or constructed such that all objects can be discerned from each other. In the former case, the function is constructed relative to an object $x \in U$ from a discernibility matrix M_A according to Eq-4.

$$f_A(x) = \prod_{y \in U} \left\{ \sum a^* : a \in M_A(x, y) \text{ and } M_A(x, y) \neq \emptyset \right\} \quad (4)$$

The function $f_A(x)$ contains $|A|$ Boolean variables, where variable a^* corresponds to attribute a . Each conjunction of $f_A(x)$ stems from an object $y \in U$ from which x can be discerned and each term within that conjunction represents an attribute that discerns between those objects. The prime implicants of $f_A(x)$ provide the minimal subsets of A that are required to discern object x from the objects in U that are associated with $R_A(x)$. Conversely, in order to discern all objects in U from each other the full discernibility function $g_A(U)$ is defined as $g_A(U) = \prod_{x \in U} f_A(x)$.

e) *Reducts*: If an attribute subset $B \subseteq A$ preserves the indiscernibility relation R_A then the attributes $A \setminus B$ are said to be *dispensable*. An information system can have several such attribute subsets B . All such subsets that do not contain any dispensable attributes (i.e. minimal subsets), are called reducts. However, the problem of computing all reducts (or a minimum-length reduct) is NP-hard, and several heuristics have been proposed [10]. This issue is discussed further in the following section.

B. Reduct Computation

Two different algorithms to compute reducts or reduct approximations are exploited in this paper. Reduct approximations generate attribute subsets that in a sense “almost” preserve the indiscernibility relation. As minimum-length reduct computation (which is of great interest in feature selection) is NP-hard, reduct approximation can therefore be found to be necessary especially when the number of attributes $|A|$ is very high.

- Johnson Algorithm

Johnson’s algorithm is a polynomial-time heuristic approach that invokes a variation of a simple greedy algorithm to compute a single reduct [12], [8]. The algorithm has a natural tendency towards finding a single prime implicant of minimal length. A step by step illustration of the algorithm is depicted below, where B corresponds to the computed reduct, and $\mathcal{S}$ denotes the set of sets corresponding to the discernibility function. Also, $w(S)$ denotes a weight for set S in $\mathcal{S}$ that is automatically computed from the data [8].

- 1) Let $B = \emptyset$
- 2) Let a denote the attribute that maximizes $\sum w(S)$, where the sum is taken over all sets S in $\mathcal{S}$ that contain a .
- 3) Add a to B .
- 4) Remove all sets S from $\mathcal{S}$ that contain a .
- 5) If $\mathcal{S} = \emptyset$ return B . Otherwise, goto step 2.

Computation of approximate reducts is performed by aborting the loop when “enough” sets have been removed from $\mathcal{S}$. This approach is computationally more efficient than exiting the loop only when $\mathcal{S}$ is entirely emptied.

- Genetic Reducer

This method relies on a genetic algorithm to compute minimal hitting sets (reducts), as depicted in [13], [14]. The technique has support for both cost information and approximate solutions. Moreover, the algorithm’s fitness function f is defined as follows:

$$f(B) = (1 - \alpha) \times \frac{\text{cost}(A) - \text{cost}(B)}{\text{cost}(A)} + \alpha \times \min\{\varepsilon, h(B)\}, \quad (5)$$

where $\mathcal{S}$ is the set of sets that correspond to the discernibility function. The parameter α defines a weighting

between subset cost and hitting fraction $h(B)$, while ε controls approximate solutions as it represents a minimal value for the hitting fraction. The hitting fraction $h(B)$ is defined as the ratio of the number of nonempty intersection of the subset B with $\mathcal{S}$ over the multiset’s ($\mathcal{S}$) cardinality as follows: $h(B) = \frac{||S \text{ in } \mathcal{S} | S \cap B \neq \emptyset||}{|\mathcal{S}|}$. The function $\text{cost}(\cdot)$ specifies the cost of an attribute subset. Moreover, subsets B of A are found through the evolutionary search driven by the fitness function, where sets that have a hitting fraction of at least ε , are collected in a “keep list”. The genetic operators crossover, mutation and inversion are used in the algorithm, whereas selection is done by universal stochastic sampling.

C. Support Vector Machine Classification

The aim of Support Vector Machine (SVM) classification [17], [18] is to derive a computationally efficient approach of learning accurate separating hyperplanes in a high dimensional feature space. Without loss of generality, given the set of training vectors belonging to two separate classes,

$$\mathcal{D} = \{(x^1, y^1), \dots, (x^l, y^l)\}, \quad x \in \mathbb{R}^n, y \in \{-1, 1\}, \quad (6)$$

the SVM requires the solution of the following optimization problem:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0. \end{aligned} \quad (7)$$

where ξ_i are slack variables that generalize the SVM approach to allow some classification errors, whereas $C > 0$ is a penalty parameter of the error term that must be chosen to reflect the knowledge of the noise present in the data. Moreover, in Eq-7, the training vectors x_i are mapped into a higher dimensional space by the function ϕ . With an appropriate nonlinear mapping $\phi(\cdot)$ to an adequate high dimension, data from two categories can always be separated by a hyperplane. In order for this mapping to be computationally feasible, a kernel function $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is typically exploited. Many commonly used functions can be used to perform this mapping, including polynomial, radial basis functions and certain sigmoid functions.

III. EXPERIMENTAL SETTINGS

In this section, the haptic-enabled virtual check application will be depicted followed with a concise description of the haptic data acquisition and preprocessing procedures.

A. Haptic-enabled Virtual Environment

The experiments are performed using the Reachin Display [6], which integrates a haptic device with stereo graphics for an immersive and high quality 3D experience. The Reachin visuo-haptic interface enables users to see and touch virtual objects at the same location in space. This approach enables a superior integration of vision and touch than a conventional

TABLE I
FEATURE REDUCTION BASED ON DATA SUBSAMPLING.

No. of features per dataset	5-fold CV on all objects	Train 60%, Test 40%		Train 80%, Test 20%	
		5-fold CV (%)	Classif. (%)	5-fold CV (%)	Classif. (%)
10000	90.8	87.2	88.5	88.5	92.3
5000	91.5	87.2	88.5	88.5	96.2
2500	92.3	87.2	88.5	88.5	96.2
2000	91.5	87.2	88.5	88.5	96.2
1000	90.8	87.2	88.5	88.5	96.2
500	90.8	87.2	86.5	88.5	96.2
250	90.8	85.9	88.5	87.5	96.2
200	88.5	88.5	84.6	85.6	92.3
150	86.2	85.9	82.7	83.7	92.3
100	86.2	83.3	80.8	84.6	88.5
80	86.9	73.1	80.8	83.7	88.5
50	80.0	75.6	84.6	75.9	92.3
20	74.6	73.1	67.3	72.1	80.8
10	67.7	67.9	61.5	72.1	69.2

2D screen-based display. The haptic stimulus is sensed using the SensAble PHANTOM Desktop force-feedback device, which is equipped with an encoder stylus that provides 6-degree-of-freedom single contact point interaction and positional sensing. In the case presented here, the visual stimuli consist of a virtual pen and a virtual check on which users can record their handwritten signature. The latter haptic-enabled virtual environment has been selected since handwritten signatures have been widely accepted as a mean to prove authenticity and authorship of a document. Conversely, the haptic stimuli are force and frictional feedback that attempt to mimic the tactile sensations felt when signing a traditional paper check. More specifically, the check is built on an elastic membrane surface with particular texture features, providing the users with a user-friendly and realistic feel of the virtual object. Moreover, the virtual check application records a wide array of haptic attributes that depict a user's physical and behavioral traits.

B. Haptic data acquisition and preprocessing

The haptic-based handwritten signatures were diligently obtained from 13 different participants, where 10 signatures are collected per individual. The data acquired depict various distinct haptic features as a function of time. A number of haptic data types are considered that characterize the instantaneous state of the haptic system, including, three-dimensional position, force (pressure exerted on the virtual check), torque, and angular orientation. Furthermore, the multi-feature and multidimensional haptic data are recorded at 100 Hz. As the data is time-varying, the resulting number of attributes per signature is in fact the number of haptic data types considered (position, force, torque, ...) times the number of samples recorded per data type during each signature acquisition. This evidently leads to significantly large feature vectors that encompass thousands of haptic-based attributes. Accordingly, the obtained feature vectors

TABLE II
CLASSIFICATION ANALYSIS OF REDUCTS GENERATED USING JOHNSON'S ALGORITHM.

No. of Features	Train 60%, Test 40%			Train 80%, Test 20%		
	No. of SF	5-fold CV (%)	Class. (%)	No. of SF	5-fold CV (%)	Class. (%)
10000	5	78.2	76.9	5	66.3	65.4
5000	5	82.1	69.2	5	67.3	65.4
2500	5	78.2	78.8	5	68.3	80.8
2000	5	61.5	57.7	5	70.2	73.1
1000	5	61.5	55.8	5	62.5	53.8

were normalized (upsampled/downsampled) to a common length of 10000. This length was selected in such a manner to minimize the information loss that is most apparent when downsampling is performed. Consequently, the computed preprocessed dataset contains 130 instances, where each consists of 10000 features. Furthermore, prior to reduct computation the preprocessed dataset is descritized using a boolean reasoning approach [19] and *equal frequency binning* [8].

IV. RESULTS

The presented results aim to analyze the performance of rough set-based feature selection in high dimensional haptic signatures (used for user identification). This is achieved while comparing the results obtained using rough set-based feature selection algorithms with those acquired using two other methodologies: (1) feature reduction through haptic data subsampling, and (2) feature selection using classical machine learning based techniques. In the former case, the time-varying and low-frequency nature of hand movements (when signing a check) is exploited to reduce the number of features. Specifically, the normalized haptic signatures

TABLE III
CLASSIFICATION ANALYSIS OF REDUCTS GENERATED USING THE
GENETIC REDUCER ALGORITHM (USING DEFAULT PARAMETERS).

No. of Features	Train 60%, Test 40%			Train 80%, Test 20%		
	No. of SF	5-fold CV (%)	Classif. (%)	No. of SF	5-fold CV (%)	Classif. (%)
10000	209	84.6	86.5	294	89.4	92.3
	190	85.9	84.6	371	89.4	92.3
5000	61	83.3	88.5	71	87.5	88.5
	66	80.8	86.5	136	90.4	96.2
2500	23	83.3	76.9	28	80.8	88.5
	58	83.3	86.5	84	86.5	88.5
2000	35	78.2	86.5	30	80.8	88.5
	28	76.9	75.0	44	81.7	88.5
1000	16	78.2	65.4	9	76.9	76.9
	24	76.9	71.2	20	78.8	76.9

discussed in Section III-B (10000 features/object) are further subsampled to reduce the number of features while possibly retaining the discrimination power in the data, despite some information loss. In the latter case, two different machine learning-based feature selection techniques are evaluated: *Best First*, and *Subset Size Forward Feature Selection*. A brief description of each method is later depicted in Section IV. In all cases, SVM classifiers are exploited to evaluate the accuracy of the selected minimal feature vectors.

A. Feature reduction based on haptic data subsampling

The results obtained using the *haptic data subsampling* feature reduction strategy are illustrated in Table I. Essentially, the original feature vectors are reduced in length using different subsampling factors (generated subsets range from 10000 features to as few as 10 features per object). Then, using an SVM classifier, a 5-fold cross-validation (CV) is initially performed on each of the generated subsets of features (on all objects). The purpose of cross-validation in this case is to provide us with an estimate of the discrimination power preserved in the reduced sets. Also, for each subset, two distinct datasets are generated: one that divides the objects (haptic signatures) into 60% training and 40% test sets, whereas the other divides the objects into 80% training and 20% test sets. The 5-fold CV accuracy on the training data and the classification accuracy of test data is then computed as reported in Table I. One interesting observation from the provided results (while inspecting the cross-validation and classification rates) is that subsets with as few as 1000 features seem to preserve the discrimination power achieved with the original haptic dataset (10000 features per object). Consequently, feature selection methods discussed in subsequent sections will be analyzed using the original dataset and generated subsets with a number of features ≥ 1000 . This should give us some insights on the possibility of combining feature vector subsampling with feature selection methods in order to further reduce the haptic feature space while maximizing discrimination between users.

B. Feature selection in haptic data using Johnson's algorithm

The feature selection results achieved using the rough set theory-based Johnson algorithm are depicted in Table II (columns labeled *No. of SF* describe the number of selected features). Reduct computation is performed on the original dataset (10000 features per object), as well as on undersampled versions of the data which include subsets containing 5000, 2500, 2000, and 1000 features per object. Also, for each subset, two distinct datasets are generated: one that divides the objects (haptic signatures) into 60% training and 40% test sets, whereas the other divides the objects into 80% training and 20% test sets. The 5-fold CV accuracy on the training data and the classification accuracy of test data is then computed as reported in Table II. Overall, approximate reducts (minimal subsets) obtained using Johnson's greedy algorithm did not demonstrate significant improvement (when inspecting the length of the reducts and the corresponding classification rates) over subsets generated when relying solely on feature vector subsampling. Nevertheless, in certain cases some improvements were apparent. For example, considering the case when reduct computation is performed on the 60% training set with 2500 features, a minimal subset of length 5 is obtained, with a classification rate of 78.8%. This suggests a feature reduction improvement by a factor of at least 4 in comparison to subsampling where even with a feature vector of length 20, the classification rate is a mere 67.3 (See Table I when 60% training data are considered).

C. Feature selection in haptic data using the Genetic Reducer algorithm

1) *Comparison between the Genetic Reducer algorithm and data subsampling*: The feature selection results achieved using the rough set theory-based Genetic Reducer technique are presented in Table III. Minimal hittings sets (reducts) are computed using the algorithm's default parameters, i.e. *mutation probability* = 0.05, *crossover probability* = 0.3, *population size* = 70. In order to compute reducts within reasonable computation time, approximate solutions were sought with a *hitting fraction* = 0.95. As in Section IV-B, reduct computation is performed on the original dataset (10000 features per object), and on subsampled subsets ranging from 5000 to as few as 1000 features per object. Also, two versions of each subset is created, one that divides the objects into 60% training and 40% test sets, whereas the second divides the objects into 80% training and 20% test sets. For each version of every subset, 20 independent runs of the GA-based reduct computation procedure was conducted and the top 10% of the reducts obtained are selected (i.e. in this case, the top 2 reducts). Reducts are selected based on the best compromise between the following two criteria: (1) maximum 5-fold CV accuracy on the training data and, (2) hitting set (reduct) is of minimal length. The top two reducts computed for each subset are reported in Table III. A comparison between the results presented in

Fig. 1. Frequency of the haptic features in subsets generated using the Genetic Reducer when (a) 60% training data, (b) 80% training data, are considered.

Tables I and III reveals that approximate reducts computed using the Genetic Reducer algorithm led to significant feature reduction improvements over subsets generated using strictly feature vector subsampling. For example, considering the case when reduct computation is performed on the 60% training set with 5000 features, a minimal subset of length 61 is obtained, with a classification rate of 88.5%. This is a feature reduction improvement by a factor of approximately 4 in comparison to subsampling, where a classification rate of 88.5% is only obtained with subsampled feature vectors of length ≥ 250 (see Table I when 60% training data are considered).

2) Relevance of haptic data types in user identification:

Feature selection using a rough set-based approach is also exploited to determine the importance of different haptic data types in discriminating between different users. As aforementioned, haptic data types include three-dimensional force, position, torque, and orientation (exceptionally, orientation is one-dimensional as it consists of the haptic device's tracker angle). In Fig. 1, box plots are provided illustrating the frequency of different haptic data types in subsets generated using the Genetic Reducer algorithm when 60% training data (Fig. 1 (a)), and 80% training data (Fig. 1 (b)), are considered. Specifically, the reducts considered are those presented in Table III, and that are associated with subsets ranging from 10000 to 2500 features. Subsets with 2000 and 1000 features per object were not considered due to their low CV performance on the training data. It can be observed that overall, haptic data types are approximately of equal importance when exploited for user identification. Specifically, position information (P_x , P_y , and P_z) appear approximately as frequently as haptic-specific data types, i.e. force and torque and orientation (F_x , F_y , F_z , T_x , T_y , T_z and O) in the generated reducts. This is an interesting result as the apparent importance of haptic-specific data types suggests that haptic-based handwritten signatures can

potentially significantly outperform traditional handwritten signatures (in user identification applications) as the latter rely solely on position information.

D. An in-depth analysis of the Genetic Reducer algorithm

In order to gain better insight into the reduct computation capabilities of the Genetic Reducer method, the experiments presented in Section IV-C were repeated while varying the most fundamental genetic parameters of the algorithm. Specifically, the parameters *crossover probability* and *mutation probability* were assigned values ranging from 0.3 to 0.7 and 0.01 to 0.05 respectively as depicted in Table IV. Moreover, to avoid much redundancy in the presented results, the analysis was performed solely on the (subsampled) subset with 2500 features per object. In fact, in Table I, it is observed that subsets with 2500 not only preserve the discrimination power of the original datasets, but also achieve a better CV result when all the objects are considered. Two versions of the subset were created, one that divides the objects into 60% training and 40% test sets, whereas the second divides the objects into 80% training and 20% test sets. Similar to Section IV-C.1, for every *crossover probability* and *mutation probability* value assignment, 20 independent runs of the GA-based reduct computation procedure was conducted and only the top two reducts were selected (see Table IV). A first inspection of the results reveals that reduct computation with higher crossover probabilities resulted in hitting sets that are shorter in length, but only at the expense of classification accuracy. This observation is more pronounced when 60% training sets are considered. Furthermore, it can be observed that generally, a higher classification accuracy is found with reducts generated using a low mutation probability of 0.01. This observation however is more pronounced when 80% training sets are considered.

TABLE IV
CLASSIFICATION ANALYSIS OF REDUCTS GENERATED USING THE GENETIC REDUCER ALGORITHM WHILE VARYING THE MUTATION AND CROSSOVER PROBABILITY PARAMETERS.

<i>GA Parameters</i>		<i>Train 60%, Test 40%</i>			<i>Train 80%, Test 20%</i>		
<i>Crossover Prob.</i>	<i>Mutation Prob.</i>	<i>No. of SF</i>	<i>5-fold CV (%)</i>	<i>Classif. (%)</i>	<i>No. of SF</i>	<i>5-fold CV (%)</i>	<i>Classif. (%)</i>
0.3	0.01	77	83.3	82.7	52	84.6	88.5
		65	82.1	84.6	97	89.4	92.3
0.3	0.03	35	82.1	76.9	38	86.5	88.5
		49	82.1	84.6	96	85.6	92.3
0.3	0.05	23	83.3	76.9	28	80.8	88.5
		58	83.3	86.5	84	86.5	88.5
0.5	0.01	50	83.3	82.7	37	86.5	96.2
		90	85.9	88.5	64	87.5	88.5
0.5	0.03	19	74.4	75.0	17	79.8	88.5
		24	80.8	75.0	23	81.7	76.9
0.5	0.05	17	82.1	75.0	16	75.0	84.6
		22	76.9	76.9	31	84.6	80.8
0.7	0.01	27	78.2	75.0	46	81.7	92.3
		44	79.5	80.8	51	83.7	92.3
0.7	0.03	18	78.2	76.9	13	80.8	84.6
		29	79.5	76.9	16	76.0	76.9
0.7	0.05	9	75.6	71.2	18	77.9	80.8
		28	76.9	75.0	31	83.7	88.5

E. Feature selection using the rough set theory Vs classical machine learning-based algorithms

In this section a comparison between rough set-based methods and classical machine learning techniques in the selection of minimal information-preserving subsets of features in high dimensional haptic datasets is presented. The machine learning methods were evaluated using the WEKA environment [20], a public data mining library in Java for feature selection and classification. WEKA contains several methods for searching through the space of feature subsets, as well as evaluation measures for features and feature subsets. Different search and evaluation methods can be combined as desired by the user. In this work, two search methods capable of handling high dimensional datasets within reasonable computation time are considered: *Best First*, and *Subset Size Forward Selection*. *Best First* Searches the space of attribute subsets by greedy hillclimbing augmented with a backtracking facility. Setting the number of consecutive non-improving nodes allowed controls the level of backtracking done. Conversely, *Subset Size Forward Selection* [16] is a recent algorithm that modifies the standard search method known as forward selection to yield computationally efficient wrapper-based feature selection technique for high-dimensional data. Explicit subset size determination in forward selection is also achieved to combat overfitting, where the search is forced to stop at a precomputed optimal subset size. In addition, *Correlation-based Feature Selection (CFS)* [15] was selected as the feature evaluation method (combined with the Best First and Subset Size Forward Selection search methods). *CFS* evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature

along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred.

Similar to Sections IV-B and IV-C, feature selection using the machine learning techniques is performed on the original dataset (10000 features per object), and on subsampled subsets ranging from 5000 to as few as 1000 features per object. Also, two versions of each subset is created, one that divides the objects into 60% training and 40% test sets, whereas the second divides the objects into 80% training and 20% test sets. First, considering the case when feature selection is performed on the 60% training sets, it can be observed that between the two techniques, *Best First* generated the subset with the highest classification rate. The minimal subset is of length 103, and achieves a classification rate of 88.5% on the test data. Conversely, the subset with the highest classification accuracy obtained with feature selection using the Genetic Reducer algorithm is a minimal approximate reduct of length 61, with a classification rate of 88.5% (see Table III). This is a feature reduction improvement by a factor of approximately 1.7 (in comparison to the machine learning techniques). Furthermore, considering the case when feature selection is performed on the 80% training sets, it can be observed that between the two techniques, *Subset Size Forward Selection* generated the subset with the highest classification rate. Specifically, the minimal subset is of length 63, and achieves an impressive classification rate of 96.2% on the test data. Conversely, the subset with the highest classification accuracy obtained with feature selection using the Genetic Reducer algorithm is a minimal approximate reduct of length 136, with a classification rate

TABLE V
CLASSIFICATION ANALYSIS OF FEATURE SUBSETS GENERATED USING CLASSICAL MACHINE LEARNING

Original Attributes	Best First						Subset Size Forward Selection					
	Train 60%, Test 40%			Train 80%, Test 20%			Train 60%, Test 40%			Train 80%, Test 20%		
	No. of SF	5-fold CV (%)	Class. (%)	No. of SF	5-F CV (%)	Class. (%)	No. of SF	5-fold CV (%)	Class. (%)	No. of SF	5-fold CV (%)	Class. (%)
10000	153	93.6	82.7	158	92.3	84.6	88	91.0	86.5	80	91.3	92.3
5000	114	92.3	86.5	130	93.3	84.6	79	89.7	84.6	79	92.3	92.3
2500	103	92.3	88.5	103	90.4	88.5	67	93.6	78.8	63	93.3	96.2
2000	91	91.0	82.7	95	92.3	92.3	58	93.6	82.7	63	91.3	96.2
1000	71	91.0	84.6	72	93.3	92.3	47	93.6	80.8	50	88.5	80.8

of 96.2%. However, if the algorithm's genetic parameters are selected such that *crossover probability* = 0.5 and *mutation probability* = 0.01, approximate reducts with as few as 37 features, and a corresponding classification rate of 96.2% can be achieved (See Table IV). However, the latter comparison is not entirely fair as results using machine learning techniques with default parameters are compared to a rough set-approach with adapted parameters. Nevertheless, the comparison is only intended to illustrate that rough-set based feature selection techniques can, in most cases, (possibly) potentially overcome machine learning techniques.

V. CONCLUSION

Feature selection using concepts from rough set theory proved to be a promising way to reduce the high dimensionality of haptic-based handwritten signatures when exploited for user identification. The results demonstrated that the combination of rough set and genetic algorithm techniques can potentially outperform well-established classical machine learning-based feature selection algorithms when analyzing haptic datasets. Furthermore, it was determined that feature selection performance using rough sets can be further improved if haptic datasets are moderately subsampled prior to the reduct computation procedure. In addition, minimal subsets generated using rough set-based approaches were exploited to determine the importance of different haptic data types (e.g. force, position, torque and orientation). Interestingly, it was observed that overall, the data types were approximately of equal importance when used to discriminate between different haptic users. As for the future work, it would be of interest to investigate the possible haptic feature selection improvements that could result if a multi-objective genetic algorithm is exploited in the reduct computation procedure. A multi-objective approach can better overcome the conflicting properties of reducts, of having small cardinality and the ability to discern among all objects.

REFERENCES

- [1] I. Czarnowski, P. Jędrzejowicz, "Data Reduction Algorithm for Machine Learning and Data Mining," New Frontiers in Applied Artificial Intelligence, Lecture Notes in Computer Science, Springer-Verlag, Heidelberg, vol. 5027, pp. 276-285, 2008.
- [2] A. El Saddik, M. Orozco, Y. Asfaw, S. Shirmohammadi, A. Adler, "A Novel Biometric System for Identification and Verification of Haptic Users," IEEE Transactions on Instrumentation and Measurement, vol. 56, no. 3, pp. 895-906, 2007.
- [3] M. Orozco, M. Graydon, S. Shirmohammadi, A. El Saddik, "Experiments in haptic-based authentication of humans," Journal of Multimedia Tools and Applications, vol. 37, no. 1, pp. 73-92, 2007.
- [4] F.A. Alsulaiman, N. Sakr, J.J. Valdés, A. El Saddik, N.D. Georganas, "Feature selection and classification in genetic programming: application to haptic-based biometric data," Proc. IEEE Symposium on Computational Intelligence for Security and Defence Applications, pp. 339-345, July 2009.
- [5] N. Sakr, F. A. Alsulaiman, J. J. Valdés, A. El Saddik, N.D. Georganas, "Relevant feature selection and generation in high dimensional haptic-based biometric data," Proc. International Conference on Data Mining (DMIN), pp. 71-77, July 2009.
- [6] Reachin Technologies ab. Reachin Display. <http://www.reachin.se/products/>.
- [7] Z. Pawlak, *Rough sets: Theoretical aspects of reasoning about data*. Kluwer Academic Publishers, Netherlands, 1991.
- [8] A. Ohrn, *Discernibility and rough sets in medicine: tools and applications*. Ph.D. Thesis. Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway, 1999.
- [9] J. Bazan, A. Skowron, P. Synak, "Dynamic reducts as a tool for extracting laws from decision tables," Proc. Eighth International Symposium on Methodologies for Intelligent Systems, Charlotte, NC, USA, Oct. 16-19, 1994. Lecture Notes in Artificial Intelligence 869, Springer-Verlag, pp. 346-355.
- [10] J. Wróblewski, "Ensembles of classifiers based on approximate reducts," Fundamenta Informaticae 47 IOS Press, pp. 351-360, 2001.
- [11] H. S. Nguyen, A. Skowron, "Quantization of real-valued attributes," In Proceedings of the Second International Joint Conference on Information Sciences, pp. 34-37, Sept. 1995.
- [12] D. S. Johnson, "Approximation algorithms for combinatorial problems," Proc. Fifth Annual ACM Symposium on Theory of Computing, pp. 38-49, 1974.
- [13] S. Vinterbo, A. Ohrn, "Minimal approximate hitting sets and rule templates," In Predictive Models in Medicine: Some Methods for Construction and Adaptation, Department of Computer and Information Science, Dec. 1999. NTNU report 1999:130. [<http://www.idi.ntnu.no/staalv/dev/thesis.ps.gz>].
- [14] S. Vinterbo, A. Ohrn, "Minimal approximate hitting sets and rule templates," International Journal of Approximate Reasoning, vol. 25, no. 2, pp. 123-143, 2000.
- [15] M.A. Hall, "Correlation-based feature selection for machine learning" Ph.D. dissertation, Dept. Comput. Sci., Waikato Univ., New Zealand, 1999.
- [16] M. Gutlein, E. Frank, M. Hall, A. Karwath, "Large-scale attribute selection using wrappers," Proc. IEEE Symposium on Computational Intelligence and Data Mining, pp. 332-339, March 2009.
- [17] B. E. Boser, I. M. Guyon, V. N. Vapnik, "A training algorithm for optimal margin classifiers," Proc. Fifth Annual ACM Workshop on Computational Learning Theory, pp. 144-152, 1992.
- [18] V. Vapnik, *Statistical Learning Theory*. Wiley, New York, 1998.
- [19] H. S. Nguyen, A. Skowron, "Quantization of real-valued attributes," Proc. Second International Joint Conference on Information Sciences, pp. 34-37, Sept. 1995.
- [20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "The WEKA data mining software: an update," SIGKDD Explorations, vol. 11, no. 1, pp. 10-18, 2009.