Carpus: A Non-Intrusive User Identification Technique for Interactive Surfaces

Raf Ramakers **Davy Vanacken** **Kris Luyten** **Karin Coninx** **Johannes Schöning**
Hasselt University - tUL - IBBT
Expertise Centre for Digital Media
Wetenschapspark 2, 3590 Diepenbeek, Belgium
firstname.lastname@uhasselt.be

ABSTRACT
Interactive surfaces have great potential for co-located collaboration because of their ability to track multiple inputs simultaneously. However, the multi-user experience on these devices could be enriched significantly if touch points could be associated with a particular user. Existing approaches to user identification are intrusive, require users to stay in a fixed position, or suffer from poor accuracy. We present a non-intrusive, high-accuracy technique for mapping touches to their corresponding user in a collaborative environment. By mounting a high-resolution camera above the interactive surface, we are able to identify touches reliably without any extra instrumentation, and users are able to move around the surface at will. Our technique, which leverages the back of users' hands as identifiers, supports walk-up-and-use situations in which multiple people interact on a shared surface.

Author Keywords
Interactive tabletops; surface computing; multi-touch interaction; multi-user applications; user identification.

ACM Classification Keywords
H.5.2 [Information interfaces and presentation]: User Interfaces: Input devices and strategies; I.4.9 [Image Processing and Computer Vision]: Applications

General Terms
Design, Experimentation, Human Factors.

INTRODUCTION

Interactive tabletops and surfaces are well suited to co-located collaboration because of their ability to track multiple inputs simultaneously. However, the multi-user experience on these devices can be enriched significantly if touch points can be associated with the user performing an action. New opportunities are, for example, customizing functionalities (e.g. multi-user undo [29]) or personalizing the appearance of the inter-

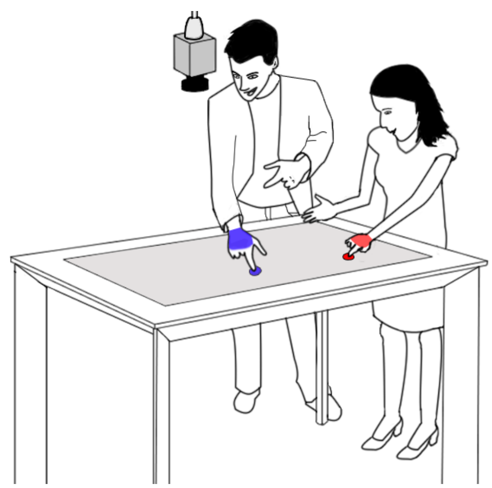


Figure 1. Carpus recognizes users by observing the dorsal region of their hands with a high-resolution camera mounted above an interactive surface.

face, such as providing automatic translations or help cues for novice users [31]. When the user's identity is readily available with every touch, it is also possible for the system to enforce social protocols in multi-user applications [19, 26].

Although several approaches exist for identifying users of a touch display, each has important limitations. Some techniques require additional instrumentation of the users at the start of each session [15, 16, 28] or the use of a mobile device [32], while others are easy to spoof because they identify the shoe that the user is currently wearing [27]. DiamondTouch [7] and Medusa [2] assign an identity to fixed positions around an interactive table, making them unsuitable for free-flow environments. In contrast to these approaches, Schmidt et al. [30] presented a system that uses a biometric technique. However, very specific hand postures are required each time the user's identity is needed, which impedes the 'naturalness' of the interaction.

In this paper, we present a novel technique to associate touch points with a user with high accuracy. Our system, Carpus, supports walk-up-and-use scenarios in which each touch can be identified transparently once users have been registered. By mounting a high-resolution camera above an interactive surface (Figure 1), we are able to extract identifying information from the back of the human hand (also known as the *dorsal* region) during traditional multi-touch interactions. As a result, our technique is non-intrusive and can be used in combination with a large range of touch technologies, including sensor-based, optical-based and vision-based hardware. Carpus supports user identification for setups intended for ad-hoc or informal collaboration [10], with users collaborating in unplanned fashion.

Among the many possible applications of Carpus, we envision new interactive setups deployed in shops and travel agencies. The system can differentiate between actions of customers searching together for information about new products, a task that frequently involves collaboration [17]. In such a situation, it is possible for the system to interpret two-handed touch gestures unambiguously while multiple users are interacting simultaneously. In addition, a profile can be created for each customer, making it possible for the system to track a user's interests in order to recommend other relevant products. We also expect that Carpus can satisfy the demand for easy to deploy techniques to identify users in groupware studies, such as studies regarding territoriality [33]. With Carpus, analysis of users' behavior will be significantly less time consuming, as each action is automatically associated with a particular user.

Our main contribution is a new technique to *identify touch points non-intrusively by observing the back of the users' hands*. Our approach is robust for hand postures that are common during the practical use of multi-touch systems. For traditional pointing, which is most common in touch interaction [29], Carpus uniquely identifies both hands of a user with more than 97% accuracy, even when 20 users are registered. As a result, it becomes possible to differentiate between actions performed by a user's right and left hand in order to support non-symmetric division of labor [9]. For more complex hand postures, which are less common, the recognition rates drop to 82% in the worst case scenario. A smaller group size is recommended to achieve higher recognition rates in these situations.

RELATED WORK

Identifying users is a challenging and prominent issue in surface computing research. In this section, we will review the previous efforts in more depth. Existing approaches to user identification have at least one of the following limitations: (1) The naturalness of the interaction is impeded due to users being required to wear additional instrumentation or because of limitations on the supported hand postures; (2) Users need to stay at a fixed position around the surface; (3) Associating an identity to a touch point can be difficult, resulting in accuracy problems. A summary of existing approaches can be found in Table 1.

Visual tags [15] can be easily used to identify users. Electronic tags such as the IdWristband [16] or the IR-Ring [28] are also able to identify users reliably, in this case by flashing a unique sequence in the infrared spectrum to the interactive surface. However, these techniques require users to wear equipment. This overhead can be a burden, especially when the system is used for short, spontaneous and unplanned interactions. Schöning et al. [32], on the other hand, require the use of a mobile device to authenticate with the system, while

	Non-intrusive	Position invariant	Unambiguous association
DiamondTouch [7]	✓	–	✓
Medusa [2]	✓	–	–
IdWristbands [16]	–	✓	–
IR-Ring [28]	–	✓	–
Tagged gloves [15]	–	✓	✓
HandsDown [30]	–	✓	✓
Mobile Phone [32]	–	✓	–
Bootstrapper [27]	✓	✓	–
Carpus	✓	✓	✓

Table 1. A summary of the strengths and weaknesses of related user identification techniques for interactive surfaces.

Schmidt et al. [30] extract biometric geometry features from the human hand. In order to measure these metrics reliably, however, a hand needs to be placed flat on the interactive surface. These approaches require either additional equipment or unusual hand postures, interrupting the 'naturalness' of multi-touch interaction.

Other researchers have presented non-intrusive user identification by associating touches to users' positions around an interactive tabletop. DiamondTouch [7] leverages a specialized capacitive surface to transmit an electrical charge through a user's body when touching it. A receiver unit inside each user's chair is used to register all touch events of the user that is currently sitting on that chair. If users swap chairs, their identity will change. Annett et al. [2] instrumented a tabletop with arrays of proximity sensors to relate each touch to a user's position. Although this technique can track users' bodies that are moving near the tabletop, the system cannot preserve identities when users move further away. These identification techniques can be practical in situations where the users' positions around the table are fixed, but they cannot be used in free-flow environments where the interaction is interrupted frequently by other activities and the users' positions around the table are likely to vary.

A few projects have demonstrated the potential of revealing identities by observing shoes or gait. Bootstrapper [27] extracts features from the top of a user's shoes, whereas Multitoe [3] observes shoe sole patterns. Orr et al. [23] took another approach, identifying users by analyzing the forces and timings while walking on custom-built floor plates. However, these extracted features are not unique, as users can wear the same shoes or walk with an abnormal gait. Furthermore, relating touches to these identified regions requires additional tracking, and is therefore challenging and error-prone.

CARPUS

Our approach uses a high-resolution overhead camera (Figure 1) and works as follows: (1) Carpus continuously captures frames from above the interactive surface and extracts the visible dorsal hand regions; (2) Only when a user performs a "touch down" event, unique features are extracted from the hand region visible at the location of the touch in the last captured frame; (3) These unique features are matched against a database of feature-user pairs that was constructed beforehand during a very short training session; (4) The resulting user identity associated with that touch becomes available to

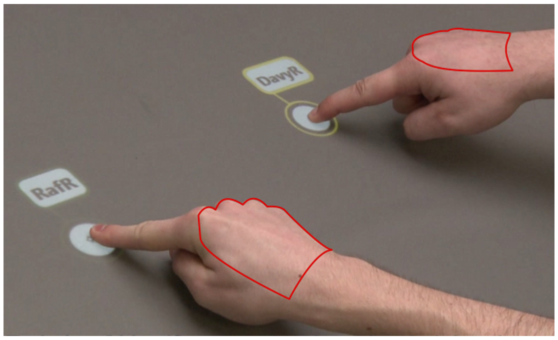


Figure 2. Unique features are extracted from the dorsal hand region. Fingers are excluded from the region.

the application running on the interactive surface. The surface then tracks the touch point to automatically identify subsequent finger movements.

The dorsal hand region (Figure 2) is well suited to user identification in a collaborative environment for several reasons. To begin with, each person's dorsal hand regions have many strongly identifying characteristics, such as lines, grooves, folds and furrows [20]. Furthermore, the dorsal hand region is quite large and is also the part of the hand that is the least flexible [20], making it possible to capture consistent patterns in the skin over many frames. The dorsal hand region is visible to an overhead camera throughout the majority of hand postures encountered during interaction with a tabletop, something that is not true of fingers (e.g. when performing vertical touches [36]).

BENEFITS AND LIMITATIONS

Carpus overcomes some drawbacks of current identification approaches (Table 1). Our technique provides transparent identification of users by extracting unique features directly from the back of the hand. Once registered with the system, Carpus supports real walk-up-and-use scenarios. The camera can be mounted above any type of interactive surface and captures the entire interaction. As a result, each touch point can be unambiguously related to an identified skin region. This enables identification of touches even when arms of multiple users overlap. Furthermore, Carpus is able to differentiate between the two hands of a user, which enables non-symmetric devision of labor [9].

Carpus is, however, subject to a few limitations. First, the dorsal hand region needs to be clearly captured by the camera. When performing very fast movements, it is possible that the camera cannot capture a sharp image of the hand. This can be addressed by using a camera with an appropriate shutter speed. Furthermore, while Carpus can handle most common hand postures encountered during multi-touch interactions, some techniques [18, 11] require the user's hand to be in such a position that the back is not visible to an overhead camera. In order to get a clear image of the dorsal region in those circumstances, we can add multiple cameras to our setup, capturing the hand from different points of view.

Secondly, Carpus has not been designed to offer authentication for privacy or security reasons. The current version of our algorithm can differentiate between users working simultaneously on an interactive surface, but has difficulties eliminating users that are not registered. As a result, the automatic detection of a 'new' unregistered user is unreliable, and thus the users' hands need to be registered beforehand to achieve higher recognition rates. However, registration is only a one-time cost compared to techniques that require configuration at the start of each session [16, 28]. We are exploring several interactive widgets for registration purposes that can be easily integrated in applications. Registration can be as simple as performing a single gesture before first usage (e.g. similar to the interactions required to unlock a smartphone).

Lastly, the reported recognition rates are only representative for groups of people with Caucasian skin. We intentionally recruited people with similar skin color for the evaluation of Carpus, because the similarity makes it harder to find unique features. Further research is needed to investigate the effect of different skin types on the accuracy of Carpus.

SKIN REGION AND IDENTITY EXTRACTION

In order to identify a hand that is visible in a captured frame, Carpus performs the following four steps: (1) extraction of the dorsal hand region; (2) feature extraction; (3) feature matching; (4) relating touches to identified regions.

Step 1: Extraction of the Dorsal Hand Region

First, the dorsal hand region needs to be detected in a captured frame. When a touch event occurs, Carpus does this in three sub-steps: (A) extraction of the skin region; (B) detection of visible fingers; (C) detection of the position of the wrist. Two iterations of this algorithm are executed for each detected skin region. During the second run, a more accurate segmentation is used in order to detect the dorsal region of the hand more precisely in particular situations, as we explain below.

Skin Detection

Skin segmentation [25] involves finding ranges of intensity values for which most skin pixels fall in a given color space. The image is first converted to the YCrCb color space [24] to obtain a decision rule that is robust under varying illumination conditions. To support various types of human skin, all pixels in a relatively large intensity range are classified as skin:

$$Y{>}20, 85{<}Cb{<}135, 135{<}Cr{<}180 \tag{1}$$

Figure 3-A1 shows areas of pixels that were selected using this static segmentation rule. Because of the large intensity range, shadowed regions between fingers are erroneously classified as skin pixels. Therefore, a second, more dynamic skin segmentation rule is used after the position of the wrist and the fingers are detected. This dynamic segmentation rule is based on the average (μ) and standard deviation (σ) of the intensity of the pixels in the detected dorsal region:

$$\begin{aligned} \mu_Y - 2\sigma_Y &< Y < \mu_Y + 2\sigma_Y \\ \mu_{Cr} - 2\sigma_{Cr} &< Cr < \mu_{Cr} + 2\sigma_{Cr} \\ \mu_{Cb} - 2\sigma_{Cb} &< Cb < \mu_{Cb} + 2\sigma_{Cb} \end{aligned} \tag{2}$$

In Figure 3-A2 the shadowed regions between the fingers are correctly excluded with dynamic segmentation. Parts of an arm or finger are sometimes erroneously classified as non-skin regions because of variations in the observed skin color. However, this does not influence the outcome of our algorithm, since only the back of the hand is used.

Finger Detection

Once the skin regions are extracted from a captured frame, fingers need to be detected in order to exclude these regions from the final contour. We first detect the tips and then the phalanges of the fingers. To detect fingertips, we use a curvature-based approach similar to Malik and Laszlo [14] and Segen and Kumar [34]. The vectors from each contour point k to $k+n$ and $k-n$ are computed (n is a fixed value and depends on the distance between two contour points). If the angle between the two vectors is below some threshold, then the contour point is marked as a fingertip (Figure 3-B). We use a relatively high angular threshold value of 60 degrees to make the fingertip detection algorithm more 'greedy'. Our algorithm even classifies the knuckles of folded fingers as fingertips (Figure 3-B), because we want to exclude all visible finger regions from our contour and are not interested in the exact position of the real fingertips. The misclassification of finger valleys is avoided by taking into account the sign of the angle between two vectors while processing the contour in counterclockwise direction. Finally, non-maximal suppression is used to avoid detecting fingertips too close to each other.

In order to extract entire finger regions, the phalanges of the fingers also need to be detected. Boreki and Zimmer [5] do this by finding the position of finger valleys between two consecutive fingertips. However, we need to support all hand postures, even when only one finger is visible (e.g. pointing). Therefore, we take a different approach. Our algorithm starts roaming the skin contour from each detected fingertip to the left and right. Each contour point is then classified as part of a finger phalange until the length of the finger has reached a maximum value or the angle between two phalanges exceeds an angular threshold value (Figure 3-B). This threshold is inversely proportional to the length of the finger and ranges between 40 and 60 degrees in our algorithm.

Wrist Detection

Before we are able to detect the position of the wrist, we need to determine its orientation. The orientation of the wrist is based on the orientation of the arm, which is given as the principal axis of inertia of the extracted skin region. The orientation of the principal axis (Figure 3-C) can be computed from the image moments up to the second order, as described by Freeman et al. [8]. When the aspect ratio of the skin region along the direction of the principal axis is almost equal to 1:1 (e.g. a user with long sleeves performing a vertical touch), this approach is less reliable. In that case, we use the principal axis of the entire arm region, obtained by doing an additional background subtraction step.

Once the orientation of the wrist is known, the precise position of the wrist can be determined. Kioke et al. [12] assume that this position is always located at a fixed distance from the top of the hand. However, we observed that the distance from

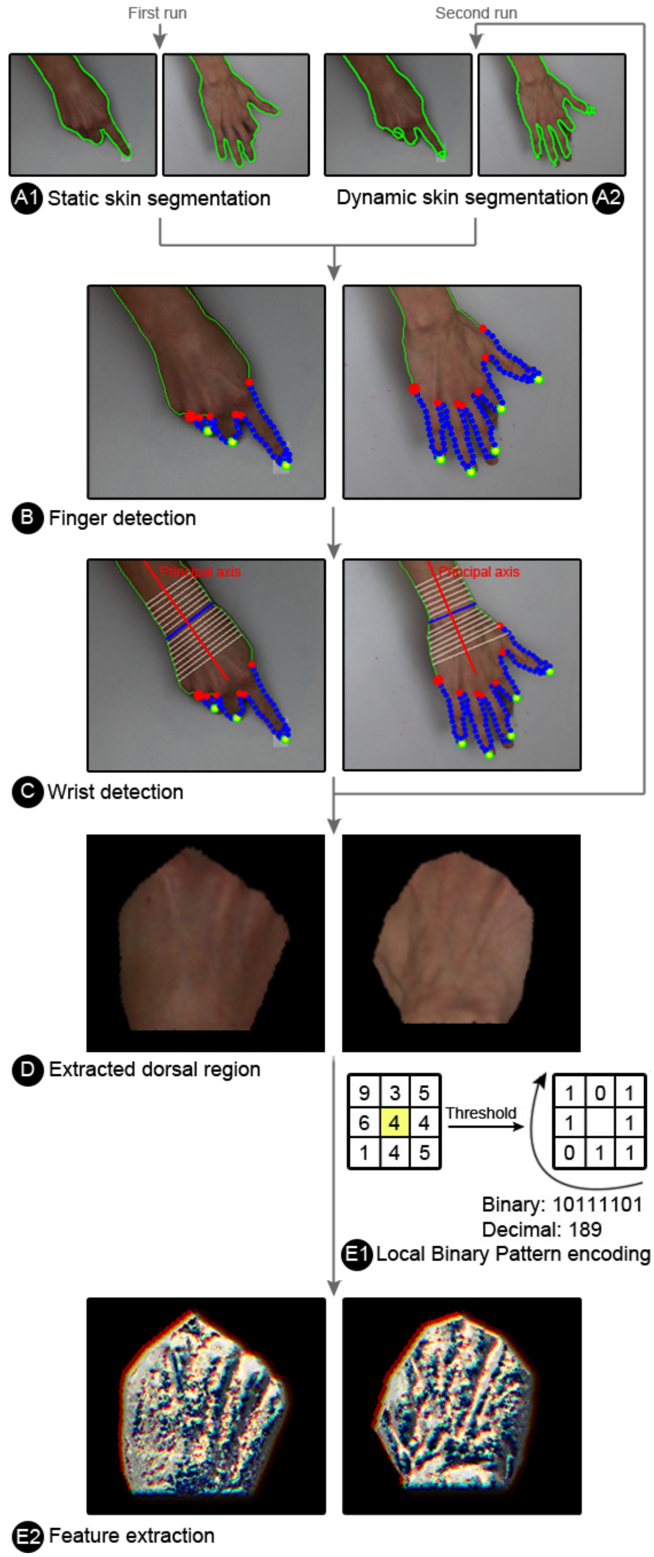


Figure 3. Extracting fine-grained details from the dorsal hand region (step 1 and step 2): (A) The skin region is detected in the image; (B) The fingers are detected; (C) The direction and position of the wrist is detected; (D) The dorsal hand region is extracted; (E) Fine-grained details are encoded using the Local Binary Pattern operator.

the wrist to the top of the hand can vary extensively, even when fingers are detected, because hands are free to move in three dimensions when working on an interactive surface. Therefore, we take an approach similar to Choi et al. [6], and evaluate the width of the contour along the direction of the principal axis from the finger tips to the arm region. The width of the contour has very specific characteristics at the position of the wrist. In the direction of the fingertips, the width will vary significantly, but in the direction of the arm, the width will be fairly constant. We consider the wrist to be located outside the finger regions, at the position where the width ratio reaches its maximum value (Figure 3-C).

Step 2: Feature Extraction

Once the dorsal hand region is detected using the algorithm described in the previous step, unique features need to be extracted from this region (Figure 3-D). There are several interest point detectors, such as SIFT [13] and SURF [4]. However, the skin of the hand is flat and does not contain a lot of these interest points. Therefore, we use the pattern of the entire skin region as a feature. Before we can compare skin patterns in different images (step 3), a texture descriptor is needed to quantize these patterns. Color histograms [35] are the most straightforward image descriptors, but this technique is very sensitive to changes in illumination. During preliminary studies, we noticed that color histograms are not able to capture unique information consistently over time, because the illumination of the hand changes significantly due to shadows cast by the user's body.

Local Binary Patterns (LBP) [21] is a technique to describe very fine-grained details of textures in images. In contrast to color histograms, LBP is invariant to any monotonic change in intensity values. The LBP-operator labels the pixels of an image by thresholding the 3x3-neighborhood of each pixel with the center value. The decimal representation of the number formed by the concatenation of all binary digits in the neighborhood is the intensity value of the central pixel in the local pattern (Figure 3-E1). Carpus uses Circular Local Binary Patterns (CLBP) [22], an extension to the basic LBP. The CLBP-operator samples the neighborhood circularly with a variable radius. If a point on the circle does not correspond to image coordinates, the point gets interpolated. The CLBP-operator is thus able to capture patterns at different scales: a small radius captures local details, whereas a larger radius captures more global information.

In our preliminary exploration, we noticed that the luminance component of the hand's skin texture (i.e. the hairs, lines, grooves, folds and furrows) contains most of the discriminative information. The color information was not very discriminative. Therefore, we only capture information in the luma (Y) component of the extracted skin regions. When capturing details at radii 2, 5 and 10 using the CLBP-operator, we can describe most of the unique features of the dorsal region of the human hand (Figure 3-E2).

Step 3: Feature Matching

To identify users, Carpus matches descriptions of textures captured using the CLBP-operator with a database of previously captured features. Before features of different images can be compared, we need to ensure that these features are captured at the same scale. Therefore, we only process a detected dorsal hand region when a "touch down" event is performed. A naive approach to comparing descriptions of textures would be a simple histogram comparison. When using this approach, however, all spatial information is lost. Through experiments, we discovered that without this spatial information, the extracted features are not very discriminative. Therefore, we use a technique similar to Ahonen et al. [1]. First, the skin region is divided into square patches of equal size (Figure 4). We currently use patches of 20 by 20 pixels. Next, histograms of these local micropatterns are computed. Histograms of the same patch, which contain information of fine-grained details at different scales, are concatenated to form a single description of the patch.

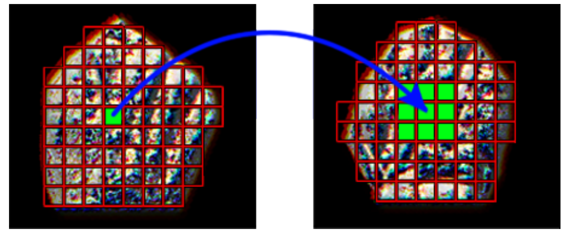


Figure 4. Feature matching (step 3): finding similar patches in a predefined neighborhood.

Opposed to the approach of Ahonen et al. [1], the local histograms of all patches in an image are not concatenated. Concatenation of these local histograms implies the integration of the hand's shape as a feature. Since the shape of the hand can vary significantly under different postures, we take another approach. Carpus compares two images by finding the best match for each patch in a predefined neighborhood around the corresponding position in the other image (Figure 4). We use a neighborhood of two times the patch size.

Experimenting with different histogram comparison metrics, we found that the histogram intersection technique provides the best matching of patches. As a result, the similarity between features extracted from two dorsal hand regions is the sum of the number of similar pixels for each of the patch's best matches. An unknown hand is assigned the identity of the dorsal hand region in the training set with the highest number of similar pixels.

Step 4: Relating Touches to Identified Regions

Once the identities of all extracted dorsal hand regions are known, each touch point needs to be related to an identified skin region. In contrast to other touch identification techniques [27, 16], Carpus is able to relate touch points to identified skin regions unambiguously. Each touch position is directly related to a single dorsal hand region by the contour of the fingers detected in step 1.

SYSTEM SPECIFICATIONS AND PERFORMANCE

For our experiments, we use a Philips BDT4225EM/06 42" multi-touch display, mounted horizontally on a table (Carpus was also informally tested in combination with other technologies, such as an FTIR tabletop). A Point Grey Grasshopper2 camera with a resolution of 1624 by 1224 pixels is mounted above the surface and transmits 22 frames per second at 45 dpi via the network interface of the camera to a PC (2.1 GHz Intel Core 2, 4 GB RAM) running Carpus. Artificial light sources are used to ensure that over different sessions of our experiment, the lighting conditions remain unchanged.

Carpus is able to identify users in real-time and scales well to higher resolutions. However, we did not focus on the performance of our algorithm, so many optimizations are possible. The extraction of the dorsal hand region (step 1) takes 28 ms on our system if only one hand is visible in a single frame with a resolution of 1100 by 790 pixels, 41 ms if two hands are visible and 52 ms if three hands are visible. When using a higher resolution camera, as in our setup, this step can still be performed at the specified lower resolution without reducing the accuracy of our identification technique. After the position of the dorsal hand region is detected, features can be extracted from the original high-resolution image. The feature extraction and matching step is only executed on a hand that is detected at the position of the "touch down" event in the last frame, to ensure that all extracted features have the same scale. This process takes on average 43 ms when the hand is captured at 45 dpi. When 16 samples are used for each registered hand, matching features of a single hand takes on average 68 ms if 4 hands are registered and 155 ms if 8 hands are registered.

UNIQUENESS OF THE DORSAL HAND REGION

In this first experiment, we evaluate the uniqueness of the dorsal hand region over different users, as well as over different hands of the same user. Here, we only consider the hands-down posture [30], because in this posture the overhead camera can capture clear images of the dorsal region. The robustness of Carpus with regard to posture changes will be tested in a second experiment, which we discuss in the next section.

Tasks

We recruited 22 participants (5 female) between 22 and 50 years old. We instructed them to take off all jewelry (jewelry could actually increase the recognition rates, because it often provides very unique features). Each participant sat down at an interactive tabletop and placed her/his left and right hand flat on the surface with fingers spread. They did this at 15 predefined positions that were evenly distributed over the entire surface area. In each position, our camera captured a single image. These images are used to train and test our system.

Procedure

From a set of 660 images, we simulated six scenarios that differed in the number of hands registered with the system (4, 10 and 20 users, each registering one or two hands). For each scenario, we generated 25 sets of randomly drawn groups. For each set, a 10-fold cross-validation with a stratified random selection of training images is performed, resulting in a total of 34000 trials (for each scenario 25 sets x 10 trials x 2 hands x number of users). A hand region is correctly identified if the system can not only match it to the correct user, but also to the correct hand of that user.

Results

Table 2 lists the recognition rates of our technique for the six scenarios. The accuracy is very high in all scenarios, although it slightly decreases when larger groups of users are registered with the system. These results demonstrate that our extracted features are unique, even for fairly large groups of users. In addition, note that if both hands of a user are registered, our system can distinguish between each hand the vast majority of the time. This enables identification of both hands and thus non-symmetric devision of labor [9].

	Group size		
	4	10	20
One hand registered	99.5%	99.4%	99.1%
Two hands registered	99.4%	99.4%	99.0%

Table 2. Recognition rates for the hands-down posture when one or both hands of 4, 10 and 20 users are registered with the system.

ROBUSTNESS AGAINST POSTURE VARIATIONS

In the first study, we showed that the dorsal hand region can be used to reliably map hands to users. The goal of this second study is to show that Carpus is sufficiently robust for postures that are common during the practical use of multi-touch systems. When interacting on a tabletop, the overhead camera often gets a clear view of the dorsal hand region, e.g. when clicking a button (Figure 5-A). When scaling or rotating an object (Figure 5-B-C), however, the captured dorsal hand region can be skewed, potentially influencing the accuracy of the recognition. To evaluate to what extent Carpus can handle posture variations, we ran a second experiment.

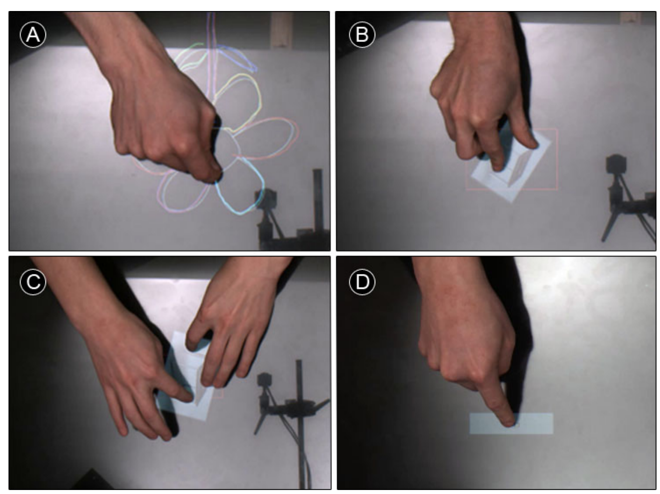


Figure 5. The four tasks in our second experiment: (A) painting a flower; (B) scaling and rotating images using one hand; (C) scaling and rotating images using two hands; (D) pressing buttons.

Tasks

We asked our 22 participants to perform four additional tasks. During these tasks, we did not give any instructions regarding hand postures with the purpose of capturing natural interaction. (1) In the first task, participants were asked to color the contours of a flower in a painting application using only their right hand (Figure 5-A). The application showed a preview of the flower, with the colors we expected the participants to use. (2) In the second task, participants scaled and rotated 10 images (evenly spread out over the surface area) to fit them in a box using traditional free transformation gestures (Figure 5-B). This task was first performed with the right hand and afterwards repeated with the left hand. The center points of the images were fixed, so that there was no need to move them. (3) In the third task, those same images were displayed and participants were asked to scale and rotate them using both their left and right hand together (Figure 5-C). (4) Finally, each participant was asked to click 15 buttons that were spread out over the surface area, first with their right hand and afterwards with their left hand (Figure 5-D).

Data Collection

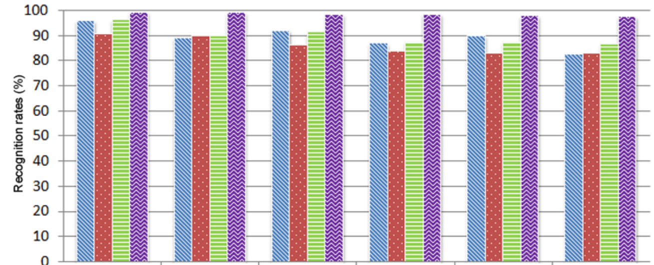
During this experiment, all interaction was captured by our overhead camera and streamed to a PC, together with all touch events ('down', 'move' and 'up'). Afterwards, all frames were extracted in which one or more 'down' events were registered, and these frames were used in simulations as training and test data. For the first, second and third task, we collected on average 35.5, 82.8 and 39.7 images per participant, respectively. For the last task, 30 images per participant were captured. In total, we collected 5009 images. For the third task, in which two hands are visible at the same time, we provided the position of the left and right hand to the system by processing those images by hand. With this data, the correctness of the result of our hand identification algorithm can be verified.

Procedure

Using our collection of captured frames, we simulated six scenarios that differed in the number of hands registered with the system (4, 10 and 20 users, each registering one or two hands). For each scenario, we generated 25 sets of randomly drawn groups. We trained the system with randomly drawn samples for each hand that needed to be registered. This training set consisted of four samples of hands in the hands-down posture from the previous experiment, and four samples of the pointing task (task 2) and the free transformation task with one (task 3) and both hands (task 4). We determined that this collection of 16 training samples per hand is needed if a large range of postures is to be supported. After this training phase, we tested our system with 10 randomly drawn samples for each hand in each of the four tasks (the randomly drawn samples were always different from the training samples), resulting in a total of 136000 trials (for each scenario 25 groups x 10 trials x 2 hands x 4 tasks x number of users). As in the first experiment, a hand region is correctly identified if the system can relate it to the correct user and the correct hand of that user.

Results

Figure 6 summarizes our findings of the second experiment. These results show that hands in a pointing posture can be identified very accurately, because the overhead camera has a clear view on the dorsal hand region. Carpus had the greatest difficulty identifying a single hand in a free transformation



	4 users - 1 hand registered	4 users - 2 hands registered	10 users - 1 hand registered	10 users - 2 hands registered	20 users - 1 hand registered	20 users - 2 hands registered
Painting task	95,7	89	91,8	86,7	89,7	82,2
Free transformation task using 1 hand	90,5	89,6	85,9	83,8	83	83
Free transformation task using 2 hands	96	89,7	91,3	87,4	87,1	86,5
Pointing task	99,1	98,9	98,2	98,3	97,6	97,3

Figure 6. The recognition rates for different group sizes for all four tasks of our second experiment.

gesture (Figure 5-B). We observed that users often rotated their hand in various directions while performing such a gesture. As a result, the dorsal hand region was highly skewed in many of the captured frames. In that case, matching features is much more difficult. We also noticed that, because of tilting of the hand, the space between two fingers was sometimes almost entirely occluded. That makes it harder for our algorithm to detect all fingers correctly, in order to exclude them. Parts of fingers still present in the feature extraction step will produce unreliable data, which can influence the accuracy of our technique. However, even for these challenging hand postures, the recognition rate for a smaller group of users is still relatively high.

The accuracy for the painting task was somewhat surprising, as we anticipated that this recognition rate would be almost identical to the pointing task because of the similarity of the expected hand posture. After analyzing our captured data, we noticed that some users tilted their hand during this task to reduce the occlusion of the active painting area, in order to paint the contour more precisely. This behavior causes the extracted dorsal hand region to be skewed, and makes feature matching more difficult. However, the results of this experiment show that, even when no instructions are given and the user can naturally interact with the surface, Carpus provides an accurate technique for identifying touches of a small group of users.

We reproduced our entire experiment, after adding four randomly drawn images of the painting task to the training set. The recognition rate of the painting task was significantly higher (97%, 96,5%, 97,8%, 96,2%, 96,1% and 95,5% for the six scenarios). The accuracy for the other tasks remained almost unchanged. This suggests that the recognition rate can be improved by training the system with samples of hand postures that are more similar to the hand postures that will occur in the application.

One can imagine many other hand postures that can be used on interactive surfaces. However, our tasks produced a wide range of postures that are very common during the practical use of multi-touch systems.

USAGE SCENARIO

Carpus enables non-intrusive and transparent identification of users on interactive surfaces. This allows an entire range of new applications in which people collaborate on a shared surface, such as buying new products [17]. We developed an interactive application that can be deployed in, for example, a mobile phone retail environment. Families or friends shopping together for a new mobile phone can collaborate on a multi-touch tabletop located inside the store to find more information about products and compare specifications.

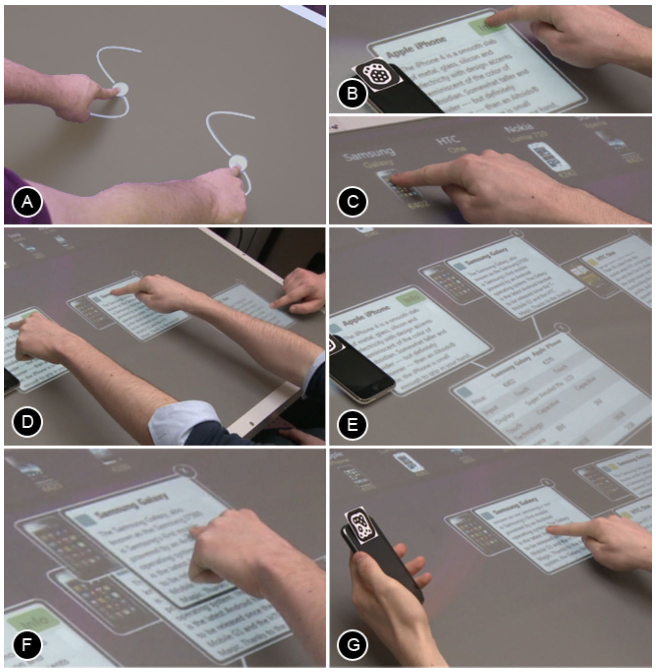


Figure 7. Carpus enables non-intrusive identification of (both hands of) users, for example in a mobile phone retail environment, allowing users to find more information about products and compare specifications.

Figure 7 illustrates a typical usage scenario. John and Jane are looking for a new mobile phone. A nice looking phone catches John's eye and they both walk to the interactive display to get more information about this product. (A) Because it is the first time that they have visited the store, a simple widget allows them to very quickly register both hands with the system. (B) John then puts the phone on the interactive display to get more information (the device is currently recognized by a visual tag). (C) Identity information from Carpus is used to create a temporary profile for each customer. This makes it possible to track the user's interests in order to provide product recommendations. (D) While Jane reads through the specifications of recommended phones, John finds another nice mobile phone and compares two products using a two-handed gesture. (E) Carpus unambiguously recognizes this gesture because the system can distinguish between touch points of both users. (F) Jane notices John's actions and also wants to take a look at the product in which he is interested. John then makes a copy of the information by moving it with his left hand to Jane. (G) John and Jane are now both interested in the phone suggested by the system and go find it in the shop. When returning back to the interactive display, they notice that other customers started a new session. However, when they touch the display, the system uses the identify information to restore their session.

DISCUSSION

Carpus extracts identity features from the back of the human hand. However, as mentioned in the benefits and limitations section, the detection of this region is difficult in two particular situations. The first situation is when identification is needed while performing a very fast gesture, such as a swipe-gesture. It is possible that the camera cannot capture a sharp image of the dorsal region. This problem can be addressed by using a camera with an appropriate shutter speed, or by adding tracking capabilities to the algorithm in order to identify the dorsal region during slow movements, and track the arm during fast movements. Secondly, researchers have investigated new types of interactions in which the dorsal hand region is not always clearly visible to the overhead camera [18, 11]. Using multiple cameras that observe the dorsal hand region from different points of view may alleviate this problem, making it much more likely that the system can capture a clear image of the dorsal hand region. Such a multi-camera setup can also be used to reduce occlusion problems that are inherent to the use of overhead cameras. This is especially the case when vertical displays or tilted tables are used in combination with Carpus.

In our first study, we found that the dorsal hand region is unique over a large group of users. In a second study, we demonstrated the robustness of our technique for a range of hand postures. Carpus has the highest recognition rate (97,3% when 20 users register both hands) when a user clicks buttons, which is a very common action in interfaces. Since we only identify a user when a "touch down" event occurs, a lot of other interactions are actually reduced to clicking (e.g. moving sliders, scrollbars, objects, etc.). The recognition rate was the lowest when identifying scale and rotate actions performed by a single hand (Figure 5-B). This is expected, because users rotate their hands in all directions when performing such gestures. As a result, the camera captures a skewed view of the dorsal region that is more difficult to match with the training set. However, user identification during these actions is only needed in very specific circumstances and the vast majority of applications will only need the user's identity during pointing actions [29].

We also found that the recognition rate is higher when training the system with samples of hand postures that are more similar to the hand postures that occur in the application. This suggest that the accuracy of our technique could be improved for some hand postures by refining the training set over time with samples of hand postures that are performed in the actual application. Future versions of Carpus will implement this by tracking users' hands during interaction with the application. Samples for which no good match is found in the training set can be added to the training data after a reliable sample has been detected during the same hand movement. Tracking was not yet integrated in Carpus because it could bias the results of our studies depending on the type of user interface (e.g. how frequently users' hands leave the surface area).

The recognition rate of Carpus is most likely sufficiently accurate for all postures when a small group of users are registered with the system. However, even in these situations, Carpus is currently not intended for applications that require real security (e.g. accessing emails, online banking). As the dorsal hand region contains no papillary ridges, our extracted features are not as persistent and immutable as fingerprints. It is, for example, possible for the skin to get a tan, and wounds can result in permanent scars. However, more long-term research is needed in this area to investigate the effects of these changes on the recognition rate of Carpus.

As already mentioned in the benefits and limitations section, our technique has some difficulties eliminating users that are not registered. In an additional study, we tested our system with four known and four unknown users, which resulted in an acceptance rate of 87% and a false acceptance rate of 7%. These results indicate that explicit registration is needed when higher reliability is required. However, registering with the system is only a one-time cost compared to systems that require configuration at the start of each session. When a reduced reliability is acceptable, for example in case of very short interactions with non-crucial applications, spontaneous registration is possible. We already demonstrated the potential use of interactive widgets as a means for transparent registration in our usage scenario. However, more experiments are needed to make sure that a sufficient number of different hand postures are captured during the interaction with this widget to get the best possible accuracy rates.

CONCLUSION

In this paper, we presented Carpus, a novel, non-intrusive technique for identifying users of interactive surfaces. Our approach relies on the extraction of unique information from the back of the human hand. This hand region is extracted from high-quality images captured by an overhead camera. Fine-grained features are encoded using a special pattern description technique. For traditional pointing tasks, Carpus can uniquely identify both hands of a user with 97.3% accuracy, even when large groups of users are registered. For more complex gestures that occur less often, smaller group sizes are recommended to achieve a higher recognition rate. Carpus enables touch identification for non-crucial applications in walk-up-and-use scenarios in which users interact frequently and in an unplanned fashion. In addition, our presented technique is able to differentiate between the two hands of a user, opening up even more interaction possibilities. Carpus is also easy to deploy and can be used in combination with all existing touch technologies. This makes our approach suitable in many situations in which touch identification can enrich the multi-user experience.

ACKNOWLEDGMENTS

We would like to thank Brent Hecht for his valuable comments and language editing.

REFERENCES

1. Ahonen, T., Hadid, A., and Pietikäinen, M. Face recognition with local binary patterns. In *Proceedings of the 8th European Conference on Computer Vision (ECCV '04)*, Springer (2004), 469–481.
2. Annett, M., Grossman, T., Wigdor, D., and Fitzmaurice, G. Medusa: a proximity-aware multi-touch tabletop. In *Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11)*, ACM (2011), 337–346.
3. Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff, T., Becker, T., Holz, C., and Baudisch, P. Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology (UIST '10)*, ACM (2010), 209–218.
4. Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. Speeded-up robust features (SURF). *Computer Vision and Image Understanding 110*, 3 (2008), 346–359.
5. Boreki, G., and Zimmer, A. Hand geometry: a new approach for feature extraction. In *Proceedings of the 4th IEEE Workshop on Automatic Identification Advanced Technologies (AUTOID '05)*, IEEE Computer Society (2005), 149–154.
6. Choi, J., Seo, B.-K., and Park, J.-I. Robust hand detection for augmented reality interface. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry (VRCAI '09)*, ACM (2009), 319–321.
7. Dietz, P., and Leigh, D. DiamondTouch: a multi-user touch technology. In *Proceedings of the 14th annual ACM symposium on User interface software and technology (UIST '01)*, ACM (2001), 219–226.
8. Freeman, W. T., Anderson, D. B., Beardsley, P. A., Dodge, C. N., Roth, M., Weissman, C. D., Yerazunis, W. S., Kage, H., Kyuma, K., Miyake, Y., and Tanaka, K. Computer vision for interactive computer graphics. *IEEE Computer Graphics and Applications 18*, 3 (1998), 42–53.
9. Guiard, Y. Asymmetric division of labor in human skilled bimanual action: the kinematic chain as a model. *Journal of Motor Behavior 19* (1987), 486–517.
10. Gutwin, C., Greenberg, S., Blum, R., Dyck, J., Tee, K., and McEwan, G. Supporting informal collaboration in shared-workspace groupware. *Journal of Universal Computer Science 14*, 9 (2008), 1411–1434.
11. Harrison, C., Schwarz, J., and Hudson, S. E. TapSense: enhancing finger interaction on touch surfaces. In *Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11)*, ACM (2011), 627–636.
12. Koike, H., Sato, Y., and Kobayashi, Y. Integrating paper and digital information on EnhancedDesk: a method for realtime finger tracking on an augmented desk system. *ACM Transactions on Computer-Human Interaction 8*, 4 (2001), 307–322.

13. Lowe, D. G. Object recognition from local scale-invariant features. In *Proceedings of the 1999 International Conference on Computer Vision (ICCV '99)*, IEEE Computer Society (1999), 1150–1157.
14. Malik, S., and Laszlo, J. Visual touchpad: a two-handed gestural input device. In *Proceedings of the 6th international conference on Multimodal interfaces (ICMI '04)*, ACM (2004), 289–296.
15. Marquardt, N., Kiemer, J., Ledo, D., Boring, S., and Greenberg, S. Designing user-, hand-, and handpart-aware tabletop interactions with the TouchID toolkit. In *Proceedings of the 2011 ACM International Conference on Interactive Tabletops and Surfaces (ITS '11)*, ACM (2011), 21–30.
16. Meyer, T., and Schmidt, D. IdWristbands: IR-based user identification on multi-touch surfaces. In *Proceedings of the 2010 ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*, ACM (2010), 277–278.
17. Morris, M. R. A survey of collaborative web search practices. In *Proceedings of the 26th annual SIGCHI conference on Human factors in computing systems (CHI '08)*, ACM (2008), 1657–1660.
18. Morris, M. R., Huang, A., Paepcke, A., and Winograd, T. Cooperative gestures: multi-user gestural interactions for co-located groupware. In *Proceedings of the 24th annual SIGCHI conference on Human Factors in computing systems (CHI '06)*, ACM (2006), 1201–1210.
19. Morris, M. R., Ryall, K., Shen, C., Forlines, C., and Vernier, F. Beyond "social protocols": multi-user coordination policies for co-located groupware. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW '04)*, ACM (2004), 262–265.
20. Napier, J., and Tuttle, R. *Hands*. Princeton Science Library. Princeton University Press, 1993.
21. Ojala, T., Pietikäinen, M., and Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition 29*, 1 (1996), 51–59.
22. Ojala, T., Pietikäinen, M., and Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 7 (2002), 971–987.
23. Orr, R. J., and Abowd, G. D. The smart floor: a mechanism for natural user identification and tracking. In *CHI '00 extended abstracts on Human factors in computing systems*, ACM (2000), 275–276.
24. Phung, S. L., Bouzerdoum, A., and Chai, D. A novel skin color model in YCbCr color space and its application to human face detection. *Proceedings of the 2002 IEEE International Conference on Image Processing (ICIP '02) 1* (2002), 289–292.
25. Phung, S. L., Bouzerdoum, A., and Chai, D. Skin segmentation using color pixel classification: analysis and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*, 1 (2005), 148–154.
26. Piper, A. M., O'Brien, E., Morris, M. R., and Winograd, T. SIDES: a cooperative tabletop computer game for social skills development. In *Proceedings of the 2006 ACM conference on Computer supported cooperative work (CSCW '06)*, ACM (2006), 1–10.
27. Richter, S., Holz, C., and Baudisch, P. Bootstrapper: recognizing tabletop users by their shoes. In *Proceedings of the 30th annual conference on Human factors in computing systems (CHI '12)*, ACM (2012), 1249–1252.
28. Roth, V., Schmidt, P., and Güldenring, B. The IR ring: authenticating users' touches on a multi-touch display. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology (UIST '10)*, ACM (2010), 259–262.
29. Ryall, K., Esenther, A., Forlines, C., Shen, C., Shipman, S., Morris, M. R., Everitt, K., and Vernier, F. D. Identity-differentiating widgets for multiuser interactive surfaces. *IEEE Computer Graphics and Applications 26*, 5 (2006), 56–64.
30. Schmidt, D., Chong, M. K., and Gellersen, H. HandsDown: hand-contour-based user identification for interactive surfaces. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction (NordiCHI '10)*, ACM (2010), 432–441.
31. Schmidt, D., Chong, M. K., and Gellersen, H. IdLenses: dynamic personal areas on shared surfaces. In *Proceedings of the 2010 ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*, ACM (2010), 131–134.
32. Schöning, J., Rohs, M., and Krüger, A. Using mobile phones to spontaneously authenticate and interact with multi-touch surfaces. *PPD '08 workshop on designing multi-touch interaction techniques for coupled public and private displays* (2008).
33. Scott, S. D., Sheelagh, M., Carpendale, T., and Inkpen, K. M. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW '04)*, ACM (2004), 294–303.
34. Segen, J., and Kumar, S. Gesture VR: vision-based 3D hand interace for spatial interaction. In *Proceedings of the 6th ACM international conference on Multimedia*, ACM (1998), 455–464.
35. Swain, M. J., and Ballard, D. H. Color indexing. *International Journal of Computer Vision 7*, 1 (1991), 11–32.
36. Wang, F., Cao, X., Ren, X., and Irani, P. Detecting and leveraging finger orientation for interaction with direct-touch surfaces. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology (UIST '09)*, ACM (2009), 23–32.