Discriminating Liquids Using a Robotic Kitchen Assistant

Christof Elbrechter and Jonathan Maycock and Robert Haschke and Helge Ritter

Abstract— A necessary skill when using liquids in the preparation of food is to be able to estimate viscosity, e.g. in order to control the pouring velocity or to determine the thickness of a sauce. We introduce a method to allow a robotic kitchen assistant discriminate between different but visually similar liquids. Using a Kinect depth camera, surface changes, induced by a simple pushing motion, are recorded and used as input to nearest neighbour and polynomial regression classification models. Results reveal that even when the classifier is trained on a relatively small dataset it generalises well to unknown containers and liquid fill rates. Furthermore, the regression model allows us to determine the approximate viscosity of unknown liquids.

I. INTRODUCTION

An important class of non-rigid objects are liquids. When we interact with liquids (using containers, funnels, stirring tools etc.), the two major distinctive properties that determine its physical behavior are its density and its viscosity. While a good prior for the density of a liquid in many situations is the density of water, the viscosity of everyday liquids (e.g., involved in food preparation in a kitchen) is surprisingly varied and needs to be anticipated, e.g. when we wish to avoid spilling or ensure proper pouring, since differences in viscosity can result in physically very different behaviors.

Humans effortlessly recognize different viscosity ranges of liquids in familiar scenarios by relying on a combination of vision and actively caused movement, e.g. by stirring or shaking the liquid in a container. Robots currently lack a similar ability, although it would be very useful for future kitchen robots designed to assist humans in food preparation tasks. Besides conveying useful information for proper adjustment of physical handling procedures, the viscosity of a liquid also can offer an important feature to discriminate between visually similar looking liquids (e.g. milk and buttermilk). Additionally, and particularly when preparing food, it also often conveys essential information about the progress and proper termination point of a preparation process (e.g. when making a sauce or preparing a jelly). This multi-faceted relevance also makes it attractive for non-robot applications, such as an intelligent kitchen agent watching a human preparing a recipe, to be able to use observed motion of liquids for discriminating hypotheses about the liquid type or the preparation stage.

This motivates the contribution of this paper: combining active motion, created through a sliding back and forth movement, with visual observation, we enable a robot to
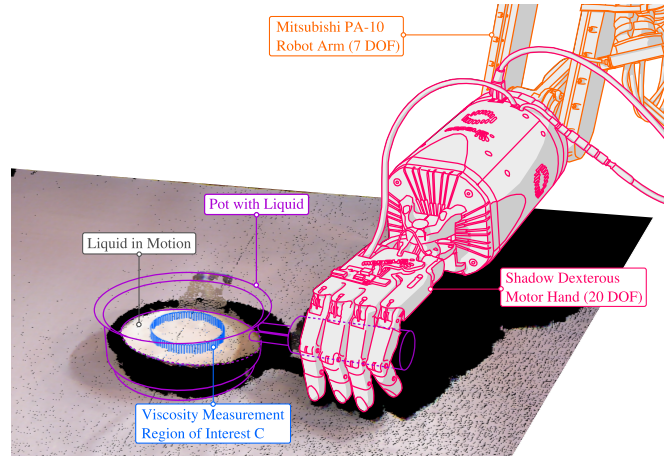
Fig. 1. Kinect image of our setup augmented with robot simulation outline overlay. Robot performs a sideways back and forth sliding motion to generate a disturbance in the liquid. An analysis of the 3D-motion of the liquid's surface, C, provides enough information to estimate a coarse viscosity value.

actively discriminate between different viscosities of visually similar looking liquids [see Figure 1]. For the vision channel, we employ the widely used Kinect sensor to provide information about the 3D-motion of the liquid's surface. Our study complements a recent approach focused on the geometric feature of filling depth [1]. A combination of both approaches would provide a very good perceptual front-end to monitor and guide functional handling of liquids in a variety of food handling tasks in a kitchen robot scenario.

II. RELATED WORK

Measurement of viscosity can be carried out using a viscometer, but this requires direct physical contact with the liquid under consideration and in a food preparation kitchen scenario this is not a desirable property. Although our vision based method cannot rival the accuracy of this device, its advantage lies in the fact that it is non-invasive, requiring as it does that just the container holding the liquid is minimally disturbed.

There are existing approaches that are also non-invasive but nonetheless strive for an accurate measurement of liquid viscosity. One such method is that of laser speckle imaging which makes use of variations in speckle contrast to infer viscosity or velocity distribution in a flow field [2], [3], [4]. We also considerd investigating the infrared channel of the Kinect 1, which produces a speckle image, but there were two reasons why we instead used the depth value channel. The first is that our technique is independent of the
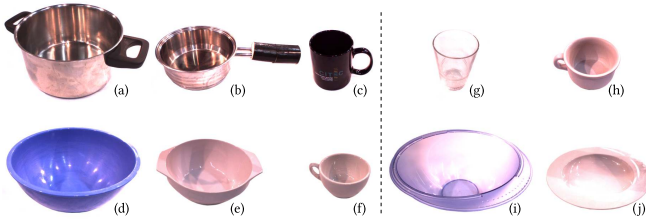
Fig. 2. Containers used for the training and for the evaluation phase. **(a-f)** The training set is subdivided into variously sized pots (a-c) and bowls (d-f). **(g-j)** For the evaluation phase, the set of containers is extended with two more pot-shaped containers (g,h) and two more bowl-shaped ones (i,j).

underlying technology used to produce the depth map and therefore can be applied to other depth cameras including the Kinect 2 (which uses the time of flight principle). The second reason is a limitation of the Kinect itself in that it is not possible to simultaneously use all three colour, infrared and depth-image modes. We note that although Yang and Choi [4] use a consumer-grade color camera, their technique requires a He-Ne laser which should be operated using safety glasses and is not suitable for a home kitchen scenario.

Analysis of the frequency shifts to determine the viscosity of liquids has also been achieved using microacoustic [5], magnetic acoustic and optical sensors [6], and ultrasound [7]. These methods employed very precise experimental setups and required an exact impulse to be applied to the unknown liquid from which inferences about its viscosity could be made. Other examples using ultrasound required sensors to be attached directly to the holding containers [8], [9]. An impressive approach which estimates viscosity in biodiesel using near infrared spectroscopy in combination with neural networks [10] requires far too complex a setup to be applicable in a home scenario. By using the Kinect camera we take advantage of a device that is cheap and readily available in most robotic setups and although our method is not as accurate as the aforementioned ones, we argue that it is sufficient to provide the robot kitchen assistant with a rudimentary knowledge of viscosity needed to be of help when preparing a meal.

III. ANALYSIS FRAMEWORK

While 6 DOFs are sufficient to completely describe the motion of a rigid object, liquids are systems with infinitely many DOFs. Although there exist motions (e.g. fully developed turbulence), where this infinity of DOFs is actually needed for a proper description of the motion, the presence of viscosity, along with a limit on the imparted energy, confines excitation of contributing movement DOFs to a sufficiently small set of eigenmodes whose velocity field depends on the geometry confining the liquid and on the fill rate of the liquid within that geometry. Even under such restrictions a detailed modeling of the motion of liquid would lead into considerable complexity and an accurate identification of the system state would be very likely strongly incompatible with the available visual information given the presence of limited resolution, noise, and incomplete information about

the container and its mechanical excitation through pushing. Similar to other situations in robotics (e.g. grasping) a workable solution should avoid any complex model and instead use a robust framework that demands only information that can be captured in an online fashion.

This motivates the following approach: we view the motion of the liquid as the superposition of a small set of damped oscillators of nearby frequencies that are primarily determined by the geometry constraints on the liquid, the fill rate and its viscosity. To first order (neglecting non-linear effects), and after the decay of the external excitation (pushing), the amplitude x of any such oscillator follows the second order equation $\ddot{x} + \gamma\dot{x} + cx = 0$, with large γ values resulting in overdamping and smaller values of γ producing underdamping. Oscillation frequency and damping of this oscillator convey information about the density and the viscosity of the liquid. Although the precise mapping will differ for different oscillation modes, our simple pushing movement often produces a single "sloshing wave" [11] whose properties allow, upon viewing of the visualized raw values or indeed the Fourier transform spectrum, a distinction between liquids to be made. This, along with the repeatability of the produced interaction patterns gives us confidence in our method.

This oscillatory mode essentially drives the observable, time-dependent depth pixel pattern delivered by the Kinect. However, as the detailed connection between x and the observable up and down motion of the liquid surface is highly complex, we first condense the depth image information to a scalar surface position variance signal $\sigma[t]$ that is captured at the Kinect's constant frame rate of 30 Hz [see Section IV]. After heuristically cropping a temporal interest window from $\sigma[t]$, we employ generic Principal Component Analysis (PCA)-based feature extraction in the spatial and frequency domains, complemented by an explicit fitting of a negative exponential decay function [see Section V-A]. The resulting feature set, which consists of two sets of PCs and the fitted negative exponential's temporal decay term are then used for classification of liquids and regression of a liquid's kinematic viscosity value. We optimized the number of used PCs and the feature weighting factors using nonlinear optimization techniques and exhaustive search [see Section V-C] on the basis of a large set of training data that was manually captured using various liquids (milk, cream and buttermilk), containers [see Figure 2 (a-f)] and fill rates (low and high fill). The generalization performance of the optimized systems was then evaluated with a test data set captured with unknown setups (new containers [see Figure 2 (g-j)] and fill rates) and also with unknown liquids. Finally, we evaluate the classification and regression system using our robot kitchen assistant.

IV. VISUAL DETECTION OF FLUID SURFACE MOTION

The first step in the processing pipeline is to detect the motion of the liquid surface over time. We used our computer-vision library ICL [12] with a calibrated Microsoft

Kinect camera pointed diagonally downwards into the scene to yield a 2D-ordered point cloud $P[\mathbf{x}] \to \mathbf{p} \in \mathbb{R}^3$, where $\mathbf{x} \in 320 \times 240$ is a point's depth image pixel coordinate. Our model-free segmentation system [13] provides a set of surfaces from which we manually choose the liquid surface patch S. By assuming that S is co-planar with the xy-plane and has a circular shape, a 2D circular region C [see Figure 1] is computed heuristically to ensure that $\approx 50\%$ of the xy-projections of S are inside C. The z-values of C are then used to create the elevation map

$$e_t[\mathbf{x}] = \begin{cases} P[\mathbf{x}]_z & \text{if } (P[\mathbf{x}]_x, P[\mathbf{x}]_y) \text{ is inside C} \\ 0 & \text{otherwise} \end{cases}$$

at each processing step t. In order to get an estimate of the temporal development of the surface movement, a variance map v_t is created. To this end, the last n elevation maps (time window $W = \{t, .., t-n-1\}$) are memorized. The current local surface motion is then defined by the variance map

$$v_t[\mathbf{x}] = \begin{cases} \text{var}(\{\mathbf{e}_{t'}[\mathbf{x}] | t' \in W\}) & \mathbf{e}_{t'} \neq 0 \; \forall t' \\ -1 & \text{otherwise.} \end{cases}$$

Missing point cloud points, which commonly occur due to either disparity gaps or due to disturbances of the internal IR-speckle pattern image are set to $(0,0,0)$, and therefore for those inside C, their corresponding elevation map entries becomes 0. By this method, we avoid using extreme variance values for point cloud points that flicker over time. The final scalar signal $\sigma[t]$ that describes the course of the surface motion over time is defined by the selective mean of the variance map:

$$\sigma[t] = \underset{\mathbf{x}}{\text{mean}}(\{v_t[\mathbf{x}] | v_t[\mathbf{x}] \geq 0\}).$$

V. CLASSIFICATION AND VISCOSITY ESTIMATION

A classification and a regression approach were implemented, optimized and evaluated. Firstly, a common NN-classifier was trained by manually generating a set of training data, which was captured with varying liquids, containers and fill-rates. Secondly, using the same data, a polynomial-regression network was trained to allow an estimation of a liquid's kinematic viscosity value be arrived at. Figure 2 gives an overview of the containers that were used for the training and the evaluation phases. In order to improve the generalization capabilities of both the classifier and the regression network, a set of features was extracted from the input signal.

A. Feature Extraction

Before features can be extracted in a meaningful fashion, a mechanism was needed to extract the decaying phase of the liquid motion. To this end, the input signal $\sigma[t]$ is tracked over time and the recording is triggered as soon as $\sigma[t]$ exceeds and then falls under a manually tuned amplitude threshold θ_{sigma}. By setting θ_{sigma} to a value that is significantly lower than the commonly occurring maximum of $\sigma[t]$, most unwanted artifacts that arise from the actual shaking motion are cropped out, leaving only the decaying
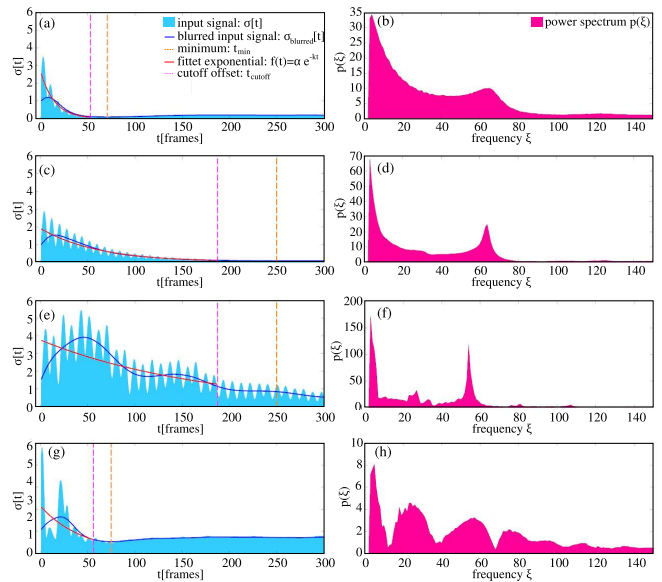

Fig. 3. Raw data (left column) and corresponding frequency (right column) plots for three selected training trials (a-f) and one robot test trails (g,h). The raw data plots are overlaid with some intermediate processing results as well as with the fitted exponential function. The raw-data plots are cropped to the first 300 of the actual 500 samples. The frequency plots are cropped to the first 150 of the actual 250 (non-mirrored) samples. **(a,b)** Buttermilk, medium bowl, high fill rate. **(c,d)** Cream, medium pot, high fill rate **(e,f)** Milk, large pot, high fill rate.**(g,h)** Buttermilk, medium pot, high fill rate.

part of the surface excitation to be used as input data. Once triggered, the following N_{raw} frames are captured. We chose $N_{\text{raw}} = 500$, which results in a trial length of ≈ 17 seconds and refer to it as the input data vector $\mathbf{d}_{\text{raw}} \in \mathbb{R}^{N_{\text{raw}}}$.

As initial tests showed that the raw input data leads to a very poor generalization performance, we performed PCA on the $\{\mathbf{d}_{\text{raw}}\}$ of the training data set. We then computed a feature vector $\mathbf{f}_d \in \mathbb{R}^{N_d}$ by projecting the $\mathbf{d}_{\text{raw}}$ into the subspace of the N_d eigenvectors of $\text{cov}(\{\mathbf{d}_{\text{raw}}\})$ that belong to the largest eigenvalues. In order to account for the fact that the underlying input signal is intrinsically created by a damped superposition of sine waves, we computed the Fourier power spectrum of the input signals and also performed PCA-based dimension reduction on these, yielding an additional feature vector $\mathbf{f}_{\text{ft}} \in \mathbb{R}^{N_{\text{ft}}}$ for a given input. While the Fourier analysis provides significant information about the nature of the underlying wave superposition patterns, the exponential decay is factored out by this method. In order to preserve this important property we explicitly add the exponential decay k to our weighted feature vector $\mathbf{f} = [\lambda_d \cdot \mathbf{f}_d, \; \lambda_{\text{ft}} \cdot \mathbf{f}_{\text{ft}}, \; k] \in \mathbb{R}^{N_d + N_{\text{ft}} + 1}$. The weighting factors λ_d and λ_{ft} are introduced to control the influence of the different feature sources during the NN-classification. For the exponential fitting, a set of heuristics was developed to find a suitable cutoff position t_{cutoff} in $\mathbf{d}_{\text{raw}}$ [see Figure 3 (left)] so that the signal from the first sample to t_{cutoff} shows the exponentially decaying behavior, but does not settle into the baseline noise of the signal. To this end, $\sigma[t]$ is smoothed (using a Gaussian kernel bandwidth of 10 samples) yielding $\sigma_{\text{blurred}}[t]$ and we find the

minimum

$$t_{\min} = \min\left(\frac{N_{\text{raw}}}{2}, \operatorname*{argmin}_{t}(\sigma_{\text{blurred}}[t])\right).$$

The signal is explicitly cropped to a max of 250 frames, which is completely sufficient for fitting. For the actual fitting carried out on the unsmoothened input signal $\sigma[t]$, a lower cutoff frame $t_{\text{cutoff}} = 0.75 t_{\min}$ is chosen to avoid a biased fitting caused by including too many baseline noise samples. The exponential fitting of the function $f(x) = \alpha e^{-kx}$ into the given set of input points $\{(t, \sigma[t]) | t \in \{1,..,t_{\text{cutoff}}\}\}$ is then performed using logarithmic linear regression. We solve $Mx = \log(\sigma)$, where the t-th row of M is $(1, t)$ using an SVD-based pseudo-inverse approach $\mathbf{x} = M^+ log(\sigma)$. The components of the resulting 2D vector $\mathbf{x}$ yield the parameters of the exponential $\mathbf{x} = (-k, \log(\alpha))^\tau$, of which only k is used, as the amplitude α is implicitly normalized by the data-recording mechanism.

Figure 3 (a-f) presents a selected input sample for each of the used liquids. Looking at the left column seems to reveal a direct correlation between the liquid viscosity and the exponential decay (e.g. buttermilk (a) which has a high viscosity decays much more quickly than milk (e)). However, a deeper analysis showed the relationship is heavily influenced by not only the viscosity of the liquid, but by other factors such as the shape of the holding container, its size and fill rate. As a general rule, we found that larger containers with a round shape (e.g., bowls), higher fill rates and less viscous liquids produced a slower decay rate. However, a simple monotonous or even linear relationship was not observed as self interference, both constructive and destructive, given slight configuration changes (container, fill rate etc.) had a larger effect on the exact nature of the decay. The same is true for the frequency peaks in the corresponding plots [see Figure 3 (right)]. While certain configurations lead to the presence of a strong fundamental frequency, a slight change to the setup parameters can yield a completely different profile. These complications and the realization that a method driven by simple heuristic rules was not feasible, guided us to the generic feature-based classification and regression approach presented here.

B. The Training Data Set

For the generation of a representative set of training samples, we manually performed a total of 360 trials. A fast acceleration/deceleration back and forth movement was used to disturb the liquid enough to exceed the threshold θ_{sigma}. Each resulting trial is characterized by a tuple of five conditions (l, t, s, r, i):

- $l \in \mathbf{L} = \{\text{milk}, \text{cream}, \text{buttermilk}\}$ is the liquid ID
- $t \in \mathbf{T} = \{\text{pot}, \text{bowl}\}$ is the container type [see Figure 2]
- $s \in \mathbf{S} = \{\text{small}, \text{medium}, \text{large}\}$ is the container size
- $r \in \mathbf{R} = \{\text{low}, \text{high}\}$ is the fill rate
- $i \in \mathbf{I} = \{\text{trial}_1, .., \text{trial}_{10}\}$ is the trial ID

We define $d : \mathbf{C} \to \mathbb{R}^{N_{\text{raw}}}$ as the raw data lookup function that returns the data vector $\mathbf{d}_{\text{raw}}$ for a given condition tuple $\mathbf{c} \in \mathbf{C} = \mathbf{L} \times \mathbf{T} \times \mathbf{S} \times \mathbf{R} \times \mathbf{I}$. Furthermore, we introduce the
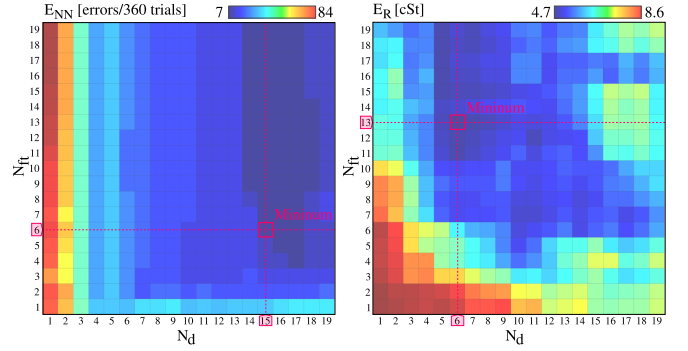

Fig. 4. Results for the exhaustive search-based optimization of the numbers of principal components N_d and N_{ft} to be used for the generation of the feature vectors $\mathbf{f}$ for both, NN-classification (left) and regression (right).

feature extraction function

$$f : \mathbb{R}^{N_{\text{raw}}} \times \mathbf{P} \to \mathbb{R}^{N_d + N_{\text{ft}} + 1}$$

that returns $\mathbf{f} = [\lambda_d \cdot \mathbf{f}_d, \ \lambda_{\text{ft}} \cdot \mathbf{f}_{\text{ft}}, \ k]$ for a given raw data vector, dependent on the also given parameter set $\mathbf{p} = (N_d, N_{\text{ft}}, \lambda_d, \lambda_{\text{ft}}) \in \mathbf{P}$. For the sake of brevity, $f(d(\mathbf{c}), \mathbf{p})$ is denoted as $f_{\mathbf{cp}}$ and $\mathbf{C_{cp}} := \{f_{\mathbf{c'p}} | \mathbf{c}' \in \mathbf{C} \setminus \mathbf{c}\}$. This formalism allows us to define the mean leave-one-out cross validation error

$$E_{\text{NN}}(\mathbf{p}) = \frac{1}{|\mathbf{C}|} \sum_{\mathbf{c} \in \mathbf{C}} \delta(c_1, \text{nn}(f_{\mathbf{cp}}, \mathbf{C_{cp}})),$$

where δ is the Kronecker delta function and nn is the nearest neighbor search. Analogously, the regression leave-one-out cross-validation RMSE is defined by

$$E_{\text{R}}(\mathbf{p}) = \sqrt{\frac{1}{|\mathbf{C}|} \sum_{\mathbf{c} \in \mathbf{C}} (v_{c_1} - \text{pr}(f_{\mathbf{cp}}, \mathbf{C_{cp}}))^2},$$

where v_l is the real kinematic viscosity coefficient of the given liquid l and $\text{pr}(a, B)$ performs polynomial regression on a using the regression coefficients optimized with B.

C. Optimization of Classification and Regression parameters

The quality of the regression with respect to the direct regression error and its generalization performance is strongly influenced by the choice of the polynomial terms used. We chose to use only linear and squared terms, as tests revealed that even the use of cubed terms resulted in a very poor generalization performance caused by extreme over-fitting. The optimization of E_{NN} with respect to $\mathbf{p}$ was carried out in two steps. We first used the Nelder-Mead simplex method [14] with different starting positions in parameter space $\mathbf{P}$. Although it was often prone to getting stuck in local minima, regardless of the starting conditions it consistently converged towards values in the vicinity of $\lambda_d = 5/N_d$ and $\lambda_{\text{ft}} = 1/N_{\text{ft}}$. After fixing these parameters , the optimum in the resulting 2D parameter subspace could be found by conducting an exhaustive search [see Figure 4 (left)]. As the polynomial regression intrinsically performs a parameter normalization, the optimization of the regression error E_{R} could directly be addressed using a 2D exhaustive search [see Figure 4 (right)].
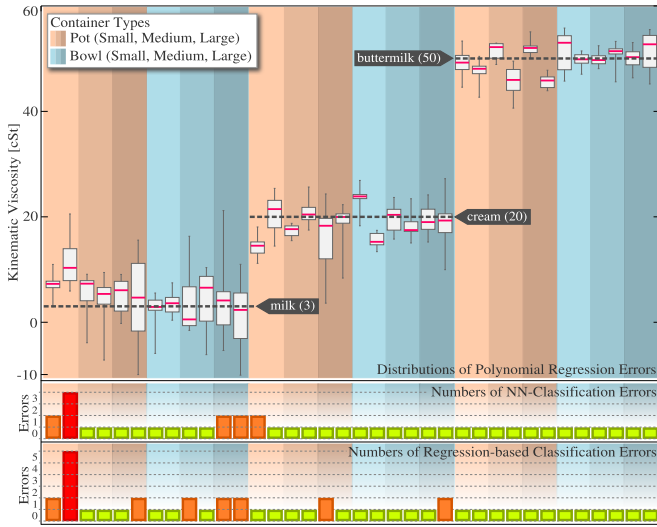
Fig. 5. Training data evaluation results for NN-classification and regression. The regression-based classification error is calculated using a closest neighbor vote in the kinematic viscosity domain.

Interestingly, the ratios of N_{ft}/N_d are flipped between both methods. The target values for the regression model reflect real kinematic viscosity values for the chosen training liquids of milk (3 cSt), cream (20 cSt) and buttermilk (50 cSt).

VI. EVALUATION

We evaluated the presented methods in three conditions: (i) using the training data, (ii) using explicitly generated test data (e.g., previously unseen liquids/containers/fill rates etc.) and (iii) using our robot kitchen assistant, which is presented in Section VII. The evaluation for all conditions was carried out using the optimized parameter sets $(15, 6, 5/15, 1/6)$ for the NN-classification and $(6, 13, 1, 1)$ for the regression.

A. Training Data Evaluation

The leave-one-out cross-validation errors and the regression statistics results for the different conditions $\mathbf{c} \in \mathbf{C}$ are presented in Figure 5. The plot shows that especially milk in the small pot container (a coffee pot), in particular with the high fill rate [2nd column from the left] was most problematic for both the classification and the regression. This effect can be attributed to the container's special geometry (height vs. width), which differed most from all the other container/fill rate combinations. The plot also reveals that lower viscous liquids yield a higher variance in their regression outcomes (e.g. milk is classified worst). The perfect classification rate for buttermilk can be attributed to this, but also to the fact that its actual kinematic viscosity is furthest away from the others. The overall training data NN-classification error rate was 7/360, i.e. < 2% and the corresponding regression-based classification error rate was 12/360, i.e. 3.33%.

B. Test Data Evaluation

The results for the test data are presented in Figure 6. For these tests 4 new containers were introduced [see Figure 2 (g-j)], 3 new liquids (yogurt, kefir and a mix of buttermilk

and cream), and unknown fill rates were employed in the medium sized pot from the training data set. It is immediately apparent that the NN-classification performed better than the regression. Upon investigation we found that in most cases large errors were caused by the data collection process being triggered too soon, i.e. during the excitation phase. This in turn led to a shift in $\mathbf{d}_{raw}$, which particularly affected k and $\mathbf{f}_d$. After manually adjusting the onset, both the classification and the regression significantly improved.

For the tests with unknown fill rates it is notable that even though the regression often performs poorly, the NN-classification is still able to provide good results except in the case of a very high fill rate (last 3 columns of (a)), where this effect is inverted. This result can only be explained by a coincidentally perfect extrapolation behavior (towards higher fill rates) of the regression in this specific situation.

In the following test trials using unknown bowls (b,c), this trend continues, whereas for the more pot-shape containers with both milk (d,e) and cream (f,g), classification and regression worked very well. For the trials in which we mixed buttermilk with cream (k) in order to simulate a liquid with an intermediate viscosity, the NN-classification results were acceptable, but the regression only provided an intermediate viscosity value in 1/3 trials. The remaining trials (l,m) also dealt with unknown liquids and while the NN-classification robustly chose the actual nearest neighbor, the regression-based extrapolation struggled.

VII. ROBOT EVALUATION

Our robotic kitchen assistant consists of two 7-DOF Mitsubishi PA-10 robot arms, two 20 DOF Shadow Dexterous Motor Hands [15] and a Kinect camera for vision. In this experiment only the left hand and arm were utilized. The NN-classification results for the robot tests [see Figure 6 (h-j)] show that for milk a classification rate of 100% was achieved, for cream this dropped to $\approx 82\%$, but for buttermilk the classifier failed every time. This can be explained by the fact that, even though the used pot [see Figure 2 (b)] was part of the training data set, the resulting liquid motion was very different. This was mainly caused by the differences between the robot and the human movement to disturb the liquid. Whereas humans have evolved to be able to easily realize "ballistic" movements, characterized by having a high acceleration and deceleration, producing a similarly abrupt movement was not realizable on the robot. The main reason for this are the acceleration and jerk limitations that are hard-coded into the robot driver to ensure safety and hardware durability. This reveals an intrinsic limitation of our setup but one which can be address by either learning better heuristic parameters for the excitation movement or adapting the hardware of our robot setup. This issue directly led to a poor choice of cropping parameters in particular for the second half of the milk trials (h) and for the buttermilk trials (j). An exemplary situation is shown in Figure 3 (g). Here, a significant portion of the excitation phase is erroneously added to the cropped data window. A slight manual shift of the cropping window for the buttermilk trials led to a perfect
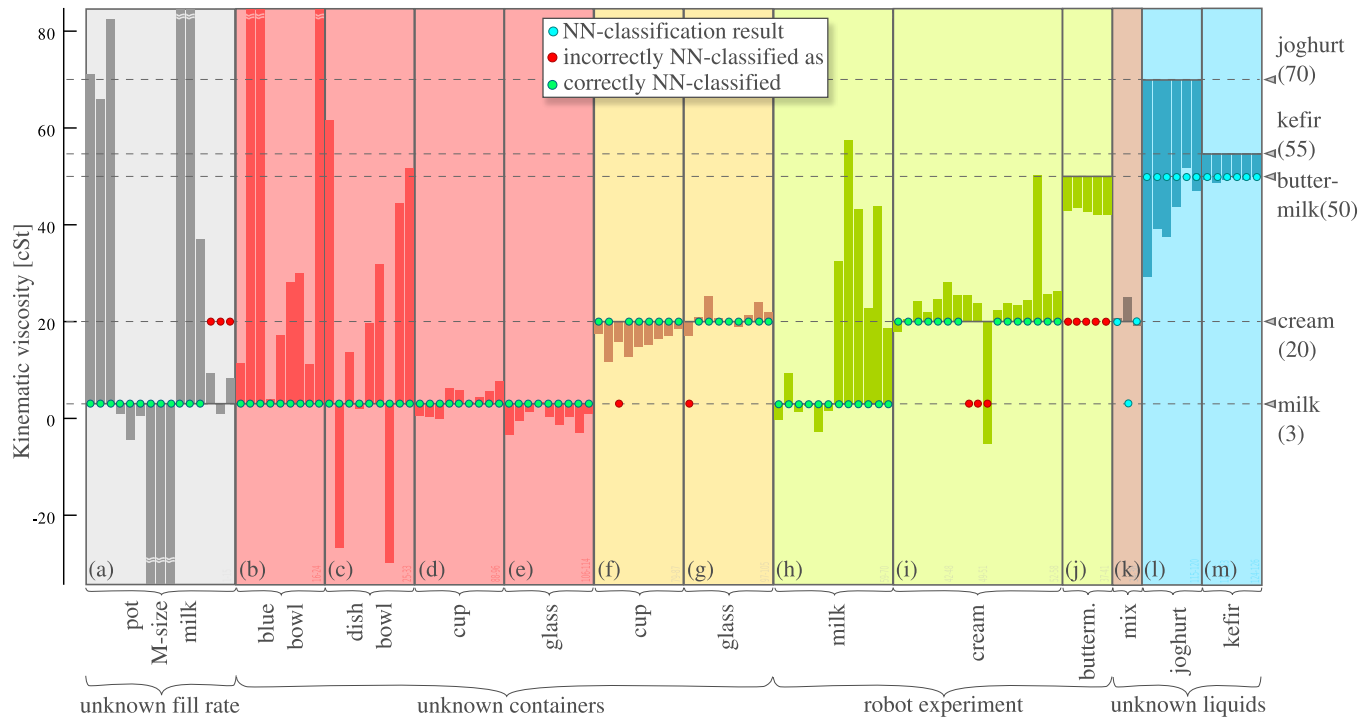
Fig. 6. NN-classification and regression results for the test data. Except for the unknown liquids and the unknown fill rate tests, scenarios contained both low and high fill examples (from left to right). The unknown liquids' NN-results are colored blue as their class was not contained in the training data.

NN-classification. Finally, the results with cream were very satisfactory.

VIII. DISCUSSION

We have presented a classification system for discriminating liquids in a variety of containers using only widely available 3D-point cloud data. The system generalizes over different fill rates and even provides remarkable extrapolation results when confronted with system unknown containers. While interpolation between two known liquids provided satisfactory results, extrapolation to previously unseen liquids suggests the need for a larger training data set. The presented methods for classification and regression and parameter optimization show promising scalability properties, but the developed heuristics for the selection of the temporal cropping window need to be made more robust. In future work, we plan to replace the heuristics by more sophisticated methods, such as dynamic-time-warping-based temporal segmentation approaches or even by automatic learning-based methods. Furthermore, we plan to use the robot to bootstrap the learning mechanism by providing triggers for the data collection onset, as well as actively exploring the interaction space of movements needed to generate more optimal excitation profiles.

REFERENCES

[1] Y. Hara, F. Honda, T. Tsubouchi, and A. Ohya, "Detection of liquids in cups based on the refraction of light with a depth camera using triangulation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014, pp. 5049–5055.

[2] A. Fercher and J. Briers, "Flow visualization by means of single-exposure speckle photography," *Opt. Comm.*, vol. 37, no. 5, pp. 326–330, 1981.

[3] Y. Cheng, S. Li, R. J. Eckersley, D. S. Elson, and M.-X. Tang, "Viscosity measurement based on shear-wave laser speckle contrast analysis," *J. Biomed. Opt.*, vol. 18, no. 12, 2013.

[4] O. Yang and B. Choi, "Laser speckle imaging using a consumer-grade color camera," *Opt. Lett.*, vol. 37, no. 19, pp. 3957–3959, Oct 2012.

[5] B. Jakoby and M. J. Vellekoop, "Viscosity sensing using a love-wave device," *Sensors and Actuators*, vol. 68, pp. 275–281, 1998.

[6] M. K. Jain, S. Schmidt, C. Mungle, K. Loiselle, and C. A. Grimes, "Measurement of temperature and liquid viscosity using wireless magneto-acoustic/magneto-optical sensors," *IEEE Transactions Magnetics*, vol. 37, no. 4, pp. 2767–2769, 2001.

[7] B. A. Martin, S. W. Wenzel, and R. W. White, "Viscosity and density sensing with ultrasonic plate waves," *Sensors and Actuators*, vol. 21-23, pp. 704–708, 1990.

[8] J. A. Bamberger and M. S. Greenwood, "Measuring fluid and slurry density and solids concentration non-invasively," *Ultrasonics*, vol. 42, no. 1-9, pp. 563–567, 2004.

[9] ——, "Non-invasive characterization of fluid foodstuffs based on ultrasonic measurements," *Food Research International*, vol. 37, no. 6, pp. 621–625, 2004.

[10] R. M. Balabina, E. I. Lomakinab, and R. Z. Safievad, "Neural network (ann) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (nir) spectroscopy," *Fuel*, vol. 90, no. 5, pp. 2007–2015, 2011.

[11] Z. C. Feng, "Transition to traveling waves from standing waves in a rectangular container subjected to horizontal excitations," *Physical Review Letters*, vol. 79, no. 3, pp. 415–418, 1997.

[12] C. Elbrechter, M. Götting, and R. Haschke, "Image Component Library (ICL)." [Online]. Available: http://iclcv.org

[13] A. Ückermann, C. Elbrechter, R. Haschke, and H. Ritter, "3D scene segmentation for autonomous robot grasping," in *IEEE/RSJ IROS*, 2012, pp. 1734–1740.

[14] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[15] Shadow Robot Company, "Shadow Dextrous Hand." [Online]. Available: http://www.shadowrobot.com/products/dexterous-hand