HeatWave: Thermal Imaging for Surface User Interaction

Eric Larson[1], Gabe Cohn[1], Sidhant Gupta[2], Xiaofeng Ren[3], Beverly Harrison[3], Dieter Fox[2,3], Shwetak N. Patel[1,2]

[1]Electrical Eng., [2]Computer Science & Eng.
UbiComp Lab, DUB Group, Univ. of Washington
Seattle, WA (USA)
{eclarson, gabecohn, sidhant, shwetak}@uw.edu

[3]Intel Labs Seattle
Seattle, WA (USA)
{xiaofeng.ren, beverly.harrison, dieter.fox}
@intel.com

ABSTRACT
We present HeatWave, a system that uses digital thermal imaging cameras to detect, track, and support user interaction on arbitrary surfaces. Thermal sensing has had limited examination in the HCI research community and is generally under-explored outside of law enforcement and energy auditing applications. We examine the role of thermal imaging as a new sensing solution for enhancing user surface interaction. In particular, we demonstrate how thermal imaging in combination with existing computer vision techniques can make segmentation and detection of routine interaction techniques possible in real-time, and can be used to complement or simplify algorithms for traditional RGB and depth cameras. Example interactions include (1) distinguishing hovering above a surface from touch events, (2) shape-based gestures similar to ink strokes, (3) pressure based gestures, and (4) multi-finger gestures. We close by discussing the practicality of thermal sensing for naturalistic user interaction and opportunities for future work.

Author Keywords
Cameras, thermal imaging, gestures, user interfaces, surface interaction, computer vision

ACM Classification Keywords
H.5.2. [**Information interfaces and presentation**]: User interfaces—*input devices and strategies*. I.5.4 [**Pattern recognition**]: Applications—*computer vision*.

General Terms
Algorithms, Design, Experimentation

INTRODUCTION AND MOTIVATION
Human-computer interface design has significant interest in natural interaction–i.e., systems that do not rely upon mediated interaction through devices such as a mouse, keyboard, or stylus. This has in part been reflected by the popularity in touch screens and surface-based systems [1, 5, 12, 22, 29, 34]. In an attempt to avoid instrumentation on users, objects, and surfaces (e.g., using RFID tags, visual glyphs), camera and imaging technologies have gained significant popularity for surface and gesture interaction. This has also largely been due to the decreasing costs, versatility, size and portability of modern cameras. Traditional (RGB) cameras have seen considerable use in the HCI community for detecting hand gestures, touch points, and object recognition [7, 12, 33, 34]. The introduction of depth cameras or pixel-mixed devices (PMDs) provides a mechanism for 3D reconstruction and depth segmentation for user interfaces [2]. However, the use of RGB and depth cameras in HCI is limited by the type of information that can be extracted from a scene, and the speed at which information can be extracted. For instance, inaccuracies or gaps in gesture detection often result if hand motion is too fast (using typical camera frame rates and real-time processing).

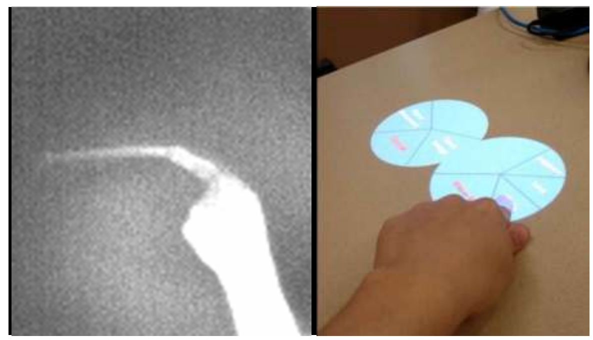

Figure 1: A thermal imaging driven projected marking menu application using the residual heat traces on a tabletop

Thermal imaging, which provides a pixel-level thermograph of anything that is in its field of view (e.g., Figure 1), has largely been under-explored in the user interface community. Recent maturation and advances in solid-state imaging technology and embedded systems have made thermal imaging more practical for consumer use in terms of size, cost and software access to video data.

In this paper, we critically examine the role of thermal imaging as a new sensing solution for enhancing user surface interaction. In particular, we demonstrate how thermal imaging and well known computer vision techniques can make segmenting and detecting certain routine interaction techniques possible in real-time and complement or simplify algorithms for traditional RGB and depth cameras. Example interactions include (1) distinguishing surface touch or target selection from hovering over surface, (2) shape-based gestures similar to ink strokes, (3) pressure based gestures, and (4) multi-finger gestures (enumerated in Figure 3). We demonstrate these in two prototype applications: (1) a pressure-aware drawing application that supports multi-touch and multi-user interactions, (2) a marking menu application using thermal traces as the "ink stroke" menu selection method (Figure 1). Both examples demonstrate the feasibility of real-time thermal traces for UI design.

In the following sections, we briefly discuss the technique of thermal imaging, how it differs from standard IR-based cameras, and the advantages gained by using thermal imaging to complement traditional RGB and depth cameras for surface user interactions. Next, we discuss the related work in surface interaction and thermal imaging, followed by details of our real-time computer vision approaches for extracting meaningful information from our thermal camera, and a collection of interaction techniques they support. Finally, we propose some challenges and future applications of thermal imaging that extend beyond surface interaction.

THERMAL IMAGING

Thermal imaging is a technique for *passively* constructing a high resolution heat map of objects appearing in a scene without using an external illumination source. This is accomplished by measuring the quantity of far-infrared (F-IR or long-wavelength infrared, LW-IR) radiation emitted by any object. Plank's law describes that the wavelength of the peak of electromagnetic radiation from any object is inversely proportional to its absolute temperature. Objects that we interact with on a daily basis are near room temperature and radiate mostly in the F-IR spectrum (Figure 2). Furthermore, the quantity of thermal, or black-body radiation, emitted by an object is directly proportional to the fourth power of its absolute temperature, as given by the Stefan-Boltzmann law [15]. Therefore, by measuring the quantity of radiation emitted in the F-IR spectrum, a thermal sensor can produce a thermographic image of anything in its field of view (e.g., Figure 1).

It is important to note that thermal imaging differs from the more well-known "IR imaging" techniques used in the HCI community. Infrared light detection and night vision devices use what is called *reflected* infrared and operate in the *near*-infrared (N-IR) spectrum. These approaches require an illumination source in order to reconstruct an image. N-IR is employed in some fairly recent interactive tabletop surfaces [11, 34] and depth cameras. Figure 2 shows the visible and infrared spectra and differentiates N-IR from F-IR. Note that we do not utilize N-IR in the present work.

Earlier sensors found in thermal imaging cameras employed a gas filled lens and required refrigeration sources. Advances in semiconductor technology have enabled the development of arrays of silicon-based bandgap detectors and photo-resistive detectors, which allow for 2D imaging planes similar to traditional CCD cameras. Thermal cameras are becoming popular for home energy auditing applications, which has created a demand for portable thermal cameras that continue to reduce in size and cost.

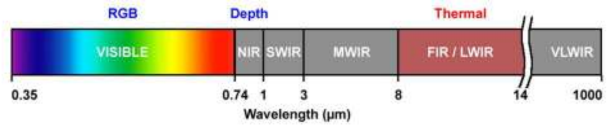

Figure 2: Infrared and visible light spectrum. Thermal imaging operates in the far-infrared (F-IR) band.

ADVANTAGES OF THERMAL IMAGING

Thermal imaging provides several distinct features that address some of the challenges faced by traditional RGB and depth cameras, and enables new applications which are difficult using traditional imaging technologies. Additionally, we believe that thermal imaging can be combined with RGB and depth to provide more robust systems for surface interactions in a variety of natural settings.

First, images produced by thermal cameras are independent of illumination and are far less susceptible to changes in light intensity than traditional RGB cameras or many IR-based depth cameras. RGB cameras do not work well in low-light scenarios, and obviously fail in complete darkness. Since thermal sensing works independently of the visible light spectrum, it works equally well in low- and no-light situations as it would under normal indoor lighting. Furthermore, thermal sensing works in direct sunlight, where some IR depth cameras do not work because their own IR illumination source is washed out by the sun. Additionally, thermal sensing is not confounded by constantly changing light sources and therefore can be used with projected systems without any special considerations.

Second, thermal sensing can detect unique features including short-lived heat transfer from one object to another, which are undetectable with traditional RGB and depth cameras. These features can be used to support a variety of user interaction techniques (Figure 3). For example, using the heat transferred from a user's hand to the surface, multiple touch points can easily be extracted as well as complicated gesture shapes. Moreover, the amount of heat transfer between the finger and the surface can indicate the *pressure* with which the user touched or grasped the object or surface. Lastly, since the transferred heat dissipates over time, a history of a user's interactions is captured in the form of residual heat traces even after the interaction is done.

Third, thermal imaging provides a distinct mechanism for easily segmenting hands (or other body parts) from the background and is independent of scene complexity, colors, and textures (i.e., it easily distinguishes heat-generating sources from inert objects, surfaces, and backgrounds). Traditional RGB sensing and algorithms rely upon computer vision techniques such as background subtraction, color and texture matching, contour detection, and/or optical flow to find target objects of interest. However, many of these features fluctuate due to changes in illumination, pose or position, and color (e.g., a person wears different clothes or color differences between skin tones). These fluctuations often make it difficult to reliably segment a person from a scene, an object from the hand that grasps it, or gestures from a scene or surface. Moreover, "warm object" segmen-

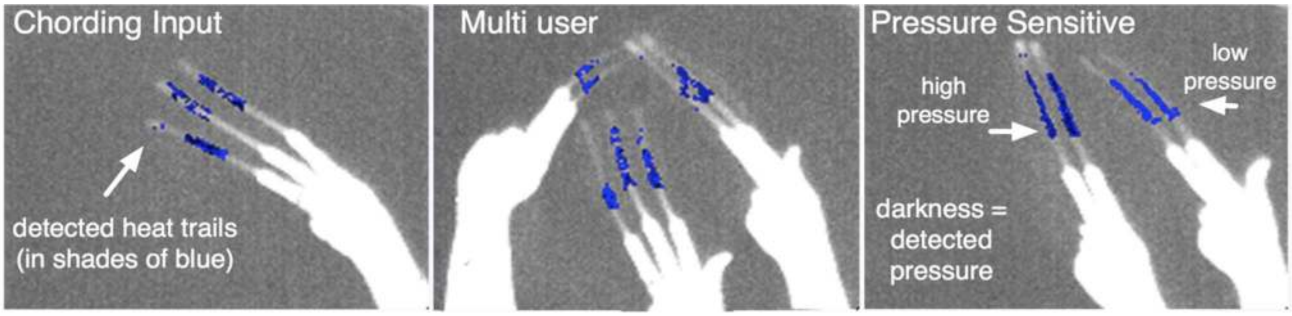

Figure 3: Top down thermal camera views of several surface interactions. Detected heat trails are shown in blue from our real-time algorithm from a single frame of video. Detected pressure differences classified by our real-time algorithm are shown using darkness of blue shades.

tation can be done easily on a real-time, frame by frame basis using well known thresholding methods, such as Otsu's method [24]. Since thermal sensing only needs a single frame for segmentation (i.e., no background image or motion from previous frame), pose and motion issues are less significant than in RGB or depth based systems.

In summary, we believe that thermal imaging provides the following added benefits over traditional imaging:

- Thermal imaging works independently of the light source and is robust to dark or sunlit environments
- Heat traces left behind enable accurate determination of hovering vs. touching without requiring under-mounted cameras or an instrumented surface Heat signatures allow for pressure-aware interaction
- The heat signatures of individual hands allow for multi-touch and multi-user discrimination
- Segmentation of people and body parts is significantly easier and faster than traditional RGB or depth

RELATED WORK

We organize the related work into three broad sections: (i) surface and gesture interaction which uses instrumented surfaces, involving traditional, depth, and thermal cameras, (ii) prior work using traditional IR sensing in the HCI community, and (iii) the use of thermal imaging in other areas of research.

Surface and Gesture Interaction

Camera-based gesture and touch detection in prior work can be roughly categorized by the type of camera sensor used: RGB, depth, and thermal. To enable multi-touch interfaces using RGB cameras [4, 12, 37] there has been substantial work in image segmentation that tracks and identifies various body parts [20]. Typically these systems use skin color matching, edge or contour detection, and motion tracking to segment fingers and hands [12]. More recently, N-IR depth cameras have also been applied alone or in concert with RGB for gesture detection and tracking [2, 8, 11, 22, 34].

Using thermal imaging, some of the initial interaction work comes from Iwai and Sato [9, 10]. They use a behind-the-surface thermal camera and rear projection for drawing upon a translucent paper surface. Users can draw with warm or cool objects in contact with the paper (such as their hands, warm water, or a hairdryer). The application deals primarily with interactive painting on a special paper, but does not investigate the problem of hand segmentation, tracking, or gesture recognition.

The work of Oka *et al.* [23] and Sato *et al.* [31] combines RGB and thermal imaging for hand segmentation and fingertip tracking to drive a gesture recognizer. They use the trajectory of extracted finger tips as input to a hidden Markov model to identify one of six different gestures. The method used does not detect finger contact with the surface, but instead makes use of the fingertip motion (whether touching a surface or in-air). The authors note that their system lags when more than one hand is used, and that their methodology may not scale well beyond tracking a few finger tips. Even so, they are able to use extracted fingertips and in-air gestures to drive an overhead projected user interface. Such a system, combined with the system we propose, could provide a pervasive vocabulary for in-air and touch gesturing.

Beyond tracking fingertips, previous work in RGB and depth imaging has attempted to identify touch pressure (for example, using finger deformation and changes in the cuticle coloring to infer how hard a finger is pressing on a surface [21]). With thermal imaging one can use not only the shape deformation of the finger but also the size and heat of the touch spot left behind to infer the pressure a user exerts on a surface (as we describe later).

Traditional IR Sensing in HCI

Aside from depth cameras, IR sensing (using N-IR) is a popular technique for surface and gesture-based interaction [8, 22, 34]. The advantage of N-IR imaging is the ability to use commodity cameras as the sensor. IR-based sensing typically requires an external illumination source, which dictates its range. A number of projects in the HCI community have used IR for tabletop interaction by detecting hand gestures using an under mounted camera and illumination source [8, 22]. Others have employed a similar approach on vertical semi-transparent surfaces [11].

IR sensing has also been used for fingertip detection and gaze tracking because the retro-reflective properties of these objects allow IR-filtered cameras to easily discern their appearance [32]. Others have used structured IR light patterns for object tracking applications [18]. As we have pointed out, IR imaging differs from thermal imaging; however, many of the computer vision techniques used in IR-based solutions parallel those that would be used in thermal imaging with minimal modification. The added value of thermal imaging is its ability to passively discern objects in view without an illumination source in addition to the hysteretic information it offers.

F-IR Thermal Imaging

Thermal imaging has largely been used for military and surveillance applications [36], where the heat signature produced by the human body is used to track individuals in arbitrary environments and conditions. Thermal imaging has also been used in hospitals and border crossings to identify individuals with a fever within a crowd of people using face temperatures [30].

The miniaturization and decreasing costs of thermal imaging cameras have more recently enabled a number of civilian applications. For example, in energy audits air filtration and insulation problems can be quickly identified by scanning the indoor space. Home inspectors have also extended the use of thermal imaging to look for hot spots near the electrical infrastructure to uncover potential arcing problems or overloaded circuits. Thermal scans of circuit boards can identify potential failure points from heat dissipation problems. The automotive industry has also used thermal imaging for similar applications, especially when temperature analysis needs to be conducted at a safe distance.

Beyond its commercial use, researchers have looked at using thermal imaging for a number of affective computing applications, such as detecting and classifying anxiety based on the minute changes in the thermal signature of one's face [13, 14, 25, 26]. Using well known techniques from the medical literature, changes in anxiety can be correlated to the blood flow on various parts of the face, such as at the forehead and cheeks. Others have extended the use of thermal imaging to infer emotional states exhibited by individuals and have used that information to enhance a user's gaming experience by altering a game's difficulty based on these sensed parameters [38]. Thermal imaging has also been used for illumination-invariant facial recognition [16].

HARDWARE

There are currently a variety of thermal cameras commercially available, and their cost varies based on the required thermal sensitivity and resolution. For instance, at the time of publication, thermal imaging cameras with super-cooled, sealed components that "see through walls" such as those used by law enforcement, cost just under 100,000 USD. HVAC auditing and general purpose thermal cameras are currently around 5,000 USD–which was the price of depth cameras less than one year ago (mass production of depth cameras for home gaming systems has recently decreased their cost significantly). As thermal sensing also gains popularity, the capabilities of these devices will surely increase while the costs decrease.

For our experimentation, we used the RazIR NANO, which contains an un-cooled Focal Plane Array (FPA) micro-bolometer sensor with 160x120 pixel resolution [28]. The thermal sensor is tuned for wavelengths in the IR spectrum between 8 and 14 μm, and captures data at a maximum rate of 30 frames per second. As an artifact of the sensor, the thermal values captured from an object of fixed temperature will drift slightly over time, and therefore a periodic re-calibration is required. We developed software to remove the effects of this drift in real-time, as described later.

Although the RazIR NANO thermal camera has an on-screen user interface and can perform some signal processing on-board, we have done all processing on an external computer using a data feed over USB. The data collected from the feed represents the raw values from the camera's analog-to-digital converter. Our algorithms operate in real-time on the 8 most significant bits of the raw data.

SOFTWARE IMPLEMENTATION

Thermal imaging has many potential applications when used independently or in conjunction with other sensors such as RGB and depth. In this work, we have focused on developing software for one of the more interesting and unique aspect of thermal imaging: the detection and extraction of heat traces. Heat traces are the residual heat left behind on a surface due to the heating of that surface by another warmer object, such as a human hand. Since traditional RGB and depth cameras cannot see signals like heat traces, there has been no other work in developing software to extract such features. This section describes our approach, using Open CV [3] on streaming thermal imaging data and demonstrates how these features can be robustly extracted in real-time using standard computer vision algorithms. Figure 4 shows the step by step processing of the algorithms with callout images for each process.

Noise Filtering

The raw thermal images returned from our camera are fairly noisy from the thermal and scattering noise around the micro-bolometer sensors in the camera. To suppress this noise we apply smoothing in both the spatial and temporal domains. We first apply a 5 pixel by 5 pixel median filter within each video frame. Then, for each pixel, we apply a 5 frame low-pass, finite impulse response (FIR) filter to smooth the signal in time.

Background Calibration

In order to accurately detect heat traces, it is extremely important to model the background signal level–there is slight drift in the hardware sensor over time and surface temperatures may change over time. For these reasons, we compute the mean background image dynamically (a moving average filter) whenever we detect that a human hand is not present in the image.

Hand Segmentation

In order to segment the hand from the image in real-time, we use Otsu's method of thresholding [24]. In this approach, second order statistics of the gray-level histogram

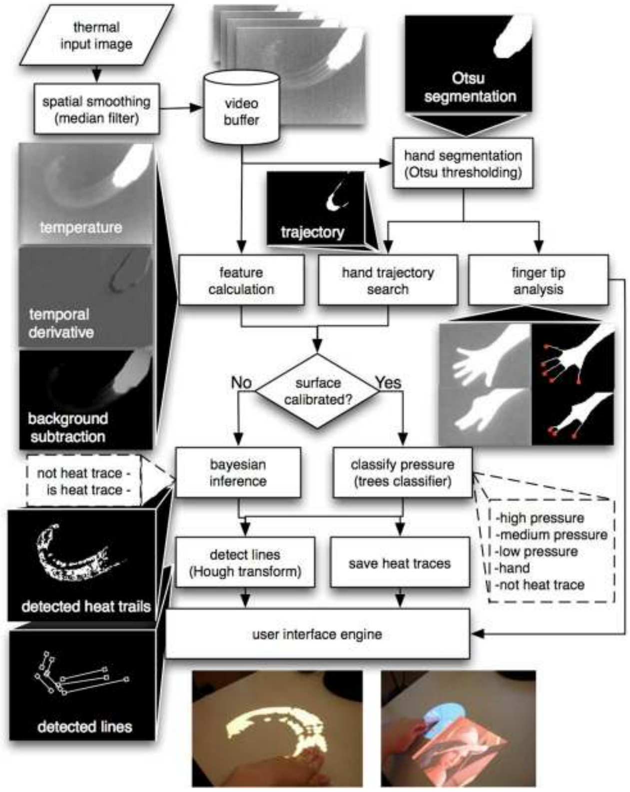

Figure 4: Flow diagram for real-time algorithm processes with callout images at each step. All callouts are captured and rendered from actual data in a real-time application.

are used to maximize the separation of pixel gray levels between two classes. This is ideal in thermal sensing because a hand's temperature is almost always distinct from the background, even when the background temperature is not uniform or has temporarily changed due to touching. This type of thresholding is widely used in computer vision applications and is highly optimized. For our image resolution, the operation takes a fraction of a millisecond. An example of segmentation is shown in Figure 4.

Fingertip Extraction

We use the segmented image of the hand together with well known morphology operations to extract multiple fingertips in a scene. We first apply a thinning operation to the binary segmented image, resulting in a skeletization of the fingers. We then iteratively apply a hit-and-miss transform to the image with a rotating, 3x3 structuring element of endpoints at each possible angle. The result tells us the endpoints of each finger. The simplicity of this extraction technique is made possible by the robust segmentation that thermal sensing provides. Ako *et al.* [23] use a more complicated form of fingertip extraction that also accounts for trajectory. However, we found endpoint extraction to be efficient and robust, especially because we only use the extracted fingertips for refining the search space in which we look for heat traces. This type of extraction is difficult using depth cameras because “thinning” is sensitive to the outline of the segmented object. Depth cameras tend to be noisy around the edges of a hand (where thermal is not) and may require further de-noising before finger extraction is possible. An example of fingertip extraction is shown in Figure 4.

Uncalibrated Heat Trace Detection

A heat trace is created when an object warmer than the background surface heats the surface enough to leave evidence of its presence. Over time, this heat trace will disappear and the surface returns to the background temperature. Figure 4 shows smoothed data in which the background has been subtracted and identifies the region corresponding to the heat decay. When the finger simply hovers over the surface, there is no heat transfer or decay, but when the finger touches the surface the heat decay is very distinctive.

We constrain our search space by "ANDing" together hand segmentations within the past one second of video (30 frames) and subtracting the current hand segmentation. In this way, we only look for heat traces in pixel locations where the hand has recently traveled. This reduces the search space significantly, and thus drastically decreases computational complexity.

We frame the detection of heat traces as a Bayesian estimation problem. In particular, we observe the likelihood of a pixel being part of a heat trace given three features: smoothed temperature, temporal derivative, and background subtracted temperature. The temporally smoothed temperature and derivative are calculated over five frames using FIR filters. This 5-frame buffering results in a system latency of 166 ms (1/6th of a second), which can be considered real-time for most interactive applications.

The likelihood distributions are assumed i.i.d. and assumed to follow a Rayleigh distribution based on empirical observations. That is, each feature is modeled well using a distribution with a single tail and the product of all these distributions is a good model of the overall posterior heat trace probability. Prior distributions are assumed to be uniform. Mathematically this is denoted,

$$P(h_p = 1|\mathbf{x}) \propto \prod_{f \in \boldsymbol{F}} \frac{x_{p,f} - \mu_f}{\sigma_f} e^{-(x_{p,f}-\mu_f)^2/2\sigma_f^2}$$

where h_p denotes whether pixel “p” is or is not a heat trace, $x_{p,f}$ denotes the value of feature “f” at pixel “p.” $\boldsymbol{F}$ is the set of all features. Each feature variance and mean threshold, σ_f and μ_f, are selected empirically using histograms of collected heat traces. When the probability of a pixel being a heat trace, $P(h_p|\boldsymbol{x})$, surpasses a global threshold, we classify the pixel as a heat trace. We found this single model to work well for a variety of surfaces and users (i.e., an out of the box working system).

In addition, we allow the system to be calibrated and adapted to each user or surface. Unlike the Bayesian system, the calibrated system attempts to classify heat traces into more than one class based on the pressure with which the user pressed on the surface (i.e., the amount of heat transferred to the surface).

Calibrated Heat Trace Classification

The calibration process consists of three stages. During each stage the user "draws" a line on the surface at increasing pressures. We save features (temperature, derivative, background subtracted temperature, and background temperature) around the trajectory of the moving fingertip using the Bayesian estimate to extract pixels that are part of a heat trace. In this way, the Bayesian estimate is used to bootstrap the training of a more complicated classifier. This is done for each pressure the user wishes to calibrate. For example, these steps can be repeated for "light", "moderate", and "hard" pressure stroke calibration.

We then train the system to identify the pressure of each heat trace using a C4.5 tree classifier [27], as implemented by Weka [35] (training is less than one second, on average). We found a number of algorithms implemented in Weka to perform well on a test set of 20,000 detected heat trace pixels. We chose the tree classifier because it provided 96% accuracy on a 10-fold cross validation of the test set and runs quickly enough to assess pressure in real-time. The test set had five classes: High, medium, and low pressure, hand and background.

In addition to pressure, this method can be used to calibrate to different surfaces. To test this hypothesis, we calibrated to four different materials and asked a user to trace a fixed projected "S" eight times on each of the four materials (32 strokes). After the experiment, we superimposed the detected traces on top of each other for each material, resulting in a set of images where brightness denotes how often a heat trace was detected on the material surface. Drawing speed, temperature of the finger, hand and surface were not controlled (similar to what one would expect in real use). The superimposed image results are shown in Figure 5. Brightness denotes classification accuracy: bright yellow is 100% correct and completely black (false negatives) is 0% detection. There were no false positives detected. Notice that paper (best), table top laminate (second best), and wooden surfaces have easily detectable traces, and that plastic is the most difficult surface to detect traces upon. Based on this initial evaluation, we hypothesize that less heat is transferred to the plastic than other materials.

Line Detection

Although heat trace detection can be used to detect arbitrary shapes and gesture patterns, it is useful to detect when heat traces are collinear. We focus on lines because detected lines can be used for chording style gestures and as input to many applications such as marking menus. To detect line gestures we buffer detected heat traces into a single image for the past one second. We then apply a binary Hough transform [6] to the buffered image to reveal heat traces that are collinear. The buffered image provides a binary history of where a user has placed ink strokes with their hands and can be used with other transformations to fit arbitrary curves and circles, not just lines.

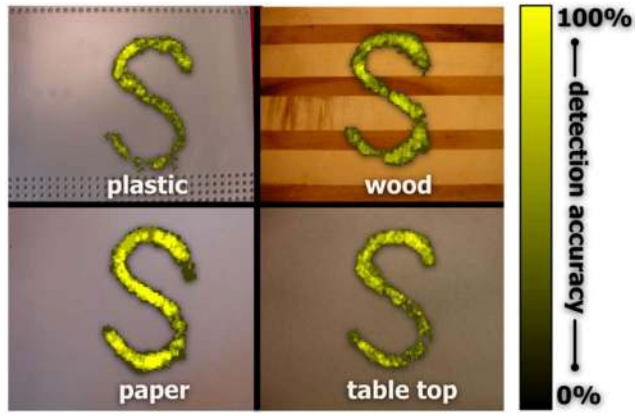

Figure 5: Heat traces (overlaid images) from a user tracing the letter S eight times on each of four materials. Each trial's detected heat trace is projected and overlaid on the previous trial. Brightness denotes accuracy.

PROTOTYPES AND INTERACTION TECHNIQUES

To test the usefulness of thermal features as driving input for user interfaces, we built two prototype applications. Each application uses an overhead projection system and an overhead-mounted thermal camera to transform an arbitrary surface into a multi-touch user interface (see Figure 6).

The system ports easily to different tables and other flat surfaces. Conversion from camera coordinates and projected coordinates is achieved using a four point calibrated homographic transform (i.e., a known mapping of four points in each space). No other instrumentation is necessary and the camera and projector can be placed at many different orientations to the surface and each other.

The first application is a multi-user and multi-touch drawing application that displays arbitrary gestures made by the users, and alters the brightness of displayed colors based on the pressure with which each user draws using three pressure levels. The second application uses line gestures for image manipulation. Images are chosen using marking menus [17], then once the images are displayed they can be translated, rotated, and scaled using thermal lines. The two applications are designed to demonstrate that thermal traces can be used as a plausible substitute for multi-touch screens and can drive typical user interfaces in real-time with naturalistic interactions. Images from interactions with each application can be seen in Figure 6.

User Interface Engine

For the drawing application no additional feature processing is necessary—it uses the unaltered heat trace locations and their corresponding pressure as the sole driving inputs. The touch positions are projected onto the surface, with their brightness representing applied pressure.

The marking menu and image editing application uses the extracted fingertip locations and heat traces (as detected by the Hough transform) as driving input. The steadiness of the fingertip is used to detect finger-down events. When a detected fingertip has been stationary for 500 ms, a marking menu is displayed at the fingertip location. After this, a

Figure 6: Example thermal camera setup with overhead projector running two prototype applications.

series of lines can be drawn by the user. The angle of the detected lines controls which sub-menu is displayed and the subsequent selections. As with any marking menu, the user can draw a line and pause to bring up another menu or, alternatively, can draw multiple lines in a single motion to navigate several menus at once. Once the user selects an image to open, they can move and scale the image using a combination of one, two, and three finger gestures. The interactions are: a single line drawn on the image translates the image; two fingers moving inward or outward from the image scales it; and three fingers in a sweeping motion dismisses the image from the surface.

We found these inputs and extraction methods to be easily implemented and the applications were straightforward additions to the system. The prototype applications provide proof of concept for systems that detect and use arbitrary gestures—using nothing more than a binary Hough transforms on detected heat traces and simple fingertip tracking.

DISCUSSION AND ANALYSIS

Although thermal imaging provides a number of advantages over traditional RGB and depth imaging solutions, there are clearly some differences and challenges. Many of the challenges with thermal imaging, such as reflection and occlusions have analogous problems in RGB and depth imaging, but there are also challenges that are unique to thermal imaging, such as surface temperature and material type. Additionally, applications driven by heat traces must be designed to minimize the effects of these limitations, but still provide an intuituve interface.

Robustness of Algorithm

In addition to the Bayesian estimation and the calibrated classifier described in this paper, we experimented with several additional methods for extracting the heat traces from a surface. These methods included temperature thresholding, change in temperature thresholding, decay template matching, hidden Markov models (HMMs), and non-probabilistic finite state machines. All of these algorithms could be tuned to work quite well in specific situations, but none of them were able to work well over a wide range of scenarios (with the exception of the HMM, which worked extremely well but was computationally too intensive to run in real-time without optmization or parallelization). The Bayesian approach described in this paper appears to be highly robust, and works well for all of the scenarios that we have tested. In addition, the approach can be used to as a bootstrap for more complicated classifiers like the C4.5 trees implemented in this paper.

Robustness of Features

During our examination of thermal imaging, we determined that the residual heat traces after touching a surface provide significant value for differentiating between hovering and touching, which is clearly a challenge for top mounted RGB cameras and can be problematic with depth cameras (e.g., noisy data due to light, reflections, or depth sensor resolution). However, two factors impact the decay rate and hence ease of detection for heat traces: material type and dwell time. Decay rate varies greatly depending upon the surface material (from a few hundred milliseconds up to five or six seconds) but can be addressed using the calibration sequence presented. Wooden and drywall surfaces exhibit the slowest decay rates due to their thermal properties and, thus, are the easiest materials to classify heat traces upon. On the other hand, metal surfaces exhibit the fastest thermal dissipation, with traces typically disappearing after only a few frames. This confounds our algorithm in many scenarios and we still consider metal surface heat trace extraction to be an open problem.

Additionally, the dwell time (i.e., how long surface contact lasts) impacts the amount of heat transfer and the size of the heated region (heat spreads). For many gesture based interactions, dwell time is not significant; however, one can easily imagine gestural vocabularies or situations where this would need to be taken into account. Our interaction techniques thus far have been fairly independent of dwell time - that is, most gestures only require the user to press the surface quickly. However, extracting reliable pressure estimates becomes more difficult with gestures that have a large variance in dwell time.

Challenges Unique to Thermal Imaging

Unlike RGB and depth cameras, the computer vision algorithms for thermal imaging must account for residual heat traces that may linger for a significant amount of time. Therefore, continually computing/updating the background model is important to avoid false positives. Similarly, it is possible for the surface to heat up due to the interaction and thus make it harder to segment out the hand if the surface temperature nears that of the hand, but we found this to be extremely rare in our experimentation where the surface was at about 20° C. Note that maintaining a background model is only important for detecting heat traces. Hand segmentation and tracking remains robust in a variety of environments and backgrounds.

Challenges with all Computer Vision Approaches

Similar to RGB and depth, occlusion is an issue for thermal imaging. Since our system relies on the heat traces left by finger contact, these traces are only visible when the finger moves away from the contact points (i.e., heat traces underneath the hand are undetectable). An angle mounted camera helps alleviate problems with occlusions when the hand covers the heat traces. Furthermore, there is an unavoidable delay in the detection of lift (mouse-up) events, and it is not possible to detect touch-down (mouse-down) events. Note that this is distinct from the algorithm processing lag of 166 ms.

An interesting solution to extracting touch-down and touch-up events with an overhead system is to use a metallic surface. Metallic surfaces tend to reflect F-IR waves (i.e., you can see thermal reflections on smooth, polished surfaces but the reflections are invisible to the user). These reflections could be used to indicate hover distance or indicate when a surface is touched (i.e., when the reflection and object meet, similar to the Wilson "shadow touching finger" approach [34]). There is a tradeoff, of course, between the reliability of extracting heat traces from polished surfaces and touch-down extraction because the reflection may be directly over the heat trace. This tradeoff is in contrast to depth cameras, which are known to have significant problems with shiny or reflective surfaces.

Limitations in Using Heat Trace Input

In our ongoing tests of multi-touch thermal drawing interfaces, it is rare that parts of the hand are not segmented correctly or heat traces are not detected. One exception occurs when someone with very cold hands uses the system. For instance, someone who was holding a cold drink moments before interacting. When this occurs, the finger tips are about the same temperature as the table top and segmentation severs a portion of the finger tip. Because we use segmentation to constrain the search space for heat traces, some heat traces are missed. Also, after long periods of being in contact with the table top (about 5-10 minutes of continual interaction) the finger tips begin to cool down and the table top begins to heat up, potentially confounding heat trace identification. The heating of the surface is magnified when many users interact. We found that letting the surface cool for about 10-20 seconds is sufficient to reset the background model and to let the fingertips return to natural temperature (alternatively users can also rub their hands together). This suggests for applications that require sustained and continuous finger strokes from the user (such as drawing and gaming applications) that thermal imaging alone may not be appropriate as an input method. Lastly, we found occlusion was not a factor for drawing applications because users almost always pull their hands away from the surface after marking a line or curve to see the visual representation, revealing the heat trace.

Our marking menu application was ideally suited for thermal line input and we were able to drive menus with selections as narrow as 15^o fairly easily. Moreover, translation and scaling of images is easily interpreted from the detected lines in near real-time. The longest delay comes from line detection because our algorithm requires buffering of detected heat traces. The buffering process adds an additional 200 ms delay onto the delay in heat trace detection, resulting in an average delay of about 400 ms between drawing a line and reaction by the interface. This delay was insignificant for our example application but could be limiting for applications that require faster driving inputs, such as gaming. For image manipulation, one limitation is when the user tries to move an image towards the camera because the hand moves directly over the heat trace path. One outcome of this is that interface feedback may appear "jerky" and delayed since we can only process segments of the heat trace that are unoccluded. We have anecdotally noticed that drawing with an index finger alone tends to naturally offset the hand orientation and thus the heat trace is less often occluded. More occlusion seems to occur with multi-finger chording-style manipulations. More investigation is required but this does suggest that gesture design choices can be optimized to reduce potential problems while preserving reasonably intuitive interaction.

Challenges not Present for Thermal Imaging

In general, varying lighting conditions did not pose any problems in extracting gestures. For instance, we observed similar system algorithm behavior across indoor, outdoor, and dark environments. This is encouraging since thermal imaging may be a viable option for use in arbitrary environments. Even more encouraging is that images projected on a surface with a digital projector also do not interfere with hand/finger segmentation. This is because the IR emitted from the projector's light source is in the N-IR band, and does not extend to the F-IR spectrum.

ADDITIONAL APPLICATIONS AND FUTURE WORK

This paper describes our initial exploration of thermal imaging as a means for detecting gestural input on surfaces to support interaction. We have additionally started investigating a wider range of possible applications, some of which are briefly described below. At present, we have collected data to illustrate the viability of each of these ideas.

Distinguishing Multiple Users

We have found that multiple users have different thermal gradients on their hands (Figure 7a). Although these gradients can change over long time intervals (e.g., coming inside from a winter's day), we believe the overall hand "heat signature" may remain unchanged for the duration of a session. Using these thermal differences among the hands of multiple users, we believe that we can uniquely identify several users within an interactive session. This would allow customized multi-user interaction on arbitrary surfaces without the need for instrumenting the surface or the users. In addition, this algorithm would not use the angle of approach, enabling users to move freely around the surface.

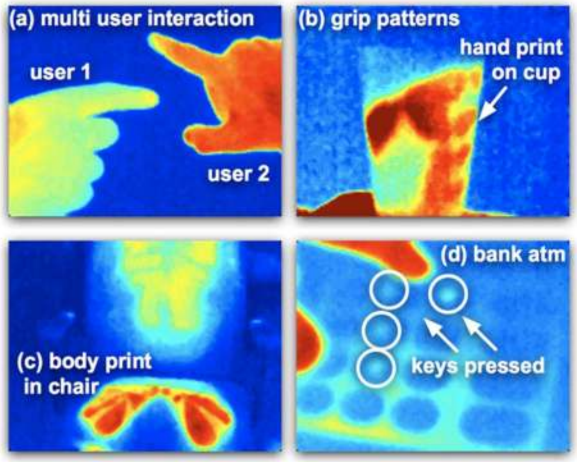

Figure 7: Additional applications of thermal imaging. Thermal data variations are mapped to different colors.

Grip Patterns

Colleagues in our lab are conducting research in personal robotics where a roving robot with a robotic arm can pick up, move or manipulate objects. A key element of this work is to understand how the robot can best grip everyday objects. Thermal imaging could help train robots about grip methods by identifying the exact positions of contact between an object and a human hand. The human hand leaves heat traces on objects, and these traces clearly show all of the points of contact between the object and the hand (Figure 7b). We plan to apply the thermal image grip patterns as a means of training robots how to grasp arbitrary objects.

Object Interaction

Similarly to the grip patterns discussed above, thermal imaging can be used to determine the points of contact between objects and any part of the human body. Figure 7c shows the heat pattern left in a chair as a result of a person sitting. This information can be used to determine elements of posture and provide feedback whenever a user stands up. Additionally, points of contact could be used to determine which chairs, household areas, and objects are used most within a space or even when these are used (e.g., eldercare or medical rehab applications).

In-Air Gestures

Since thermal imaging makes human bodies and body parts easily distinguishable from scene elements, we can easily detect and track airborne gestures and movements. We plan to continue this work, developing real-time algorithms to robustly segment and track people and airborne gestures on non-planar surfaces, arbitrary objects, and in scenes where the user stands in front of the camera.

Determining Surface Material

Our current implementation is designed to work on surfaces that are made from a variety of materials. We believe that thermal imaging can be used to identify the surface material based solely on its thermal properties. For example, each material will dissipate heat at a different rate, and the spatial spread of the initial heating point will also be different. We imagine a system in which the user would simply touch the surface (which would likely be part of a training or calibration procedure) and the system would detect and measure the spatial and temporal properties of the surface. Using a database of such parameters, the surface material could potentially be determined. This would provide mobile surface interaction systems with additional context regarding which surface the user is interacting with.

Security

The concept of heat trace detection also has interesting security ramifications. For instance, thermal heat trace detection can be used to view a password or bank PIN that a user types on a keyboard or keypad. A residual heat trace is left behind on the keys even *after* the password has been entered. Figure 7d shows the heat traces on the keys of an actual bank ATM.

CONCLUSION

We have described how thermal imaging technology can complement or augment more traditional RGB or depth cameras for surface gesture interaction. Thermal imaging is more robust under circumstances where RGB or depth would fail and thus could provide more robust solutions for the variability that occurs in natural settings. We have demonstrated that well-known computer vision techniques can provide good models for extracting the heat traces that human interaction with surfaces (and objects) leaves behind in real-time. We also demonstrated this approach on a variety of different surfaces and offered a technique for surface calibration. We have demonstrated several traditional user interface techniques can be driven in real-time based on thermal heat trace input. Finally, we have outlined a number of interesting new opportunities beyond gesture-based surface interactions where thermal imaging provides unique data to enable new applications.

ACKNOWLEDGEMENTS

We would like to acknowledge Ryder Ziola for his advice and help in the extension of this work to include an overhead projected system.

REFERENCES

1. Arai, T., Machii, K. and Kuzunuki, S. Retrieving electronic documents with real-world objects on InteractiveDESK. In *Proc. of UIST '95*. New York: ACM Press, 1995, pp. 37-38.
2. Barras, C. Microsoft's body-sensing, button-busting controller. *New Scientist: Technology*. 7 Jan 2010.
3. Bradski, G.; Kaehler, A. (2008), Learning OpenCV: Computer Vision with the OpenCV Library
4. Cohen, C.J., Beach, G.J. and Foulk, G.A. Basic hand gesture control system for PC applications. In *Proc. of AIPR '01*. IEEE Computer Society, 2001, pp. 74.
5. Dietz, P. and Leigh, D. DiamondTouch: A multi-user touch technology. In *Proc. of UIST '01*. New York: ACM Press, 2001, pp. 219-226.
6. Duda, R. and Hart, P "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Comm. ACM*, Vol. 15, pp. 11–15, Jan., 1972

7. Felzenszwalb, P., McAllester, D., and Ramanan, D. A Discriminatively Trained, Multiscale, Deformable Part Model. In *Proc. of IEEE CVPR '08*, 2008.
8. Hilliges, O., Izadi, S., Wilson, A.D., Hodges, S., Garcia-Mendoza, A., Butz, A. Interactions in the air: adding further depth to interactive tabletops. In *Proc. of UIST '09*. New York: ACM Press, 2009, pp. 139-148.
9. Iwai, D. and Sato, K. "Heat Sensation in Image Creation with Thermal Vision", *ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (ACE2005), pp.213-216, Jun. 2005.
10. Iwai, D. and Sato K, "Limpid Desk: theoretical papers of transparent projection of Mixed Reality", *IPSJ*, vol.48, no.3, pp.1294-1306, 2007.
11. Izadi, S., Hodges, S. Taylor, S., Rosenfeld, D., Villar, N., Butler, A., Westhues, J. Going beyond the display: a surface technology with an electronically switchable diffuser. In *Proc. of UIST '09*. New York: ACM Press, 2009, pp. 269-278.
12. Kane, S., Avrahami, D., Wobbrock, J.O., Harrison, B., Rea, A.D., Philipose, M., LaMarca, A. Bonfire: a nomadic systems for hybrid laptop-tabletop interaction. In *Proc. of UIST '09*. New York: ACM Press, 2009.
13. Khan, M.M., Ingleby, M., and Ward, R.D. Automated Facial Expression Classification and affect interpretation using infrared measurement of facial skin temperature variations. *ACM Trans. on Autonomous and Adaptive Systems,* 2006.
14. Khan, M.M., Ward, R.D., and Ingleby, M. Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature. *ACM Trans. on Applied Perception,* 2009.
15. Kittel, C. and Kroemer, H. *Thermal Physics*. 2nd ed. New York: W.H. Freeman, 1980.
16. Kong, S.G., Heo, J., Boughorbel, F., Zheng, Y., Abidi, B.R., Koschan, A., Yi, M., and Abidi, M.A. Multiscale Fusion of Visible and Thermal IR Images for Illumination-Invariant Face Recognition. *International Journal of Computer Vision,* 2007.
17. Kurtenbach, G. The Design and Evaluation of Marking Menus, Ph.D. Thesis, Department of Computer Science, University of Toronto. May 1993.
18. Lee, J., Hudson, S., Summet, J., and Dietz, P. Moveable interactive projected displays using projector based tracking. In *Proc. of UIST '05*. New York: ACM Press, 2005, pp. 63-72.
19. Letessier, J. and Bérard, F. Visual tracking of bare fingers for interactive surfaces. In *Proc. of UIST '04*. New York: ACM Press, 2004, pp. 119-122.
20. Manresa, C., Varona, J., Mas, R. and Perales, F. Hand tracking and gesture recognition for humancomputer interaction. *Electronic Letters on Computer Vision and Image Analysis*, 5 (3), 2005, pp. 96-104.
21. Marshall, J., Pridmore, T., Pound, M., Benford, S., Koleva, B. Pressing the Flesh: Sincing Multiple Touch and Finger Pressure on Arbitrary Surfaces. In *Proc. of Pervasive '08*. Berlin: Springer-Verlag, 2008, pp. 38-55.
22. Microsoft Surface. http://www.microsoft.com/surface
23. Oka, K., Sato, Y. and Koike, H. "Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems." *IEEE Computer Graphics and Applications*. Vol. 22, No. 6, pp. 64-71, 2002.
24. Otsu, N, "A threshold selection method from gray-level histograms". *IEEE Trans. Systems, Man, and Cybernetics.* 9: 62–66, 1972
25. Pavlidis, I., Dowdall, J., Sun, N., Puri, C., Fei, J., and Garbey, M. Interacting with human physiology. *Computer Vision and Image Understanding,* 2007.
26. Puri, C., Olson, L., Pavlidis, I., Levine, J., and Starren, J. StressCam: non-contact measurement of users' emotional states through thermal imaging. In *Proc. of CHI '05,* 2005.
27. Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
28. RazIR. NANO Thermal Camera, http://raz-ir.com/raz-ir-nano-thermal-infrared-camera.html.
29. Rekimoto, J. SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In *Proc.of CHI '02.* New York: ACM Press, 2002, pp. 113-120.
30. Saletan, W. Heat Check: Swine flu, body heat, and airport scanners. http://www.slate.com/id/2217148/, 28 April 2009.
31. Sato, Y., Kobayashi, Y., and Koike, H. "Fast tracking of hands and fingertips in infrared images for augmented desk interface," *Proc. 2000 IEEE International Conference on Automatic Face and Gesture Recognition* (FGR 2000), pp. 462-467, March 2000.
32. Tanriverdi, V. and Jacob R.J.K. Interacting with eye movements in virtual environments. In *Proc. of CHI '00*. New York: ACM Press, 2000, pp. 265-272.
33. Wellner, P. Interacting with paper on the DigitalDesk. *Communications of the ACM, 36* (7), 1993, pp. 87-96.
34. Wilson, A. D. PlayAnywhere: a compact interactive tabletop projection-vision system. In *Proc. of UIST '05.* New York: ACM Press, 2005, pp. 83-92.
35. Witten, I. and Frank, E.: "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco (2005).
36. Wong, W.K., Tan, P.N., Loo, C.K., and Lim, W.S. An Effective Surveillance System Using Thermal Camera. In *Proc. of International Conference on Signal Acquisition and Processing,* 2009.
37. Wu, M. and Balakrishnan, R. Multi-finger and whole hand gestural interaction techniques for multiuser tabletop displays. In *Proc. of UIST '03*. New York: ACM Press, 2003, pp. 193-202.
38. Yun, C., Shastri, D., Pavlidis, I., and Deng, Z. O' game, can you feel my frustration?: improving user's gaming experience via stresscam. In *Proc. of CHI '09,* 2009.