ABSTRACT
Frustrated Total Internal Reflection (FTIR) is a key technology for the design of multi-touch systems. With respect to other solutions, such as Diffused Illumination (DI) and Diffused Surface Illumination (DSI), FTIR based sensors suffer less from ambient IR noise, and is, thus, more robust to variable lighting conditions. However, FTIR does not provide (or is weak on) some desirable features, such as finger proximity and tracking quick gestures. This paper presents an improvement for FTIR based multi-touch sensing that partly addresses the above issues exploiting the shadows projected on the surface by the hands to improve the quality of the tracking system. The proposed solution exploits natural uncontrolled light to improve the tracking algorithm: it takes advantage of the natural IR noise to aid tracking, thus turning one of the main issues of MT sensors into a useful quality, making it possible to implement pre-contact feedback and enhance tracking precision.

ACM Classification Keywords
H.5.2 Information interfaces and presentation: User Interfaces. - Graphical user interfaces

Author Keywords
Multi-touch, FTIR, Tracking

INTRODUCTION
Multi-touch displays represent an intriguing research field that, recently, has gained new attention. Following seminal work from, among others, Buxton [3], and up to the recent developments of Han [5, 6], multi-touch systems offer a suitable working environment for computer supported co-operative work, leveraging the exploration of new frontiers of social computing.

A key technology for the design of multi-touch systems is Frustrated Total Internal Reflection (FTIR). Common FTIR setups [5] have a transparent acrylic pane with a frame of LEDs around the side injecting infrared light. When the user touches the acrylic, the light escapes and is reflected at the
finger's point of contact. The infrared sensitive camera at the back of the pane can clearly see these reflections. As the acrylic is transparent a projector can be located behind the surface (near to the camera) yielding a back-projected touch sensitive display. The software part consists in a basic set of computer vision algorithms applied to the camera image to determine the location of the contact point. An advantage of FTIR based sensors over competing solutions (such as DI, DSI [8]) is that this technology suffers less from ambient IR noise, and is thus more robust to changing lighting conditions. On the other hand, it is well known that FTIR has some disadvantages:

- it does not sense finger proximity, the user must touch the surface;
- it is difficult to track the fingers during movements;
- though more robust to changes in ambient light, it still relies on a control over lighting conditions.

To partly address such issues we propose to take advantage of the shadows that the hands of the user project on the interaction surface. Our experiments show that such solution allows to effectively sense user interaction in an uncontrolled environment, and without the need of screening the sides of the multi-touch table (see Figure 1). IR shadow tracking is described in depth in Section 2. In order to help other researchers and practitioners to duplicate our results the complete image processing pipeline is described in Section 3. References to related work and state of the art are given throughout the text where appropriate.

**TRACKING IR SHADOWS**

Tracking infrared shadows to improve the quality of multi-touch interaction has been studied before. Echtler and co-workers [4] describe a system to sense hovering on the surface, and thus provide pre-contact feedback in order to improve the precision of touch on the user's part. However the system they describe is based on a controlled IR lighting source above the table. In this sense their system exploits an additional artificial lighting source, increasing the dependence on the lighting conditions.

Our solution, as further described below, exploits natural uncontrolled light to improve the tracking algorithm. We take advantage of the natural IR noise to aid tracking, thus turning one of the main issues of MT sensors into a useful quality, making it possible to enhance tracking precision and implement pre-contact feedback.

The proposed technology exploits the shadows projected on the surface by the hands of the users to improve the quality of the tracking system. As said above, ambient light has a negative impact on the IR based sensors when the light coming from the IR LEDs is not bright enough to prevail on the background noise. However, the hands of the user project a shadow on the surface (that will appear as a dark area in the noisy background). Such dark area is easily tracked because it is almost completely free of noise.

Furthermore, fingertips correspond to the darker parts of the shadow, and can be recognized with good accuracy. Note that tracking the shadow is more and more effective as the
ambient light increases (as opposite from IR blobs tracking), thus IR tracking and shadow tracking tend to complement each other, the former working better in full darkness, the latter in full daylight. A second useful feature, consists in the ability of the shadow tracking system to sense objects that are only close (i.e., don’t actually touch) the surface, thus allowing the sensor to recognize a richer collection of gestures.

Finally, a well known problem of FTIR based systems is that blob brightness decreases as the user moves her hands fast. This problem is typically addressed covering the screen with compliant surface and silicon rubber. Shadow tracking does not suffer from this issue, and can thus be exploited to improve finger tracking during sharp movements. Such complementarities are key aspects of our work: it allows the system to work in less controlled environments, and to be more robust to changing lighting condition, as may easily happen in real world, off-lab installations. This latter is, as known, one of the major issues for computer vision based interactive systems.

Our implementation, based on OpenCV [7] for computer vision algorithms, shows significant improvements in the effectiveness of the sensor and, as a consequence, on the quality of interaction.

Figure 2 shows some frames from the image processing pipeline. Frames (1a-4a) are raw images as captured from the IR camera. The hand of the user is moving from top left to bottom right. Frames (1b-4b) are the output of the IR light tracking. Frames (1c-4c) are the output of IR shadow tracking. At (1a) the user has just touched the surface in an area relatively free of noise. The fingertips adhere well to the surface and the FTIR effect works perfectly as the result of IR tracking displayed in (1b) shows.

At (2a) the user is beginning to move her hand. As known, the IR light blobs tend to dim, but are still clear and trackable (2b). This is due to the fact that (i) the finger adhere less effectively to the surface while moving, and (ii) the hand is entering a noisy area. However the latter is partially counterbalanced by the IR shadow tracking (2c).

At (3a) the hand of the user is moving very fast and is within an area of high IR noise. The IR light blobs are invisible (3b), but the IR shadow appears clear and is easily tracked (3c).

Finally, at (4a) the user has completed the interaction phase and holds her hand still. Again the IR light blobs prevail on the noisy background and can be tracked with great precision (4b).

At this point, combining the two input sources (light blobs and infrared shadows) is a straightforward task; details are given in the next section (Tracking).

**IMAGE PROCESSING PIPELINE**

As known, the process of finger tracking for CV based multi-touch sensors is typically modeled as a pipeline consisting of several stages: from image acquisition to preprocessing, finger detection and tracking. All transformations are implemented by means of convolution matrices. The steps through which our implementation passes are as following.

**Smoothing**

A blur filter is applied to smooth the image removing the Gaussian noise, thus getting rid of pixel size spots (see Equation 1 and Figures 3b and 4b).

$$G(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}}$$  

**Enhancement**

A rectification filter enhances the luminosity of each pixel (see Equation 2 and Figures 3c and 4c).
\[ \text{img}(x, y) = \frac{(\text{img}(x, y))^2}{(\text{max}(\text{img}(x, y)))^2} \tag{2} \]

**Background Removal Filter**

The picture is filtered in order to find the areas of the screen on which an interaction is happening. To this purpose a \(7 \times 7\) matrix with Gaussian distribution was empirically determined. The result is thresholded in order to select relevant areas. This operation in practice finds local maxima in the captured image. However the resulting image still presents some noise and must be further processed. Note that this same filter, applied to the negative image, is used in shadow tracking (see Figures 3d and 4d).

**Opening**

An opening filter erodes spots whose size is smaller than a given value, often referred to as salt and pepper noise (see Equation 3 and Figures 3e and 4e).

\[ \text{img} \circ m = (\text{img} \ominus m) \oplus m \tag{3} \]

**Lens Distortion Removal**

The image is processed in order to compensate radial and tangential distortion due to the camera. Radial (Equation 4) and tangential (Equation 5) distortion correction require parameters \(p\) and \(k\) that can be computed by identifying distortions of images containing known regular patterns [2] (see Figure 5). Note that OpenCV provides black-box functions to this purpose.

\[
\begin{align*}
x_{\text{corrected}} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\
y_{\text{corrected}} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \tag{4}
\end{align*}
\]

\[
\begin{align*}
x_{\text{corrected}} &= x + [2p_1 y + p_2 (r^2 + 2x^2)] \\
y_{\text{corrected}} &= y + [p_1 (r^2 + 2y^2) + 2p_2 x] \tag{5}
\end{align*}
\]

**Perspective Distortion Correction**

This last stage aims at transforming between capture coordinates and display coordinates and getting rid of perspective when (as often happens) the camera is not placed perfectly perpendicular against the plane of interaction. This operation requires four points on the screen to be matched against 4 points in the capture. Usually this is performed manually (during an initial calibration phase). Such transformation is efficiently computed as an inverse mapping between triangular meshes [1].

To do so, the position of a point to be mapped from camera space to display space can be expressed in barycentric coordinates: if \(A, B\) and \(C\) are the vertices of a triangle, a point \(P\) inside the triangle is uniquely identified by \(P = \lambda_1 A + \lambda_2 B + \lambda_3 C\), where \(\lambda_1 + \lambda_2 + \lambda_3 = 1\). Any deformation applied to the triangle does not change the barycentric coordinates of the point \(P\), then since the coordinates of points \(A, B\) and \(C\) in the display are known from the calibration phase it’s easy to compute the coordinates of point \(P\) on the display.

CONCLUSIONS AND FUTURE WORK

Summarizing, we have shown how the performances of FTIR based multi-touch sensors can be improved by tracking the shadows that user hands project on the screen. The value of such improvement becomes evident considering that the efficacy of shadows tracking is higher just in those conditions that are more critical for IR blob tracking.

This allows to develop a sensor based on a combination of the two strategies, that is more robust to changing lighting conditions. Additional benefits include the ability of shadow tracking to sense proximity to the surface, where blob tracking is only sensitive to finger contact.

Further development will be aimed at exploiting shadows (tracing them back to the body of the person) to discriminate user action. The ability to associate the gestures sensed to the user that executes them is a key aspect in the development of multiuser collaborative (and even more for competitive) applications.

Figure 5. Lens distortion (pincushion and barrel) must be corrected.
REFERENCES


