PastVision: Exploring "Seeing" into the Near Past with Thermal Touch Sensing and Object Detection

For Robot Monitoring of Medicine Intake by Dementia Patients

Martin Cooney, Josef Bigun

School of Information Technology, Halmstad University, Halmstad, Halland, Sweden, martin.daniel.cooney@gmail.com

Abstract

We present PastVision, a proof-of-concept approach that explores combining thermal touch sensing and object detection to infer recent actions by a person which have not been directly observed by a system. Inferring such past actions has received little attention yet in the literature, but would be highly useful in scenarios in which sensing can fail (e.g., due to occlusions) and the cost of not recognizing an action is high. In particular, we focus on one such application, involving a robot which should monitor if an elderly person with dementia has taken medicine. For this application, we explore how to combine detection of touches and objects, as well as how heat traces vary based on materials and a person's grip, and how robot motions and activity models can be leveraged. The observed results indicate promise for the proposed approach.

Keywords

Thermal Sensing, Action Recognition, Home Robots, Monitoring, Medicine Adherence.

1 INTRODUCTION

This paper explores a concept for how to infer a person's recent past actions using thermal touch sensing and object detection, as shown in Figure 1. We focus on the case of a home robot which should monitor medicine intake, toward potentially supporting health and well-being in elderly persons with dementia.

An "action", sometimes called a sub-activity or actionlet, refers to a basic human behavior; actions can be combined to form more complex "activities". A "past" action here refers to an action which was not sensed by the system at the time the action was performed. A "thermal camera" outputs grids of pixels whose intensities depend on the temperatures of remote objects passively sensed through radiated long-wavelength infrared light; we refer to spots where heat has been transferred through touch as "heat traces" or "thermal touches". "Object detection" involves both finding and recognizing objects. "Well-being" refers to a subjective feeling of being well, related to happiness and good quality of life.

Inferring past actions will be useful in scenarios when sensing can fail and the cost of not recognizing an action is high (e.g., when occlusions could interfere with monitoring for terrorist attacks or acute health problems). The scenario we focus on in the current paper, home robots intended to care for elderly persons with dementia, is characterized by both of these properties. A robot might not be able to sense a person if

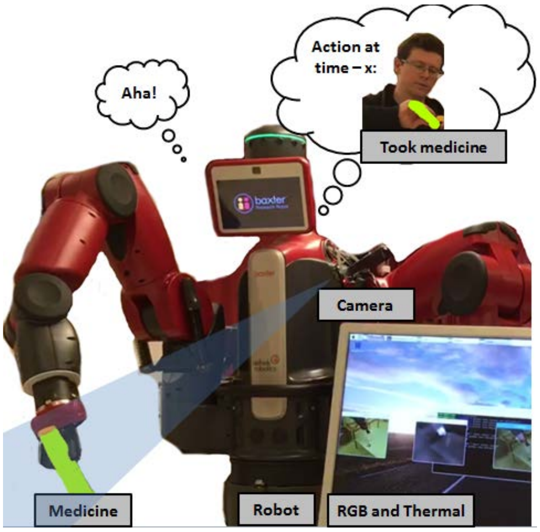

Figure 1. PastVision: basic concept

it is in a different part of a home to conduct a task or to provide a person with some time alone, or if there is a temporary sensor failure. Yet monitoring could help to save lives by allowing a robot to intervene in emergencies; e.g., by sending out an alarm and providing health care professionals with information. Emergencies can arise from various causes such as dehydration, wandering at night, burns from fires or hot water, or non-adherence to medicine regiments. The latter problem has been reported as particularly serious and widespread. Taking too little, too much, or the wrong medication can have disastrous consequences, including death;

forgetting is a problem for dementia patients; and in some situations, such as when doses are frequent, compliance by only half of patients with a regiment has been reported [22]. Thus, we focus on monitoring of medicine intake.

Various approaches could be used by a machine to infer if a person has recently taken medicine. For example, cameras, RFID tags, and weighing scales could be used to determine, e.g., if a medicine package has been moved or weighs less, sounds of opening and swallowing could be detected, or a robot could directly ask a person what they have done. Some shortcomings of such approaches include that medicine might have been moved absentmindedly or replaced in a similar position after intake, frequent weighing and scrutiny could make a person feel little peace of mind, sounds can be easily muffled by noise (e.g., from a television, phone call, or alarm), constantly being asked could be irritating, and a person, especially with dementia, might not remember everything they have done.

Here we turn our attention to one possibility for inferring past actions, based on thermal touch sensing and object detection. Our proposal rests on an assumption that activities of daily living often involve touches and objects: we touch appliances when cooking, eat with cutlery and dishes, dress ourselves with clothes taken from cupboards, clean ourselves and our environments with brushes and cloths, and touch packages to take medicine. Furthermore we guessed that the result in a typical scenario of a home heated to around 20 degrees [36] would be some heat traces on objects which could be sensed after actions had occurred.

The challenge was that how or if thermal sensing can be combined with object detection was unclear. Furthermore, we did not know what can be sensed thermally in a typical scenario and how a robot could move to facilitate inference.

To address this challenge we propose an approach which we call PastVision, based on building a prototype, while also exploring some typic thermal properties of materials and grips, and strategies for a robot to move to reduce uncertainty via locomotion and manipulation.

To assist others interested in exploring this promising area, source code for an implementation of PastVision, as well as a video, will also be made available online (from martin-cooney.com).

2 RELATED WORK

The current work relates to recognition and inference of (past) actions, thermal touch sensing, and motion planning for healthcare robots dealing with medications.

2.1 Past action inference

Various innovative work has been conducted on autonomous inference, also of past events, actions, touches, and activities. Timing was used by a robot to infer who was doing an action, by correlating motion commands and visually sensed movements [13]. Adaptation of another robot's model of itself based on observed motions was used to infer a person's goal of pressing buttons [14]. And, a robot was able to infer the rules of games such as hide-and-go-seek and tag, by correlating its own motor commands and the actions of salient detected objects [10].

Inference of the past is also common in various fields. Looking up, evidence can be seen in the night sky of events such as star formations and deaths which have occurred millions of years ago. Looking down, fossils and rocks describe various events in the evolutionary history of species and the history of our world which we have not directly witnessed.

As well, many approaches have been used toward ascribing meaning to actions (e.g. [7]). For example, sounds were used to recognize some actions in homes [31]. And, similar to our approach, information from objects has been used to recognize actions like reading a book [27].

Also, touch-based actions have been recognized in various ways [30, 16]. For example, Lee demonstrated a way to separate spatiotemporally co-occurring touches using spatial Independent Component Analysis (ICA) and time series clustering [19]. Typical grips people use for opening packages have also been described (e.g., "spherical", "cylindrical" or "lateral" grips for bottles), which could be recognized [3].

Furthermore, some recent work has tackled "predicting" actions, which is typically framed as a problem of early recognition. For example, a human's next actions have been predicted by modelling object interactions using a conditional random field (CRF) on RGB-D data (for some activities such as taking medicine) and generating likely next moves [17]. Another system leveraged action sequences and objects [20]. And, a "memory" model was used to associate observed actions with previously learned ones [35].

What has received little attention is the combination of past inference and action recognition (possibly because actions cannot be sensed after completion by typical sensors such as cameras and microphones); furthermore, we are not aware of a previous work which proposed a way to ascribe meaning to touches by combining object detection and thermal touch sensing. (On the side, we also suggest how the basic approach of early recognition for action prediction can be inverted to see deeper into the past).

2.2 Thermal touch sensing

Thermal sensing has been used for many applications, some also related to healthcare and detecting touches.

In healthcare, use of infrared thermography has been described for detection of Raynaud's, fever, injury, breast cancer, diabetes neuropathy, dental and brain problems [25], apnea/hypopnea [21] and potentially skin cancer [5]. Moreover, thermal cameras have been used, along with an RGB camera and GPS, to find reclining people in emergencies [26], and to detect falls in a home [37].

Thermal sensing has also been studied in the area of human interfaces as a way to detect touches. Described as a concept by Iwai and Sato for a table-shaped interface [15], Benko et al. reported how touches could also be detected on a spherical display by normalizing, binarizing, and detecting and tracking connected components [4]. Larson and colleagues, similar to the current paper, sought to explore various facets of thermal touch sensing for their application [18]. The authors described a video-based approach for detecting heat traces which searches within a region in which a person's hands have moved recently; Bayesian estimation is conducted per pixel based on spatiotemporally smoothed temperatures, changes over time, and background subtraction, and belief is compared to a threshold for binary classification. Furthermore, the authors distinguished touches from hovering, recognized shapes of gestures and touch pressures, also for touches with multiple fingers, and suggested that interfaces could also take advantage of thermal reflections. This latter suggestion was investigated by Shirazi and colleagues [29], who also examined properties of some typical materials: glass, tile, MDF, and aluminum [1].

For the current study, it was considered that it would be useful to have an indication for how feasible it is to sense thermal touches to monitor medicine intake: e.g., how long touches remain visible on typical packaging materials, and how people typically touch medication packages.

2.3 Motion planning for robotic medication reminder systems

Various robotic systems have been proposed to remind people about medicine, bring medication, or check if a person has taken medication; and usability studies have confirmed that elderly see the usefulness in such systems [34]. One early work reported an aim for the nurse robot Pearl to remind in such a way that people comply without feeling annoyed or becoming overly reliant [23]. Also, some other robots capable of providing medication reminders have been designed to care for persons with mild dementia or Chronic Obstructive Pulmonary Disease (COPD) in smart environments; moreover some robots have been designed to fetch medication when users are indisposed [11]. Similarly, a small robot was proposed for keeping track of medicine and navigating to an elderly person to provide medicine [6]. For checking medicine intake, RFID tags were proposed for detecting if a person extracted and replaced a sheet of tablets from a tray on top of a small mobile robot [33]; and, some work outside of robotics has examined detecting swallowing by sound or muscle activation [2].

Additionally, much work has shown how robots can move to achieve various goals. For example, social force models have been used to describe how robots can locomote in a human-like, safe way (e.g., [28]). Another strategy was proposed for how a robot can try to move to detect people using a thermal and RGB camera [9]. Some other robots have manipulated objects to better segment them visually for grasping [12] and learn affordances [32]. The current work proposes how locomotion and manipulation can be used by a robot to facilitate past action inference for medication intake monitoring.

2.4 Contribution

The contribution of the current paper is exploring the concept and feasibility of combining thermal touch sensing and object detection to infer past actions, which we call PastVision, in the context of robotic monitoring of medicine intake; along the way, some typical thermal touch parameters and strategies for robot motion were also explored for this context.

3 METHODS

3.1 PastVision: making basic sense of thermal touches via object detection

This section describes step by step our approach for inferring past actions based on thermal touch sensing and object detection, illustrated in Figures 2, 3, and 4.

In general, the precondition for our algorithm is that a system has inferred that there might have been some missed data. The input to the algorithm is one thermal and one RGB image. The algorithm then outputs labels of touched objects. The postcondition is that the system has inferred what it has missed (a person's recent interactions with objects).

In particular, we focus here on a simplified basic scenario in which one robot is monitoring one person's medication intake; the scenario can be further simplified if the robot has been constrained to surveil a designated area in which only objects of interest are placed. During monitoring, some problem occurs with the robot's sensing, which is then fixed several seconds later. For example, this could be due to a person temporarily blocking the robot's view. The robot in front of some medicine packages then seeks to infer what the person has done. Here we consider some typical objects related to medicine intake: pill bottles, medicine boxes, flat sheets of medicine, water bottles, cups, and glasses (for oral medicine), as well as creams and syringes (for topical medicines and injections).

The process for inferring past actions targeting such objects is shown in Figure 2:

- (a-b) a thermal and RGB image are recorded
- (c-d) registration is conducted using a simple mapping found ahead of time
- (e) the thermal image is thresholded to detect warm regions
- (f) objects are detected within the RGB image, yielding bounding boxes and predicted classes

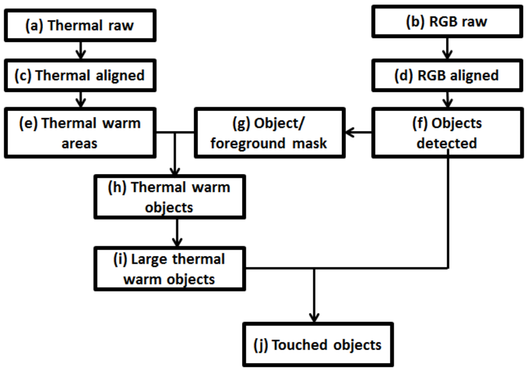

Figure 2. PastVision: process flow

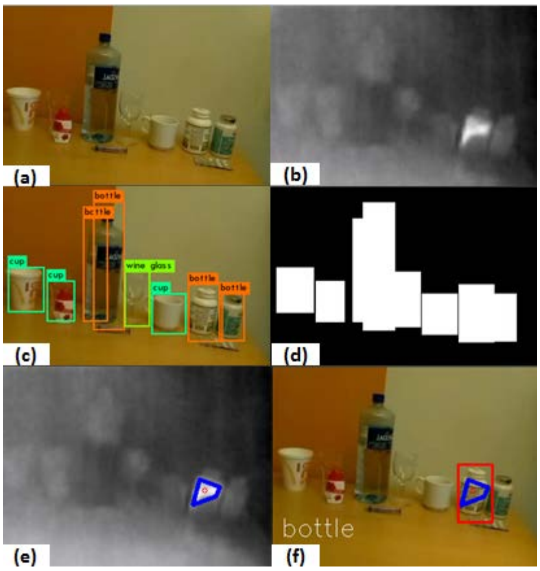

Figure 3. PastVision: (a) RGB image, (b) thermal image, (c) objects detected, with a few false positives and negatives, (d) mask image to reduce noise, (e) thermal image with touched region and contour centroid drawn, (f) touched object identified

- (g) The bounding boxes are used to prepare a mask image to extract foreground regions in the thermal image.
- (h) The intersection (the bitwise and) of the object mask and thresholded thermal image is computed to remove noise, which can arise due to heat from light sources, thermal reflections or unintended touches.
- (i) Connected components (contours) are detected. A threshold is used on contour size to further remove small noise.
- (j) For each contour, the centroid of the contour is calculated, and distances are found to centers of detected objects. The algorithm ouputs the labels of detected objects with bounding box centers closest to the centroids of detected heat traces.

Figure 4. Example of a simplified approach for extracting heat traces and not humans via thresholds, morphology, and a basic shape model for touching: (a) thermal image, (b) RGB image with heat traces drawn by the algorithm

We note some considerations in using this approach:

- Action inference. There is no guarantee that inference is correct. For example, if a person touches a water bottle their intention could be to drink but could also be to dilute paint, water a plant, or cool their forehead. For this reason, our algorithm also approximates detected contours with the Ramer-Douglas-Peucker algorithm, in the expectation that shape information will provide useful information for inferring what a person has done. More information about context (e.g., current actions, timing models) will also be useful.
- Videos containing humans. Our approach is not limited to single images of objects; we also offer an example of the algorithm used in video, shown in Figure 4; we note that due to the simplified scenario hot objects such as humans can be ignored by the algorithm using some simple processing. More complex approaches can be used to enhance robustness of sensing for more challenging scenarios (possibly at the cost of incurring some waiting time). For example, motion estimation can be used to detect pixels with continuously changing heat, with a precalculated model of heat decay fit to every pixel to eliminate false pixels and confirm correct pixels. The approach used by Larson and colleagues of detecting heat traces near where a person's hands move could also be applied, although this approach was designed for a user interface with assumptions which might not hold for monitoring medicine intake (a robot might not have a background model, and a person's hands might not be visible). Also, cues like skin color could be used for detecting people.
- Feasibility for the application. For monitoring medicine intake, one problem could be if current object detection methods cannot recognize required categories. To gain insight, we checked the degree to which typical medication dosage forms are found within a common object recognition dataset, used for the ImageNet Large Scale Recognition Challenge (ILSVRC). We found pill bottle (ImageNet category number n03937543), water bottle (n04557648), cup (n07930864), goblet (n03443371), beer glass (n02823750), sunscreen (n04357314), lotion (n03690938), Band Aid (n02786058), face powder (n03314780), and syringe (n04376876). Two categories which appeared not to be represented

were paper boxes and flat sheets of lozenges. (As a check, an image of a paper box submitted to an ImageNet classifier was perceived to be of an “eraser” and lozenges were also not detected in the example image in Figure 3). However, it seemed many categories of interest for monitoring medicine intake can already be recognized, suggesting the promise of this approach for the application.

- Proof-of-concept. We note that the examples in Figures 3 and 4 show a highly simplified case chosen for initial proof-of-concept investigation: images were recorded in a lab environment at regular temperature with few objects and a simple background.
- Seeing further into the past via activity recognition models. We propose that, in the same way that early recognition of activities allows action prediction, recognizing the last action in an activity could allow inference into the more distant past. Other potential benefits of considering activities include increased certainty of inference if multiple related actions are observed, and possible ability to infer difficult-to-sense actions involving metals, for which heat traces disappear quickly. We provide two examples below of potential action sequences for medicine intake, while noting that many variations are possible: oral intake might include fetching objects, pouring water into a glass, opening a medicine bottle, taking out a pill, swallowing it, and drinking water; injections or topical applications might include fetching objects, baring a body part, washing and/or disinfecting, applying or injecting, and covering. Automatic modeling of such activities is possible, e.g., through “fluent learning” combining interval calculus and co-occurrence frequencies [8].

Although our approach and scenario are highly simplified, we verified that our system was able to detect unobserved object interactions which would be highly difficult for previous systems using only thermal touch detection or object detection.

3.2 Typical properties of heat traces

We proposed an approach for how to infer which objects have been recently touched by correlating extracted locations of heat traces and objects, but some questions remained with regard to what can be recognized. In regard to timing, if only a few milliseconds into the past can be recognized our approach might not be very useful. As well, we did not know to what degree it might be possible to recognize different kinds of touches, like typical ways of opening a medicine package or just picking up a package and putting it down.

To answer the questions, two simplified tests were conducted. First we touched some typical materials for medicine packages for a set time (2 seconds) lightly with the pad of one finger and checked how visible the heat trace was after 30 seconds. Materials tested were polyethylene terephthalate (PET, for a pill bottle), high-density polyethylene plastic (HD-PE, for a sunscreen lotion bottle), paper (for a medicine box), glass (for a drinking glass), and ceramic (for a cup). Ambient room temperature was measured to be 24.7 degrees, and hand temperature to be 36.8 degrees.

For the second test we recorded some data of opening a pill bottle with spherical, cylindrical or lateral grips.

As a result of the first test, we were surprised by the extent that short touches to all materials tested except ceramic resulted in heat traces which could be seen for over thirty seconds, as shown in Figures 5 and 6. We think this indicates the feasibility for our proposed approach because we expect that people will often touch longer than two seconds to open a medicine package, especially elderly persons with dementia.

For the second test, patterns for spherical and lateral grips, shown in Figure 7, appeared to be similar. This is because, although the fingers on the top are positioned differently, in unscrewing the bottle, fingers are removed and replaced in different positions several times, leading to unclear smudging at the top for both. Heat traces for the cylindrical grip had a different appearance, due to some smudging from the lower half of a hand brushing against the bottle. The possibility was also suggested that these grips could be distinguished from some absent-minded touching if the latter only affects one region of the bottle.

3.3 Maximizing information gain via robot motion tion planning

In the last section, our investigation suggested that being able to see a package from various angles could yield useful information. Based on this, we propose that robots can enhance past action recognition by (1) locomoting and (2) manipulating objects.

(1) Locomotion: A robot could seek to find a position which allows better observation if its view becomes occluded and is not expected to improve immediately. Here we consider an example of a person moving in front of the robot to interact with some objects, who then becomes stationary. We propose that the robot should move to:

- minimize occlusions by the person on objects in its view
- be close enough to see objects clearly and appear socially positive to a person
- be far enough not to prevent object interactions or bother a person
- minimize work

To achieve such requirements one way to calculate how a robot should move is to use a social force model [28]. Social force models seek to model natural locomotion via forces pulling a robot toward a goal state and pushing it away from obstacles. Individual forces, coinciding or conflicting, are summed to calculate a net force acting on

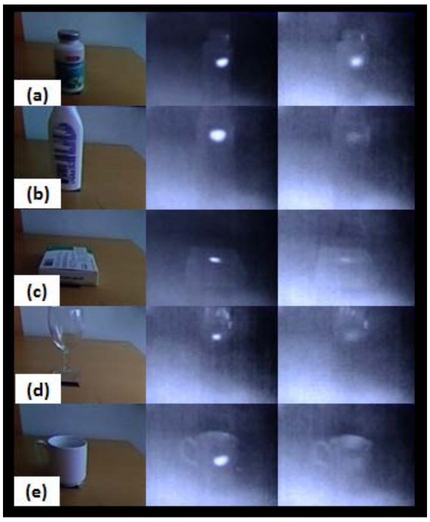

Figure 5. Heat trace decay over time for some typical materials for medicine intake: for (a) PET, (b) HD-PE, (c) paper, (d) glass, and (e) ceramic; (left) RGB data, (center) thermal data soon after touching, (right) thermal data after 30s

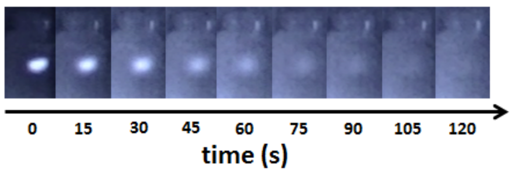

Figure 6. Heat trace decay over time for PET. After two minutes the heat trace was difficult to see

the robot. All of the proposed requirements above except the first are common in robot social force models. We propose to model such a requirement via torque (a force applied to the robot position with the object position as an axis of rotation). Similar to the Collision Prediction (CP) specification, which computes a hypothetical time when pedestrians would become closest, the direction of force can be calculated by translating the robot position along the lever arm connecting it to the objects until the distance between robot and human is minimal. Intuitively, this means a robot could try to move circularly left if a human is in front and to its right, and circularly to the right if a human is standing in front and to the left. The proposed model is shown in Figure 8.

A more complex approach might involve predicting when and where a human will move next to determine how

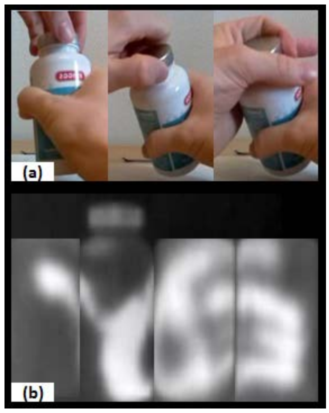

Figure 7. Typical grips for opening packages: (a) spherical, lateral, and cylindrical; (b) an example of a heat trace for a spherical grip

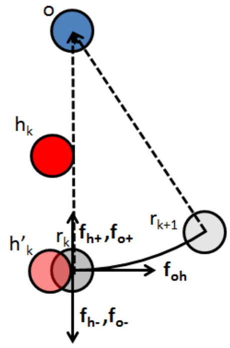

Figure 8. A social force model can be used to model how a robot can move to monitor medicine intake when occluded. Symbols: *o*: object position, h_k and r_k human and robot positions at time *k*, h'_k hallucinated human position to calculate torque on the robot (and force f_{oh}); f_{h+} and f_{h-} attractive and repulsive forces toward a human, f_{o+} and f_{o-} attractive and repulsive forces toward objects; the dashed line represents the robot's orientation

a robot should move. For example, people's bodies might start to turn as they shift their attention to some new object. A robot could also use knowledge of typical activities; e.g., a person picking up some bread might next move to a toaster.

(2) Manipulation: We propose that a robot can pick up objects to better determine if they have been touched recently and how. This addresses the problem that touches might not be fully visible from the robot's initial perspective.

To gain some insight, we acquired some data by conducting some actions (either pretending to take medicine, or just picking up and putting down packages) and then commanding our robot to pick up packages and observe them from different angles. (A more complex approach could involve checking that the robot's gripper is not covering a heat trace, and use image stitching for cylindrical or spherical objects and segmentation.) Figure 9 demonstrates how inference can be facilitated by gaining extra information from viewing objects close-up and from various perspectives.

3.4 Implementation: Software and Hardware

To explore the PastVision concept, some tools were used, which are described below for reproducibility. For software, we used Robot Operating System (ROS) to send commands and data, OpenCV 3.2.0 for image processing, and Darknet code for YOLO (You only look once) object detection. For the latter, we used YOLO9000 version 2 because it was easy to use (predicting bounding boxes and class probabilities at once), fast, and provided excellent performance [24]. YOLO uses a single convolutional neural network with many layers (deep learning) trained on a mixture of detection and recognition data, and non-maximum suppression with a confidence threshold (0.25).

For hardware, we used a thermal camera and small computer attached to a robot, and a remote desktop for control and processing. An inexpensive 80 x 60 forward looking infrared (FLIR) camera was used, which is capable of detecting long wave infrared wavelengths from 8 to 14 microns with a 51-degree horizontal field of view (63.5 degree diagonal) and thermal sensitivity less than 0.05K (we consider this to be sufficient for the context of detecting touches to medicine packages, as touches to typical materials can result in much larger changes in surface temperature [1]). Thermal images were read over SPI by a Raspberry Pi 3, with RGB images obtained over CSI. Unoptimized code, showing thermal and RGB streams both independently and overlaid while recording data, ran at approximately 8.6fps. Processing was conducted on a desktop (i5 2400 CPU @ 3.1GHZ).

For a robot we used a Baxter humanoid upper body on a Ridgeback mobile base (approx. 100 x 80 x 180 cm (l x w x h), weight approx. 210 kg). The robot was equipped with various actuators: two seven degree of freedom (DOF) arms incorporating springs for safety with a reach of 1.2m capable of lifting approx. 2kg (25kg with safety disabled), four independent omnidirectional wheels with maximum speed 1.1 m/s, and a touch display screen showing a face. For sensors, the robot had a LIDAR, a 360 degree ring of 12 ultrasonic range sensors, three additional RGB cameras, two infrared range sensors, force sensors and accelerometers in each arm joint, base encoders, and an inertial measurement unit (IMU). Processing on the robot took place on a i7-3770 Processor (8MB, 3.4GHz) with HD4000 Graphics driver.

We note that, although we found these tools to be useful for our exploration, the proposed approach is not dependent on this implementation and can be applied with different sensors and robots: the thermal and RGB cameras used are standard off-the-shelf components, and various robots are capable of locomotion and manipulation.

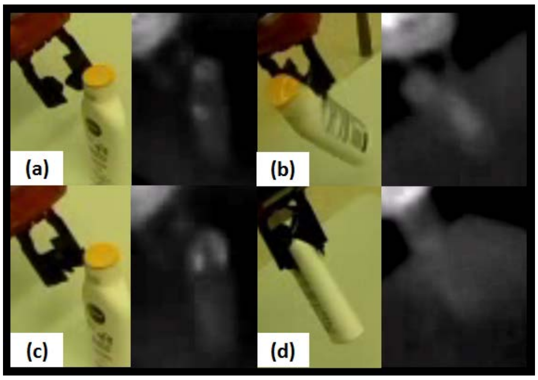

Figure 9. Robotic manipulation of touched objects can yield additional information to aid inference: (a-b) depict the front and backside of a bottle which has been opened, and (c-d) the front and backside of a bottle which has only been picked up and replaced; the thermal images in (a) and (c) appear similar, with minor signs of touching, but (b) shows heavy touching on the backside compared to (d).

4 DISCUSSION

In the current paper, we presented PastVision, an approach to infer past actions by correlating thermally sensed touches and object detection, with a focus on facilitating robust robotic monitoring of medicine intake in dementia patients. We furthermore explored the concept, checking how long touches on typical materials used for medication packaging persisted and how heat traces from typical grips related, and suggesting how robot motion planning can be used to enhance inference results.

4.1 Ethics

We believe that privacy is an important consideration and we do not suggest that monitoring is for everyone; it should only be provided for those who want it and foresee benefits which outweigh the downsides. Also, we advocate that monitoring should not lead to control being taken away from people, but rather that it should enhance people's confidence and increase their control over conditions such as dementia. For example, a person with dementia who loves cooking might feel better about using a hot stove if it is believed that someone is there who can remind them if they forget to turn off the stove at the right time.

4.2 Limitations

There are many challenges affecting thermal and RGB cameras. Thermal cameras must deal, inter alia, with high variances in ambient temperatures (e.g., in winter or summer), heat contamination when objects are touched multiple times, and thermal reections from remote heat sources. RGB cameras can be affected by illumination and shadows, and we consider general object detection against complex backgrounds (e.g. with wallpaper or art) to still be an open problem in computer vision. Thus, we emphasize that the current work is a feasability study, whose results are limited to a highly simplified scenario chosen for initial investigation: images were recorded in a lab environment with controlled temperature, few objects and a simple background. We intend to extend these results in future work.

4.3 Future Work

Our next step will be to conduct further tests and obtain some quantative results, e.g., for system accuracy in inferring touched objects and discriminating touch types (medicine intake vs. just touching), as well as durations for which traces can be detected. Also, the surface has only been scraped in terms of what kinds of past action inference can be conducted. For example, for thermal inference, temperature changes not only in objects but in people will facilitate inference (e.g., cooling of a person's mouth and hands might provide extra confirmation that the person has taken some oral medication). A variation of the intersection over union (IOU) metric could also be used for detecting which object has been touched, since centroid distances do not take into account bounding box locations. And, how to make sense of spatiotemporally co-occurring thermal touches will be an interesting problem.

Other modalities such as sound and olfaction will also be useful. For example, sounds such as nose-blowing might indicate a person has caught a cold, and hair-drying might indicate a person has taken a shower recently or been out in the rain. Also, by touching objects people can leave scents; the promise of olfaction as an inference modality is suggested by the amazing abilities of some animals to track and locate objects people have touched via smell. We believe such work will contribute to the ability of robots to monitor and care for humans, toward supporting well-being.

5 REFERENCES

[1] Abdelrahman Y, Shirazi AS, Henze N, Schmidt A. 2015. Investigation of Material Properties for Thermal Imaging-Based Interaction. CHI 2015. http://dx.doi.org/10.1145/2702123.2702290

[2] Amft O, Troster G. 2006. Methods for Detection and Classification of Normal Swallowing from Muscle Activation and Sound. Pervasive Health Conference and Workshops.

[3] Bell AF, Walton K, Chevis JS, Davies K, Manson C, Wypych A, Yoxall A, Kirkby J, Alexander N. 2013. Accessing packaged food and beverages in hospital. Exploring experiences of patients and staff. Appetite 60, 231-238.

[4] Benko H, Wilson AD, Balakrishnan R. 2008. Sphere: Multi-Touch Interactions on a Spherical Display. Proceedings of the 21st annual ACM symposium on User interface software and technology (UIST 2008), 77-86.

[5] Bhowmik A, Repaka R, Mulaveesala R, Mishra S. 2015. Suitability of frequency modulated thermal wave imaging for skin cancer detection: A theoretical prediction. Journal of Thermal Biology 51(2015) 6582.

[6] Chelvama YK, Zamina N, Steeleb GS. 2014. A Preliminary Investigation of M3DITRACK3R: A Medicine Dispensing Mobile Robot for Senior Citizens. Procedia Computer Science 42, 240 246.

[7] Chen L, Hoey J, Nugent CD, Cook DJ, Yu Z. 2012. Sensor-Based Activity Recognition. Transactions on Systems, Man, and Cybernetics. Part C: Applications and Reviews, Vol. 42, No. 6.

[8] Cohen PR, Sutton C, Burns B. 2002. Learning Effects of Robot Actions using Temporal Associations. Proceedings of the 2nd International Conference on development and Learning. DOI: 10.1109/DEVLRN.2002.1011807

[9] Correa M, Hermosilla G, Verschae R, Ruiz-del-Solar J. 2012. Human Detection and Identification by Robots Using Thermal and Visual Information in Domestic Environments. J Intell Robot Syst, 66:223243. DOI 10.1007/s10846-011-9612-2

[10] Crick C, Scassellati B. 2008. Inferring Narrative and Intention from Playground Games. Proceedings of the 12th IEEE Conference on Development and Learning.

[11] Dragone M, Saunders J, Dautenhahn K. 2015. On the integration of adaptive and interactive robotic smart spaces. Paladyn, Journal of Behavioral Robotics 6(1):165179

[12] Fitzpatrick P. 2003. First Contact: an Active Vision Approach to Segmentation. Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003). DOI: 10.1109/IROS.2003.1249191

[13] Gold K, Scassellati B. 2007. A Bayesian Robot That Distinguishes Self from Other. 29th Annual Meeting of the Cognitive Science Society.

[14] Gray J, Breazeal C, Berlin M, Brooks A, Lieberman J. 2005. Action parsing and goal inference using self as simulator. RO-MAN 2005: 202-209

[15] Iwai D, Sato K. 2005. Heat sensation in image creation with thermal vision. In Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology (ACE '05), 213216.

[16] Jung MM, Poel M, Poppe R, Heylen DKJ. 2017. Automatic recognition of touch gestures in the corpus of social touch. J Multimodal User Interfaces, 11:8196. DOI 10.1007/s12193-016-0232-9

[17] Koppula HS, Saxena A. 2013. Learning Spatio-Temporal Structure from RGB-D Videos for Human Activity Detection and Anticipation. Proceedings of the 30th International Conference on Machine Learning.

[18] Larson E, Cohn G, Gupta S, Ren X, Harrison B, Fox D, Patel SN. 2011. HeatWave: Thermal Imaging for Surface User Interaction. CHI 2011, Session: Touch 3: Sensing, 2011.

[19] Lee K, Ikeda T, Miyashita T, Ishiguro H, Hagita N. 2011. Separation of Tactile Information from Multiple Sources Based on Spatial ICA and Time Series Clustering. IEEE/SICE Int. Symposium on System Integration (SII), pp. 791-796

[20] Li K, Fu Y. 2014. Prediction of Human Activity by Discovering Temporal Sequence Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, No. 8.

[21] Murthy JN, van Jaarsveld J, Fei J, Pavlidis I, Harrykissoon RI, Lucke JF, Faiz S, Castriotta RJ. 2009. Thermal Infrared Imaging: A Novel Method to Monitor Airflow During Polysomnography. SLEEP, Vol. 32, No. 11.

[22] Osterberg L, Blaschke T. 2005. Adherence to Medication. N Engl J Med, 353:487-97.

[23] Pollack M, Engberg S, Thrun S, Brown L, Colbry D, Orosz C, Peintner B, Ramakrishnan S, Dunbar-Jacob J, McCarthy C. 2002. Pearl: A Mobile Robotic Assistant for the Elderly, in AAAI Workshop on Automatino as Caregiver.

[24] Redmon J, Divvala S, Girshick R, Farhadi A. 2016. You only look once: Unified, real-time object detection. In:CVPR.

[25] Ring EFJ, Ammer K. 2012. Infrared thermal imaging in medicine. Physiol. Meas. 33(3), R33.

[26] Rudol P., Doherty P. 2008. Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery. IEEEAC Paper 1274.

[27] Shiga Y, Dengel A, Toyama T, Kise K, Utsumi Y. 2014. Daily activity recognition combining gaze motion and visual features. UbiComp Adjunct: 1103-1111

[28] Shiomi M, Zanlungo F, Hayashi K. et al. 2014. Towards a Socially Acceptable Collision Avoidance for a Mobile Robot Navigating Among Pedestrians Using a Pedestrian Model. Int J of Soc Robotics 6(3): 443-455. doi:10.1007/s12369-014-0238-y

[29] Shirazi AS, Abdelrahman Y, Henze N, Schneegass S, Khalilbeigiy M, Schmidt A. 2014. Exploiting Thermal Reflection for Interactive Systems. Session: Novel Mobile Displays and Devices, CHI 2014.

[30] Silvera-Tawil D, Rye D, Velonaki M. 2015. Artificial skin and tactile sensing for socially interactive robots: A review. Robotics and Autonomous Systems 63 230-43.

[31] Stork JA, Spinello L, Silva J, Arras KO. 2012. Audio-Based Human Activity Recognition Using Non-Markovian Ensemble Voting. IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication.

[32] Stoytchev A. 2005. Behavior-Grounded Representation of Tool Affordances, ICRA.

[33] Takacs B, Hanak D. 2008. A prototype home robot with an ambient facial interface to improve drug compliance. Journal of Telemedicine and Telecare 2008; 14: 393395.

[34] Tiwari P, Warren J, Day K, MacDonald B, Jayawardena C, Kuo IH, Igic A, Datta C. 2011. Feasibility study of a robotic medication assistant for the elderly (2011). Conference: Proceedings of the Twelfth Australasian User Interface Conference - Volume 117.

[35] Wang L, Zhao X, Si Y, Cao L, Liu Y. 2017. Context-Associative Hierarchical Memory Model for Human Activity Recognition and Prediction. IEEE Transactions on Multimedia, Vol. 19, No. 3.

[36] Wang Z. 2006. A field study of the thermal comfort in residential buildings in Harbin. 2005. Building and Environment 41, 10341039. doi:10.1016/j.buildenv.2005.04.020

[37] Wong WK, Chew ZY, Lim HL, Loo CK, Lim WS. 2011. Omnidirectional Thermal Imaging Surveillance System Featuring Trespasser and Faint Detection. International Journal of Image Processing (IJIP), Volume 4, Issue 6. version 32.

6 ACKNOWLEDGEMENT

We thank those who helped us. We received support from the Swedish Knowledge Foundation for the SIDUS AIR project, and the first author is part of the REMIND project on medicine reminder systems.