Interactive Computational Imaging for Deformable Object Analysis

Donald G. Dansereau[1], Surya P. N. Singh[2], Jürgen Leitner[1]

***Abstract*— We describe an interactive approach for visual object analysis which exploits the ability of a robot to manipulate its environment. Knowledge of objects' mechanical properties is important in a host of robotics tasks, but their measurement can be impractical due to perceptual or mechanical limitations. By applying a periodic stimulus and matched video filtering and analysis pipeline, we show that even stiff, fragile, or low-texture objects can be distinguished based on their mechanical behaviours. We construct a novel, linear filter exploiting periodicity of the stimulus to reduce noise, enhance contrast, and amplify motion by a selectable gain – the proposed filter is significantly simpler than previous approaches to motion amplification. We further propose a set of statistics based on dense optical flow derived from the filtered video, and demonstrate visual object analysis based on these statistics for objects offering low contrast and limited deflection. Finally, we analyze 7 object types over 59 trials under varying illumination and pose, demonstrating that objects are linearly distinguishable under this approach, and establish the viability of estimating fluid level in a cup from the same statistics.**

I. Introduction

The need to interact with deformable objects is legion in robotics, from domestic applications like shopping and food preparation to agricultural, manufacturing and industrial applications [1]–[3]. Compliance challenges classical approaches and strains rigid-body assumptions. A rigid geometric view, while deeply informative, neglects the rich spectrum of mechanical behaviours found in physical objects.

Measurement through mechanical sensing can be expensive or prohibitive, as appropriate sensors are not universally deployed, and contact forces for fragile or highly compliant objects can be difficult to capture. Cameras, on the other hand are widely deployed, making the incremental cost of visual object analysis negligible. When an object vibrates or deflects, its motion reveals details of its mechanical properties. Motion magnification provides a mechanism for capturing even very small displacements [4]–[6]. This makes visual sensing appropriate for analyzing a wide range of objects, from soft foams to stiff plastics.

We propose a method for analyzing deformable objects through mechanical stimulation and visual sensing. The bulk of the complexity in this approach is algorithmic, limiting hardware requirements to a camera and any system capable of periodic stimulation of the scene – a capability which most robots that interact with their environments have. Fig. 1 depicts a Baxter robot applying the proposed method to analyze a set of deformable objects.

Fig. 1. **Interactive, Visual Analysis of an Object's Material Properties**. Baxter squeezing a range of items (top-left), with an external camera observing (top-right). By applying a known, periodic stimulus, we show that matched video filtering and simple statistics can be used to distinguish objects based on material properties, even for stiff, fragile, or low-texture objects. (bottom) Example input frames for three beverage container types: C1 has little texture, C3 shows only small deformations, and C1 and C2 are easily crushed.

The key contributions of this work are: combining controlled stimulus with computational imaging to analyze the mechanical properties of even challenging stiff, fragile or low-texture objects, without the use of specialized hardware; a linear noise reducing and motion amplifying filter which represents a significant simplification over prior approaches to motion magnification; and a set of statistics derived from optical flow which allow the mechanical properties of objects to be distinguished based on visual observations.

The proposed approach has a range of potential applications including informing prehensile perception of a compliant object of unknown physical state. An illustrative case is the problem of determining the fill level of a non-transparent cup. While the cup's static appearance remains unchanged, its global stiffness changes with fluid level. The mode shape of the cup will also vary, much like the resonant cavity of a musical instrument, and this change can be observed via the ratio of longitudinal to lateral visual deformation, akin to Poisson's ratio. We demonstrate that this statistic can be used to visually determine the cup's fill level, which can in turn inform a traditional manipulation strategy [7].

Our method is compatible with force sensing and can provide an estimate of inertial state before lifting, supporting grasping strategies that try to "feel" the grip [8]. It can also help compensate for drift/error in cases where less precise

*This research was partly supported by the Australian Research Council (ARC) Centre of Excellence for Robotic Vision (CE140100016) and the ARC Biomechanics Meets Robotics Project (DP160100714).

[1]D. Dansereau and J. Leitner are with the Australian Centre for Robotic Vision (ACRV), Queensland University of Technology (QUT), Brisbane, Australia {donald.dansereau, j.leitner}@qut.edu.au

[2]Surya P. N. Singh is with The Robotics Design Laboratory (RDL) at The University of Queensland, Brisbane, Australia spns@uq.edu.au

instruments are available, e.g. where force sensing resistors are employed, or when the finger/wrist is obstructed and unable to make adequate contact for load transference and sensing. Our approach is also compatible with a decision-making framework for varying excitation motions in the exploration and exploitation of an object [9].

II. Related Work

Interactive object recognition is a well studied task with a range of applications. In addition to direct sensing (e.g. *Syntouch BioTacs*), two main threads have emerged for handling deformable objects: learning approaches [1] and proprioceptive system identification [10]. Proprioceptive methods typically employ exploratory behaviours in combination with sensing to analyze objects. While such approaches can identify an object's mass and moment of inertia, many methods assume near-rigid body objects [11], [12].

Data-driven approaches have been extended to discern relationships between known deformable objects and observed reactions. Approaches include that of Gemici and Saxena [2], who perform predesignated actions with a tool (fork and knife) to learn a haptic representation of the mechanical properties of a set of known deformable objects using Dirichlet processes. Other avenues employ machine learning to augment low-resolution, time-varying tactile sensors. For example, Bhattacharjee, *et al.* [13] estimate object properties from incidental touch using Hidden Markov Models. These methods require extensive prior data and careful training to allow generalization across a diversity of novel cases.

Other approaches employ sound through contact or manipulation such as dropping, shaking, or crushing [13]–[15]. Inference is drawn in the frequency domain, e.g. between sound characteristics, and a discrete set of objects as classified using machine learning [16].

Deformable object analysis also has parallels with non-destructive testing and material property estimation [6]. Employing finite-element or point-mass and spring models [8], [17], displacement and compliance can be related, ultimately yielding metric estimates. In general, these methods are batch (off-line), designed for uncluttered environments, and seek a strong prior.

While existing strategies are often motivated by discrete manipulation tasks such as counting, arranging, sorting, and pile singulation [18], [19], they can also sometimes inform the handing of deformable 3D objects. In general this is more nuanced than planar grasping, for while the compliance can simplify some operations by deforming to compensate for errors, precision grasping and pickup of soft materials requires at least approximate object properties for anticipating gravity loading [7], [8].

Our approach brings together dynamic interaction and computational imaging to estimate the properties of deformable objects. This combination is powerful because video analysis can leverage the strong prior information afforded by intentional manipulation of the scene. In contrast to prior work, this allows targeted and isolated inspection of an object in the presence of clutter and interference.

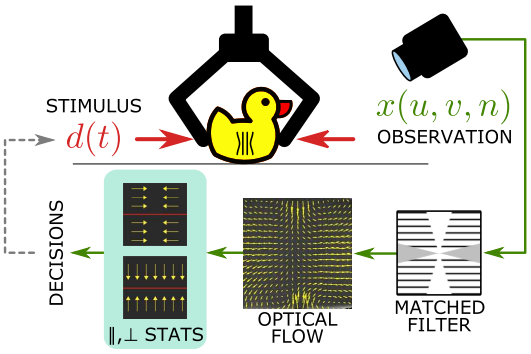


Fig. 2. The robot applies a time-varying stimulus $d(t)$ to a compliant object while observing the video sequence $x(u,v,n)$. The video is filtered to enhance contrast and motion, allowing fragile, stiff and low-contrast objects to be analyzed. The optical flow of the filtered sequence is decomposed into orthogonal components which yield a set of flow statistics. These in turn drive object and material discriminations, and ultimately decisions on manipulation strategy. Feedback to $d(t)$ indicates a potential to modify the stimulus to yield more information, e.g. starting gently and increasing as appropriate.

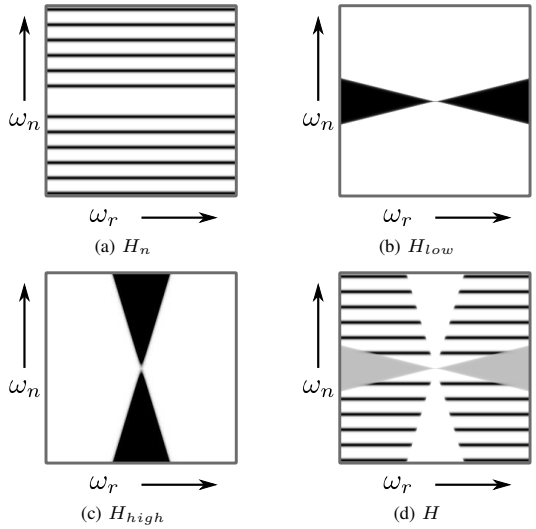
We build on recent motion magnification approaches [4], [20], which employ sophisticated filter banks for estimating motion. Our approach is simpler, employing a single linear filter that increases signal-to-noise ratio (SNR) and amplifies motion. The proposed filter is easily parallelized, making it appropriate for real-time embedded deployment, and we show it to be very effective despite its simplicity.

III. Interactive Computational Imaging

In computational imaging light is often considered in terms of the plenoptic function, which describes rays as a function of position, direction, and time, as seen by a passive observer [21]. We build on this by allowing the observer to interact with the scene, yielding a time-varying response to intentional stimulus. This effectively adds a dimension of mechanical interaction to the plenoptic function.

Fig. 2 summarizes the proposed approach. The robot applies a stimulus $d(t)$, observing an object's response using a camera. This yields a video $x(u,v,n)$, where u,v are pixel indices and n is the frame index. Some objects are fragile and will only tolerate small deformations, e.g. paper cups that are easily crushed. Other objects are stiff, and show little deformation even under strong stimulus. Finally, some objects have little contrast, making their deformation difficult to observe. These factors can make video analysis challenging, and so we propose a spatio-temporal filter, matched to the stimulus, that enhances contrast and amplifies motion while rejecting noise. The stimulus and matched filter are described in Sect. IV.

Following filtering we estimate optical flow and extract a set of statistics based on decompositions of flow into orthogonal components. This process is described in Sect. V. Finally, the optical flow statistics drive decisions about how the object should be manipulated. In this work we employ principal component analysis (PCA) to distinguish materials and object properties, as seen in Sect. VI. We leave as future work the formulation of manipulation strategies based on these discriminations.



(a) H_n (b) H_{low} (c) H_{high} (d) H

Fig. 3. Constructing the motion amplification filter. All figures are rotationally symmetric about ω_n; darker represents higher magnitude: (a) H_n contains all periodic motion with the prescribed period; (b) H_{low} contains all motion slower than a prescribed rate; (c) H_{high} contains all motion faster than a prescribed rate; and (d) H, the overall filter removing fast motion, passing slow motion, and amplifying periodic motion.

We envision the proposed method fitting into a broader strategy in which the robot begins with a gentle stimulus, varying it over time within predetermined safety limits, until informative deformation is observed. Once sufficient information is collected to characterize an object, a manipulation strategy is formulated and applied.

IV. Modulated stimulus and matched filtering

Visual characterization of material properties requires the perception of both texture and motion. Consequently, materials that allow little deflection due to their stiffness or fragility are difficult to analyze, as are those offering low textural contrast. These scenarios offer too weak a signal to overcome the camera's resolution and noise limits.

Modulated stimulus and filtering allows us to overcome these limitations. Assuming elastic material behaviour, a cyclic, stationary stimulus will yield a cyclic, stationary response at the same frequency, despite amplitude variations associated with material behaviours such as natural frequency and damping ratio. This allows us to design a filter which increases the SNR, drawing out subtle textures, and at the same time exaggerates small deflections through motion amplification.

A. Periodic Motion

Our simplified approach to motion amplification is built on a few key observations about periodic video. Assume a sampled monochrome video sequence $x(u, v, n)$, where u, v are pixel indices and n is the frame index. For linear material behaviours, a periodic stimulus with integer period T frames will yield a periodic video,

$$x(u, v, n + iT) = x(u, v, n),\ i \in \mathbb{N}_0. \tag{1}$$

The requirement for integer period T is easily approximated for high frame rates, e.g. tens of frames per cycle or more.

A periodic video has a discrete Fourier transform (DFT) $X(\omega_u, \omega_v, \omega_n)$ which lies on a periodic frequency-domain region of support (ROS)

$$\omega_n = \pm k\omega_0,\ k \in \mathbb{N}, \tag{2}$$

where $\omega_0 = 1/T$. This periodic ROS is sufficient to effect a significant improvement in the SNR of periodic video: noise will in general occupy the entire frequency spectrum, so by building a filter H_n matching the periodic ROS, we can reject much of this noise without affecting the dynamic scene content.

An example periodic H_n is depicted in Fig. 3(a). In this figure, darker regions correspond to higher amplitudes. Because all the frequency spectra shown here are rotationally symmetric about ω_n, we denote the horizontal axis $\omega_r = \sqrt{\omega_u^2 + \omega_v^2}$.

B. Smooth Motion

It is well established that an image patch moving with constant projected velocity appears on a tilted frequency-planar ROS [22]. For projected velocity $\boldsymbol{v} = [v_u, v_v]^\intercal$ pixels/frame, the ROS is given by

$$v_u\omega_u + v_v\omega_v + \omega_n = 0. \tag{3}$$

Stationary patches lie on the plane $\omega_n = 0$, and patches moving with small velocities lie on a family of planes forming small angles with $\omega_n = 0$. In fact, this family of planes can be described as the outside of a wide cone aligned with the ω_n axis. Such a family, which we denote H_{low} to indicate low-velocity scene content, is shown in cross-section in Fig. 3(b). By superposition, a scene comprising many elements moving with small but different projected velocities will also appear within H_{low}.

A final observation is that patches with very high velocities, beyond what is possible for a given scene stimulus, will lie on a frequency-cone ROS near the ω_n axis, as depicted in Fig. 3(c). This is useful as an additional tool for rejecting noise, as we can exclude this H_{high} region from the signal without adversely affecting the desired signal.

C. Combined Filter

Combining these observations we build a single-step linear motion-amplifying and noise-rejecting filter:

$$H = H_{low} + GH_n(1 - H_{high} - H_{low}), \tag{4}$$

where G represents the motion amplification gain, and the term that it multiplies contains everything moving faster than the minimum defined by H_{low}, slower than the maximum defined by H_{high}, and periodically as defined by H_n.

An example H is depicted in Fig. 3(d), for $G = 4$. The ratio between the lighter H_{low} and darker H_n reflects the gain in motion. The tuneable parameters of the filter are the minimum and maximum passband velocities, the period of the modulated manipulation, the sharpness of the rolloff of each component, and the gain G.

Note that by including H_{low} we have retained but not amplified slow motions. An alternative approach would be

to discard H_{low}, passing perfectly stationary content in its place. In practice we have found this unnecessary, but leave its exploration as future work.

D. Gain Selection

A crucial parameter of the motion amplifying filter is the gain G. In this work we manually fix G, but in practice we expect that a robot exploring its environment would take a more dynamic approach, for example slowly varying G until a desired minimum level of periodic motion is observed. If the gain is set too low, small and untextured deformations will not be observable. If the gain is too high, motion will suffer from saturation artefacts, ultimately yielding inaccurate statistics.

V. Dynamic Visual Material Properties

Having filtered the video sequence and amplified its motion to a perceptible level, we wish to estimate a set of properties which are useful in robotics tasks. These properties might be calibrated, metric, physically grounded quantities such as stiffness k, Young's modulus E, Poisson's ratio ν, and so on. Alternatively, raw visual quantities might directly drive decision or estimation processes: "Is that paper cup stiff enough to pick up?", "How much deflection should be applied when picking it up?", and "What's the best place to grip it?" are all questions that might be addressed using this method, without needing to establish calibrated quantities.

As proof-of-concept, we propose the use of well-established machine learning techniques to associate visual features with material or object behaviours. To this end, we extract a set of features from video of the stimulated scene, and employ these to cluster objects in a way that can drive decisions. Although the visual features we extract are not calibrated metric values, they are inspired by commonly employed physical quantities such as Poisson's ratio.

A. Optical Flow Components

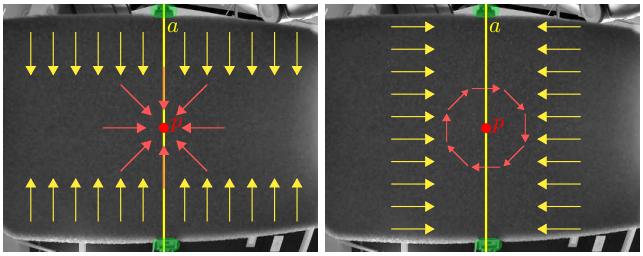
Following the filtering and motion amplification described in Sect. IV, we apply Matlab's implementation of Lucas-Kanade optical flow with Gaussian smoothing to obtain a dense flow field $\boldsymbol{V}$ over the object's visible surface. We then extract a set of visual features based on decompositions of the flow field into orthogonal components, as depicted in Fig. 4. The first of these decompositions is relative to the point $\boldsymbol{p}$ at the center of compression, and the other is relative to the axis of compression $\boldsymbol{a}$. In both cases, the flow is decomposed into "parallel" flow, towards $\boldsymbol{p}$ or towards $\boldsymbol{p}$ along $\boldsymbol{a}$, as in

$$V_{\parallel a} = \boldsymbol{V} \cdot \boldsymbol{F}_{\parallel a}, \quad V_{\parallel p} = \boldsymbol{V} \cdot \boldsymbol{F}_{\parallel p}, \tag{5}$$

and "perpendicular" flow circulating around $\boldsymbol{p}$ or moving perpendicular to $\boldsymbol{a}$, as in

$$V_{\perp a} = \boldsymbol{V} \cdot \boldsymbol{F}_{\perp a}, \quad V_{\perp p} = \boldsymbol{V} \cdot \boldsymbol{F}_{\perp p}, \tag{6}$$

where $\boldsymbol{F}$ are the parallel and perpendicular flow fields depicted in Fig. 4. Note that the decomposed flows are scalar fields, while the optical flow is a vector field.



(a) (b)

Fig. 4. Flow is decomposed relative to the point at the center of compression $\boldsymbol{p}$, and the axis of compression $\boldsymbol{a}$. (a) $\boldsymbol{F}_{\parallel p}$: flow towards the point $\boldsymbol{p}$ (red) and $\boldsymbol{F}_{\parallel a}$: along $\boldsymbol{a}$ (yellow); (b) $\boldsymbol{F}_{\perp p}$: flow around the point $\boldsymbol{p}$ (red) and $\boldsymbol{F}_{\perp a}$: perpendicular to $\boldsymbol{a}$ (yellow). The gripper is visible at the top and bottom of each frame, highlighted in green.

The decomposed flow components contain information about the dynamics of the scene, in both temporal and spatial dimensions. As a first pass we discard spatial information by taking the mean flow over each frame. We then compute running temporal means based on both the signed and absolute values of the per-frame sums. Examples of these statistics are included in Fig. 8.

Inspecting Fig. 8 we note that while the means of the signed values all converge to zero, as expected for a stationary cyclic signal, the means of the absolute values converge to values that depend on the material properties of the object being inspected. The absolute flow, shown in Fig. 8(i), reveals the extent to which the object is deformed for a given stimulus, a measure of softness. Distinct from this are the ratios of parallel to perpendicular components, which reveal how compressible or, conversely, fluid-like the object is. We expect highly compressible objects to exhibit small perpendicular flow, and more fluid-like objects to exhibit larger perpendicular flow, approaching the magnitude of the parallel flow.

B. Normalized Statistics

Following the arguments above, we propose a statistic inspired by Poisson's ratio, based on the proportion of flow parallel to the axis $\boldsymbol{a}$,

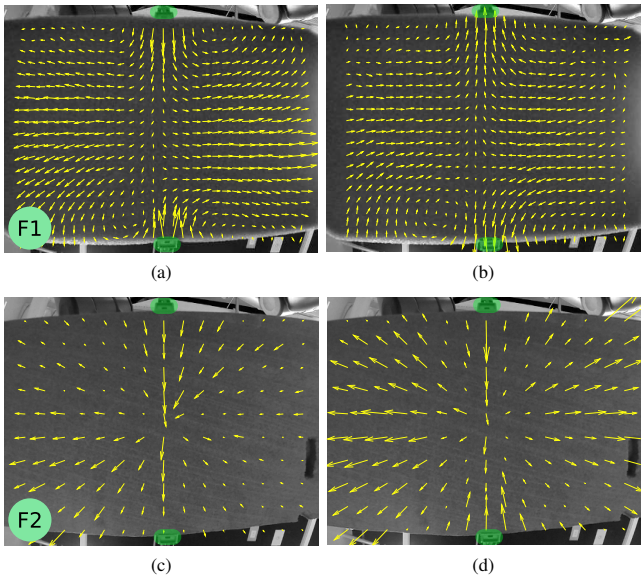
$$\nu = \sum |V_{\parallel a}| / \sum |\boldsymbol{V}| \tag{7}$$

which we call the visual compressibility ratio. This is an attractive feature because it is not tied directly to the magnitude of deflection, but rather to the shape of the flow.

Similarly, we propose the normalization of all four flow components by dividing by the total flow magnitude, as in (7). This makes these quantities at least partially immune to variations in magnitude due to changes in stimulus or positioning of the camera or object under investigation.

VI. Experiments

A Rethink Robotics Baxter robot was used to squeeze a series of deformable objects: foam blocks of varied stiffness and textural detail, paper and plastic cups with differing textures and filled to different levels, rubber ducks, squishy toys, plastic bottles and other beverage containers. Baxter is shown in Fig. 1 along with some of the objects we tested.



(a) (b) (c) (d)

Fig. 5. Estimated optical flow for (a) a relatively soft foam F1 shown at the moment of compression and (b) release; and (c) a relatively stiff foam F2 showing little deflection or texture. Unfiltered video of this foam yields a poor, inconsistent flow estimate – note how all flow is down and to the left – while (d) filtered and motion-amplified video shows a consistent flow field similar to that of the softer foam. The gripper is visible at the top and bottom of each frame, highlighted in green.

For the foam block experiments a GoPro camera was used at 1280×720×50fps, while the remainder of the experiments used a monochrome Point Grey Firefly MV at 640×480×60fps. Frames were manually cropped to include only the object under investigation, and the center and axis of compression $\boldsymbol{p}, \boldsymbol{a}$ were manually specified. We expect these manual steps to be automated in future.

A. Foam Blocks

Estimated flow fields for two foam blocks, a softer block F1 and a stiffer block F2, are shown in Fig. 5. For these experiments Baxter was set to repeatedly squeeze the blocks through a deflection of a few mm at a rate of about 4 Hz. The soft foam responded as expected and yielded consistent, coherent flow estimates, shown in Figs. 5(a) and (b) at the points of maximum compression and release, respectively.

The stiff foam only deflected by a fraction of a mm and yielded poor, inconsistent flow estimates. Low SNR due to lack of textural detail, together with the very small deflections, prevented accurate estimation of optical flow. The filtering process, shown in more detail below, yielded motion-amplified video which allowed for a much clearer motion estimate, shown in Fig. 5(d) for a constant motion gain of 10. A rotational component has been removed from the flow estimates for block F2 to facilitate visualisation only, this is not used in any computations.

The filtering process for the stiff foam F2 is visualized in Fig. 6. A single frame of the input is shown in Fig. 6(a), and a spatio-temporal slice of the video is shown in Fig. 6(c). The later was produced by selecting a vertical slice near the center of u, shown in red in (a), and plotting this slice as a function of n. Fig. 6(d) shows how histogram equalization brings out the texture of the foam, but also amplifies noise. Filtering alone reduces much of this noise, as seen in Fig. 6(e), and motion amplification reveals substantial motion throughout the foam, as seen in Fig. 6(f) for $G = 4$.

TABLE I

FILTER SETTINGS FOR CUPS EXPERIMENT

Duration	243 frames
Period	16 / 243 cycles/frame
Max Velocity	none
Min Velocity	1/30 pix/frame
H_n BW	0.25 / 243
H_{low}, H_{high} BW	2 / 243
Motion Gain	1

Decomposition of the flow field for foam F1, as described in Sect. V-A, is shown in Fig. 7, while Fig. 8 depicts per-frame running mean values for each of the decomposed fields and their absolute values. The sum of the absolute flow magnitude is also shown. Values are depicted over about five cycles of stimulus, 60 frames of video, during which all statistics have converged to stable values.

The compressibility ratio for both F1 and F2 are depicted in Fig. 9. While these two foams are very easy to distinguish on the basis of the magnitude of deflection alone, distinguishing them using the compressibility ratio indicates the power of this statistic to distinguish objects without exact knowledge of the magnitude of the flow. This is important where different manipulators are in use, or where positioning of the object under investigation can vary.

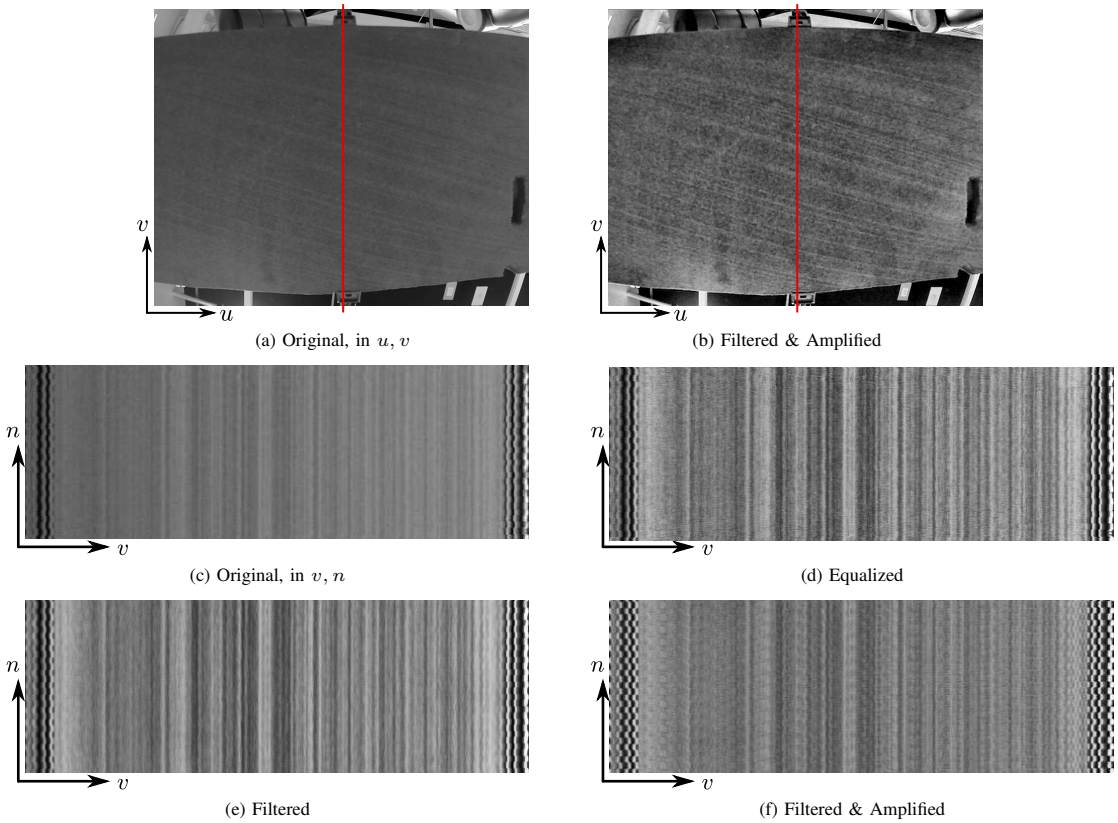
B. Cups, Bottles and Squishy Toys

In a larger experiment we investigated the behaviour of seven deformable objects over 59 videos, including variations in pose and illumination. The dataset included four beverage containers C1–C4, a water bottle B1 and plastic milk bottle B2, and a stress toy T1.

C1–C3 are depicted in Fig. 1: C1 and C2 are paper coffee cups, C3 is a soda can in a foam insulating sleeve, and C4 (not shown) is a rigid plastic cup. We tested each container and bottle at multiple fill levels by adding water in 94 mL increments, yielding 7 fill levels for B1, B2 and C3, 5 for C1 and C2, and 4 for C4. Baxter was set to squeeze each object through a deflection of a few mm, at a rate of about 4 Hz, which at 60 fps resulted in about 15 frames per cycle. For C1–C3, the experiment was repeated over two trials.

Because some objects showed little texture, e.g. C1, and others offered little deflection, e.g. C3, the filter described in Sect. IV was necessary to obtain usable optical flow estimates. The filter settings were selected manually, and are shown in Table I. The video duration was selected to contain an approximately integer number of cycles. Note that no motion gain was necessary for this experiment, as even small deflections were sufficient for detection by dense optical flow, once the noise-reducing filter was applied.

To establish our ability to distinguish object types, we applied PCA to the flow field statistics for all 59 videos. We employed the normalized statistics described in Sect. V-B, for their immunity to variations in absolute flow magnitude

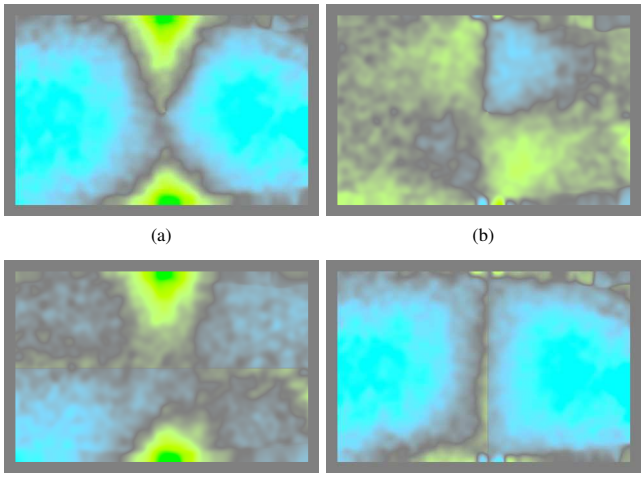


(a) Original, in u, v (b) Filtered & Amplified

(c) Original, in v, n (d) Equalized

(e) Filtered (f) Filtered & Amplified

Fig. 6. The stiff foam F2 (a) shown as a single frame in u, v, and (c) as a vertical slice near the center of u, as it varies with frame index n. Note the lack of motion and textural detail. (d) Adaptive equalization emphasizes both texture and noise, while (e) filtering reduces much of the noise, and (f),(b) filtering and motion amplification reveal texture and motion.



(a) (b)

(c) (d)

Fig. 7. Decomposing the flow field; (a) $V_{\|p}$: magnitude of flow toward the center of compression $\boldsymbol{p}$ and (b) $V_{\perp p}$: around $\boldsymbol{p}$; (c) $V_{\|a}$: flow inward along the axis of compression $\boldsymbol{a}$, and (d) $V_{\perp a}$: perpendicular to a. Green represents positive flow and cyan negative. Different materials show different characteristic flow field decompositions.

and object position. Figs. 10(a) and (b) show the dataset as seen through the first two principal components. Note that a single PCA was performed, but the results are divided into two figures for visualization purposes. From the figures it is clear that the objects show different behaviours, and their statistics cluster such that many object types are linearly distinguishable from the others. The bottles and toy, for example, are mutually separable, as are the four cup types.

Observe from Fig. 10(b) that the statistics for C2 appear more spread out than for the other cup types. We hypothesize this is due to C2 responding more strongly to its fluid level than the other objects. To test this hypothesis, we plotted the visual compressibility ratio of C2 as a function of fluid level, as shown in Fig. 11. Although only repeated over two trials, it's clear from the resulting plot that there is a strong potential to estimate fill level based solely on the compressibility ratio for this cup type, with higher fill levels yielding more repeatable results. Why the other container types show less variation, and how their fluid levels might be estimated, are left as future work.

C. Limitations

The proposed method and experiments have a few important limitations. The apparent brightness of a deforming object will change as surfaces face towards and away from illumination sources. Though our statistics focus on normalized ratios of textural flow, which should be invariant to such factors, further experimentation is called for to confirm invariance to illumination conditions.

In analyzing an object's behaviour we assume elastic deformation. Plastic deformation or rigid materials with zero deformation will not yield the cyclic motion required by the filtering and analysis stages. This calls for higher-

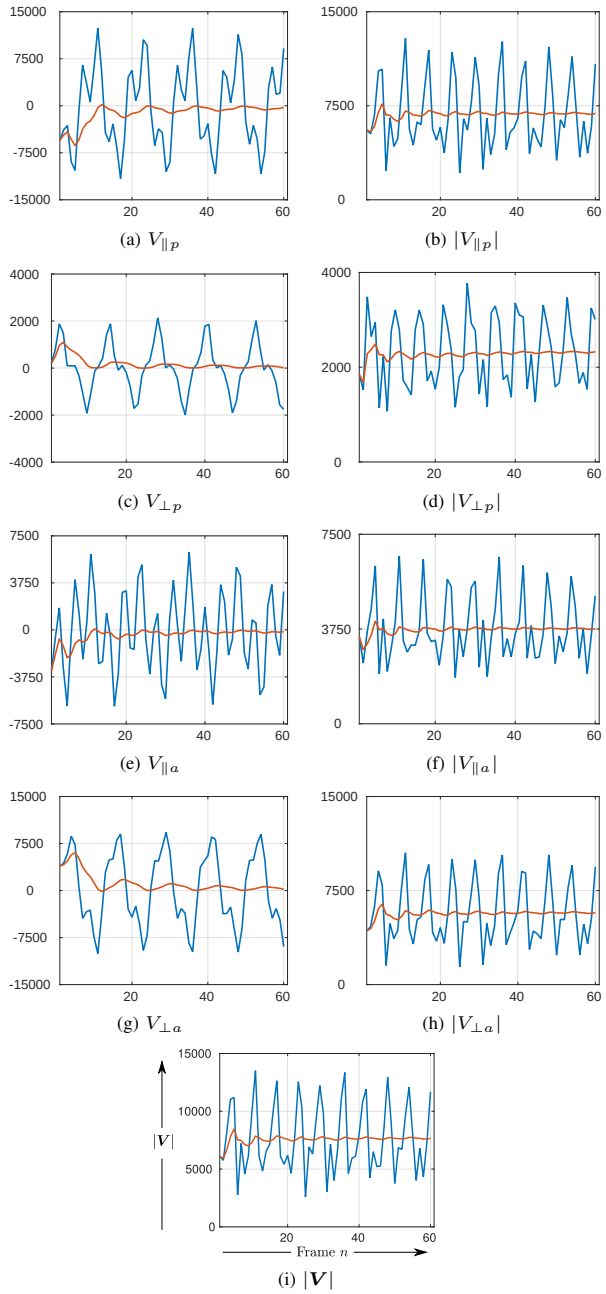


Fig. 8. Features based on decomposed flow fields, taken as (left) signed and (right) absolute values: (a),(b) flow into $\boldsymbol{p}$ and (c),(d) around $\boldsymbol{p}$; (e),(f) compression along $\boldsymbol{a}$ and (g),(h) perpendicular to $\boldsymbol{a}$; and (i) absolute flow. Signed values converge to zero as expected, while absolute values converge differently, revealing the characteristics of the object under inspection.

level reasoning, in which the robot begins with a gentle stimulus and increases this as needed, detecting the failure modes of inelastic or zero deformation. Note that there is no requirement for the material to be linear, in that it follows Hooke's law, provided deformation is elastic.

Decomposition of the flow field requires knowledge of the center and axis of compression. For the present work we have manually specified these values, and automating this task is left as future work, as is analysis of sensitivity to errors in these values.

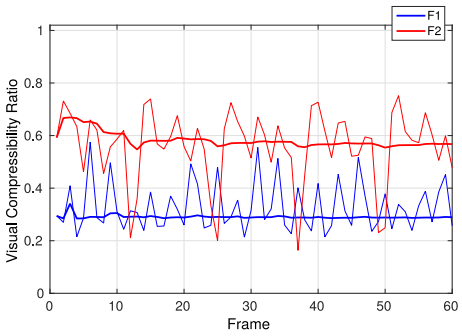


Fig. 9. Estimating a quantity similar to Poisson's ratio, the proportion of flow parallel to the axis of compression, for the soft and stiff foams F1 and F2, respectively. Note that this is not the same as softness, and in this case the softer foam has a lower compressibility ratio.

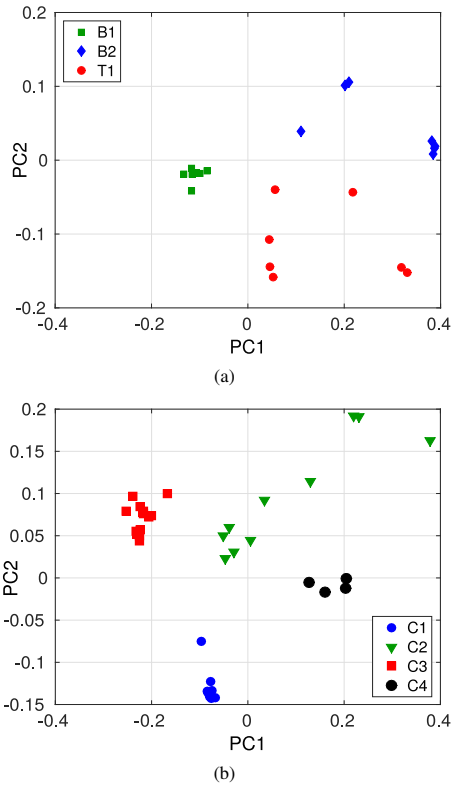


Fig. 10. The first two components of PCA conducted on normalized flow statistics for 59 videos of 7 object types: clearly clustered are (a) two bottle types and a squishy toy, and (b) four cup types.

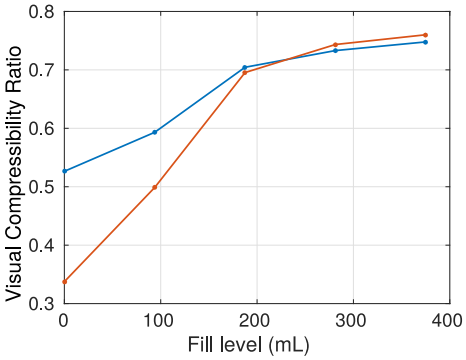


Fig. 11. Plotting the visual compressibility ratio for cup type C2 over five fill levels, over two trials, reveals a strong potential to estimate fill level from flow statistics alone.

We manually identified the spatial and temporal extents of the object under analysis. Given that we have control over the stimulus, automatically choosing an appropriate temporal extent is trivial. Segmenting out the appropriate spatial extent can be automated based on the magnitude of the observed flow field $|\boldsymbol{V}|$, which is near zero for background elements.

Throughout the experiments Baxter's gripper introduced small undesired motions to the object under analysis. Though the filtering stage is designed to ignore such small, non-periodic motions, a detailed analysis of its effectiveness at doing so would be appropriate.

VII. Conclusions and Future Work

We presented a method for visually analyzing objects through modulated manipulation and matched filtering and motion amplification. The algorithmic components are linear and easily parallelized, making them appropriate for real-time embedded deployment. No specialized hardware is required, only a camera and any means of applying periodic stimulus to the scene.

The proposed approach can be seen as an extension of computational imaging to exploit the ability of robots to interact with their environments. Thinking of robotic manipulation in this way, we anticipate generalizing beyond simple periodic stimuli into more complex schemes. Periodic motion in dynamic scenes may interfere with periodic stimulus, motivating extension to *coded* stimulus, in which extended patterns ("codes") drive the manipulator, and are chosen so as to occur rarely in natural scenes.

This result intuits from a mechanical perspective. For this may be thought of as an online approach that carefully tracks displacements and leverages the mapping between them and (unspecified) driving forces applied, to help identify object properties (such as stiffness), in essence giving a spring law (e.g. Hooke's Law, $F = kx$). In this way, this approach may be extended to identifying changing mechanical properties more quickly than traditional system identification.

Other extensions include dynamic gain and code selection and their incorporation into the resulting feature values, and an exploration of features based on spatial flow distributions and dynamics. We anticipate automating the collection of large datasets allowing the application of modern, high-performance learning algorithms and, ultimately, using these to drive manipulation decisions. Finally, we anticipate installing a force sensor to allow a robot to learn the visual estimation of physically grounded quantities.

References

[1] E. Klingbeil, D. Rao, B. Carpenter, V. Ganapathi, A. Y. Ng, and O. Khatib, "Grasping with application to an autonomous checkout robot," in *Robotics and Automation (ICRA)*, 2011, pp. 2837–2844.

[2] M. Gemici and A. Saxena, "Learning haptic representation for manipulating deformable food objects," in *Intelligent Robots and Systems (IROS)*, Sept 2014, pp. 638–645.

[3] S. Singh, M. Bergerman, J. Cannons, B. Grocholsky, B. Hamner, G. Holguin, L. Hull, V. Jones, G. Kantor, H. Koselka, G. Li, J. Owen, J. Park, W. Shi, and J. Teza, "Comprehensive automation for specialty crops: Year 1 results and lessons learned," *Intelligent Service Robotics*, vol. 3, no. 4, pp. 245–262, 2010.

[4] H. Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 4, p. 65, 2012.

[5] J. G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk, "Structural modal identification through high speed camera video: Motion magnification," in *Topics in Modal Analysis I, Volume 7*. Springer, 2014, pp. 191–197.

[6] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman, "Visual vibrometry: Estimating material properties from small motions in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5335–5343.

[7] A. Bicchi and V. Kumar, "Robotic grasping and contact: A review," in *Robotics and Automation (ICRA)*, 2000, pp. 348–353.

[8] H. Lin, F. Guo, F. Wang, and Y.-B. Jia, "Picking up a soft 3D object by "feeling" the grip," *Intl. Journal of Robotics Research*, vol. 34, no. 11, pp. 1361–1384, September 2015.

[9] K. M. Seiler, H. Kurniawati, and S. P. N. Singh, "GPS-ABT: An online and approximate solver for POMDPs with continuous action space," in *Robotics and Automation (ICRA)*, 2015, pp. 2290–2297.

[10] A. M. Okamura, M. L. Turner, and M. R. Cutkosky, "Haptic exploration of objects with rolling and sliding," in *Robotics and Automation (ICRA)*, vol. 3, 1997, pp. 2485–2490.

[11] T. Nakamura, T. Nagai, and N. Iwahashi, "Multimodal object categorization by a robot," in *Intelligent Robots and Systems (IROS)*, 2007, pp. 2415–2420.

[12] O. Khatib, "Inertial properties in robotic manipulation: An object-level framework," *The International Journal of Robotics Research*, vol. 14, no. 1, pp. 19–36, 1995.

[13] T. Bhattacharjee, J. M. Rehg, and C. C. Kemp, "Inferring object properties from incidental contact with a tactile sensing forearm," *arXiv preprint arXiv:1409.4972*, 2014.

[14] E. Torres-Jara, L. Natale, and P. Fitzpatrick, "Tapping into touch," in *Proc. of the Fifth Intl. Workshop on Epigenetic Robotics*, L. Berthouze, F. Kaplan, H. Kozima, H. Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojanov, and C. Balkenius, Eds., vol. 123. Lund University Cognitive Studies, 2005, pp. 79–86.

[15] J. Sinapov, T. Bergquist, C. Schenck, U. Ohiri, S. Griffith, and A. Stoytchev, "Interactive object recognition using proprioceptive and auditory feedback," *Intl. Journal of Robotics Research*, vol. 30, no. 10, pp. 1250–1262, 2011.

[16] J. M. Romano, J. P. Brindza, and K. J. Kuchenbecker, "ROS open-source audio recognizer: ROAR environmental sound detection tools for robot programming," *Autonomous Robots*, vol. 34, no. 3, pp. 207–215, 2013.

[17] A. Shabana, *Vibration of discrete and continuous systems*. Springer Science & Business Media, 2012.

[18] M. T. Mason, *Mechanics of Robotic Manipulation*. Cambridge, MA: MIT Press, 2001.

[19] L.-Y. Chang, J. R. Smith, and D. Fox, "Interactive singulation of objects from a pile," in *Robotics and Automation (ICRA)*, 2012, pp. 3875–3882.

[20] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Phase-based video motion processing," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, p. 80, 2013.

[21] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," *Computational models of visual processing*, vol. 91, no. 1, pp. 3–20, 1991.

[22] L. Bruton and N. Bartley, "Three-dimensional image processing using the concept of network resonance," *IEEE Transactions on Circuits and Systems*, vol. 32, no. 7, pp. 664–672, 1985.