# A Novel Dynamic-Vision-Based Approach for Tactile Sensing Applications

Fariborz Baghaei Naeini [iD], Aamna Alali [iD], Raghad Al-Husari [iD], Amin Rigi [iD], Mohammad K. AlSharman [iD], Dimitrios Makris [iD], and Yahya Zweiri [iD]

*Abstract*—In this paper, a novel Vision-Based Measurement (VBM) approach is proposed to estimate the contact force and classify materials in a single grasp. This approach is the first event-based tactile sensor which utilizes the recent technology of neuromorphic cameras. This novel approach provides a higher sensitivity, a lower latency, and less computational and power consumption compared to other conventional vision-based techniques. Moreover, Dynamic Vision Sensor (DVS) has a higher dynamic range which increases the sensor sensitivity and performance in poor lighting conditions. Two time-series machine learning methods, namely, Time Delay Neural Network (TDNN) and Gaussian Process (GP) are developed to estimate the contact force in a grasp. A Deep Neural Network (DNN) is proposed to classify the object materials. Forty-eight experiments are conducted for four different materials to validate the proposed methods and compare them against a piezoresistive force sensor measurements. A leave-one-out cross-validation technique is implemented to evaluate and analyze the performance of the proposed machine learning methods. The contact force is successfully estimated with a mean squared error of 0.16 N and 0.17 N for TDNN and GP respectively. Four materials are classified with an average accuracy of 79.17% using unseen experimental data. The results show the applicability of event-based sensors for grasping applications.

*Index Terms*—Vision-Based Measurements, Force Estimation, Material Classification, Haptics, Dynamic Vision Sensor.

## I. Introduction

**H**UMAN sense of touch comprises different receptors to acquire a rich information about objects' properties. The receptors in the human skin evaluate the contact area between the objects and body with respect to the applied force, temperature, and pain sensation [1]. Besides the skin receptors, other senses such as vision and audition assist human to extract further properties such as shape, material and hardness. Combining all of these acquired information enables humans to perform gripping tasks for unknown objects with a high degree of robustness. For instance, human can grasp objects robustly without a prior knowledge of the object properties such as weight, friction coefficient and materials [2]. Research works on tactile sensing involves measuring the

contact force magnitude and direction, mapping the force distribution, classifying the materials, and detecting slippage by monitoring the physical properties of the contact area [3].

Different types of tactile sensors were developed to simulate humans sense of touch for robotic hands [4]. A remarkable progress has been made for various types of tactile sensors. A capacitive sensor in [5] considers four capacitors to measure the horizontal and vertical forces. Although the sensor can measure the displacement in different directions, electronic interference and sensor hysteresis are remained unsolved. An application of object's localization and orientation estimation is demonstrated in [6] using piezoresistive force sensors. The sensor is capable of localizing the objects with a high accuracy while the estimation of the object orientation has a poor resolution. Another piezoresistive sensor with high durability and low hysteresis is developed in [7] to measure pressure under cyclic loading. A magnetic tactile sensor is proposed in [8] with capability of slip detection as well as estimation of the contact force in three dimensions. Some other approaches are utilizing the tactile sensors to extract further information about the objects properties and classify the materials. For instance, a piezoelectric multifunctional sensor is used to acquire objects' hardness by rolling over the sensor on the objects surfaces [9].

A hybrid sensor in [10] composes of piezoelectric transducers, force sensor and inclinometer in order to classify six different materials. An artificial finger with embedded PolyVinyliDene Fluoride (PVDF) membrane and strain gauge sensors are used to classify various materials [11]. Another research is presented in [12], whereas two piezoresistive tactile sensors are utilized to classify softness of vegetable using a decision-tree machine learning technique. Other applications of object classifications using different types of force sensors and traditional machine learning algorithms such as Support Vector Machines (SVM) are presented in [13], [14]. However, most of the aforementioned tactile sensors have limited resolution, considerable hysteresis and high sensitivity to the electromagnetic disturbances. This research focuses on optical tactile sensing techniques including camera-based methods that provide higher resolution, low hysteresis and resistant to electromagnetic disturbances.

Optical sensors with transparent elastomer or rubber, wave emitters and receivers have been developed for precise tactile sensing applications [15], [16]. Wave emitters scatter the light to the surface of the elastomer while receivers capture the back-scattered beams from the surface. Study of the reflected

beams regarding the interruption, phase, and magnitude apprises the distribution of force on the surface. The main advantages of this approach are immunity of the optical sensors to high electromagnetic disturbances, providing a high spatial resolution, flexibility and durability with high speed of signal transmission. In earlier work in [17], the optical tactile sensor is developed to measure the displacement and surface roughness with a high spatial resolution using artificial neural Networks. One of the main approaches in optical sensing is to place optical fibers in the finger membrane and employ techniques like intensity modulation, Bragg grating, and specklegram. In [18], a Fiber Bragg Grating (FBG) sensor is proposed which has high sensitivity, but low spatial resolution of 5mm. Another FBG-based tactile instrument is suggested in [19] to map the force distribution with a minimum weight sensitivity of 0.05 kg. A new class of optical tactile sensors are presented in [20] considering PolyDiMethylSiloxane (PDMS). A high sensitivity for measuring minimum weight of 0.005 kg is demonstrated practically. Furthermore, a technique is offered to detect an objects' shape and surface roughness with the sensors. Most of the optical tactile sensors have a lower spatial resolution compare to vision-based techniques.

Recent developments in visual technologies made cameras available in smaller sizes, lower cost, and higher resolution. Furthermore, advancements of processors and computational devices enabled the cameras to be considered as a VBM instrument to measure physical properties, localization of the objects, and to classify materials [21]. Camera-based tactile sensors observe the contact area, object, and elastomer surface to detect slippage and estimate the applied force. In earlier work in [22], a camera is combined with a force sensor to estimate deformation of the elastic object. A camera-based tactile sensor is introduced in [23], which measures three-dimensional force vectors on the contact area. Similarly, a marked elastomer is embedded in the silicone membrane in [24], to estimate force as well as friction coefficient of the surface. Further, a vision-based sensor is developed in [25] to evaluate multi-dimensional force vectors and the object stick ratio. In [26], three states for the sensor have been considered for non-contact detection, stick and slippage of the object. Moreover, a camera, a textured elastomer and a light diode are used in [27] to compute force magnitude and find directions for several rigid and soft contacts. Other sensors demonstrated the capability of vision-based sensors to deal with deformable objects and detect slippage to feedback the force controllers of the grippers [28]. In another approach [29], a conventional camera is used to estimate force magnitude and direction utilizing hemispherical markers. The markers are located inside the elastomer with different colors to estimate force magnitude and direction based on markers displacement. Later on, this approach became more popular which reflects the significant progress in camera-based tactile sensors [30]–[33].

A different approach to reconstruct a 3D geometry map in order to estimate force using Recurrent Neural Networks (RNN) with a pair of cameras was presented in [34], [35]. Further, a stereo vision-based sensor is suggested in [36] to estimate the contact force between tools and body tissues for surgical applications. Another stereo vision-based measurement instrument [37] inspects the automotive rubber profiles. An interesting application of visuo-haptic sensors in [38] allows the robot to measure the contact force and shape considering a foam rod.

A hybrid method is developed in [39] which considers a camera and a tactile sensor to classify materials considering multimodal learning. Two methods are implemented to classify 17 different materials considering both supervised and unsupervised learning. The best performance in both supervised and unsupervised techniques are achieved through the Mean Maximum Covariance Analysis ($\mu$MCA) with a high accuracy.

Most of the tactile sensors focus on force measurement under stable and static conditions, i.e. without dynamic variation of the applied forces. However, in many applications including robotic grasping, applied forces may vary significantly and a fast response is required to properly handle the grasped object. Even-though many VBM instruments and hybrid techniques have been contributing significantly in the field of tactile sensing, no attention has been paid to utilize neuromorphic vision sensors in this field. For the first time, in [40], we proposed a vision-based sensor to detect incipient slip using DVS which provides a low latency with low power consumption. It is demonstrated that the sensor can detect incipient slip in grasping applications with an average of 44.1 ms using traditional image processing methods. The sensor successfully detects the incipient slippage without a prior knowledge of the objects properties or friction coefficients. Table I lists a different techniques for tactile sensing and measurements applications.

In this paper, a new dynamic-vision-based tactile sensor is introduced. The sensor is based on a neuromorphic camera which provides a higher time-resolution, a lower latency, less computational cost and power consumption compared to other vision-based techniques.
The contributions of this paper are as follow:

1) To the best of our knowledge, it is for the first time, a Dynamic-Vision-Based approach is proposed to measure the contact force and classify materials in a grasp using a neuromorphic camera (DVS).
2) A Time Delay Neural Network (TDNN) and a Gaussian Process (GP) are developed to find the correlation between the triggered events and the contact force.
3) A Deep Neural Network (DNN) model is implemented to classify materials based on the triggered events in a single grasp.

The rest of the paper is structured as follows. Section II describes the proposed event-driven sensor prototype. In Section III, time-series models are designed to estimate force in a single grasp. In Section IV, a DNN is developed to classify different materials which is followed by the validation and results in Section V. The results are analyzed and discussed in Section VI. Finally, Section VII concludes the paper and suggest future works.

## II. EVENT-DRIVEN TACTILE SENSOR

Vision-based tactile sensors estimate the applied force on the surface of the fingertip (silicone membrane) by observing

Table I: A list of different tactile sensing techniques for various applications

| Reference | Sensors | Purpose | Specifications |
|---|---|---|---|
| [5] | Capacitive | Measure the the force vector (3D) | •Low parasitic capacitance effect<br>•Resolution of 12.5 $\mu m$<br>•A considerable hysteresis |
| [13] | Capacitive and actuators | Object classification | •Objects are varied in shape, size and stiffness<br>•High accuracy of object recognition<br>•Non-time series machine learning technique |
| [10] | Piezoelectric transducers | Classify materials | •Time and frequency domain analysis<br>•Multiple environments performance<br>•High accuracy with static applied force threshold |
| [7] | Piezoresistive | Measure the pressure | •High linearity factor<br>•100 cycles hysteresis<br>•Time analysis of the measurements |
| [6] | Laser and piezoresistive | Object recognition and orientation detection | •High classification accuracy<br>•Limited orientation measurements<br>•Non-time series machine learning technique |
| [17] | Optical fiber | Measure displacement and surface roughness | •Measuring range of $\pm(0.8\ mm)$<br>•Displacement error of $\pm(0.5\ \mu m)$ |
| [20] | Optical fiber | Measure the force and shape detection | •High sensitivity of 0.005 $kg$<br>•Latency of 600 $ms$<br>•Limited range of few grams |
| [30] | Conventional camera | Measure the force vector (3D) | •High spatial resolution (4 $mm$)<br>•Low sampling rate of 30 FPS<br>•Static approach based on markers displacement |
| [31] | Conventional camera | Estimation of slippage angle and stick ratio | •High spatial resolution<br>•Low sampling rate of 30 FPS<br>•Static approach based on markers displacement |
| [40] | DVS | Incipient slip detection | •High spatial resolution<br>•Angular view of camera with 44.1 ms response time<br>•Traditional image processing algorithms |
| This work | DVS | Force estimation and material classification | •High spatial resolution of 0.04 $mm^2$<br>•Force measurement range of 0.15-3.7 N<br>•Sensitivity of 0.01 N<br>•Low computational cost and logical response time of 21 ms<br>•Maximum size of the contact region: 48x36 mm |

the contact area directly or indirectly. Direct methods monitor the contact area between the object and fingertip whereas indirect approaches observe extra elements inside the fingertip. Some of the indirect methods use the markers inside the silicone membrane and track the displacement of the markers [30], [31], [41]. Due to the flexibility of the elastomer, the markers are displaced when a force is applied to the object. The contact force vector can be calculated based on the elasticity theory considering the displacement of markers. In [42], the force vector is formulated by assuming linearity and uniformity of the elastomer.

Although camera-based tactile sensors provide a high resolution of the contact area, a low sampling rate (normally 30 FPS) and limited dynamic range reduce the sensor performance in unknown environments. To satisfy requirements for fast grasping applications, it is essential to consider a sensor with a higher sampling rate and sensitivity. Furthermore, processing images is often involved with a lot of redundant pixels which adds a further computational and memory requirements to the system. Thus, the research in this paper considers an event-based sensor with a high dynamic range and a low latency to observe the contact area.

Asynchronous event-based cameras are bio-inspired sensors that consider intensity changes (events) in the scene with precise timestamps and a high dynamic range. DVS is one of the well known frame-free sensors with a high temporal resolution of few microseconds [43], [44] which is significantly faster than ordinary cameras. This vision sensor captures intensity changes logarithmic ally at each pixel rather than capturing the whole scene in a fixed interval. Each pixel compares the current intensity value with the previous value repeatedly (in microseconds). If the compared value exceeds a threshold level, then the sensor fires either a positive or a negative event. DVS has a resolution of $240 \times 180$ pixels with a latency of 12 microseconds for the mean of 20 events. The sensor streams positive and negative events with precise timestamps and pixel location $(x, y)$. Moreover, DVS requires a lower power (4-15 mW) and memory compared to conventional cameras.

The threshold level of the events is a crucial parameter for filtering noise and changing the sensitivity of the sensor. In this work, several threshold levels are examined to balance the noise level and sensitivity of the sensor. The relationship between the triggered events and intensity changes is logarithmic which is formulated in [43]. Equation 1 presents the correlation of temporal contrast (TCON) and photo-current ($I$). A threshold is considered for the temporal contrast to fire

positive (higher intensity) or negative (lower intensity) which can be modified to change the sensor sensitivity and filter the noise.

$$TCON = \frac{d(ln(I(t)))}{dt} \qquad (1)$$

To establish the event-driven tactile sensor, a semi-transparent silicone fingertip is located between the object and gripper. When a force is applied on the object, the silicone deforms due to the elasticity. Therefore, DVS captures the changes within the contact area and triggers both positive and negative events. The silicone membrane is molded with dimension of $4.0 \times 2.0 \times 0.2$ cm to cover the contact area between the gripper plane and the object. The silicon properties and depth of the fingertip have a significant impact on the sensor sensitivity and the range of force estimation. Figure 1 illustrates a diagram of the sensor including transparent grippers, a semi-transparent silicone fingertip, DVS, and the object.
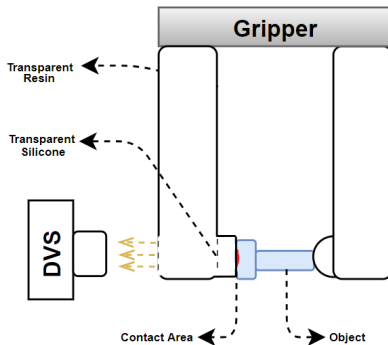


Figure 1: Event-based tactile sensor diagram

## III. FORCE ESTIMATION

Camera-based approaches consider different techniques to estimate magnitude and direction of the contact force. For instance, the force magnitude and direction are estimated by detecting the markers displacement in the elastomer using a single frame-based camera [30], [31], [41]. On the other hand, a number of researches investigate the force estimation using geometrical reconstruction techniques [34]–[36]. Frame-based stereo cameras are used to reconstruct a depth map within the contact area to estimate magnitude and direction of the applied force. In this paper, a novel event-driven (dynamic) method is presented to estimate force by observing the intensity changes within the contact area.

### A. Concept

As mentioned in Section II, DVS fires either positive or negative events depending on the intensity changes in the scene. Since a semi-transparent silicone has an opaque surface, the contact area is barely visible prior to the contact of an object to the membrane. Due to the deformation of the silicone membrane, the visible part of the contact area becomes larger by applying more force, and intensity of the contact area increases significantly.

Accordingly, an increase of the applied force triggers the negative events while a decrease of the applied force triggers positive events. In this paper, positive and negative events are presented in green and red respectively.

Deformation of silicone under a pressure is highly non-linear which depends on the type and size of the membrane as well as the range of the applied force. Other factors such as direction of force, shape of the contact area and temperature can affect this relationship. Consequently, the correlation between events and the contact force is highly non-linear considering the following parameters: (i) The deformation of silicone membrane [45]; (ii) The logarithmic relation between changes in intensity and triggered events which is presented in (Equation 1).

To visualize the correlation of triggered events to the contact force, events are accumulated over a time interval whereas the applied force increases significantly. Figure 2 represents the triggered events and image of the contact area where the contact force is increased.
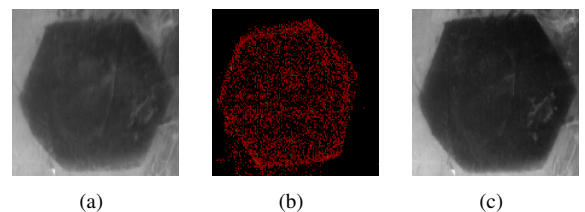


Figure 2: (a) Image of the contact area when a low amount of force is applied. (b) Accumulation of events over a 40 ms time window during a grip. (c) Image of the contact area when a high amount of force is applied.

As it can be observed in Figure 2(b), events are triggered in most regions of the contact area by applying force to the object. On right bottom of the contact region, a number of events are triggered due to the noise and a slight displacement of the silicone membrane.

### B. Grasping Procedure

The contact force estimation from DVS events can be approached as a time series regression problem. A single grasp can be divided into three main phases: (i) Grasping phase; (ii) Holding phase; (iii) Releasing phase. The contact force changes significantly in the grasping and releasing phases while in the holding phase the force variation is related to the vibration. This research aims to estimate the contact force in both grasping and releasing phases without consideration of the holding phase (vibration). The grasping phase includes the first instance where the object touches the silicone membrane and the applied force is increased until it reaches a constant value (holding phase).At the first instance when the object touches the membrane, a lot of negative events are triggered due to the intensity changes in all of the correspondent pixels of the contact area. Therefore, a first touch is determined when the first significant number of negative events are triggered.

Once the contact is obtained, negative and positive events represent the changes in the applied force and vibration of the object. It is noteworthy to mention that after a certain

amount of increase in force, the silicone membrane reaches a saturation point where further increase of the applied force does not deform the membrane. Hence, the sensor can estimate a limited range of force which depends on the silicone membrane properties which is a case for all the camera-based tactile sensors with a silicone membrane.

In the holding phase, the applied force varies due to vibration and noise which are not considered in this research. Finally, in the releasing phase, the applied force is decreased which leads to trigger positive events within the contact area. Figure 3 illustrates the normalized value of events and the measured force in a single grasp over a time.

Both number of events and the measured force are framed over 7 ms intervals. The framing process helps to differentiate meaningful events and reduce the impact of noise over a longer period. In an ideal grip, the grasping phase must include only negative events. However, the object vibrates slightly in a short amount of time to reach stability which causes triggering the positive events. There is a trade-off between the filtering of the unwanted events and the sensor sensitivity which can be adjusted by changing the DVS threshold.
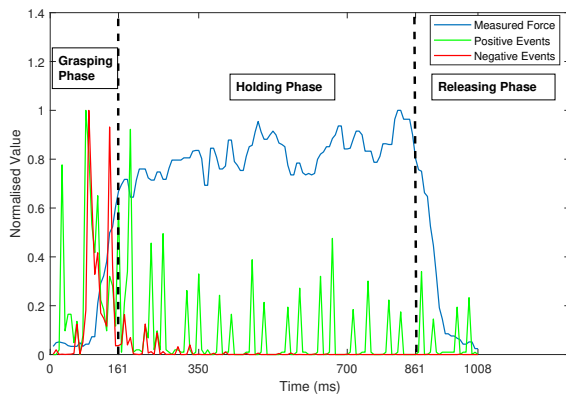


Figure 3: Normalized value of the force which is measured by a piezoresisitve force sensor (blue), number of negative events (red), number of positive events (green) in a single grasp.

In Figure 3, the first peak in the negative events represents the first touch of the object and the fingertip. The peaks of the negative events indicate a significant increase of the contact force. The threshold of the positive events are lower than the negative events which makes the sensor to have a higher sensitivity for the decrease of the contact force. In an ideal grasp, only negative events are expected to trigger by applying more amount of force. However, vibration and instability of the grasp results in triggering positive events as well as negative events. The first significant spike in the releasing phase demonstrates a loss in the contact area which leads to the object slippage.

To correlate the triggered events and the contact force, a robust time-series learning technique is required to capture the non-linear relationship over a time. In this work, TDNN and GP models are chosen since these models are able to model time-series data with non-linear relationship between variables. Accordingly, the force values are measured by a

piezoresistive force sensor at each time interval to train and test the machine learning methods. It should be noted that the measured force values are used to train the models and the accumulation of triggered events are considered as the inputs to the models.

### C. Time Delay Neural Networks

One of the well-known time-series machine learning models is TDNN which is widely applied on speech recognition, text recognition and regression problems [46]. The main advantage of TDNNs is the ability to relate temporal sequences to each other, enabled by their main characteristic, the delay nodes. Number of delay nodes is a crucial parameter in TDNN network which specifies a time interval to capture patterns of a signal.

Another advantage of TDDNs is a faster training time compared to Recurrent Neural Networks (RNNs) due to the constant value of the delay nodes. The delay node in TDDNs is often assigned to a short time to capture the relationship of the current point with the previous observations. A fixed and small number of time delay nodes helps to avoid vanishing gradient problem which is a well-known problem in traditional RNNs with long-term dependencies [47].

In this research, a variety of networks with different number of hidden layers and neurons are tested to find the best architecture. The events are accumulated over time to provide a full memory of the sequences to the network which improves the detection of global features. The accumulation of positive and negative events are passed to the network separately to identify decrease and increase of the contact force respectively. Followed by the input layer, $k$ fully-connected hidden layers with $n$ neurons in each layer are considered to capture the non-linear relationship between the events and the applied force.

The sigmoid activation function is assigned to all hidden layers after some initial experimentation. A variety of experiments are performed to find appropriate parameters for the model to achieve a good performance. In Section V, a variety of network architectures are analyzed comprehensively to investigate the impact of the number of neurons and hidden layers on the network performance. Figure 4 demonstrates the deep TDNN network for the force estimation. A cost function is defined based on the error of the estimated force from events in comparison to the measured force using a piezoresistive force sensor. To optimize the error of the network, the cost function can be approached as a minimization of the sum of squares of a non-linear function (Equation 2) whereas $F(x)$ represents the error function. One common approach to solve a minimization problem is the Levenberg-Marquardt (Equation 3). The Levenberg-Marquardt uses the Gauss-Newton method to search for a direction in order to decrease $F(x)$ at each iteration. In Equation 3, the Jacobian matrix and damping factor (non-negative scalars) are denoted as ($J$) and ($\lambda$) respectively. Damping factor is considered as 0.01 which is multiplied by an identity matrix ($I$) to vectorize the parameter. The Levenberg-Marquardt method searches the directions which is given by a solution
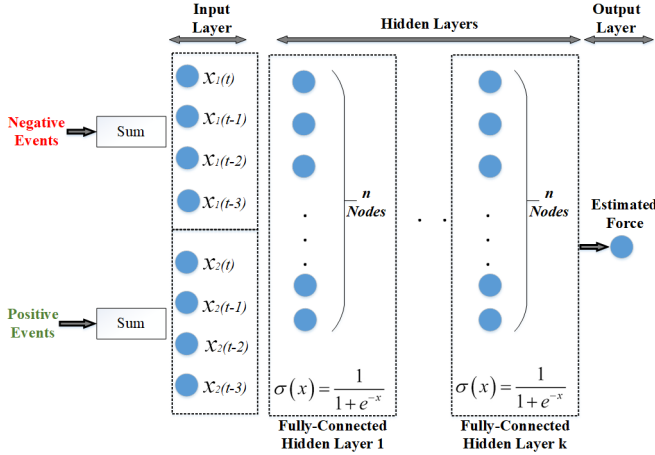
Figure 4: A deep TDNN model with a time delay of three nodes (21ms) to estimate force from accumulative events.

$(p_k)$.

$$min(f(x)) = \sum_{i=1} F_i^2(x) \quad (2)$$

$$(J(x_k)^T J(x_k) + \lambda_k I)p_k = -J(x_k)^T F(x_k) \quad (3)$$

### D. Gaussian Process

Gaussian Process (GP) is a stochastic modeling method to predict and forecast variables based on combination of random variables over the data points. A variety of kernels can be considered to fit a function with random variables corresponding to the multivariate normal distribution [48]. The choice of the kernels and hyper-parameters have a significant impact on the model to estimate the function between inputs and outputs. As mentioned in Section II, the triggered events and intensity changes in the scene have a logarithmic relationship. On the other hand, the silicon membrane behaves non-linearly over different contact forces. Therefore, we consider the Automatic Relevance Determination (ARD) squared exponential covariance kernel to build a robust model in order to find a highly non-linear correlation between events and the contact force. Equation 4 presents the kernel function whereas $x_i, x_j$ are two inputs, $\sigma_f$ is the signal standard deviation, and $\theta$ represents the parameterized version of the covariance function. Each predictor ($m$) can have a different length scale ($\sigma_m$) whereas $m = 1, 2, \cdots, d$.

$$k(x_i, x_j | \theta) = \sigma_f^2 exp[\frac{-1}{2} \sum_{m=1}^{d} \frac{(x_{im} - x_{jm})}{\sigma_m^2}] \quad (4)$$

One of the most important hyper-parameters in a GP model is the length scale ($\sigma_m$) which affects the model performance significantly. In order to optimize the length scale, a Bayesian optimization technique is performed over 10 iterations to find the best length scale. Afterwards, the length scale with the best performance is selected and replaced in the kernel function. To improve the GP performance to estimate force in a grasp, timestamps are also passed to the GP model as an input. In other words, the GP model considers positive events, negative events, and time to estimate the contact force.

## IV. MATERIAL CLASSIFICATION

Acquiring information about the object properties such as material, friction coefficient, stiffness and weight facilitate the grasping process. This research proposes a novel technique to classify objects materials using DVS events from the grasping and the releasing phases. In the proposed classification method, the input consists of two features (accumulation of positive and negative events) for all the sequences of each grip for different materials. Figure (5) illustrates accumulation of events for different materials in a single grasp considering a similar range of applied force. It can be observed from Figure



(a)



(b)
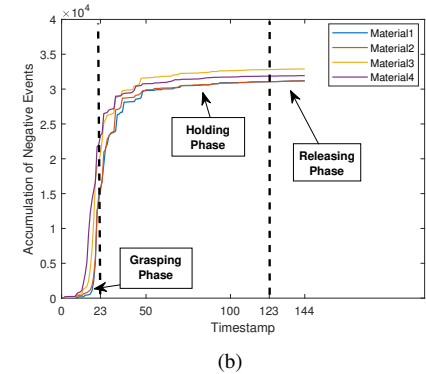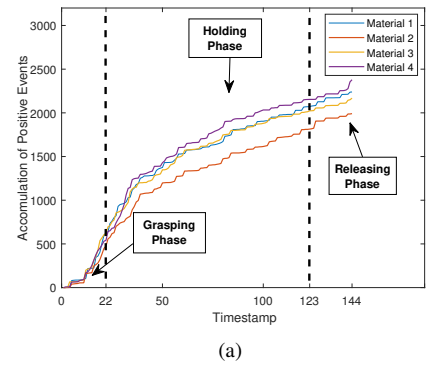
Figure 5: Four different materials are considered in a single grasp with a similar range of the contact force. (a) Accumulation of positive events. (b) Accumulation of negative events.

(5) that the accumulation of negative and positive events are distinguishable for different stiffness in a similar range of the applied force. The objects and the silicone membrane deform differently for each material during the grasping phase and the releasing phase. The number of positive and negative events follow different patterns for each object. Other factors such as background noise and shape of the contact area affect the number of events.

A Deep Neural Network (DNN) model is developed to classify materials considering the grasping and releasing phases. The network consists of $k$ fully-connected hidden layers and $n$ neurons in each layer. The sigmoid activation function is selected for all the layers after initial experiments. Furthermore,

a soft-max function is used in the output layer to classify different materials. The Scaled Conjugate Gradient (SCG) back-propagation method is used to train the network. Figure 6 demonstrates the architecture of the proposed network for the material classification.


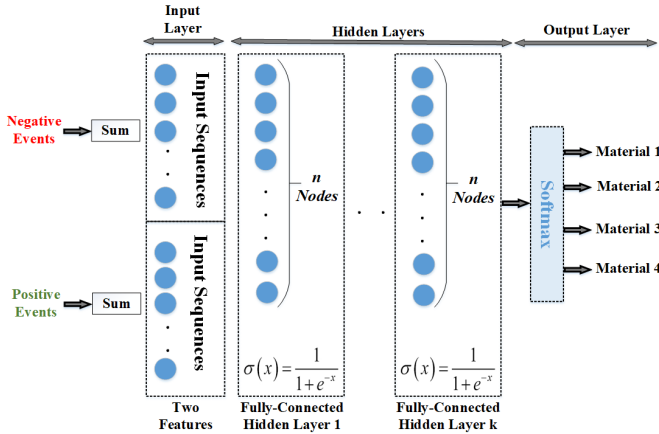
Figure 6: A DNN model for material classification

## V. VALIDATION AND RESULTS

This section describes the validation of the proposed models for force estimation and material classification. Detailed information on the experimental setup, data collection, synchronization process, as well as experimental results are provided in the following sections.

### A. Experimental Setup And Data Collection

The experiments are designed to grasp four objects with different Young's modulus: (i) Foam; (ii) Rubber; (iii) Silicon; (iv) Steel. All the objects are formed in a same shape (hexagon) and dimension $(0.75 \times 0.65 \times 3.55)$ cm. To increase the contrast with an opaque surface of the silicone membrane and eliminate the effect of light reflection for different materials, all the objects are colored in black. Since the objects are in the same shape and color, the classification method only relies on the elasticity of the objects rather than the objects texture or color.

The gripper consists of a static and a dynamic transparent finger. In each experiment, each object is gripped and a constant pressure is applied to hold the object for 700 ms. Then, the gripper returns to the starting position to release the objects. The DVS sensor is located in a distance of 5 cm from the static finger to minimise the noise and capture the changes in the contact area. A lens with 4.5 mm focal length is mounted on the camera which can be adjusted regarding the size of the objects. The linear horizontal field of view of the lens corresponds to 9.8 cm in 10 cm distance. In this setup, each pixel of the scene corresponds to 0.04 mm$^2$ area on the silicon surface. Consequently, the maximum sensing region is 48x36 mm which can be changed by use of different lens, changing the camera position and silicone thickness. On the dynamic finger, a piezoresistive force sensor (FlexiForce-A201) is located to measure the contact force.

The dynamic finger is controlled by a servo motor (AX-12A Dynamixel) using a micro-controller (Arduino) to control the gripper acceleration and position. The force is applied to the object by the dynamic finger with an angle of 15° with respect to the $z$-axis. Figure 7(a) demonstrates the experimental setup: The DVS observes the contact area through a static finger of the gripper. Figure 7(b) illustrates the contact area from the view of the DVS.
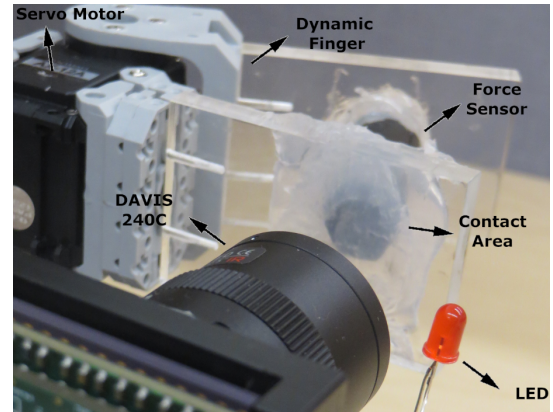


(a)
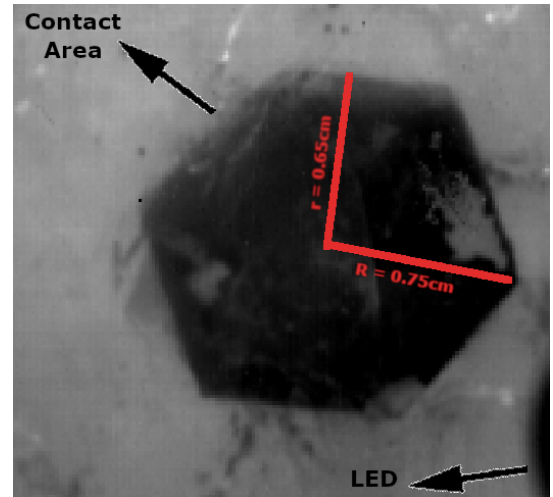


(b)

Figure 7: (a) Experimental setup includes the piezoresistive force sensor, DAVIS 240C, AX-12A Dynamixel servo motor, and two transparent finger. (b) The image of the contact area from the DVS point of view.

### B. Force Sensor and Synchronization

A piezoresistive force sensor of FlexiForce A201 type is used as a tactile sensor to validate the proposed event-based sensor. The force sensor has a response time $< 5$ $\mu$s, percentage error $\pm < 3\%$, hysteresis $< 4.5\%$ of full scale and is adjusted to measure forces from $0 - 111$ N. Moreover, experiments are performed for a range from 0 to 3.7 N. This range is selected based on the saturation of silicone deformation. The force sensor is covered by a silicone layer in order to mimic the same friction coefficient on both sides

of the objects.

A Light-Emitting Diode (LED) is used to synchronize the DVS camera and the force sensor. In each experiment, the LED is turned on and after few milliseconds off prior to the grasp. When the LED is turned off, the time is recorded by the micro controller to start recording the force measurements. This time is also detected by the DVS by finding a significant spike of negative events (LED OFF) in the scene. Afterwards, the artificial frames are constructed by accumulation of positive and negative events during 7 ms window. Each experiment is divided into the grasping phase (from the $1^{st}$ frame to $22^{nd}$ frame), the holding phase (from $23^{rd}$ frame to $122^{nd}$) and the releasing phase (from $123^{rd}$ frame to $144^{th}$ frame).

The measured force varies significantly due to the vibration and movement of the dynamic finger. A third order median filter is applied to smoothen the force values and filter the noise. Figure 8 illustrates the distribution of force over 48 experiments each of them is captured for all timestamps.

As shown in Figure 8, the contact force decreases significantly
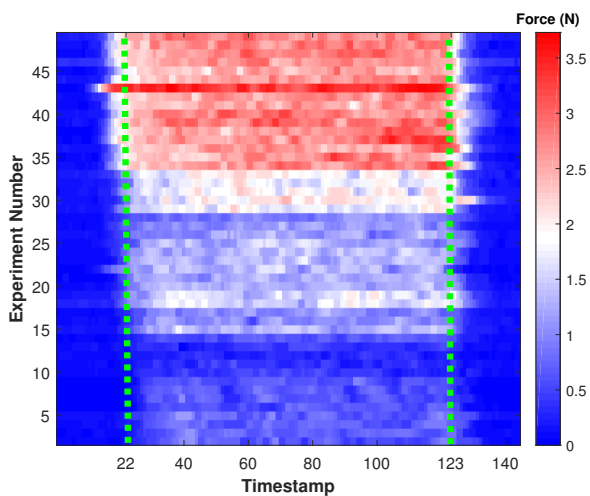


Figure 8: Each row represents an experiment over 144 timestamps whereas the color indicates the contact force at each point. The dotted green lines show the signal clipping boundaries for the grasping, holding and releasing phases

during the releasing phase. Since the force sensor is mounted on the dynamic plane and it is covered by a silicone layer, the measured contact force is small but non-zero.

## C. Force Estimation

This section presents and analyzes the results of the proposed models for the force estimation. The most common approach to evaluate a machine learning method is to partition the data into training and test subsets. The model design and hyper-parameters are tuned to achieve the highest performance on the test set. The training set is given to the machine learning method to find appropriate hyper-parameters in order to minimize the error. The test subset (unseen data) is not involved in the training process. A significant disadvantage of this approach is that researchers changes the model design and hyper-parameters based on the assessment on the test subset.

Therefore, the test subset is in-directly involved in the design of the method which makes a bias in this process.

Another approach is to divide the dataset into three different partitions (training, validation and test). The validation set assists the training process to stop when the network reaches the saturation point, and therefore, it reduces the time of the training process. Afterwards, the hyper-parameters are optimized on the validation set. An appropriate machine learning model and kernels can be selected by considering the method performance on the validation set. Finally, the test subset only is used to report the performance of the network rather than finding the optimum model and hyper-parameters.

In this paper, the data is divided into three subsets: 87.5% for training (forty-two experiments), 10.4% for validation (five experiments) and 2.1% for test (one experiment). The five experiments in the validation set are selected randomly from a wide range of forces to make sure that all possible values of the applied force are covered. Furthermore, an exhaustive leave-one-out cross-validation method is deployed to test each experiment individually over 48 folds. The leave-one-out method provides a comprehensive evaluation by testing the models on all of the experiments individually.

The TDNN error is calculated over all the folds (48 folds) and the average of Mean Squared Error (MSE) is calculated to compare different network architectures. To find the optimal architecture, number of neurons and hidden layers are varied. All the weights in the network are initialized randomly and biases are set to zero at the first place. The networks are trained in parallel on a CPU with double precision (Corei7-8700 6cores) using MATLAB neural network toolbox. Table II demonstrates the average MSE over all folds for different number of hidden layers ($k$) and neurons ($n$). The lowest

| $k/n$ | $n=5$ | $n=10$ | $n=15$ | $n=20$ | $n=25$ | $n=30$ | $n=35$ | $n=40$ |
|---|---|---|---|---|---|---|---|---|
| $k=1$ | 0.23 | 0.19 | 0.18 | 0.19 | 0.20 | 0.18 | 0.16 | 0.16 |
| $k=2$ | 0.21 | 0.19 | 0.16 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 |
| $k=3$ | 0.19 | 0.20 | 0.18 | 0.18 | 0.16 | 0.17 | 0.17 | 0.17 |
| $k=4$ | 0.20 | 0.19 | 0.16 | 0.17 | 0.18 | 0.17 | 0.17 | **0.15** |
| $k=5$ | 0.20 | 0.16 | 0.20 | 0.18 | 0.18 | 0.16 | 0.17 | 0.16 |

Table II: Mean Squared Error of the estimated force(N) on the validation set whereas the lowest error is highlighted in bold.

validation error (0.15 N) is achieved through a network with 4 hidden layers and 40 nodes. Since the validation experiments are chosen from a wide range of forces, it is expected to achieve a generalized model for the force estimation. Note that choosing different experiments for the validation partition changes the performance of the network. Table III presents the average MSE over 48 folds for the sequences of the unseen experiments. The average MSE is highlighted for the proposed

Table III: Mean Squared Error of the estimated force(N) on the test set whereas the error of the proposed network architecture is illustrated in bold.

| $k/n$ | $n=5$ | $n=10$ | $n=15$ | $n=20$ | $n=25$ | $n=30$ | $n=35$ | $n=40$ |
|---|---|---|---|---|---|---|---|---|
| $k=1$ | 0.16 | 0.24 | 0.15 | 0.22 | 0.19 | 0.16 | 0.24 | 0.15 |
| $k=2$ | 0.17 | 0.16 | 0.16 | 0.16 | 0.17 | 0.18 | 0.15 | 0.35 |
| $k=3$ | 0.18 | 0.18 | 0.16 | 0.17 | 0.16 | 0.16 | 0.16 | 0.16 |
| $k=4$ | 0.18 | 0.17 | 0.16 | 0.16 | 0.16 | 0.17 | 0.19 | **0.16** |
| $k=5$ | 0.18 | 0.16 | 0.24 | 0.17 | 0.20 | 0.15 | 0.24 | 0.18 |

network architecture which is the second best accuracy overall. Similar performance of the network for both validation and test partitions indicates a good generalization of the network.

Figure 9 illustrates the multilayer TDNN response (red) and the measured force (ground truth) for an experiment tested on unseen data considering leave-one-out cross-validation method. As it can be observed, the estimated force follows the measured force pattern with a high accuracy during the grasping phase whereas the estimated force drops to a steady level. In the beginning of the releasing phase, the object loses all of the contact area with the fingertip which leads to a significant spike in number of triggered positive events. After this moment, a slight number of events are fired which indicates environment noise. Therefore, the network recognizes the frames that the object is not in contact with the fingertip and it remains steady. Since the force sensor is mounted on the dynamic plane of the gripper, the measured force is affected by noise due to the motion of the gripper in the releasing phase. Moreover, the force sensor hysteresis adds a further delay to the measured force over the time. Consequently, the amount of measured force is decreasing slower over the time rather than a sharp drop at the first frame of the releasing phase. To evaluate the proposed GP model, the same folds as
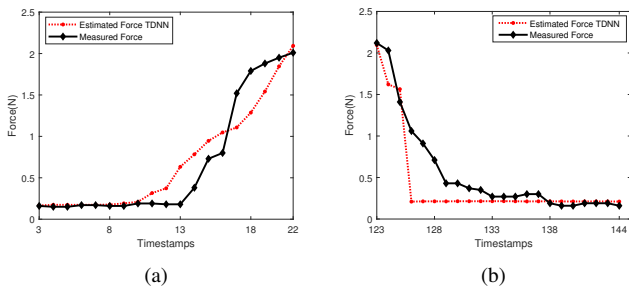


(a)          (b)

Figure 9: Measured force and estimated force for the TDNN model on the unseen experiment during the grasping phase (a) and the releasing phase (b).

TDNN are considered to allow the comparison of models. The Bayesian optimization is performed on each fold individually over ten iterations to tune the hyper-parameters. Figure 10 illustrates the estimated force by the GP model for the unseen experiment in one of the folds. The avaraged MSE of 0.17 N is achieved through the time-series GP method. The response of this technique appears to be able to estimate the force in the grasping phase with a high accuracy. In the releasing phase, the GP response decreases with a slight slope compare to the measured force in this selected fold. Since the number of triggered events are close to zero in the releasing phase, the GP method learns to estimate the measured force by considering the force values as a function of time. Figure 11 illustrates the averaged MSE and standard deviation for the estimated force on the all folds at each timestamp for both TDNN and GP.

To calculate the sensitivity of the sensor, the estimated force values on the test experiments are considered over all folds. Flexible piezoresistive sensors often have a lower accuracy and high level of noise within the low range of the applied
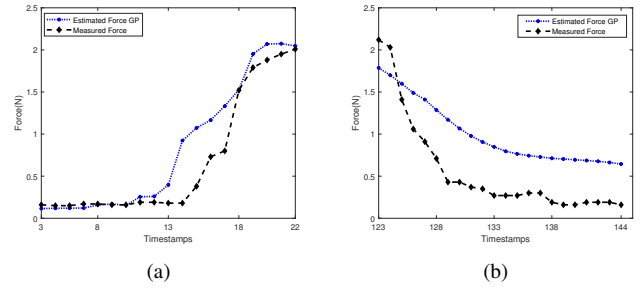


(a)          (b)

Figure 10: Measured force and estimated force by GP model on the unseen experiment during the grasping phase (a) and the releasing phase (b).
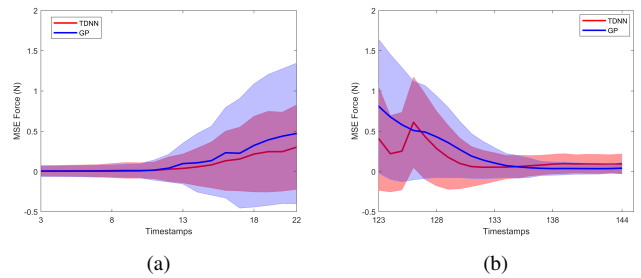


(a)          (b)

Figure 11: Red and blue lines present the average MSE of the estimated force over all folds at each timestamp during the grasping phase (a) and the releasing phase (b). The standard deviation of MSE is presented by a highlighted area over the average of MSE.

force [49]. Therefore, all the data points with the force value of less than 0.2 N are eliminated for the purpose of sensitivity calculation. The minimum value of changes in the estimated force for all the points over 0.2 N is determined for each fold individually. Finally, the minimum value of changes in the estimated force is averaged over all folds. The TDNN and GP models provide the sensitivity of 0.01 N and 0.02 N respectively. The TDNN model provides a slightly better sensitivity due to the high number of parameters which allow the network to learn a highly non-linear relationship between the contact force and events.

### D. Material Classification

Obtaining further information about the objects facilitates the grasping tasks. The elasticity of the objects is one of the key factors in differentiate objects. The proposed classification model classifies the objects with different Young's modulus considering the grasping and releasing phases. As mentioned in Section V-A, the objects are considered with same size and shape to eliminate these features in the classification process. Furthermore, all the objects are covered with a layer of black colour to minimize effect of light reflection and objects texture during the experiments. Therefore, our classifier is fed with fewer features which makes the classification task harder to learn a highly non-linear relationship between objects materials and the events independent of size, shape, location,

texture and color. The experiments are repeated for each material as the following: Foam (11), Rubber (9), Silicone (14), and Steel (14).

Four experiments are chosen for the validation set and leave-one-out cross validation is implemented to evaluate the classification network accuracy. Table IV represents accuracy of the network for different numbers of hidden layers and nodes. The

Table IV: Accuracy of the material classification on the validation data

| $k/n$ | $n$=5 | $n$=10 | $n$=15 | $n$=20 | $n$=25 | $n$=30 |
|---|---|---|---|---|---|---|
| $k$=1 | 80.21 | 76.56 | 83.33 | 84.38 | 85.94 | 83.85 |
| $k$=2 | 70.31 | 81.25 | 84.38 | 85.94 | 86.98 | **90.10** |
| $k$=3 | 64.06 | 77.60 | 81.77 | 85.42 | 86.98 | 88.54 |
| $k$=4 | 67.19 | 77.60 | 78.65 | 84.38 | 84.90 | 89.06 |
| $k$=5 | 60.42 | 68.75 | 76.56 | 81.25 | 86.46 | 90.10 |
| $k$=6 | 60.94 | 69.79 | 74.48 | 81.77 | 85.42 | 86.46 |

highest accuracy for the validation set is achieved through two models with 30 nodes. A higher number of neurons and hidden layers might lead to achieve a better result while increases the training and testing time significantly. Therefore, the network with 2 hidden layers (k=2) and 30 neurons (n) is selected to classify materials. Table V illustrates the accuracy of the proposed network for the unseen experiments over 48 folds.

Table V: Accuracy of the material classification model on the unseen data (test set)

| $k/n$ | $n$=5 | $n$=10 | $n$=15 | $n$=20 | $n$=25 | $n$=30 |
|---|---|---|---|---|---|---|
| $k$=1 | 70.83 | 68.75 | 62.50 | 70.83 | 62.50 | 68.75 |
| $k$=2 | 58.33 | 62.50 | 68.75 | 75.00 | 60.42 | **79.17** |
| $k$=3 | 50.00 | 58.33 | 56.25 | 72.92 | 72.92 | 77.08 |
| $k$=4 | 60.42 | 58.33 | 72.92 | 68.75 | 60.42 | 75.00 |
| $k$=5 | 50.00 | 45.83 | 54.17 | 62.50 | 68.75 | 72.92 |
| $k$=6 | 39.58 | 58.33 | 66.67 | 58.33 | 70.83 | 75.00 |

The highest accuracy (79.17%) for the unseen test data stands for the proposed network. Figure 12 demonstrates the confusion matrix for these experiments considering leave-one-out cross-validation method.

As observed in Figure 12, the rigid material (Steel) has the highest accuracy with only one error over all folds. The classification of soft materials with a closer Young's modulus is a more challenging process. The results indicate an average accuracy of 73.3% for Foam, Rubber and Silicone. As mentioned in Section I, many approaches consider multiple force sensors to classify materials and objects. A similar neural network approach is implemented over all folds to classify materials with different Young's modulus using the force sensor measurements instead. The best network indicates accuracy of 50% on the unseen data which is 29.17% lower than the accuracy of the event-based proposed sensor.

## VI. Discussion

The proposed TDNN achieved slightly higher accuracy than the GP model. The delay nodes in TDNN enable the modelling of temporal coherence of the sequences through a time window. As regards to Figure 11, both TDNN and
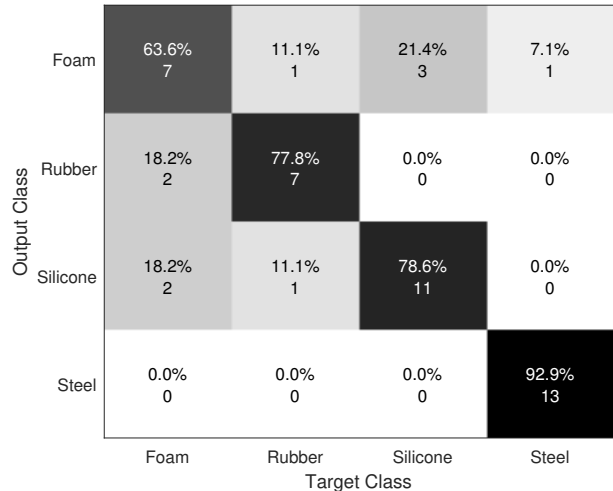


Figure 12: Confusion matrix for the material classification over the all folds with overall accuracy of 79.17%

.

GP methods identify the start of the the grasping and the releasing phases faster than the tactile sensor due to the force sensor hysteresis and experimental setup. The estimated force by TDNN drops rapidly to a low steady level, indicating the low latency of this method. Interestingly, this phenomenon is evident through most of the experiments. Figure 13 illustrates the same behaviour of the TDNN in a different fold.

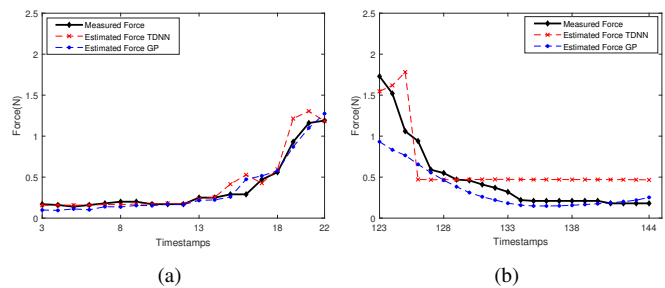

(a)                                          (b)

Figure 13: Responses of the estimated force and the measured force (ground truth) considering two different folds.

Unlike the TDNN, the GP response shows a slight decrease during the releasing phase. Even though the triggered events in this phase is close to zero, the GP model estimates the force as a function of time. Therefore, the GP model has a lower error than the TDNN where the number of the triggered events are low. However, the actual contact force must drop rapidly when the object releases. Since the force sensor is mounted on the moving gripper, the measured force includes noises until the gripper stops. Moreover, the force sensor hysteresis leads to have a considerable delay to measure the real contact force when the force varies significantly in a short amount of time. As presented in Figure 11, the MSE of TDNN is lower than the GP model in the grasping phase. Although the GP model has a better response for the last 8 timestamps whereas the measured force are not correlated with the triggered events,

this part of the releasing phase represents only noise since the object is not in contact with the gripper.

In this work, we consider a measurement range between 0.15-3.7 N for our sensor based on observations and initial experiments. We initially conducted experiments for 0-6 N and the proposed models are performed to estimate the force. Based on the selected silicone membrane, it is observed that the sensor was able to capture contact area changes for the contact forces up to 3.7 N. The silicone membrane properties such as stiffness, shape, and thickness as well as DVS threshold can be altered in order to change the measurement range of the sensor.

In order to compare our sensor with others, we consider main parameters of the sensors such as sensitivity, range and latency. In [27], a resolution of 0.05 N for a maximum range of 7 N is presented whereas each frame takes 55 ms to capture. Even though our proposed sensor currently has a lower range of measurements, we achieved a lower latency 21 ms which makes our sensor suitable for real-time applications. It is worthy to mention that some vision-based force measurement sensors such in [50], are designed for precise applications with high resolution in µN-mN range. However, the sampling rate of 30 FPS and high computational cost of processing coloured images will increase the response of time of the vision-based sensors.

The proposed classification model achieves a accuracy of 79.17% on the unseen test data considering the leave-one-out cross-validation method. Most of the miss-classifications are between the soft materials with a close Young's modulus. In [11], the average accuracy of 95%is presented to classify different materials. The sensor includes strain gauges and PVDF sensors embedded in the fingertip. The materials have a different texture which provides further information to sensors in order to classify the materials.

The proposed event-based sensor is the first-ever attempt to utilize event-based cameras to estimate the contact force and classify materials. There is clearly a high potential to achieve better accuracy results, similar to the mentioned vision-based sensors. The main advantages of the proposed sensor compared to other vision-based techniques as follows:

1) Events are captured in a high time-resolution of few microseconds. In contrast, the sampling rate of conventional cameras is normally 30 ms which is significantly higher than event-based cameras. Therefore, our proposed event-driven sensor demonstrates shorter latency than what could be achieved by conventional cameras.

2) Event-driven cameras have a lower power consumption compare to the conventional camera. Moreover, capturing and processing images require high computational power and memory requirements whereas neuromorphic cameras provide intensity changes in binary representation.

3) The proposed sensor does not require any markers inside the silicone membrane. Therefore, the fingertip can be easily changed by replacing a new transparent silicone membrane.

In addition, our proposed TDNN solution for the event-based force estimation seems to have a much lower hysteresis than the piezoresistive force sensor, despite the fact that it has been trained using data from the force sensor.

## VII. Conclusion and Future Work

In this paper, a novel event-driven tactile sensor is proposed to estimate the contact force and classify different materials in a grasp. A deep TDNN model with time delay of 3 nodes, four fully-connected hidden layers, and 40 neurons at each layer is implemented to estimate force measurements. The TDNN estimates the contact force with the averaged MSE of 0.16 N during the grasping and the releasing phases of an unseen grip. Moreover, a time-series GP model is developed which achieves the averaged MSE of 0.17 N. The results indicate a promising relation between the triggered events and the contact force variation, especially if one takes into account that the source of the estimated errors may come from the hysteresis of the piezoresistive force sensor that was used to provide the ground truth.

Forty-eight experiments are performed on four different materials with a similar dimension and different Young's modulus. A multilayer neural network is suggested to classify materials in a single grasp using events only. The proposed network achieves a accuracy of 79.17% on the completely unseen experiments, almost 30% higher accuracy compared to the piezoresistive sensor.

For future work, we aim to increase the sensor sensitivity and accuracy by performing advanced time-series machine learning algorithms and using different types of force sensor with a higher accuracy and sensitivity.

## Acknowledgment

## References

[1] J. Tegin and J. Wikander, "Tactile sensing in intelligent robotic manipulation a review," *Industrial Robot: An International Journal*, vol. 32, no. 1, pp. 64–70, 2005.

[2] G. Westling and R. S. Johansson, "Factors influencing the force control during precision grip," *Experimental Brain Research*, vol. 53, no. 2, pp. 277–284, 1984.

[3] H. Yousef, M. Boukallel, and K. Althoefer, "Tactile sensing for dexterous in-hand manipulation in robotics - A review," *Sensors and Actuators, A: Physical*, vol. 167, no. 2, pp. 171–187, 2011.

[4] Y. Wan, Y. Wang, and C. F. Guo, "Recent progresses on flexible tactile sensors," *Materials Today Physics*, vol. 1, pp. 61–73, 2017.

[5] J. G. V. da Rocha, P. F. A. da Rocha, and S. Lanceros-Mendez, "Capacitive sensor for three-axis force measurements and its readout electronics," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 8, pp. 2830–2836, 2009.

[6] P. Payeur, C. Pasca, A. M. Cretu, and E. M. Petriu, "Intelligent haptic sensor system for robotic manipulation," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 4, pp. 1583–1592, 2005.

[7] A. Sanli, R. Ramalingame, and O. Kanoun, "Piezoresistive pressure sensor based on carbon nanotubes/epoxy composite under cyclic loading," *I2MTC 2018 - 2018 IEEE International Instrumentation and Measurement Technology Conference: Discovering New Horizons in Instrumentation and Measurement, Proceedings*, pp. 1–5, 2018.

[8] S. Takenawa, "A magnetic type tactile sensor using a two-dimensional array of inductors," *IEEE International Conference on Robotics and Automation*, no. 1, pp. 3295–3300, 2009.

[9] A. Kimoto and Y. Matsue, "A new multifunctional tactile sensor for detection of material hardness," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 4, pp. 1334–1339, 2011.

[10] S. Baglio, L. Cantelli, F. Giusa, and G. Muscato, "Intelligent prodder: Implementation of measurement methodologies for material recognition and classification with humanitarian demining applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 8, pp. 2217–2226, 2015.

[11] N. Jamali and C. Sammut, "Majority Voting: Material Classification by Tactile Sensing Using Surface Texture," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 508–521, 2011.

[12] I. Bandyopadhyaya, D. Babu, A. Kumar, and J. Roychowdhury, "Tactile sensing based softness classification using machine learning," *2014 IEEE International Advance Computing Conference (IACC)*, pp. 1231–1236, 2014.

[13] M. V. Liarokapis, B. Calli, A. Spiers, and A. M. Dollar, "Unplanned , model-free , single grasp object classification with underactuated hands and force sensors," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, no. August, pp. 5073–5080, 2015.

[14] J. Konstantinova, G. Cotugno, A. Stilli, Y. Noh, and K. Althoefer, "Object classification using hybrid fiber optical force/proximity sensor," *Proceedings of IEEE Sensors*, vol. 2017-Decem, pp. 7–9, 2017.

[15] R. Ahmadi, M. Packirisamy, J. Dargahi, and R. Cecere, "Discretely loaded beam-type optical fiber tactile sensor for tissue manipulation and palpation in minimally invasive robotic surgery," *IEEE Sensors Journal*, vol. 12, no. 1, pp. 22–32, 2012.

[16] N. F. Lepora and B. Ward-Cherrier, "Superresolution with an optical tactile sensor," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-Decem, pp. 2686–2691, 2015.

[17] K. Zhang, C. Butler, Q. Yang, and Y. Lu, "A fiber optic sensor for the measurement of surface roughness and displacement using artificial neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 46, no. 4, pp. 899–902, 1997.

[18] J. S. Heo, J. H. Chung, and J. J. Lee, "Tactile sensor arrays using fiber Bragg grating sensors," *Sensors and Actuators, A: Physical*, vol. 126, no. 2, pp. 312–327, 2006.

[19] L. H. Negri, A. S. Paterno, M. Muller, and J. L. Fabris, "Sparse Force Mapping System Based on Compressive Sensing," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 4, pp. 830–836, 2017.

[20] A. Massaro, F. Spano, A. Lay-Ekuakille, P. Cazzato, R. Cingolani, and A. Athanassiou, "Design and characterization of a nanocomposite pressure sensor implemented in a tactile robotic system," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 8, pp. 2967–2975, 2011.

[21] S. Shirmohammadi and A. Ferrero, "Camera as the Instrument: The Rising Trend of Vision Based Measurement," *IEEE Instrumentation & Measurement Magazine*, vol. 17, pp. 41–47, 2013.

[22] A. Ikeda, Y. Kurita, J. Ueda, Y. Matsumoto, and T. Ogasawara, "Grip force control for an elastic finger using vision-based incipient slip feedback," *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 1, pp. 810–815, 2004.

[23] K. Vlack, T. Mizota, N. Kawakami, K. Kamiyama, H. Kajimoto, and S. Tachi, "GelForce: a vision-based traction field computer interface," *Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems*, vol. 2, pp. 1154–1155, 2005.

[24] G. Obinata, A. Dutta, N. Watanabe, and N. Moriyama, "Vision Based Tactile Sensor Using Transparent Elastic Fingertip for Dexterous Handling," *Mobile Robotics - Perception & Navigation -*, no. February, pp. 137–148, 2007.

[25] Y. Ito, Y. W. Kim, and G. Obinata, "Slippage degree estimation by using vision-based tactile sensor for dexterous handling," *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 42, no. 16, pp. 281–286, 2009.

[26] Y. Ito, Y. Kim, and G. Obinata, "Contact region estimation based on a vision-based tactile sensor using a deformable touchpad," *Sensors (Basel, Switzerland)*, vol. 14, no. 4, pp. 5805–5822, 2014.

[27] A. Kolker, M. Jokesch, and U. Thomas, "An optical tactile sensor for measuring force values and directions for several soft and rigid contacts," *47th International Symposium on Robotics, ISR 2016*, vol. 2016, pp. 63–68, 2016.

[28] K. Zhao, X. Li, C. Lu, G. Lu, and Y. Wang, "Video-based slip sensor for multidimensional information detecting in deformable object grasp," *Robotics and Autonomous Systems*, vol. 91, pp. 71–82, 2017.

[29] K. Kamiyama, H. Kajimoto, N. Kawakami, and S. Tachi, "Evaluation of a vision-based tactile sensor," *IEEE International Conference on Robotics and Automation, 2004.*, vol. 2, pp. 1542–1547, 2004.

[30] K. Kamiyama, K. Vlack, T. Mizota, H. Kajimoto, N. Kawakami, and S. Tachi, "Vision-based sensor for real-time measuring of surface traction fields," *IEEE Computer Graphics and Applications*, vol. 25, no. 1, pp. 68–75, 2005.

[31] Y. Ito, Y. Kim, and G. Obinata, "Robust slippage degree estimation based on reference update of vision-based tactile sensor," *IEEE Sensors Journal*, vol. 11, no. 9, pp. 2037–2047, 2011.

[32] S. Luo, W. Yuan, E. Adelson, A. G. Cohn, and R. Fuentes, "ViTac: Feature Sharing between Vision and Tactile Sensing for Cloth Texture Recognition," 2018.

[33] W. Chen, H. Khamis, I. Birznieks, N. F. Lepora, and S. J. Redmond, "Tactile Sensors for Friction Estimation and Incipient Slip Detection - Towards Dexterous Robotic Manipulation: A Review," *IEEE Sensors Journal*, 2018.

[34] A. I. Aviles, A. Marban, P. Sobrevilla, J. Fernandez, and A. Casals, "A recurrent neural network approach for 3D vision-based force estimation," *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6, 2014.

[35] A. I. Aviles, S. Alsaleh, P. Sobrevilla, and A. Casals, "Sensorless force estimation using a neuro-vision-based approach for robotic-assisted surgery," *International IEEE/EMBS Conference on Neural Engineering, NER*, vol. 2015-July, pp. 86–89, 2015.

[36] N. Haouchine, W. Kuang, S. Cotin, and M. Yip, "Vision-Based Force Feedback Estimation for Robot-Assisted Surgery Using Instrument-Constrained Biomechanical Three-Dimensional Maps," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2160–2165, 2018.

[37] R. Anchini, G. Di Leo, C. Liguori, and A. Paolillo, "Metrological characterization of a vision-based measurement system for the online inspection of automotive rubber profile," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 1, pp. 4–13, 2009.

[38] N. Alt and E. Steinbach, "Navigation and manipulation planning using a visuo-haptic sensor on a mobile platform," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 11, pp. 2570–2582, 2014.

[39] O. Kroemer, C. H. Lampert, and J. Peters, "Learning dynamic tactile sensing with robust vision-based training," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 545–557, 2011.

[40] A. Rigi, F. Baghaei Naeini, D. Makris, and Y. Zweiri, "A Novel Event-Based Incipient Slip Detection Using Dynamic Active-Pixel Vision Sensor (DAVIS)," *Sensors*, vol. 18, no. 2, p. 333, 1 2018.

[41] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, "Measurement of shear and slip with a GelSight tactile sensor," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2015-June, no. June, pp. 304–311, 2015.

[42] K. Kamiyama, H. Kajimoto, M. Inami, N. Kawakami, and S. Tachi, "A Vision-based Tactile Sensor," in *International Conference on Artificial Reality and Telexistence*, 2001.

[43] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 x 128 120 dB 15 $\mu$s latency asynchronous temporal contrast vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[44] M. Yang, S. C. Liu, and T. Delbruck, "As dynamic vision sensor with 1% temporal contrast sensitivity and in-pixel asynchronous delta modulator for event encoding," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 9, pp. 2149–2160, 2015.

[45] M. A. Hopcroft, W. D. Nix, and T. W. Kenny, "What is the Young's modulus of silicon?" *Journal of Microelectromechanical Systems*, vol. 19, no. 2, pp. 229–238, 2010.

[46] A. Weibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks." *IEEE Transactions On Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 328–339, 1989.

[47] Y. Bengio, S. Patrice, and F. Paolo, "Learning Long-Term Dependecies with Gradient Descent is Difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[48] M. Seeger, "Gaussian Processes for Machine Learning," *International Journal of Neural Systems*, vol. 14, no. 02, pp. 69–106, 2004.

[49] S. Parmar, I. Khodasevych, and O. Troynikov, "Evaluation of flexible force sensors for pressure monitoring in treatment of chronic venous disorders," *Sensors (Switzerland)*, vol. 17, no. 8, pp. 1–18, 2017.

[50] F. Karimirad, S. Chauhan, and B. Shirinzadeh, "Vision-based force measurement using neural networks for biological cell microinjection," *Journal of Biomechanics*, vol. 47, no. 5, pp. 1157–1163, 2014.