User-Independent Detection of Swipe Pressure using a Thermal Camera for Natural Surface Interaction

Tim Dunn, Sean Banerjee, Natasha Kholgade Banerjee
Clarkson University, Potsdam, NY
{dunntj, sbanerje, nbanerje} @clarkson.edu

***Abstract*—In this paper, we use a thermal camera to distinguish hard and soft swipes performed by a user interacting with a natural surface by detecting differences in the thermal signature of the surface due to heat transferred by the user. Unlike prior work, our approach provides swipe pressure classifiers that are user-agnostic, i.e., that recognize the swipe pressure of a novel user not present in the training set, enabling our work to be ported into natural user interfaces without user-specific calibration. Our approach generates average classification accuracy of 76% using random forest classifiers trained on a test set of 9 subjects interacting with paper and wood, with 8 hard and 8 soft test swipes per user. We compare results of the user-agnostic classification to user-aware classification with classifiers trained by including training samples from the user. We obtain average user-aware classification accuracy of 82% by adding up to 8 hard and 8 soft training swipes for each test user. Our approach enables seamless adaptation of generic pressure classification systems based on thermal data to the specific behavior of users interacting with natural user interfaces.**

***Index Terms*—thermal, swipe pressure, hard, soft, natural user interface, natural surface interaction**

I. Introduction

Natural user interfaces involving users interacting with content projected on real-world surfaces and objects are gaining significant interest due to their impact in ubiquitizing augmented reality (AR) applications. While traditional touch devices consist of sensor-equipped touch surfaces, the key idea behind natural surface interaction is to avoid instrumentation of the surfaces interacted with, as surface-intrusive instrumentation is expensive, has the potential to mar the aesthetics of the surface, and may not integrate seamlessly with the surface materials and construction. As a result, approaches in natural user interaction use non-intrusive cameras in the environment to recognize user interactions with projected content. While several approaches already exist to use RGB or depth cameras to perform swipe detection and tracking [1], [2], [3], there is currently a growing interest in using thermal cameras for detecting user gestures on surfaces [4], [5], [6], [7], [8], [9], as the change in thermal signature of the surface by heat transfer from the fingers and hands of a user introduces a strong contrast in a thermal image which can be used to extract the gesture pattern.

In this work, we use thermal cameras to detect differences in pressure applied by the fingers on a surface for pressure-based interactions in natural user interfaces. Using pressure-based interactions, actions with the same motions such as swipe-based screen change and drag-and-drop can be distinguished based on differences in pressure. A swipe-based screen change may be performed using a soft swipe, while a drag-and-drop may be performed using a hard swipe, thereby reducing the latency induced by actions such as selecting the item to be dragged, and double or long tapping. With increasing pressure from the fingers of a user, the amount of heat transferred to the surface from the finger rises, enabling pressure changes to be identified by detecting differences in thermal intensity. While approaches exist to provide pressure-based interactions in multi-touch surfaces [6], [7], these approaches require the user to pre-provide hard, medium, and soft samples to calibrate the pressure-sensitivity of the interface to that user by training a user-specific pressure classifier. Given the vast diversity in the thermal behavior of natural surfaces due to differences in ambient temperature and heat transfer rate, performing a user-specific calibration for every surface interacted with can prove infeasible for natural user interaction.

We investigate the ability of classification algorithms trained on hard and soft swipes from a group of subjects to be user-agnostic, i.e., to recognize the swipe pressure of a novel user not present in the training set, for materials such as wood and paper. The end goal of our work is to enable seamless natural surface interaction without prior user-specific calibration. Our dataset consists of 9 users with 16 hard and 16 soft swipes per user. We use leave-one-out cross-validation to train random forest classifiers using all 16 hard and 16 soft swipes of 8 training users during each training fold, and we run the classifier on a test set of 8 hard and 8 soft swipes of the left out user. Our work provides an average classification accuracy of 76% for random forests trained using swipes on paper and wood together. When the two materials are treated separately, our approach provides nearly the same accuracy for paper, i.e., 77%, and an accuracy of 88% for wood. We also perform partial swipe classification and reach 77% classification accuracy using random forests trained on both materials within 40% of the swipe enabling early detection of pressure for low latency in interaction.

As a user interacts with the natural user interface, it is desirable to personalize the interface to the user's behavior over time. We perform comparisons of the user-agnostic classification discussed above to user-aware classification using random forests trained by adding increasing numbers of samples from a training set of 8 hard and 8 soft swipes for the test user. The training set and test set for the test user are mutually exclusive. We reach an average accuracy of 82% for both paper and

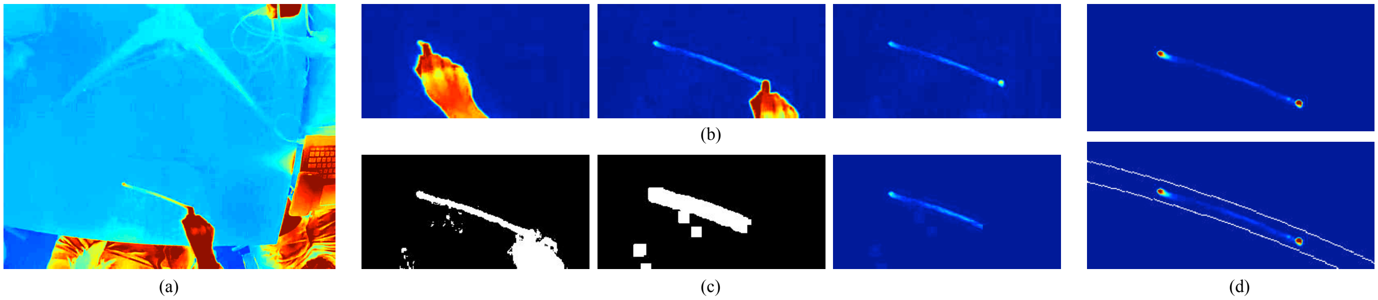

Fig. 1. (a) Thermal intensity image of a user performing a swipe recorded by an overhead FLIR Vue Pro 640 thermal camera. (b) Sample frames from a video of a swipe, showing the start of the swipe (left), end of the swipe (center), and swipe after the user removes the hand (right). (c) Thresholded finger swipe image (left), thresholded image with hand removed using morphological operations (center), and masked finger swipe (right). (d) Accumulative intensity image (top) and quadratic band fit to accumulative image (bottom).

wood, 87% for paper alone and 91% for wood alone using 8 hard and 8 soft training swipes. Our approach demonstrates that we can reach comparable accuracies of 81% for paper and wood, 84% for paper, and 89% for wood by adding 8 hard swipes alone, enabling a user-targeting of agnostic classifiers by solely probing hard swipes. By adding increasing amounts of data, our approach can be used in natural user interfaces to perform adaptation of a generically trained system to the behavior of the specific user without interfering with the interactions of the user.

II. Related Work

There exists a growing body of work on using thermal cameras to perform swipe tracking in natural surface interactions. These methods do not face the challenge of tracking fingers overlaid by projected content that approaches using color cameras (e.g., [1], [2], [3]) need to address, as they rely on the strong heat signature of humans in comparison with the ambient surrounding. Oka et al. [4] use an overhead thermal camera to detect the hand, arm, and fingertips of a user as high intensity regions in the thermal image, to track the fingertip trajectories using Kalman filtering, and to recognize gestures. Iwai and Sato [5] use a thermal camera installed behind a thin paper-based interaction surface to detect swipes using the heat transferred by the user to the paper. Kurz [8] uses a thermal camera attached to a tablet to detect interactions from users on natural surfaces for augmented reality. Palovuori and Rakkolainen [9] use a low-cost thermal imager to track fingertips near the imager for immaterial interaction. These approaches do not perform detection of swipe pressure.

While the approach of Larson et al. [6], also used by Saba et al. [7], performs swipe pressure classification, it trains a decision tree based pressure detector on the user prior to the user's interactions. Their approach requires re-calibration of the pressure detection to every user who uses their system. In contrast, our classification approach recognizes the swipe pressure of novel users not present in the training set.

III. Capture Setup and Data Collection

We use an FLIR Vue Pro 640×512 30fps thermal camera with a 9mm lens mounted overhead on a tripod and facing downwards toward the surface of interest to capture the user swipes. We use two surfaces—a sheet of white paper and a plywood board. We record each user swipe as a video using the MJPEG format and a linear white-hot temperature encoding. For our experiments, we recruited 9 test subjects, 7 male and 2 female, with ages ranging between 18 and 37 years, and with prior experience in using touch based devices. All 9 test subjects used their right hand as the dominant hand. During the data collection phase, we asked each user to perform 16 soft (i.e. low pressure) and 16 hard (i.e. high pressure) swipes on the surface of interest. We asked users to wait 5 seconds between each soft swipe and 10 seconds between each hard swipe to allow any remnant thermal signature to dissipate from the surface. To replicate real world natural user interfaces, we asked users to use any finger and perform swipes of any length or direction. Given the user data, we perform manual spatial cropping of the recorded video to retain the region around the swipe, and we automatically segment the user data into 32 independent samples, each containing a single swipe. With 32 samples per user for 9 users on 2 materials, we obtain a total of 576 videos. Figure 1(a) shows the thermal image captured by the FLIR Vue Pro camera encoded using the jet color map, and Figure 1(b) shows three frames from a user swipe cropped around the region of the swipe.

IV. Swipe Path Estimation

A. Swipe Extraction by Hand Removal

We perform static background subtraction to remove motionless objects and we threshold the thermal image as shown in the left image of Figure 1(c). We use a threshold of 4 on a scale of 0 to 255. We then remove the hand as shown in Figure 1(c) using morphological image processing as the hand represents the hottest object in the scene and overshadows the thermal intensity of the swipes. First, we use a 3×3 square structural element to apply opening and closing operations to the image for salt-and-pepper noise removal. Then, we dilate the image using an 11×11 square structural element to fill in the gaps between the fingers of an open hand so that long narrow structures only correspond to swipes as opposed to fingers. We remove the hand by first computing an image

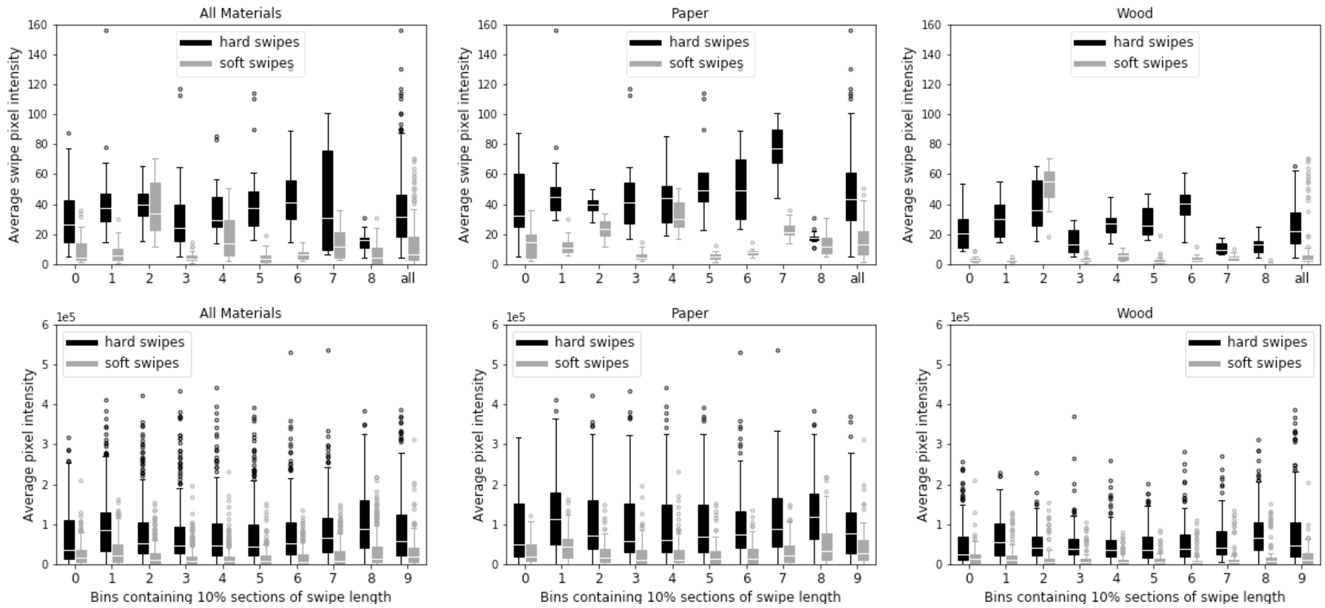

Fig. 2. Box plots for each user of average pixel intensity over a swipe, created using 16 hard and 16 soft swipes per user for all materials, paper only, and wood only.

that eliminates narrow lines corresponding to swipes using a 25×1 and a 1×25 structural element to retain the hand, and then subtracting the hand only image from the prior image. The difference image contains the swipe devoid of the hand. We apply a final opening operation with a 7×7 structural element to remove noise occurring near the edge of the user's hand, and we consider all remaining white pixels to be the region of the image containing the swipe. In case background noise contaminates the resulting mask, we perform manual adjustment of the binary threshold. We find that 48 of the 576 videos require a manual adjustment of the threshold from 4 to 8. We process each video frame independently and we apply the resulting binary images as masks over the corresponding original video frames to generate masked swipes. An example masked swipe shown in the right image of Figure 1(c).

B. Swipe Path Extraction using Curve Fitting

We determine the path of the swipe by fitting a curve to an accumulative intensity image shown at the top of Figure 1(d). The accumulative image is obtained by summing pixel intensities in the masked video over time. Since the majority of the user swipes were simple arcs, we approximate each swipe by a translated and rotated quadratic function. For each finger swipe, we first estimate the line of best fit to the pixel intensities. We then transform the coordinate system to orient the best fit line along the positive x axis. Next, we estimate the best fitting quadratic to the transformed pixels. The region of interest describing the swipe is defined by pixels which when transformed lie within 10 pixels of the quadratic fit. We show an example polynomial swipe fit at the bottom of Figure 1(d). Only pixels within the quadratic band and within the masked region are retained as part of the swipe.

V. Swipe Classification using Random Forests

We train a 500-tree random forest classifier [10] using 30 features, with each split using 6 features. We use the default maximum voting scheme when determining the predicted class label of the hard or soft swipe. We generate features from the pixel values in the masked video and the accumulative image within the region of interest. Using the masked video, we generate a 10-bin histogram of average pixel intensities as features, with each bin representing $1/2$ second intervals of time over the video length. Using the accumulative image, we generate a 10-bin histogram of average pixel intensities and a 10-bin histogram of maximum pixel intensities as features, with each bin comprising 10% of the total swipe length over the quadratic.

To perform the user-agnostic classification, we use leave-one-out cross-validation to train random forest classifiers using all 16 hard and 16 soft swipes of 8 training users during each training fold, and we run the classifier on a test set of 8 hard and 8 soft swipes of the left out user. For the user-aware classification, we add in increasing levels training swipes from each test user. We use three methods to add in the user-specific training data: (a) we add only hard training swipes from 1 hard swipe to 8 hard swipes, (b) we add only soft training swipes from 1 soft swipe to 8 soft swipes, and (c) we add both hard and soft training swipes from 1 hard and 1 soft to 8 hard and 8 soft swipes. All training swipes in user-aware classification are distinct from the swipes in the test set for the user. We train random forests for each material, as well as random forests that combine both materials.

All Materials (Predicted Class)

Actual Class	Hard	Soft
Hard	0.82	0.18
Soft	0.30	0.70

Paper Only (Predicted Class)

Actual Class	Hard	Soft
Hard	0.71	0.29
Soft	0.18	0.82

Wood Only (Predicted Class)

Actual Class	Hard	Soft
Hard	0.93	0.07
Soft	0.16	0.84

Fig. 3. Overall confusion matrix for user-agnostic classification using random forests trained and tested on all materials, paper only, and wood only.

VI. Results

A. User Data Analysis

The top row of Figure 2 shows box plots for the average intensity over all pixels in a swipe over the 16 hard and 16 swipes obtained for each of the 9 users in our experiments. As expected, on average, soft swipes have lower average intensities than hard swipes. The graphs depict a lower variance in intensities of soft swipes overall in comparison to the hard swipes. One reason for the high variance in intensity for hard swipes may be that users are not accustomed to hard swipes from everyday interactions. The graphs also depict some dependence of separation on user, however, for several users a constant separation boundary may be drawn enabling user-agnostic classifier training, especially for individual materials.

For most users, there is a separation between the hard and soft swipes, enabling user-aware re-training of user-agnostic classifiers. Separation based on average intensities is weak for user 2 interacting with wood and user 4 interacting with paper, which also influences the spread in the box plots for both the materials. The variance in intensities for user 7 may be explained by the differences in heat transfer and retention of paper versus wood, combined by the differences in hard swipe interaction by user 7.

The average pixel intensity trends for each user from the top row of Figure 2 are replicated in the 10-bin histograms from Section V used in random forest classification. The bottom row of Figure 2 shows overall mean pixel intensities averaged over all 9 users accumulated into 10 bins over the length of the the swipes. On average, the graphs show a difference between hard and soft swipes even in the early stages of the swipe, enabling swipe pressure detection in the beginning of the swipe for low latency in response to user interaction.

B. Results of Classification

Figure 3 provides confusion matrices averaged across the 9 users for classifying hard and soft swipes using random forests trained on all materials, paper only, and wood only. Our approach shows high classification accuracy for wood due to the ability of wood to retain heat transferred by the subject for longer periods of time. When trained by mixing all materials, the classifier is biased toward improving the accuracy of hard swipes. Table I gives accuracy and precision of user-agnostic classification for all materials, paper only, and wood only.

TABLE I
Overall accuracy, hard swipe precision, and soft swipe precision for user-agnostic random forest classifiers trained using all materials, paper only, and wood only.

	Accuracy	Hard Precision	Soft Precision
All Materials	0.76	0.73	0.80
Paper Only	0.77	0.80	0.74
Wood Only	0.88	0.85	0.92

TABLE II
Overall accuracy, hard swipe precision, and soft swipe precision for user-aware classification by adding increasing levels of user-specific swipe data.

		Accuracy	Hard Swipe Precision	Soft Swipe Precision
Add 1 Hard	All Materials	0.77	0.74	0.82
	Paper Only	0.79	0.82	0.77
	Wood Only	0.88	0.85	0.92
Add 1 Soft	All Materials	0.76	0.73	0.80
	Paper Only	0.78	0.84	0.75
	Wood Only	0.88	0.85	0.91
Add 1 Hard, 1 Soft	All Materials	0.77	0.74	0.82
	Paper Only	0.81	0.84	0.78
	Wood Only	0.89	0.85	0.92
Add 4 Hard	All Materials	0.80	0.75	0.86
	Paper Only	0.83	0.82	0.85
	Wood Only	0.89	0.85	0.94
Add 4 Soft	All Materials	0.77	0.75	0.79
	Paper Only	0.79	0.86	0.75
	Wood Only	0.87	0.86	0.88
Add 4 Hard, 4 Soft	All Materials	0.80	0.76	0.85
	Paper Only	0.85	0.86	0.85
	Wood Only	0.89	0.86	0.93
Add 8 Hard	All Materials	0.81	0.75	0.89
	Paper Only	0.84	0.82	0.88
	Wood Only	0.89	0.85	0.94
Add 8 Soft	All Materials	0.78	0.77	0.79
	Paper Only	0.80	0.86	0.75
	Wood Only	0.88	0.89	0.87
Add 8 Hard, 8 Soft	All Materials	0.82	0.78	0.88
	Paper Only	0.87	0.87	0.87
	Wood Only	0.91	0.88	0.93

Table II provides accuracy and precision for adding increasing numbers of user-specific training swipes for user-aware classification. Each table shows the result of adding n hard only, n soft only, and n hard and n soft swipes, where n is 1, 4, or 8. The overall average accuracy of classification reaches 82% for all materials on adding 8 hard and 8 soft swipes. Figure 4 shows the change in accuracy in going from user-agnostic classification to increasing levels of user-awareness in the classification. In the top row, we show classification accuracies for adding hard swipes only. We notice that the addition of hard swipes improves the classification for most users, with the maximum change occurring for wood with users 6 and 8. We notice that the classification improvement introduced by soft swipes is minimal. These trends may be attributed to the fact that most users experience soft swiping in everyday interactions with mobile devices, however, hard swiping is rare, and may be tuned to user-specific behavior.

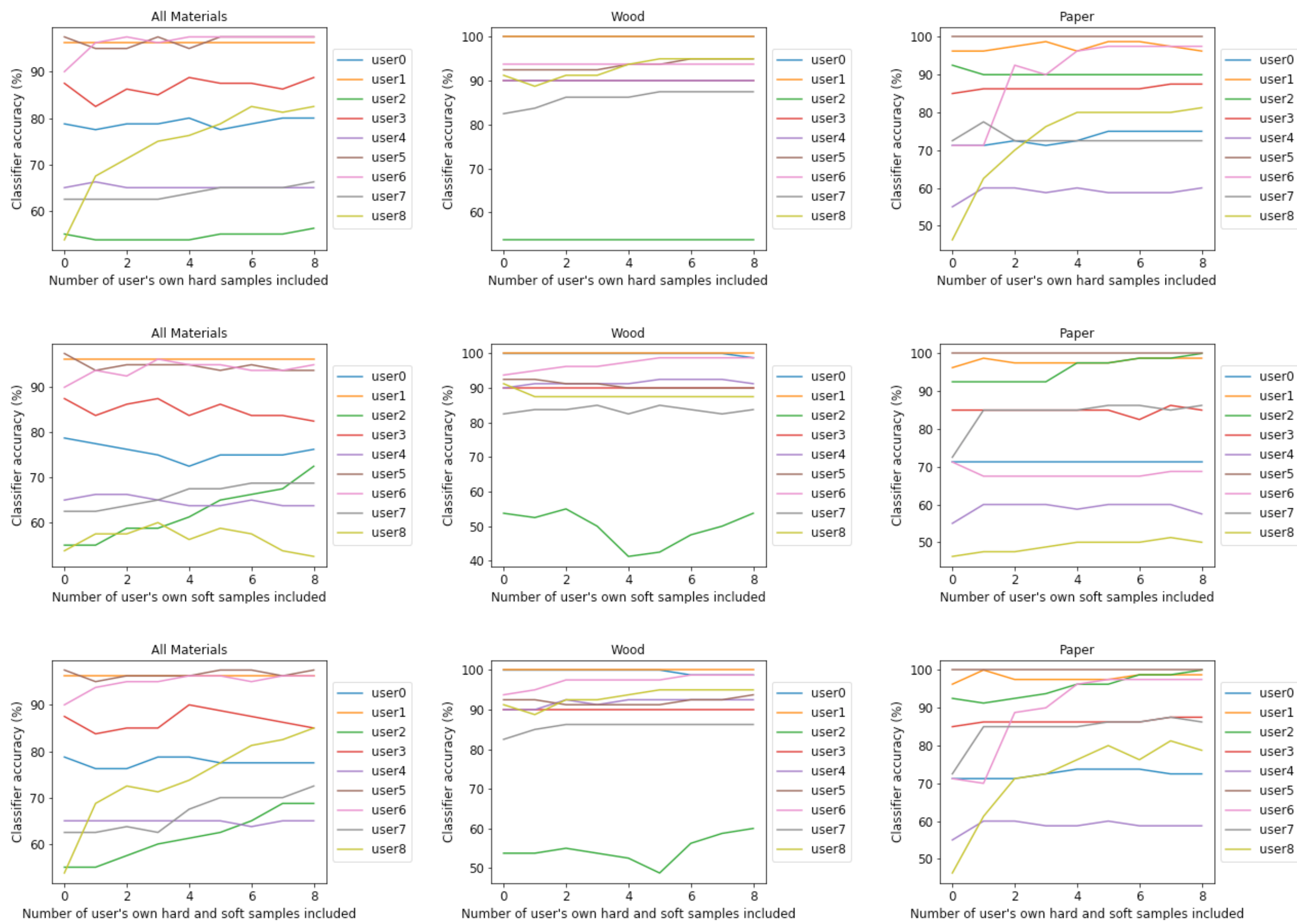

Fig. 4. Change in user-aware classification with addition of increasing numbers of hard swipes only from the user (top), soft swipes only from the user (middle), and hard and soft swipes from the user (bottom) for all materials (left), wood (center), and paper (right). The horizontal axis for the bottom row represents the number of hard and number of soft swipes used, e.g., 4 indicates that 4 hard and 4 swipes are used, i.e., a total of 8 swipes. The accuracy at 0 on the horizontal axis represents user-agnostic classification.

The trends of adding hard only and soft only swipes are reflected in the addition of both hard and soft swipes, lending further support for user-aware training with hard swipes.

C. *Results of Partial Swipe Classification*

To provide low-latency response to users as they perform swipes, we also provide results of user-aware and user-agnostic classification of swipe pressure over partial swipe data. We train random forest classifiers using 10%, 20%, 40% and 70% of the swipe data by using the first 1, 2, 4, and 7 bins of the 10-bin video and accumulative image histograms. Accuracy and precision for the two types of classification are shown in Table III. The user-aware classification corresponds to adding all 8 hard and 8 soft training swipes for the test user. As shown by the table, by 40% of the swipe length, the classification results approach the results at the full length of the swipe, indicating that swipe pressure classification can be performed early in the swipe path.

VII. Discussion

We provide user-agnostic and user-aware approaches to distinguish hard and soft swipe pressures for users interacting with natural surfaces using data recorded by a thermal camera. Our analysis provides an average accuracy of 76% using random forest classifiers trained on paper and wood collectively solely using intensity data over the video. Our results on performing user-aware classification indicate with sparse sets of swipes, a higher improvement in classification can be obtained by adding hard swipes only.

In this work, we choose not to use temporal data such as the amount of time involved in performing a swipe, as the time taken is often task dependent, e.g., a soft swipe may be performed slower for scrolling down a page as opposed to for sifting through pages. However, it may be hypothesized that due to the effort involved in performing a hard swipe, the time taken to perform hard swipes is longer for the same distance than the time taken to perform a soft swipes. Furthermore, the effort involved may also impact the shape of the swipes, causing hard swipes to be straighter than soft swipes. As part

TABLE III
OVERALL ACCURACY, HARD SWIPES PRECISION, AND SOFT SWIPES PRECISION FOR PARTIAL SWIPE DATA CLASSIFICATION.

		Accuracy	Hard Swipe Precision	Soft Swipe Precision
User-Agnostic, 10%	All Materials	0.60	0.60	0.59
	Paper Only	0.59	0.60	0.57
	Wood Only	0.64	0.63	0.65
User-Agnostic, 20%	All Materials	0.73	0.72	0.73
	Paper Only	0.65	0.66	0.64
	Wood Only	0.79	0.76	0.83
User-Agnostic, 40%	All Materials	0.77	0.74	0.80
	Paper Only	0.73	0.74	0.71
	Wood Only	0.87	0.84	0.91
User-Agnostic, 70%	All Materials	0.79	0.76	0.84
	Paper Only	0.74	0.75	0.73
	Wood Only	0.88	0.85	0.91
User-Aware, 10%	All Materials	0.65	0.66	0.64
	Paper Only	0.65	0.68	0.62
	Wood Only	0.73	0.72	0.74
User-Aware, 20%	All Materials	0.77	0.78	0.77
	Paper Only	0.75	0.76	0.74
	Wood Only	0.82	0.81	0.84
User-Aware, 40%	All Materials	0.81	0.77	0.86
	Paper Only	0.84	0.84	0.85
	Wood Only	0.88	0.86	0.91
User-Aware, 70%	All Materials	0.83	0.79	0.89
	Paper Only	0.83	0.82	0.84
	Wood Only	0.89	0.86	0.92

of future work, we will perform task-driven analysis of the duration and shapes of hard and soft swipes, as potential features for classification especially in earlier portions of the swipe for faster response.

Our approach indicates that due to potentially limited experience of users with hard swiping, user-aware training may show improvements for hard swipe classification. Using our work, an approach to perform online re-training of user-agnostic swipe pressure classifiers to be user-aware may be to probe the user at intervals about the desire of their swipe based on factors such as the length of time to perform a swipe. The response of the user may be used to re-train the classifier if the intended swipe was hard. It may be desirable in everyday applications to limit the use of hard swipes for less frequent interactions such as drag-and-drop, which may provide the scope to restrict user probing to maintain non-invasive user experience. In future work, we will investigate techniques to perform seamless probing of the desired swipe using the human-in-the-loop to perform user-aware re-training of user-agnostic classifiers. Future work will also include investigating the effect of changes in ambient temperature, expanding the dataset of materials tested, and investigating the effect of the heat capacity of various materials on swipe pressure detection.

VIII. ACKNOWLEDGEMENTS

This work was partially supported by the National Science Foundation (NSF) grant #1730183.

REFERENCES

[1] Pierre Wellner, "Interacting with paper on the digitaldesk," *Communications of the ACM*, vol. 36, no. 7, pp. 87–96, 1993.

[2] Andrew D. Wilson, "Playanywhere: A compact interactive tabletop projection-vision system," in *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*, 2005, pp. 83–92.

[3] Pranav Mistry and Pattie Maes, "Sixthsense: a wearable gestural interface," in *ACM SIGGRAPH ASIA 2009 Sketches*. ACM, 2009, p. 11.

[4] Kenji Oka, Yoichi Sato, and Hideki Koike, "Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. IEEE, 2002, pp. 429–434.

[5] Daisuke Iwai and Kosuke Sato, "Heat sensation in image creation with thermal vision," in *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. ACM, 2005, pp. 213–216.

[6] Eric Larson, Gabe Cohn, Sidhant Gupta, Xiaofeng Ren, Beverly Harrison, Dieter Fox, and Shwetak Patel, "Heatwave: thermal imaging for surface user interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 2565–2574.

[7] Elliot N Saba, Eric C Larson, and Shwetak N Patel, "Dante vision: In-air and touch gesture sensing for natural surface interaction with combined depth and thermal cameras," in *Emerging Signal Processing Applications (ESPA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 167–170.

[8] Daniel Kurz, "Thermal touch: Thermography-enabled everywhere touch interfaces for mobile augmented reality applications," in *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 9–16.

[9] Karri Palovuori and Ismo Rakkolainen, "The heat is on: thermal input for immaterial interaction," in *Proceedings of the 19th International Academic Mindtrek Conference*. ACM, 2015, pp. 152–154.

[10] Leo Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.