Abstract—This paper presents new methods for the recognition and categorization of object properties such as surface texture, weight, and compliance using a multi-modal artificial skin mounted on both arms of a humanoid. In addition, it introduces two novel feature descriptors, which are useful for providing high-level information to learning algorithms. The artificial skin has built-in 3-axis accelerometer, normal force, proximity, and temperature sensors. To explore different surface textures and weights, objects were left sliding between the NAO humanoid’s arms. The caused vibration was detected by accelerometers. Surface texture and weight recognition models were learned from the extracted features of the vibration signals thanks to two learning algorithms, namely the support vector machine (SVM) and the Expectation Maximization (EM). In order to recognize objects having different compliances, SVM and EM took into account total amount of forces applied by the arms to hold the object firmly. The experimental results show that the humanoid can distinguish between different objects having different surface textures and weights with a recognition rate of 100%. Furthermore, it can categorize objects with hard and soft surfaces and classify objects having similar compliance with 100% and 70% accuracy rates respectively.

I. INTRODUCTION AND RELATED WORKS

Developing technological expectations are changing the face of robotics. If humanoids are meant to interact with the environment and real world objects, they need to be equipped with cognitive skills such as perception and learning [1]. To meet this prerequisite, humanoids need to be provided with different sensing modalities and learning techniques. In this study, the perception medium used is a multi-modal artificial skin and can be used to access the contact parameters such as texture surface, compliance, and weight of manipulated objects. The aim of this study is to investigate the robust feature descriptors in order to select and decipher relevant tactile features and to provide optimum strategy for data fusion. The further goal is to construct object patterns that humanoids could learn while manipulating the objects. In previous studies, many researchers employed different customized tools or end effectors equipped with uni-modal tactile sensors such as force sensors or accelerometers to execute a single or multiple exploratory action/s such as scratching [2], [3], rubbing and pushing [4], tapping and squeezing [5], [6], and knocking [7] on objects in order to recognize their fundamental properties, such as surface texture, compliance, and weight. Since the experimental results might have been affected by the properties of that surface, resulting in a biased collection of data. In these experiments, the velocity of the tools/end effectors used to perform the exploratory actions, as well as the applied normal force on the objects remained constant. Moreover, they used the Fourier transform technique [2], [3], [8]–[11], intensity, variance, skewness, and kurtosis of the signals [12]–[14] for interpreting the tactile information in frequency or time domains. However, the Fourier transform method is not appropriate for analyzing non-stationary signals, and the magnitude of the signals is highly sensitive to noise. In this respect, the short time Fourier transform (STFT) or wavelet transform may be more useful techniques [11], [15]–[17] for analyzing non-stationary signals [18], [19]. These techniques analyze the signal locally by windowing in the time domain. However, these methods deal with a large number of data points, thereby causing difficulties during the classification step. More features require more training samples, which result in an increased computational complexity as well as the risk of over-fitting. To overcome the problems mentioned above, we propose the use of a multi-modal skin coupled with robust feature extraction methods and adapted learning algorithms. The robot can execute non-sliding exploratory behaviors and sliding exploratory behaviors (see section (III-A)) to identify useful characteristics of the objects. In this study, we introduce two novel feature descriptors, which represent statistical properties of the signal in the time domain to extract informative and abstract high level information for the learning algorithms. Since the calculation

---

*This work is supported by European CONTEST Project.

1Mohsen Kaboli, Philipp Mittendorfer, and Gordon Cheng are with the Institute for Cognitive System, Faculty of Electrical Engineering and Information Technology, Technical University of Munich (TUM)-Germany. Email: mohsen.kaboli@tum.de (http://www.ics.ei.tum.de)

2Vincent Hugel is with the Robotics Dept. Uni. of Toulon-Var, France.
of our proposed feature descriptors is based on variance and correlation coefficient of signals, the computational cost of this method is extremely low. This makes it appropriate descriptors for the real-time task.

II. SYSTEM DESCRIPTION

A. Skin Patches on a Humanoid’s Arms

In our daily life, we use our arms even our upper part of the body to grasp, lift, and hold unknown large objects. In this case, a large area of skin will be used for the object properties exploration. Having no prior information of object properties, such as weight, surface texture, and softness or hardness, makes object manipulation not only difficult but also almost impossible, especially when there is no visual feedback available. Knowledge about object properties, therefore, is crucial for humans and also humanoids, in order to be able to interact with real world objects precisely. In our current study, we aim to provide humanoids with the ability of human-like tactile sensing (biological inspired skills) in order to interact with unknown large objects.

B. Multi-Modal Artificial Skin

The artificial skin used in the experiment is designed and manufactured in our lab. It is a hexagonal shaped multi-modal sensory cell, called Hex-o-Skin [20]. Each skin cell has a microcontroller on the back side and a set of multi-modal tactile sensors on the front side, which includes a 3-axis accelerometer for object surface texture and weight classification; three normal-force sensors to estimate the required force to hold an object between the humanoid’s hands; a proximity sensor and a temperature sensor which are not used in this current study.

C. Humanoid Robot

We employed an existing autonomous humanoid robot from Aldebaran Robotics which is called NAO. With respect to the sensing device, we covered NAO’s arms with a thin layer of flexible and stretchable foam. Then we mounted one skin patch on each arm of the robot (Fig.1), which included 7 skin cells. Thus, each arm had 7 three-axis accelerometer sensors, 21 normal-force sensors, and 7 proximity sensors.

D. Properties Of Experimental Objects

Taking into account NAO’s size and weight, we selected 10 large objects having the same dimension from two different classes of weight \((w \in W)\). This means that one half of the objects belonged to a class of 1500 g and the second half belonged to a class of 500 g. Five common everyday materials were deliberately chosen for both classes of weight, such as card box/paper, glass, bubble plastic, sponge, and rough texture/sand paper \((t \in T)\) as shown in Fig.(2). It is noteworthy to mention that the sponge and rough textures are irregular and non-uniform textures.

III. TACTILE PERCEPTION

A. Exploratory Behaviors

In this work, in order to provide more descriptive, realistic, and human-like exploratory behaviors during data collection, we present a new exploratory interactions (Fig.1). In our approach, NAO collects tactile data while an unknown object is sliding or non-sliding between its arms which were already covered with multi-modal artificial skin. NAO uses sliding behavior to classify weight and surface texture and non-sliding behavior to detect stiffness of an object. Moreover, using the non-sliding behavior, NAO can estimate the required amount of force in order to keep an unknown object from sliding between its arms and skin cells. Contrary to previous works, sliding behavior depends in our study only on the intrinsic object properties, for instance that a heavy object slides faster than a lighter one. Since NAO uses its arms to interact with objects, the collected data will not be affected by any other surrounding object properties.

B. Data Collection

Each of the 2 behavioral interactions \((b \in B)\), sliding and non-sliding was performed 15 times on each of the 10 selected objects placed between NAO’s arms, resulting in a total of \(2 \times 15 \times 10 = 300\) trials.

1) Sliding Behavioral Interaction: In this interaction approach, NAO’s arms were kept fixed. They were straight and parallel to each other and were able to open and close only in the horizontal direction. While collecting the exploratory data, during each \(i\)th trial, an object was placed between NAO’s arms, which were close enough for NAO to firmly grasp the object. However, the orientation of the object while grasping varied across trials. As the experiment proceeded, NAO slowly opened its arms until the object started to slide between them. During the process, tactile information was recorded via multi-modal artificial skin mounted on its arms that involved acceleration and force data used to detect object properties. However, the sensory data from each tactile sensor was recorded for 1000 ms or 250 data points (since data were recorded at 250 Hz). The recording time started slightly before the object started to slide and continued until the object completely left NAO’s arms.
2) Non-Sliding Behavioral Interaction: In this interaction approach, the initial position of NAO’s arms was similar to the situation of the sliding behavioral approach. For each trial, an object was placed between NAO’s arms, and the arms were brought closer until the humanoid was able to grasp the object between its arms with a minimum amount of force required to prevent sliding. In this position, NAO then started to record each tactile sensory data for 1000 ms.

IV. PROPOSED ROBUST FEATURE DESCRIPTORS

In earlier works, researchers used the Fourier transform technique [11], [15]–[17] in the frequency domain and magnitude of tactile signals in the time domain for interpreting tactile information such as surface texture. However, the Fourier transform method is not appropriate for analyzing non-stationary signals, particularly in the case of surface texture detection where texture is irregular and non-uniform. The Fourier transform presents the relative power of each frequency and calculates frequency responses based on specific time. In addition, the magnitude of the signal is highly sensitive to noise. Wavelet transform may be the best technique [18], [19] for analyzing non-stationary signals. However, this method deals with a large number of data points, thereby causing difficulties at the classification step. More features require more training samples which will result in increased computational complexity as well as the risk of over-fitting. To overcome these issues, we propose novel feature extraction techniques, inspired by the Hjorth parameters.

A. Hjorth Parameters

Hjorth [21] presented a set of parameters for real-time biological signal analyses (Electroencephalography) which represented statistical properties of the signal in the time domain. These parameters are called Activity, Mobility, and Complexity. Although these parameters are defined in the time domain they can be interpreted in the frequency domain as well. The first parameter (Eq.(1)) is the total power of a signal. It is also the surface of the power spectrum in the frequency domain (Parseval’s relation). The Mobility parameter, defined in Eq.(2), is determined as the square root of the ratio of the variance of the first derivative of the signal to that of the signal. This parameter has a proportion of standard deviation of the power spectrum. It is an estimate of the mean frequency [22]. The last parameter (Eq.(3)) gives an estimate of the bandwidth of the signal, which indicates how the shape of the signal is similar to a pure sine wave. If the shape of the signal becomes more similar to a pure sine wave, the value of complexity converges to one [23]. Equation (2) and equation (3) are specified by means of the first and second derivatives of variance, so they called *normalized slope descriptors* [23]. Since the calculation of the Hjorth parameters is based on variance, the computational cost of this method is extremely low, which makes it appropriate descriptors for the real-time task.

\[\text{Activity} = \text{Var}(s(t)) = \frac{1}{N} \sum_{t=1}^{N} (S_t - \bar{S})^2\]  \hspace{1cm} (1)

\[\text{Mobility} = \sqrt{\frac{\text{Var}(ds(t)/dt)}{\text{Var}(s(t))}}\] \hspace{1cm} (2)

\[\text{Complexity} = \frac{\text{mobility}(ds(t)/dt)}{\text{mobility}(s(t))}\] \hspace{1cm} (3)

B. The Proposed Feature Extraction Techniques

We proposed two feature methods, essentially inspired by Hjorth parameters, to get higher level and abstract information from tactile signals, called Inter-Hybrid and Intra-Hybrid feature descriptors (Tab.I).

1) Inter-Hybrid Feature Descriptor: An object sliding between NAO’s arms generates vibration on NAO’s skin. The caused vibration was measured by each 3-axis accelerometer sensor existing in every skin cell. Since the three accelerometer components are highly correlated during the measurement, our first proposed feature includes the correlation coefficient between each of the two axes of the accelerometer (Eq.4), namely \(\text{corr}(a_x,a_y), \text{corr}(a_x,a_z), \text{and} \text{corr}(a_y,a_z)\). The feature also includes the Activity, Mobility, or Complexity parameter related to each of the three acceleration signal components (see Tab.I). The feature dimensionality was reduced to six parameters for each accelerometer output instead of \(3 \times 250 = 750\) data points (# of accelerometer axis)\(# of accelerometer axis) \times (# data sample recorded from each axis in 1000 ms) = 750). For instance [Complexity,Correlation] is equal to \([\text{comp}(a_x), \text{comp}(a_y), \text{comp} (a_z), \text{corr}(a_x,a_y), \text{corr}(a_x,a_z), \text{corr}(a_y,a_z)]\)

\[\text{corr}(a_x,a_y) = \frac{\sum_{i=1}^{N} (a_{xi} - \bar{a}_x) (a_{yi} - \bar{a}_y)}{\sqrt{\sigma(a_x) \sigma(a_y)}}\] \hspace{1cm} (4)

2) Intra-Hybrid Feature Descriptor: Our second proposed feature includes the composition of each of the two computed Hjorth parameters from each of the three acceleration signals. The feature dimensionality is then similar to that of the Inter-Hybrid Feature. In addition, the super composition of all three computed Hjorth parameters from acceleration data was considered as another alternative feature descriptor. The dimensionality of this feature is therefore nine instead of \(3 \times 250 = 750\) data points.

<table>
<thead>
<tr>
<th>TABLE I PROPOSED FEATURE DESCRIPTOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-Hybrid Features</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Intra-Hybrid Features</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
V. LEARNING METHODOLOGY

A. Texture Recognition Methodology

The primary task of NAO was to classify the five selected objects having different surface textures \((t \in T)\). The first algorithm that was evaluated was Support Vector Machine (SVM) [24], which is a discriminating classifier formally defined by a separating hyper-plane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper-plane which categorizes new examples.

B. Weight Classification Methodology

The other task of NAO was to recognize different weights \((w \in W)\) of an object having the same surface texture \((t \in T)\) using pure mobility, and the proposed \([\text{mobility, complexity}]\) and \([\text{mobility, correlation}]\) features, while employing sliding exploratory behavior. NAO used these features and SVM to learn a recognition model which is then evaluated on a new test data in order to predict the correct weight class.

C. Weight Categorization Methodology

In addition to weight classification, an important task for NAO was to qualitatively differentiate between varying categories of weight. This means that objects having the same surface texture and almost similar weights tend to be in the same cluster or category. Therefore, in this scenario, NAO clustered the two unlabeled categories of weight using Expectation Maximization (EM) algorithm [25] which is an unsupervised learning approach. So that, when a test data is presented in a real time environment whose weight does not need to be exactly similar to the two categories, it gets assigned to one of the two clusters. For example, in our experiment, NAO clustered the two categories of weights, i.e. 500 g and 1500 g, so that when an unknown test data, say 700 g is presented to NAO, it can give an estimate of its weight via nearest neighbor measure and assign it to one of the clusters, which in this case would be the class of 500 g.

D. Hardness/Softness Detection Methodology

In addition to the previous tasks, NAO was asked to differentiate hard and soft objects having different textures \((t \in T)\) from non-sliding behaviour. In order to do this, the SVM classifier was trained with obtained force data to build a learning model. Moreover, NAO used the EM algorithm to categorize the objects \((t \in T)\) having similar harness and softness. In this scenario, NAO added all obtained force information over contact positions to get the total force applied to an object.

VI. RESULTS

A. Texture Recognition Results

The robot learning was evaluated by randomly partitioning the collected data set into 70% and 30% for training and testing respectively. However, to find the optimal radial basis kernel parameter \((\gamma)\) and regularizer value \((C)\) for SVM, 10-fold cross validation (CV) was used. In this respect, the data set was randomly split into 10 folds and during each evaluation, 9 of those were used for training and one was used for testing. This process was repeated 10 times to obtain an average performance on the evaluation sets. This entire process was repeated 50 times using different values for kernel parameter and regularizer in order to find the one with the lowest CV error. The SVM with optimal parameters was then re-trained on the entire training data set to obtain classification models. These classification models were used by NAO to predict the surface textures for the test data. The prediction results are reported in terms of recognition accuracy. The results show the surface texture recognition rates corresponding to each of the feature descriptors for the multi-class classification problem. However, NAO was able to achieve recognition accuracy substantially better than chance, using all proposed feature descriptors. The classification of five different surface textures resulted in a recognition accuracy of 86% and 91% using mobility and complexity features respectively. By employing \([\text{mobility, correlation}]\) as well as \([\text{activity, mobility, complexity}]\) from the proposed Intra-Hybrid features, NAO was able to reach 100% surface texture classification accuracy. Although NAO achieved the same result, the former was preferred since it has a lower number of features compared to the latter. This is because a lower dimensionality of the feature vector implies a lower computational complexity while computing the kernel in SVM. NAO was also able to reach surface texture recognition accuracy of 100% employing the proposed \([\text{mobility, correlation}]\) Inter-Hybrid feature. Thus, the results of surface texture recognition show that our both sets of proposed feature descriptors have improved the recognition accuracy in comparison to using pure Hjorth parameters. We attribute this to the fact that our proposed features provide more discriminative and informative data for the classifier. Since the activity parameter is not informative enough, its composition with the other two pure Hjorth parameters as well as with correlation did not improve the recognition accuracy.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Activity</th>
<th>Mobility</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hjorth Parameters</td>
<td>[Activity, Mobility, Complexity]</td>
<td>0.4</td>
<td>86%</td>
</tr>
<tr>
<td>Inter-Hybrid Features</td>
<td>[Activity, Mobility, Complexity]</td>
<td>0.7</td>
<td>96%</td>
</tr>
<tr>
<td>Intra-Hybrid Features</td>
<td>[Activity, Mobility, Complexity]</td>
<td>0.4</td>
<td>93%</td>
</tr>
</tbody>
</table>
explained above. In order to evaluate the classification model, the collected data set was randomly split into 80% and 20% for training and testing respectively. The SVM classifier along with the optimal kernel parameters was then trained using the entire training data set. Finally, this trained model was used by NAO in order to recognize the weight of a new test data. The recognition accuracy of weights corresponding to each of the five different textures are presented in Tab.(III).

The results shows that the NAO is able to classify weights perfectly using the [mobility, correlation] feature. It also shows that [mobility, correlation] feature clearly outperforms the [mobility, complexity] feature. In the case of Sponge and Paper, using the [mobility, complexity] feature is still better than a chance.

C. Weight Categorization Results

In this experiment, NAO employed EM as a probabilistic learning approach to generate two weight categories. In order to do this, EM was trained with the entire unsupervised data set. A class to clustering approach was used to evaluate how well NAO can recognize the correct category of a novel test data. In this approach, classes were assigned to the categories, based on the majority value of the class attribute within each category. Later on, these assignments were employed to compute the value classification performance. Fig.(VI-D) shows the results of this experiment where [Activity,Mobility], [Activity,Complexity], and [Activity,Correlation] were used as hybrid features. From the results, it is clear that NAO managed to recognize the categories of weights with an accuracy significantly higher than chance. It was able to categorize weights of objects having glass and paper texture perfectly. In the case of interacting with an object having more distinct weight values than 500 g and 1500 g, humanoids could achieve a better clustering accuracy rate.

D. Hardness/Softness Estimation Results

In this experiment, the entire data set including normal force information was randomly split into 5 folds and at each evaluation, 4 of those were used for training and one for testing. The entire process was repeated 100 times using $C = 0.001$ and $\gamma = 0.4$. In this case the overall recognition accuracy was 70%. Therefore, it was worth investigating the type of recognition error that NAO made. Table (IV) shows the confusion matrix obtained from the classification procedure. The confusion matrix indicates how often a given surface ($t \in T$) was mis-classified as another surface.

<table>
<thead>
<tr>
<th>[Mobility, Correlation]</th>
<th>Glass</th>
<th>Sponge</th>
<th>Paper</th>
<th>Rough Texture</th>
<th>Bubble Plastic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-Hybrid Features</td>
<td>0.2</td>
<td>100%</td>
<td>0.3</td>
<td>100%</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>100%</td>
<td>0.3</td>
<td>100%</td>
<td>0.5</td>
</tr>
<tr>
<td>Intra-Hybrid Features</td>
<td>0.2</td>
<td>100%</td>
<td>0.4</td>
<td>84%</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>100%</td>
<td>0.3</td>
<td>84%</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>100%</td>
<td>0.8</td>
<td>100%</td>
<td>0.6</td>
</tr>
</tbody>
</table>

A perfect classification would result in a diagonally-filled table. However, table (IV) shows that most errors involve textures having similar compliance. The sponge and glass have the least errors and the highest accuracy. This is to be expected, however, as the sponge is the softest and glass is the hardest of all selected textures. The bubble plastic is often confused with sponge, paper, and rough texture which shows that it is not definitely hard but almost close to soft class. The rough texture is confused with three different textures but mostly with the sponge which means that it is almost close to objects in the soft class. Overall, the confusion matrix shows that the errors of the NAO’s learning model were not random in nature. Instead, whenever an error was made the predicted texture was often somewhat similar to the actual one in terms of material and/or texture. This suggests that the learning model could be used to estimate a measure of similarity between hard and soft surfaces based on force information. In this part of the experiment, the EM algorithm was used to evaluate how well NAO can recognize the correct categories of objects having different soft or hard textures ($t \in T$). To do this, NAO used unlabeled data set to cluster each two surface texture. Table (V) shows that NAO could perfectly categorize both sponge and glass with 100% accuracy. Moreover, it could successfully cluster each of paper, rough texture, and bubble plastic from glass texture, which means that they are pretty far from category of hard texture. Furthermore, imperfection clustering between sponge and each of paper, rough texture, and bubble plastic texture is due to existing similarities between them. In this situation NAO considered bubble plastic and specially paper and rough texture more closer to soft categories. Finally, the result of clustering between rough texture and bubble plastic shows that these two textures are very similar to each other in terms of stiffness.

![Fig. 3. Weight Categorization Result Using EM and three of proposed feature descriptors.](image)
TABLE IV
CONFUSION MATRIX FOR SVM CLASSIFICATION

<table>
<thead>
<tr>
<th>Class</th>
<th>Sponge</th>
<th>Glass</th>
<th>Paper</th>
<th>Bubble Plastic</th>
<th>Rough</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sponge</td>
<td>13</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Glass</td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Paper</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bubble Plastic</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Rough</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

TABLE V
HARDNESS/SOFTNESS CLUSTERING RESULTS

<table>
<thead>
<tr>
<th>Class to Cluster</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glass &amp; Sponge</td>
<td>100 %</td>
</tr>
<tr>
<td>Paper Rough Texture &amp; Sponge</td>
<td>74 %</td>
</tr>
<tr>
<td>Rough Texture &amp; Glass &amp; Bubble Plastic &amp; Sponge</td>
<td>90 %</td>
</tr>
<tr>
<td>Paper Rough Texture &amp; Glass &amp; Bubble Plastic &amp; Glass</td>
<td>94.3 %</td>
</tr>
<tr>
<td>Rough Texture &amp; Paper &amp; Bubble Plastic &amp; Rough Texture</td>
<td>90 %</td>
</tr>
</tbody>
</table>

VII. CONCLUSION

In order to recognize objects from their physical properties by a humanoid and via multi-modal artificial skin, we propose two groups of biologically inspired and robust feature descriptors. These universal descriptors provide the learners with high informative and very abstract information. By employing them, there is no need to reduce the dimensionality of data with further data processing (i.e. using Principal Component Analysis, PCA). Therefore, they are appropriate for the real-time task. Moreover, these feature extractors can provide robust information from non-stationary tactile signals where an object has a non-uniform or irregular texture which other methods like the Fourier transform fail to identify. The experiments conducted on the NAO humanoid adopting sliding and non-sliding exploratory behaviors showed promising results. Actually the robot was able to distinguish unknown objects having five different everyday surface textures and two different weights, both with 100% accuracy. In addition it could classify and cluster unseen objects of various compliance successfully.

ACKNOWLEDGMENT

This work is supported by the European Commission under grant agreements PITN-GA-2012-317488-CONTEST.

REFERENCES