Hand tracking for enhanced gesture recognition on interactive multi-touch surfaces

Florian Echtler, Manuel Huber, Gudrun Klinker, PhD
{echtler,huberma,klinker}@in.tum.de

Technische Universität München - Institut für Informatik
Technical Report TUM-I-07-21

Abstract

Recently, interactive surfaces with multi-touch sensors based on frustrated total internal reflection (FTIR) have seen increased attention in research and commerce. In this paper, we present a new method of gathering data about the users' gestures on an interactive table beyond simple binary touch information.

In addition to the infrared light emitters at the rim of the interaction surface, a second infrared light source is placed above an interactive table to create shadows of hands and arms. By tracking these shadows with the same rear-mounted camera, several consecutive and disjoint surface contacts can be traced back to the same user, thereby enabling new interaction techniques. We demonstrate this approach at the example of a virtual whiteboard with persistent color assignments for each participant.

1 Introduction

In HCI research, multi-touch screens have recently seen growing attention. They offer new possibilities for the development of interaction techniques towards a more direct manipulation of data. Moreover, in the form of multi-touch tables, these systems enable new forms of collaboration between several concurrent users which were previously difficult to realize.

However, the sensing hardware still has limitations. For example, currently available systems are unable to distinguish the fingers on one hand from those on another one. This limits the types of gestures which can be detected by such a system.

We present a method to alleviate this problem. By combining a shadow tracker with an interactive table based on FTIR, we allow the system to associate different touch points on the surface with disjoint shadows. This enables determination whether two fingers touching the surface belong to the same hand.

2 Related Work

The major part of research is currently concentrated on three established multi-touch systems. These are DiamondTouch [4], SmartSkin [13] and FTIR [7]. We will give a short summary of each in this section.

DiamondTouch works by emitting high-frequency signals through antennas embedded in the opaque interaction surface. Every user is connected to a separate receiver which allows the system to determine to which user each surface contact belongs. However, no differentiation between fingers is possible, though the system is able to record multiple concurrent touches. Users also have to remain seated on their receiver-equipped chair for the system to work. Feedback for the user(s) is given through front projection on the interaction surface. Objects on the surface are not registered, even if conductive.

One exemplary system which builds on the DiamondTouch is the DigiTable project [3]. It combines a visible-light camera above the table with the interaction surface to allow remote gesture visualization.

SmartSkin is based on capacitance measurements through a wire grid. When a conductive object like a user's hand comes into close proximity of the surface, the capacitance between two perpendicular wires changes. A waveform which is transmitted from the vertical wires to the horizontal ones is therefore reduced in amplitude. This change can be detected by a receiver connected to the horizontal wires. Note that no direct touch is necessary, a hand hovering closely over the surface can also be

detected. Electrically conductive objects on the surface can be tracked as long as the feedback loop to the receiver is closed through a user's touch.

FTIR-based systems have been pioneered by Jeff Han. He has adapted the underlying physical principles, which have originally been used for fingerprint scanning, to provide multi-touch information. A detailed description of FTIR will be given in the next section, as our system is based on this method.

One of the main appeals of FTIR is that no customized hardware like antenna or capacitor sheets are necessary. This has boosted acceptance in the research community, as it allows construction of a multi-touch screen from common off-the-shelf hardware with moderate skills. Since they do not contain opaque layers, back projection is also possible, thereby improving the user experience over front-projected systems.

FTIR has also spread to commercial applications. Two companies, Perceptive Pixel [1] and Microsoft [10], have announced ready-to-run systems.

3 System Description

In this section, we will give an overview of the hardware and software setup, with emphasis on our additions to a plain FTIR-based system. A more in-depth description of the FTIR principle provides background information.

3.1 FTIR: physical principles and limitations

These systems work because touching a transparent surface changes the optical properties of the contact spot. Infrared light emitters (usually LEDs) are placed around the rim of the interaction area, e.g., a sheet of acrylic glass, and radiate into the material.

At the surface, total reflection occurs due to the difference in refractive index between air and substrate (see Figure 1). Therefore, a large percentage of the emitted infrared light is reflected back and transported through the material, similar to an optical fiber. However, once a denser material, such as skin, touches the surface, the change in refractive index prevents total reflection.

Most of the light illuminates the skin, which is then visible on the back side of the interaction area as a bright spot. An infrared camera behind the screen records these spots, which are then processed to extract touch information. Feedback to the user is presented on a projection screen behind the interaction surface. Note that almost no other materials except some soft plastics show this behaviour. Other objects, e.g., a mug standing on the surface, do not create any noticeable reflections.

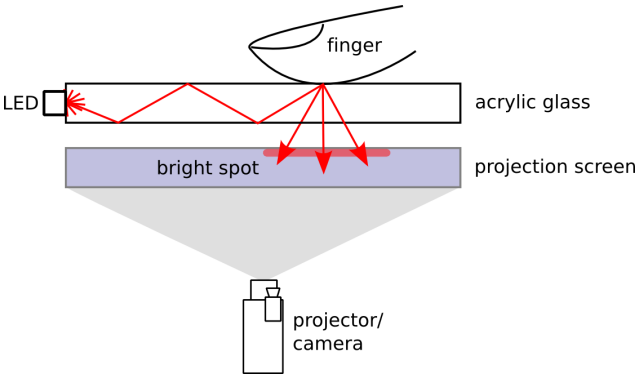


Figure 1. FTIR principle

However, one of the limitations of the FTIR principle is that it is impossible to differentiate between different fingers on the same hand and fingers from different hands, as they both appear as bright spots without structure.

This limits the variety of gestures that can be recognized by such a system. For example, a three-finger gesture can not be distinguished from a two-finger gesture with one hand and an additional finger from another hand (as seen in Figure 2). We aim to remove this limitation with our approach.

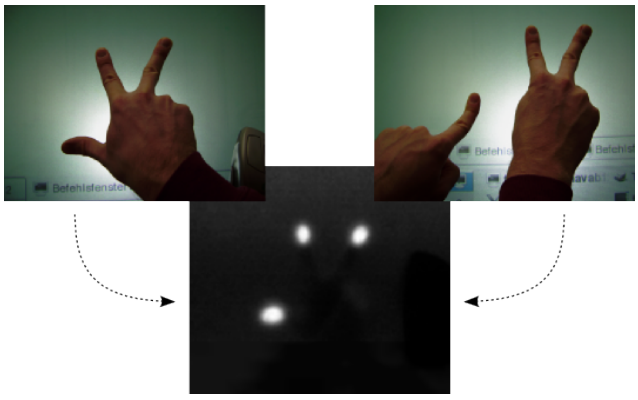


Figure 2. ambiguous gestures

3.2 Multi-Touch Table

The centerpiece of our setup is the **T**angible **I**nteraction **S**urface for **C**ollaboration between **H**umans (TISCH), an interaction device which provides room for 4 to 6 concurrent users (see Figure 3). It consists of a table of hardened frosted glass (1.10 x 0.7 m) which is used as backprojection

surface. On top of this table, an acrylic glass plate of the same size is placed which carries 70 infrared LEDs (Osram SFH4250 SMD) around its rim as described above.

To improve light transmission into the plate, the LEDs are mounted on the rim with instant glue. This creates a seamless, transparent bond and therefore, the rim of the acrylic glass does not have to be polished. The upper surface can be treated with silicone spray to decrease friction and improve the user experience, especially when dragging a finger across the table.

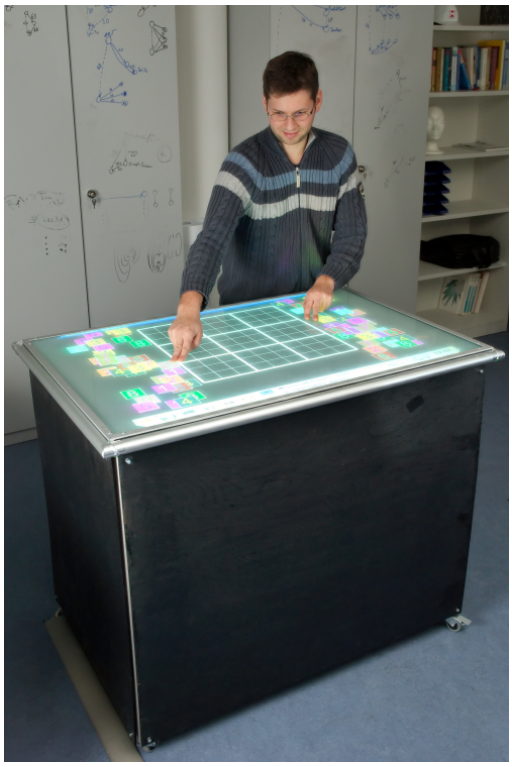


Figure 3. TISCH: multi-user interaction device

Below the table, two mirrors provide the necessary optical path length between the rear-mounted projector-camera setup and the surface. We use a USB camera (Quickcam 5000 Pro) which has been modified with a daylight filter to provide infrared video at 30 FPS. The entire hardware is carried by an aluminium frame and protected by front and side panels. For mobility, the frame has been mounted on small wheels. This robust construction has already proven quite valuable when exposing the device to visitors, especially to children.

3.3 Shadow Tracking Hardware

In order to overcome the limitations of a pure FTIR-based setup as described in section 3.1, we have placed a second infrared light source above the table at a height of about 2.50 m to create shadows of objects over the surface (see Figure 4).

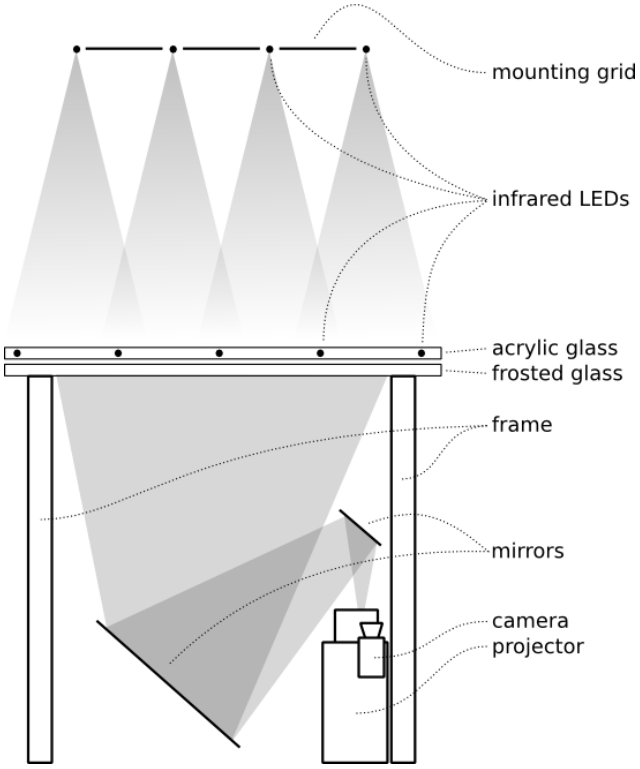
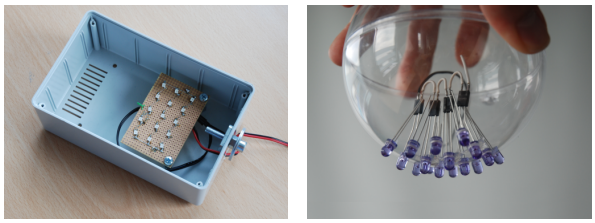


Figure 4. hardware setup

We alternate between illuminating the table sideways from the LEDs in the rim of the surface and from above in odd and even frames. Two consecutive frames can thus be used to track finger contacts as well as hand shapes. These two images will be referred to as contact and shadow image, respectively (see Figure 9 for an example). This results in a reduction of the overall frame rate by a factor of two. It also limits the maximum detectable speed of action. If the movement of the hand or fingers is too fast, the spatial offset between corresponding points in the two images grows too large. However, due to friction, the highest speed at which a finger can be dragged over the table surface is about $1\frac{m}{s}$ (even when using silicone spray), which translates to an offset of at most 3 cm per frame.

We have evaluated several options for top-down lighting. For optimal operation of the shadow tracker, hard shadows are desirable, so the ideal light source would be a point light.

We have therefore first attempted to create such a point light with two different hardware setups. The first one consists of 15 LEDs of the same kind as used in the table itself on a small circuit board, which have a very broad radiation pattern of 50% intensity at 60° angle (see Figure 5(a)). The second one consists of 16 Osram SFH485 LEDs, which have a standard 5 mm casing for easier handling and a significantly more narrow radiation pattern with 50% intensity at 20° angle. The LEDs are placed inside a sphere at an angle of 15°, thereby creating overlap between the cones of light for a more even pattern (see Figure 5(b)).



(a) with SMD LEDs (b) with 5 mm LEDs

Figure 5. point light sources

Unfortunately, due to the reflection properties of the two polished glass plates, both setups illuminate only a small part of the table directly below the light source.

The reason for this behaviour is that light from outside has to pass a total of four material-air interfaces (see Figure 6). If each layer reflects only 15% of incoming light (a conservative assumption), the total intensity arriving at the camera already drops to approximately $(1 - 0.15)^4 \approx 52\%$ of emitted light. The reflected percentage increases with decreasing angle of incidence according to Fresnel's equations (see [6]). Below the critical angle of approximately 41°, the light transmission even drops to zero. This is simply because at this angle, total reflection starts to occur and all light is captured in the upper plate.

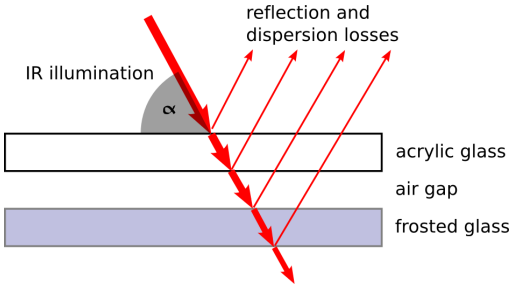


Figure 6. reflection of external light

Therefore, a large percentage of infrared light does not reach the camera when it does not hit the table almost perpendicular to the surface. The contrast between lit and unlit areas thus drops sharply after a small distance. Discernible shadows only appear in this central area.

Due to this effect, a different approach is necessary. We have constructed a metal grid on which 28 LEDs are mounted at a distance of 25 cm in each direction (see Figure 7). As the grid is suspended from the ceiling and stays parallel to the table, the cones of light hit the surface at the desired angle of about 90° and the reflection losses are kept small. This setup provided far better results with only one drawback. It is possible to unintentionally create additional shadows with head and upper body when bending over the table. This may cause undesirable merging of shadows.

Figure 7. LED mounting grid

Both sets of LEDs operate at a current of about 1 A with short pulses of 20 μs at 5% duty cycle to increase total light output. To avoid interference between the light from the overhead LEDs and the light from the LEDs in the table surface, they are synchronized to the 30 Hz frame clock of the camera, resulting in up to 83 flashes per frame. The different signals are shown in Figure 8. Note that an adjustable timeout t_0 after each frame is used to prevent "crosstalk" between the two modes due to inherent system delays on the USB bus.

The LEDs are connected to a control circuit based on a PIC18F microcontroller. It generates the pulse pattern, synchronizes the two light sources to the camera frame clock and allows user control through a serial link, e.g., for setting the frame change timeout. The camera frame clock

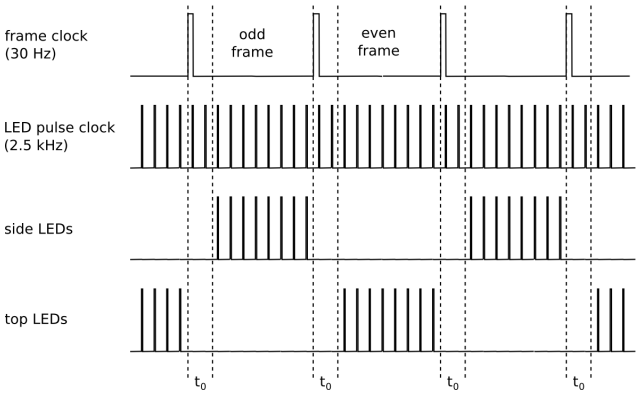


Figure 8. lighting control signals

is extracted from the video signal through a LM1881 sync separator IC and passed to the microcontroller.

3.4 Software Architecture

In order to present an abstraction of user input to applications, both images have to be processed and correlated. In Figure 9, an overview of our architecture is given.

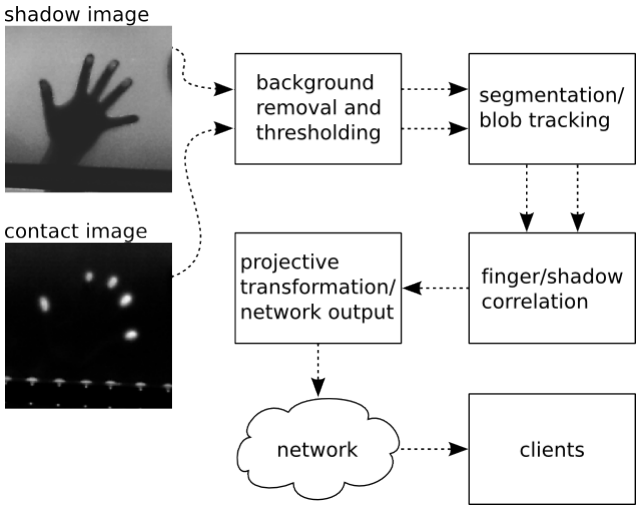


Figure 9. video processing pipeline

After two consecutive images (one for each light source) have been acquired from the camera, a static background image is subtracted from each. This background image is taken at system start, thereby adapting to the current environmental lighting conditions. It can be refreshed through a user command if the infrared brightness level changes, e.g., after opening the window blinds.

The resulting image is thresholded and segmented into disjoint blobs. Blobs below a certain pixel count are regarded as noise and discarded. For larger blobs, the centroids are calculated and are assigned an numeric identifier to track blobs across consecutive images. A linear motion prediction model is used to determine the most likely blob position in the current image in order to keep the identifiers consistent. This process happens separately for each of the two images.

For each blob in the contact image, a circular area with increasing radius around the centroid is searched in the corresponding shadow image. When a shadow blob is found, its identifier is assigned to the respective contact spot, thereby creating a relationship between shadows and surface contacts. This process is illustrated in Figure 10. As mentioned above, it can be impaired by very fast hand movements, however, when the upper limit for the scan radius is chosen large enough, this does not pose a problem.

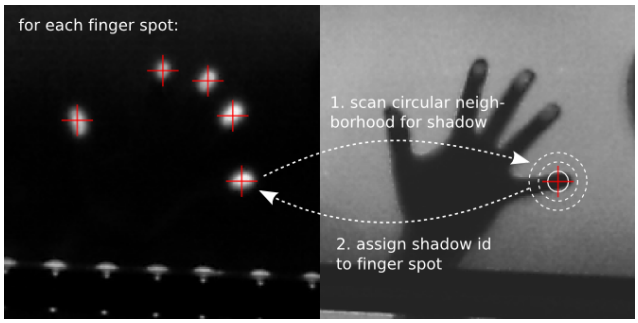


Figure 10. shadow processing

After all contact blobs have been checked for corresponding shadows, the projective distortion and different resolutions between the projector image and the camera image have to be accounted for.

A separate calibration tool is provided which guides the user through the process of touching 4 crosshairs on the screen in order to calculate a homography between camera and screen coordinates. The resulting matrix is stored in a file and read by the software in order to provide applications with input data in screen coordinates. An advantage of this method is that one single calibration homography can be used for both tracking modalities.

The list of contacts and shadows is now ready to be sent out over the network. Both lists are transmitted as UDP packets to the broadcast network address on port 31409 (0x7AB1) in a human-readable text format for easy debugging. Every packet starts with an identifier. Currently, there are 3 possible identifiers: `frame` which is sent when the processing of a new image pair starts and `contact` or `shadow` which describe a blob in the contact or shadow image, respectively. Apart from the identifier, a frame

packet contains the current frame counter value, while a contact/shadow packet contains

- the x/y screen coordinates of the blob's centroid
- the size in pixels of the blob
- the tracking id of the blob
- the id of the corresponding shadow (only for contacts)

An example datastream for a screen with a resolution of 800x600 pixels looks like this (every line is a single UDP packet):

```
frame 115
contact 312.349 483.932 50 73 10
contact 651.964 576.277 43 74 12
contact 612.432 580.304 27 81 12
shadow 115.693 510.481 539 10
shadow 715.846 439.234 683 12
frame 116
contact 313.959 483.735 51 73 10
contact 651.626 576.515 34 74 12
shadow 106.827 493.931 439 10
shadow 715.337 440.344 674 12
```

Note that the third contact with id 81 has been removed from the table between frame 115 and 116. By looking at the shadow id, it is now possible to determine whether the fingers belong to the same or to different hands.

4 Example Application: Virtual Whiteboard

As a proof of concept, we have developed a virtual whiteboard which can be used by several participants simultaneously. No additional setup is necessary, users can simply walk up to the table and start drawing with their fingers.

As opposed to comparable applications like CollabDraw [11], every user can select their current color themselves without having to interfere with the work of others. The color is assigned to a user's hand, i.e., every finger from one hand draws with the same color. At first, no color is assigned when a hand appears over the table. The user has to choose a color by means of a selector, which is accessed through a simple gesture: touching the table surface with all 5 fingers of one hand simultaneously opens the menu.

This menu offers a choice of 4 drawing colors (see Figure 11). Touching one closes the menu and assigns this color to the hand. From now on, every surface touch with a finger from this hand will draw with the selected color. The color assignment can be changed by opening the menu again or removed by withdrawing the hand from the table - as soon as the shadow vanishes, the color assignment becomes invalid, too.

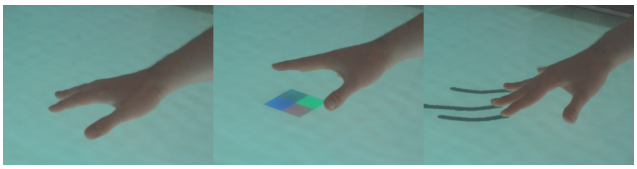


Figure 11. selecting a color on the Virtual Whiteboard

The application itself is based on OpenGL and opens a UDP socket to receive the data packets mentioned in the previous section. A list of contacts and shadows and their respective color assignments is maintained for each frame. When one of both items in a shadow-contact combination does not have an assigned color, but the other one does, the color is propagated to the other item. This makes the system more robust with respect to short dropouts. As every shadow has a color assignment independent of the others, several users can draw with both hands simultaneously, even while using a different color for each hand (see Figure 12).

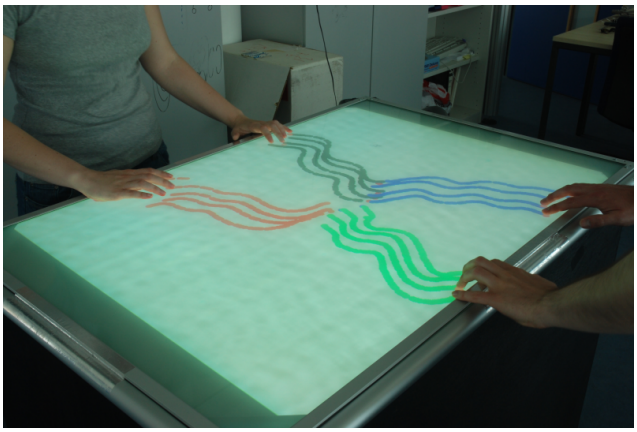


Figure 12. multi-handed drawing with different colors

5 Discussion and Future Work

In order to achieve better mobility for the whole system, the overhead light source could be fixed to a boom mounted on the frame. This would have the drawback of creating a physical obstacle on one side of the interaction surface,

but the entire system could be transported as a single object.

A promising next step is to apply a robust hand model to the shadow images in order to determine which finger corresponds to which touch point. Several options are presented in [2] and [9]. This offers the possibility to transfer well-known interface metaphors like left and right mouse click to multi-touch screens by assigning different actions to, e.g., thumb and index finger. Of course, a thorough user study is needed to evaluate the usability of such an interface.

As mentioned in the previous section, support for tangible user interfaces is also possible. There are several potential methods to achieve this.

Tangible user interfaces could be identified by their shadows. A distinctive shape, e.g., a pentagon, might be helpful to identify the tangible objects. Clearly, this is somewhat limited by the number of shapes and might need additional support by one of the methods mentioned below. However, in a casual setting similar to the Philips CafeTable [12], objects like mugs on the table could easily be identified and tracked by their round shadows.

It seems feasible to extend our approach with other tracking modes which are also interleaved between successive camera frames. For example, all external light sources could be disabled for one such mode. Tracked objects, e.g., tangible user interfaces on the table surface, can then be equipped with active infrared markers which are tracked by the camera. These would also have to be synchronized to the frame rate, presumably with a wireless connection in order to avoid distracting cables.

A third light source which illuminates the table surface from below could also be used to support tracked objects. Fiducial markers like those described in [8] can then be seen in the video image through the frosted glass. For the example image in Figure 13, both LED light sources were disabled. An opaque fiducial marker was placed face down on the table and illuminated from below with a separate infrared light source. This is works best when the markers in question are as close as possible to the frosted glass, i.e., are placed flat on the surface.

When considering additional tracking modes, a camera with higher frame rate would also be desirable. The effective frame rate results from the actual frame rate divided by the number of modes. Therefore, at least 60 frames per second would be necessary to provide smooth interaction with three different modes.

6 Conclusions

We have extended FTIR-based multi-touch screens with a shadow tracker. Our approach allows to differentiate between surface touches from different hands and opens the possibility to identify different fingers, thereby offering a wide range of additional options for user interface design.

Figure 13. camera view with marker on surface

Only a small amount of additional hardware is required (second light source with infrared LEDs and controller), but no additional calibration. This allows mobile systems which can be moved from their current location and back without extra setup work.

Our approach shows potential for further extensions like tangible user interfaces or a fully mobile version. For a short video demonstrating its use, please see [5].

References

[1] Perceptive Pixel. `http://www.perceptivepixel.com/`.

[2] A. Burns and B. Mazzarino. Finger tracking methods using EyesWeb. In *Proceedings of the International Gestures Workshop 2005*, pages 156–167, 2005.

[3] J. Coldefy and S. Louis-dit Picard. DigiTable: an interactive multiuser table for collocated and remote collaboration enabling remote gesture visualization. In *ProCams '07: Proceedings of the IEEE international workshop on projector-camera systems*, Washington, DC, USA, June 2007. IEEE Computer Society. `http://web.mit.edu/~ashdownm/www/procams2007/papers/paper9_Coldefy.pdf`.

[4] P. Dietz and D. Leigh. DiamondTouch: a multi-user touch technology. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 219–226, New York, NY, USA, 2001. ACM Press.

[5] F. Echtler. FTIR/Handtracking Demonstration Video. `http://campar.in.tum.de/personal/echtler/tum-i0721-handtrack.avi`.

[6] C. Gerthsen and D. Meschede. *Gehrtsen Physik*. Springer, 2005.

[7] J. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118, New York, NY, USA, 2005. ACM Press.

[8] H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. *2nd IEEE and ACM International Workshop on Augmented Reality*, page 85, 1999.

[9] J. Letessier and F. Berard. Visual tracking of bare fingers for interactive surfaces. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 119–122, New York, NY, USA, 2004. ACM Press.

[10] Microsoft. Surface. `http://www.microsoft.com/surface/`.

[11] M. Morris, A. Huang, A. Paepcke, and T. Winograd. Cooperative gestures: multi-user gestural interactions for co-located groupware. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1201–1210, New York, NY, USA, 2006. ACM Press.

[12] Philips. CafeTable. `http://www.design.philips.com/about/design/section-13507/index.html`, 2004-7.

[13] J. Rekimoto. SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 113–120, New York, NY, USA, 2002. ACM Press.