CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions Through a Fisheye Ring

Liwei Chan[†‡] **Yi-Ling Chen**[‡] **Chi-Hao Hsieh**[‡] **Rong-Hao Liang**[‡] **Bing-Yu Chen**[‡]
[†]Graduate School of Media Design, Keio University [‡]National Taiwan University
liwei@kmd.keio.ac.jp [‡]{yilingchenntu, r02922072, rhliang, robin}@ntu.edu.tw

ABSTRACT
This paper presents CyclopsRing, a ring-style fisheye imaging wearable device that can be worn on hand webbings to enable whole-hand and context-aware interactions. Observing from a central position of the hand through a fisheye perspective, CyclopsRing sees not only the operating hand, but also the environmental contexts that involve with the hand-based interactions. Since CyclopsRing is a finger-worn device, it also allows users to fully preserve skin feedback of the hands. This paper demonstrates a proof-of-concept device, reports the performance in hand-gesture recognition using random decision forest (RDF) method, and, upon the gesture recognizer, presents a set of interaction techniques including on-finger pinch-and-slide input, in-air pinch-and-motion input, palm-writing input, and their interactions with the environmental contexts. The experiment obtained an 84.75% recognition rate of hand gesture input from a database of seven hand gestures collected from 15 participants. To our knowledge, CyclopsRing is the first ring-wearable device that supports whole-hand and context-aware interactions.

Author Keywords
Wearable devices, ring, whole-hand interaction, palm touch input, finger touch input, hand gesture input

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Designing wearable interfaces for hand-based interactions is an intriguing challenge. Previous research has demonstrated interfaces in various wearable forms worn at various body positions, such as accessories set on heads [29], chests [3], shoulders [12] or hands [17]. The two key measures for assessing these interfaces, which should be considered simultaneously, are their forms and emerging interactions.

Ring wearables are generally considered both attractive and challenging, because of their miniature forms as well as their aim to augment the most powerful interface - human hands.

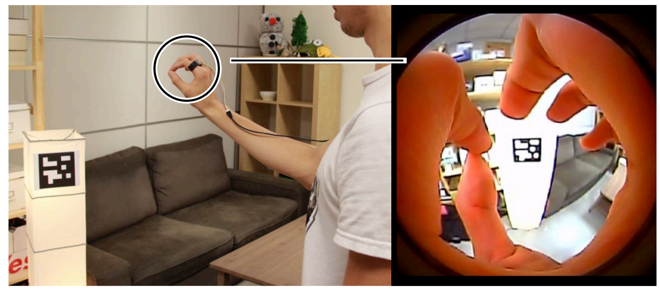

Figure 1. By supporting whole-hand and context-aware interactions, CyclopsRing allows users to interact with real-world objects such as the smart lamp by pinch-and-motion input.

Unfortunately, most existing ring wearables were exclusively designed for augmenting the wearing fingers, lowering the opportunity to take the full advantages of the hands.

CyclopsRing
This paper presents CyclopsRing, a ring-style fisheye imaging wearable device worn on the hand webbing. By observing from a central position of the hand through the hand-centric fisheye perspective, CyclopsRing can capture the whole frontal skin region of the hand. Thus, this region of the hand becomes a useful interactive surface and allows for performing hand gesture recognition. A further benefit of its extremely wide field of view is that CyclopsRing can incorporate environmental contexts into hand-based interactions. Finally, the form factor of CyclopsRing is a ring, which preserves skin haptic feedback of the hands.

Figure 1 illustrates a scenario in which CyclopsRing is used to control a smart lamp by pinch-and-motion input. The proof-of-concept prototype demonstrated the feasibility of gesture recognition using RDF-based algorithms, and a set of interactive techniques combining whole-hand and context-aware interactions. In the experiment, RDF-based pixel classification achieved a 84.75% recognition rate for hand gesture input from a database of seven hand gestures collected from 15 participants, which indicated the effectiveness of the fisheye images for rich hand-based interactions.

Contribution
The main contribution of this paper is the concept of using a fisheye hand-centric view of the hand to enable whole-hand and context-aware interactions. This work (1) developed a proof-of-concept prototype, (2) demonstrated the effectiveness of CyclopsRing for hand gesture input using random decision forest (RDF) method, and (3) presented a set of whole-hand and context-aware interactions.

UIST'15, November 08–11, 2015, Charlotte, NC, USA

DOI: http://dx.doi.org/10.1145/2807442.2807450

RELATED WORK

This paper is related to wearable interfaces for hand-based interactions, particularly those in which a finger ring is used as the interface.

Hand-Based Interactions Through Body Wearables

Because they can detect interactions at a distance, cameras can be mounted on different body locations to enable hand-based interactions. Depending on the mounting position (e.g., heads [6, 20, 29], chests [3, 11, 18, 19], shoulders [12], foot [1] and wrists [17, 24]), hand-based interactions have different benefits. For example, head-mounted cameras can potentially augment any objects under users' perspectives with interactive functions, including users' hands, but require users' visual attentions. Chest- and shoulder-mounted cameras require users moving and facing the operating hands to the camera's field-of-view. Wrist-mounted cameras are generally considered effective for hand-based interactions because they are near the hands. However, the limitation is occlusion at the angles of wrists. CyclopsRing avoids this limitation, but limits to gestures where sufficient parts of fingers are visible to the camera.

Unlike cameras, low-level sensing approaches are benefited from direct contact with the skin. The wrist is a common location for implementing such sensing techniques. Acoustic sensors [14, 21], for example, have been used to detect tapping positions on the hand. Hand gestures can be recognized by swept frequency capacitive sensing [26], forearm electromyography [25], and by measuring wrist contours [10] or pressures made to the wrist [8]. Low-level sensing techniques are beneficial for lightweight power and computational consumption but suffer from limited input capabilities (*e.g.*, discrete input), particularly when the devices are worn and designed in ring-like forms [23]. In contrast, the selected form factor for CyclopsRing is a finger ring to allow rich and continuous hand-based interactions.

Hand-Based Interaction Through Finger Wearables

Finger wearables are attractive forms, but their challenge is the limited area for implementing functions. Thereby, the inputs for most existing finger wearables are exclusively associated with the instrumented fingers. For example, iRing [23] detected tapping on the body of the ring. Using magnetic localization, Abracadabra [13] enabled around-device input with the magnet-instrumented finger, uTrack [5] enabled touch interaction on the palm, and FingerPad [4] turned fingertips into touch interface. LightRing [16] enabled the instrumented fingers to perform touch interaction on any surface. MagicFinger [15] and EyeRing [22] proposed camera-instrumented fingers to interact with environmental contexts that are in association with the finger through the camera view. In comparison, CyclopsRing demonstrates the potentials to augment the whole hand of the instrumented finger and support context-aware interactions.

HARDWARE PROTOTYPE

The main component of CyclopsRing is a miniature fisheye camera mounted on a ring capable of observing the entire frontal skin region of the user's hand. Here, we demonstrate the components of our hardware prototype, and present the variation of wearing the device at different hand webbings and the benefits of our choice to wear it in index finger.

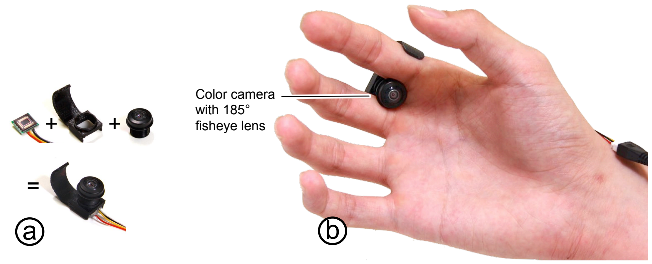

Figure 2. (a) CyclopsRing comprises a miniature 185-degree fisheye camera mounted on the edge of a 3D printed ring. (b) Wearing it on fingers can position the device at certain hand webbings.

Fisheye Ring Devices

Figure 2 shows that the hardware prototype consists of a miniature camera with a mini-fisheye lens, which has a focal length of 1.2 mm and an aperture of F1.8 to enable a 185-degree field-of-view. The camera is 14 mm in diameter and 15 mm in height. A 1/3" SONY CCD sensor was used to acquire 640×480 images at 30 fps. A special ring was fabricated to hold the fisheye camera by the edge such that the fisheye camera is positioned at a hand webbing while users wearing it as shown in Figure 2b.

The prototype acquired this unusual design to compensate for the relatively large body of the fisheye lens (i.e., 14 mm in diameter) while allowing the lens to be positioned as close as possible to a central location of the hand. The size of the fisheye lenses is expected to decrease with future advances in camera optics, which will compensate for design compromises in the prototype. For reference, the NanEye camera manufactured by AWAIBA[1], despite does not provide the required viewing angles, provides 120-degree field-of-view and measures 1.0 mm $\times$ 1.0 mm $\times$ 1.7 mm in three dimensions.

Placement of the ring device

Figure 3 displays the observed images of a user performing two gestures taken from different hand webbings by wearing the ring on the thumb, index finger, middle finger, and ring finger. Different webbings obtain considerably different views of the same hand gestures.

For the PINCH gesture (Figure 3a), for example, the views from the index, middle, and ring fingers clearly reveal the circle formed by the index finger and thumb at various distances, but the straight fingers are barely visible. In comparison, the device at the thumb cannot view the circle but can clearly view the straight fingers as if it could count the fingers for knowing the gesture types. In the GUN gesture (Figure 3b), the device at the middle and ring fingers are mostly occluded by the curling fingers. The views from the index finger and thumb are also very different.

[1] http://www.awaiba.com/product/naneye/

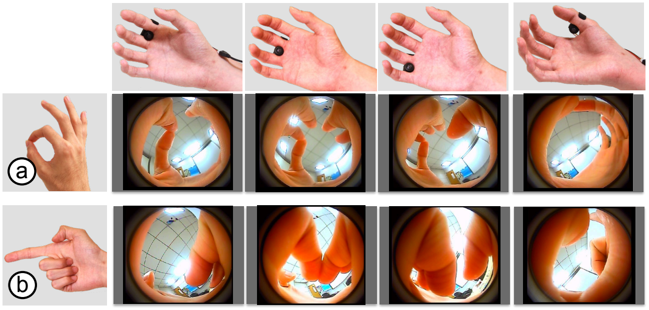

Figure 3. Placement of the ring in different hand webbings.

The current implementation still requires an excessively large area (*e.g.*, 1 square cm) to accommodate the device on the fingers. The device is positioned on the webbing of the index and middle fingers by wearing the device on the index finger (Figure 2b) because the index finger is more flexible to yield a space in the webbing for the device without preventing users from performing gestures.

RECOGNIZING HAND GESTURES

To perform gesture recognition, the images captured by the fisheye camera were first cropped by a 480×480 window to exclude some irrelevant regions. For efficiency, the images were further resized to 60×60 to extract the skin color regions. The binary mask images were then used for RDF gesture recognizer training and testing.

Foreground Extraction Using Skin Color

The foreground region was extracted by using the method described in [32] to detect skin color. The observed image is first converted into YCbCr color space, and a skin color pixel is extracted if its Cr and Cb values fall into the ranges $[145, 185]$ and $[85, 115]$, respectively. Small background noises are removed by morphological operations.

Owing to the unique placement of CyclopsRing, the real hand regions always correspond to blobs of skin colors that penetrate the peripheral boundary of the fisheye image. Specifically, the "floating" components in the central region of the image may be filtered out to obtain a clean mask of hand regions. Notably, the blobs other than hand regions may still correspond to relevant contextual objects, *e.g.*, human faces. CyclopsRing also retains these blobs for context-aware interaction. Figure 4 illustrates an example of skin color detection and the corresponding binary mask.

Random Decision Forest for Gesture Classification

The RDF is a generic data-driven learning algorithm that is widely used for computer vision [7, 27]. Recent HCI research has also used RDF methods for gesture recognition (e.g., in-air hand gestures [28] and hand gestures on and above keyboards [30]) and for full body posture recognition via a chest-mounted fisheye camera [3].

As described in [27], hand gesture recognition is formulated as a pixel classification problem by using RDF classifiers. Similar to [3, 30], the RDF classifiers are trained from intensity difference features derived from pairs of randomly generated offset vectors from the signature images. Instead of the

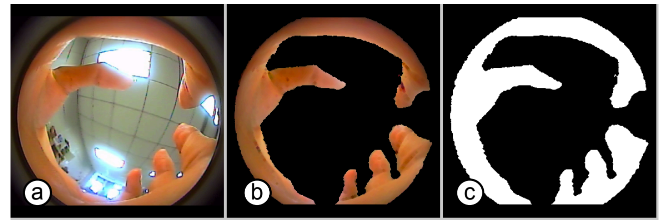

Figure 4. Foreground Extraction Using Skin Color.

motion history images (MHIs) used in previous works, this study directly employed binary images obtained by performing skin color detection in static gestural images (Figure 4c). Unlike [3], additional sensor data are not used to facilitate the recognition tasks. More details of RDFs are referred to the relevant works.

Evaluation: 84.75% Rate of Hand Gesture Input

Experimental Settings

As presented in Figure 5, the experiment included seven hand gestures, including four pinch gestures and three application-related gestures. Notably, each pinch gesture includes two tapping positions: one at the fingertip and the other at the middle of the finger.

Participants

Fifteen participants (7 females) were recruited from our department. They were aged between 23 and 27. We attempted to recruit participants with widely varying height (mean = 167cm, std = 7.6) to increase the variance in palm size. The length of their hands from the tip of the middle finger to the wrist were recorded (mean = 17.79 cm, std = 1.21 cm). All participants were right handed.

Training Data Acquisition

Data were acquired in two sessions with a 3-day interval. After helping the participants put on the device, the experimenter explained the gestures to be performed. The experimenter then asked the participants to practice making all gestures to ensure that they fully understood the details. During the study, a monitor prompted the participants with a photograph of each trial gesture. Each participant performed four rounds of the entire gesture set. Each participant performed 28 labeled gesture images (7 gestures $\times$ 4 repetitions). Finally, we recorded a short video of non-gesture hand motions by asking users to casually stretch or curl their fingers.

Performance Evaluation

A technical evaluation was performed to determine the recognition rates achieved by applying RDF classifier. Parameter settings are summarized as follows: tree number $T = 3$ and maximal depth $D = 19$; for each training image, 2000 pixels were randomly sampled as data points; 2000 candidate features and 50 candidate thresholds were evaluated for splitting an intermediate node. Leave-one-person-out cross-validation was performed in the evaluation. The RDF classifier achieves the recognition rates to nearly 84.75% for hand gesture recognition. Figure 5c shows the confusion matrices. Generally, the gesture classifier worked well for most classes. The cases

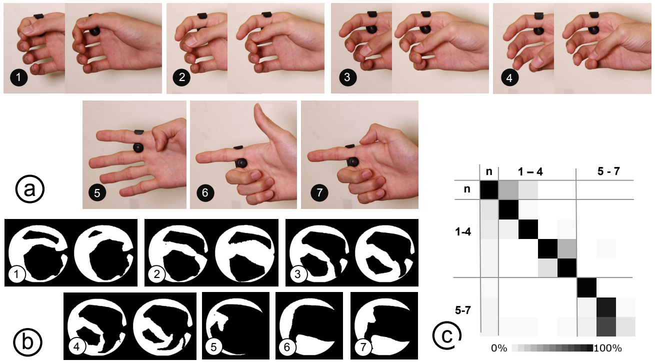

Figure 5. Experimental results of hand gesture recognition using RDF method. (a) Seven hand gestures could be recognized. (b) Examples of the skin foreground region. (c) Confusion matrix of gesture recognition.

of false recognition mainly come from the two handgun gestures, which are visually similar in their binary representations and thus difficult to be distinguished from each other.

LIMITATIONS

While the fisheye view allows a rich set of hand gestures, CyclopsRing cannot differentiate gestures in which the dissimilar parts of the gestures are beyond its field of view. For example, experiments showed that CyclopsRing has difficulty using the thumb to recognize GUN and GUNFIRE gestures.

This limitation can be resolved by wearing multiple CyclopsRings. For example, an alternative fisheye view from the thumb with a second CyclopsRing considerably complements the field of view of the first CyclopsRing. In addition, our current implementation only adopted color information. A future version of CyclopsRing with depth sensing capability will further improve recognition of natural hand interactions.

INTERACTION TECHNIQUES

The hand gesture recognition capability of CyclopsRing was further exploited by implementing other interaction techniques, including on-finger, on-palm touch interactions, and hand gesture interaction with real-world objects. Each interaction technique is realized by applying heuristics for hand-centric fisheye images.

The interaction design included a *activation/deactivation* gesture framework to enable users to switch among various interaction techniques. By detecting pre-defined hand gestures, CyclopsRing activates a specific interaction technique and either applies the corresponding heuristics or disengages from the current interaction. The deactivation gesture OPENHAND was selected because it is naturally performed when users complete an input and relax. To avoid accidental inputs, CyclopsRing switches to a new interaction technique only when it is in inactive state.

On-Finger Pinch-and-Slide Input

Users transform their fingers into functional sliders by first pinching a finger (on either fingertips or the middle of the finger) to trigger and then sliding along the specific finger. The four PINCH gestures and OPENHAND are exploited activate and deactivate the interaction, respectively.

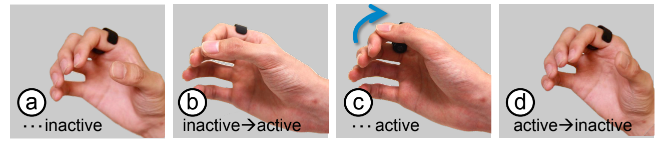

Figure 6. The state transition of on-finger pinch-and-slide input. From left, the user performs the interaction on the index finger (from inactive to active state) and then releases the pinch (return to inactive state).

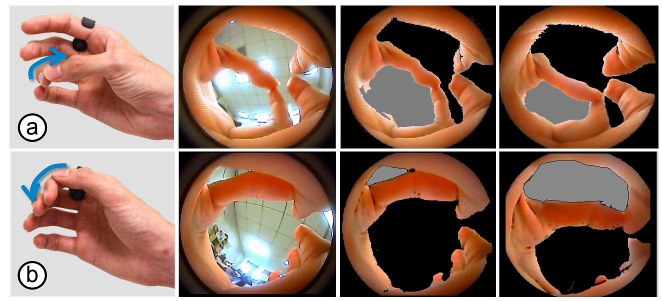

Figure 7. Interaction with multiple on-finger sliders by varying the size of the closed area formed by a pinch gesture. Effects of (a) decreasing the value of the ring-finger slider, and (b) increasing the value of the index-finger slider.

Figure 6 illustrates the state transition of the pinch-and-slide interaction performed with the index finger. When CyclopsRing is in inactive state (OPENHAND), it continuously detects the occurrence of pinch gestures. When it detects a pinch gesture, it activates the slider associated with the corresponding fingers. In this example, the interaction is triggered on the index finger as shown in Figure 6b. In active state (Figure 6c), finger-slider input is enabled by finding the relatively large enclosed component formed by the pinch gesture in the fisheye images. The variation in the component area is used to alter the slider values (Figure 7). When the user releases the pinch gesture, the slider corresponding to the index finger is deactivated. CyclopsRing then detects the next activation gesture (Figure 6d).

Palm-Writing Input

The fisheye view enables partial observation of the skin region of the palm in the flat hand posture and full observation of that in the half-curved hand posture. The palm can then be used as a touchpad for writing with fingers or pens. This interaction involves one activation gesture: the user bends his thumb to enable writing input mode.

Figure 8 illustrates the state transition of palm-writing input. Typically, the index finger of one hand is the input device used to write on the palm of the other hand. Therefore, after entering the active state, CyclopsRing takes the first captured image as the reference image and extracts the skin color region as background.

Figure 9 shows that, in active state, the finger or a color-capped pen can be used to write on the palm. We exploit the well-known Viola-Jones object detector [31] to localize the fingernail (Figure 9a-d) or color blob extraction and tracking to identify the cap of the pen (Figure 9e-h) so as to estimate the stroke trajectory. The fingernail detector was trained with 800 positive and 1600 negative training samples. False posi-

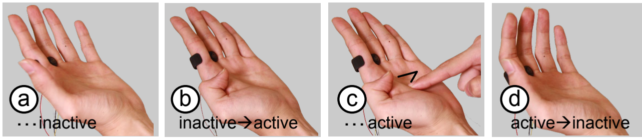

Figure 8. The state transition of palm-writing input. From left, the user bends his thumb to activate the interaction, and relaxes his thumb to end the interaction.

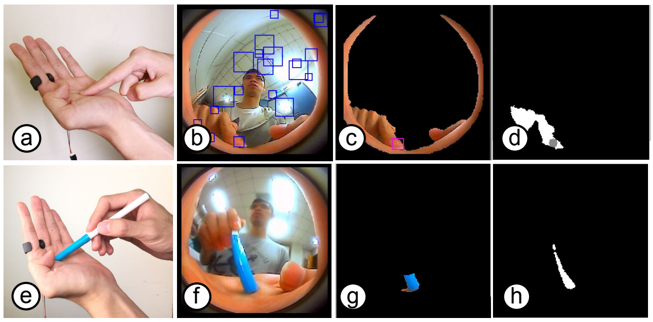

Figure 9. Palm-writing input. (a) The user writes on the palm with the index finger. (c) The user writes in different colors using color pens.

tive results were filtered out by excluding windows detected outside of the background region. Size/temporal coherence was also considered when determining the most likely results (Figure 9c). The lowest part of the foreground skin color region (Figure 9d) was then used to cue the selection of the final detection window.

The strokes entered by the user were reconstructed by exploiting the following heuristic to map the positions of the detected fingernails or color blobs to a local coordinate system (x-y) defined on the palm of the user. The X coordinates of the detected fingernails or color blobs in the fisheye image are mapped to the x coordinates by simply applying a scaling function. The y coordinates are obtained by using a mapping function, which is inversely proportional to window size or blob area. Intuitively, as the fingernail or pen cap moves closer to the camera, the perceived window size or blob area increases, which decreases the Y coordinate.

In-Air Pinch-and-Motion Input

Figure 10 shows how CyclopsRing uses finger pinch gestures for in-air pinch-and-motion input. This interaction is enabled by four activation gestures: the thumb touches the tip of one of the four fingers. Notably, pinching with different fingers increases the input modality.

Figure 11 shows that, once CyclopsRing activates the interaction, motion input is enabled by estimating hand movements from displacements of matched SURF features [2] in the consecutive undistorted fisheye images. A robust estimation of global camera motion is obtained by discarding the motion vectors of outliers that are too far from the mean motion vector in magnitude or orientation. The motion trajectory is formed by concatenating all the mean motions until the user releases the pinch gesture.

Context-aware Interaction Techniques

The fisheye camera enables CyclopsRing to perceive the environment and enables the user to interact with real world

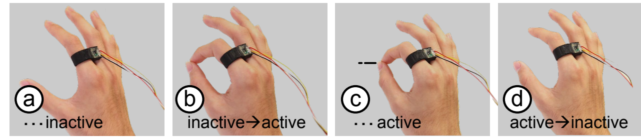

Figure 10. The state transition of in-air pinch-and-motion input. Starting from the left, the user performs pinch gesture (from inactive to active state), moves to draw an in-air stroke, and releases the pinch (return to inactive state) to complete the interaction.

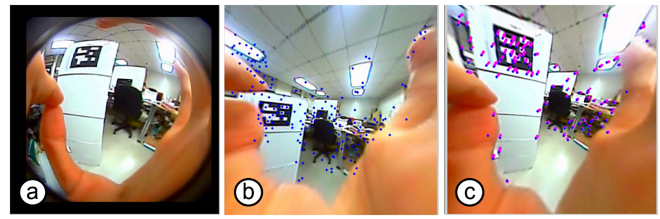

Figure 11. Interaction using in-air pinch-and-motion input. (a) The user performs the pinch gesture on the index finger. (b) SURF features are computed in the undistorted image. (c) The displacements of matched features from the last frame.

objects such as smart appliances visible to the fisheye images during interaction. The interaction with real world objects can be realized by common computer vision techniques such as object and face recognition.

Pinch-and-motion input with object recognition

Figure 12a shows how the user performs the pinch-and-motion technique described above to interact with a smart lamp. In this example, the user triggers the selection of device by performing a pinch gesture on the smart lamp. CyclopsRing then detects the device type by recognizing the AR tag [9] on the lamp, which is visible in the fisheye image. To turn on the lamp, the user performs a corresponding operation (*e.g.*, drawing a tick in the air) and then releases the pinch to issue the command. For object recognition, feature matching techniques such as SURF can be used instead of the AR tag.

Human interaction with face recognition

Combining face detection and recognition enable the use of pinch gestures to interact with nearby users. Figure 12b shows a possible scenario in which CyclopsRing sends digital files from one user to another user. When the user triggers a selection action by performing a pinch gesture, CyclopsRing detects faces in its field of view.

From a head-mounted display, the user sees an undistorted version of CyclopsRings view in which the detected faces are highlighted. The cross displayed at the center of the image helps the user to aim at a face when moving his hand with the pinch gesture. When the face of the receiving user is verified, the sending user releases the pinch gesture, and the digital content is delivered.

Finger copy functions

Another interesting capability is bridging the physical and cyber worlds by digitizing physical content. For example, Figure 13 shows how CyclopsRing can be used to copy a printed image by dragging the index finger across the image. The index finger is used to specify the region of interest and CyclopsRing perform "scan and copy" operations. This interaction

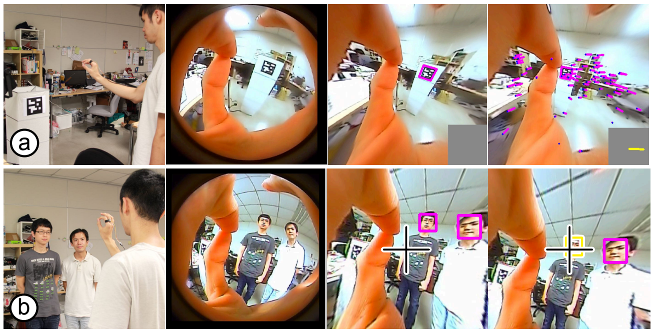

Figure 12. Additional context related functions can be incorporated with in-air pinch-and-motion input. (a) Direct interaction with a smart lamp through an AR tag. (b) Pick-and-drop interaction with a person via face detection.

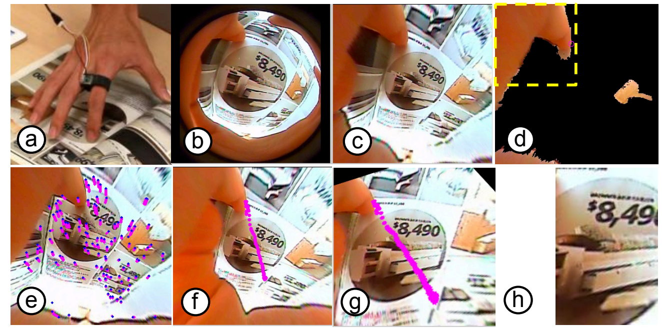

Figure 13. Finger copy of an image on the physical paper. (a) Photograph taken from third-person view. (b) The raw fisheye image. (c) Undistorted image. (d) Fingertip position. (e) Matched SURF features. (f) The finger stroke. (g) Rectified image. (h) Cropped image.

is realized by using a simple heuristic to identify the tip of the index finger. The consistent spatial relationship of fingers with respect to the fisheye camera simplifies the search for the pixel of the greatest x coordinate in the skin region of the left-upper quarter in the undistorted images (Figure 13cd).

The next step is to determine the finger stroke on the paper. For this purpose, homography transformations between every two consecutive undistorted images (Figure 13e) are computed by identifying at least four pairs of SURF features therein. In this case, a global homography is sufficient to model the projective transformation between two images since the physical object being scanned is of a planar surface. The homographies are then used to re-project all the fingertip positions onto the first undistorted image where the stroke is initiated to form the stroke trajectory (Figure 13f).

While the stroke roughly captures the diagonal line of the content-of-interest, we can not simply take the rectangle defined by the diagonal line unless the CyclopsRing perpendicularly looks at the paper. We thus rectify the image by applying a predefined homography transformation H_o, to calibrate the relative orientation between the CyclopsRing and the scanned object, which is assumed to be a planar surface (*e.g.*, a catalog in this example). Given that users perform the interaction with similar inclined angles, H_o can be estimated once by reorienting the image to an upright angle. Figure 13g shows an example of orientation rectification and Figure 13h is the corresponding cropped region by using the stroke as the diagonal of a minimal bounding box.

EXAMPLE APPLICATIONS

In this section, we demonstrate several applications to showcase the capability of CyclopsRing to realize a variety of whole-hand and context-aware interactions.

Whole-hand interactions

Gestural interaction for virtual reality

Hand gestures in VR gaming allow players to perform various hand gestures that are metaphorically related functions. In the first-person shooting game in Figure 14a, the player performs a GUN gesture to switch weapon to the pistol barrel. Bending down the thumb fires the gun, pinching with the thumb and index finger retrieves a grenade. Interactivity is enhanced by augmenting CyclopsRing with a 6 DOF IMU to aim a pistol or to detect a grenade being thrown.

Single-handed interaction for smart watches

Touch interaction on smart watches is subject to the display size, yet the touch interaction requires users' both hands. On-finger pinch-and-slider input (Figure 14b) can be used to support single-handedly interaction with smart watches.

Palm-Touch Interaction for head-mounted displays

CyclopsRing transforms the palm of the hand into a remote touch pad for glass displays. Instead of pressing buttons on the glasses or relying on an external touch pad for input, the user simply writes on the palm to perform touch control of the head-mounted displays. Figure 14c shows how palm-writing input is used to navigate and to reply to messages.

Context-aware interactions

Cyber-world Clipboard

The cyber-world clipboard of CyclopsRing allows to digitize physical content. In Figure 14d, the user drags an index finger across the region of interest, *e.g.*, a photo on an actual piece of paper, the content of the specified region is digitized and directly passed to the neighboring tablet. Here, this interaction is implemented with finger-copy functions.

Pinch into the Context

For contextual interactions, CyclopsRing quickly determines the intention of the user because its camera is typically aimed at the object with which the user is interacting. Therefore, CyclopsRing enables intuitive interaction with real-world objects by pinch gestures. For example, in a smart home, users may *select* a smart device with the pinch gesture and issue a command by drawing in-air strokes (Figure 14e). Figure 14f shows another example of file sharing with other users. A pinch gesture is used to pick up digital content in the smart phone. The content is then dropped by releasing the pinch gesture, which triggers the send command.

CONCLUSION

This study presented CyclopsRing, which is the first finger ring device that supports both whole-hand and context-aware interactions. Hand gesture recognition experiments showed 84.75% accuracy in recognizing seven hand gestures across

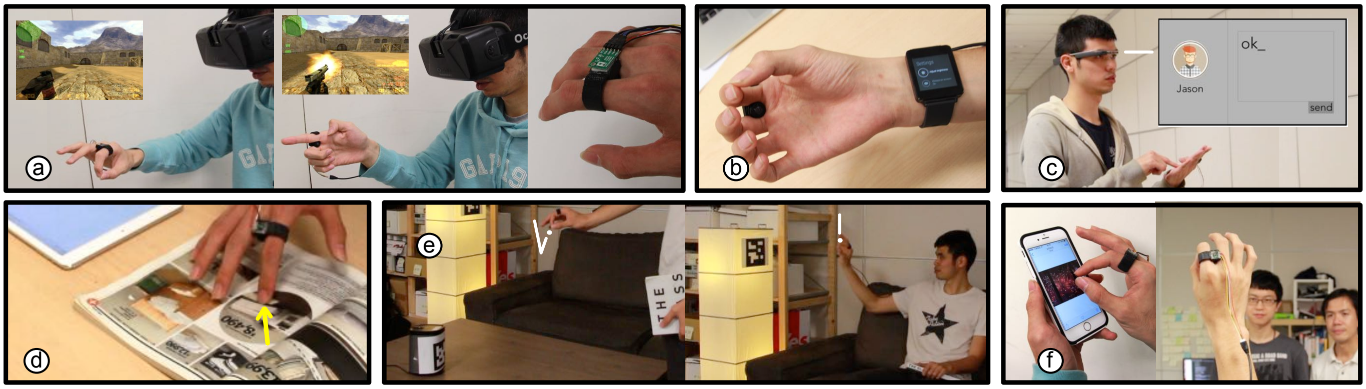

Figure 14. Examples of applications of CyclopsRing for gestural interaction in virtual reality, hand based input for smart devices and various context-aware interactions.

15 participants. Built upon gesture recognition, this work demonstrated a set of hand-based interactions including on-finger slider input, in-air pinch-and-motion input, and palm-writing input, and their interactions with the environmental context. In future works, we will investigate the use of RDF methods for continuous tracking of hand skeletons. Additional objectives include using micro lens technology to miniaturize CyclopsRing and increasing the field of view by accommodating multiple miniaturized wide-angle lens in the design for a single finger ring.

ACKNOWLEDGMENTS

This work was supported in part by Ministry of Science and Technology, Taiwan, National Taiwan University and Intel Corporation under Grants MOST103-2911-I-002-001, MOST103-2218-E-002-014, MOST103-2218-E-002-024-MY3, NTU-ICRP-104R7501, NTU-ICRP-104R7501-1, and also JST ACCEL Embodied Media Project.

REFERENCES

1. Bailly, G., Müller, J., Rohs, M., Wigdor, D., and Kratz, S. ShoeSense: A new perspective on gestural interaction and wearable applications. In *Proc. ACM CHI '12* (2012), 1239–1248.
2. Bay, H., Tuytelaars, T., and Van Gool, L. Surf: Speeded up robust features. In *Computer Vision ECCV 2006*, vol. 3951 of *Lecture Notes in Computer Science*. 2006, 404–417.
3. Chan, L., Hsieh, C.-H., Chen, Y.-L., Yang, S., Huang, D.-Y., Liang, R.-H., and Chen, B.-Y. Cyclops: Wearable and single-piece full-body gesture input devices. In *Proc. ACM CHI '15* (2015), 3001–3009.
4. Chan, L., Liang, R.-H., Tsai, M.-C., Cheng, K.-Y., Su, C.-H., Chen, M. Y., Cheng, W.-H., and Chen, B.-Y. FingerPad: Private and subtle interaction using fingertips. In *Proc. ACM UIST '13* (2013), 255–260.
5. Chen, K.-Y., Lyons, K., White, S., and Patel, S. uTrack: 3D input using two magnetic sensors. In *Proc. ACM UIST '13* (2013), 237–244.
6. Colaço, A., Kirmani, A., Yang, H. S., Gong, N.-W., Schmandt, C., and Goyal, V. K. Mime: Compact, low power 3D gesture sensing for interaction with head mounted displays. In *Proc. ACM UIST '13* (2013), 227–236.
7. Criminisi, A., and Shotton, J. *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, 2013.
8. Dementyev, A., and Paradiso, J. A. WristFlex: Low-power gesture input with wrist-worn pressure sensors. In *Proc. ACM UIST '14* (2014), 161–166.
9. Fiala, M. ARTag, a fiducial marker system using digital techniques. In *Proc. IEEE CVPR '05*, vol. 2 (June 2005), 590–596 vol. 2.
10. Fukui, R., Watanabe, M., Gyota, T., Shimosaka, M., and Sato, T. Hand shape classification with a wrist contour sensor: Development of a prototype device. In *Proc. UbiComp '11* (2011), 311–314.
11. Gustafson, S., Bierwirth, D., and Baudisch, P. Imaginary interfaces: Spatial interaction with empty hands and without visual feedback. In *Proc. ACM UIST '10* (2010), 3–12.
12. Harrison, C., Benko, H., and Wilson, A. D. OmniTouch: Wearable multitouch interaction everywhere. In *Proc. ACM UIST '11* (2011), 441–450.
13. Harrison, C., and Hudson, S. E. Abracadabra: Wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proc. ACM UIST '09* (2009), 121–124.
14. Harrison, C., Tan, D., and Morris, D. Skinput: Appropriating the body as an input surface. In *Proc. ACM CHI '10* (2010), 453–462.
15. Jing, L., Cheng, Z., Zhou, Y., Wang, J., and Huang, T. Magic ring: A self-contained gesture input device on finger. In *Proc. ACM MUM '13* (2013), 39:1–39:4.
16. Kienzle, W., and Hinckley, K. Lightring: Always-available 2D input on any surface. In *Proc. ACM UIST '14* (2014), 157–160.

17. Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I., and Olivier, P. Digits: Freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proc. ACM UIST '12* (2012), 167–176.

18. Loclair, Christian, G. S., and Baudisch, P. Pinchwatch: A wearable device for one-handed microinteractions.

19. Mistry, P., and Maes, P. SixthSense: A wearable gestural interface. In *Proc. ACM SIGGRAPH Asia '09 Sketches* (2009), 11:1–11:1.

20. Mistry, P., Maes, P., and Chang, L. WUW - wear ur world: A wearable gestural interface. In *Proc. ACM CHI EA '09* (2009), 4111–4116.

21. Mujibiya, A., Cao, X., Tan, D. S., Morris, D., Patel, S. N., and Rekimoto, J. The sound of touch: On-body touch and gesture sensing based on transdermal ultrasound propagation. In *Proc. ACM ITS '13* (2013), 189–198.

22. Nanayakkara, S., Shilkrot, R., Yeo, K. P., and Maes, P. Eyering: A finger-worn input device for seamless interactions with our surroundings. In *Proc. ACM AH '13* (2013), 13–20.

23. Ogata, M., Sugiura, Y., Osawa, H., and Imai, M. iRing: Intelligent ring using infrared reflection. In *Proc. ACM UIST '12* (2012), 131–136.

24. Prätorius, M., Valkov, D., Burgbacher, U., and Hinrichs, K. DigiTap: An eyes-free VR/AR symbolic input device. In *Proc. ACM VRST '14* (2014), 9–18.

25. Saponas, T. S., Tan, D. S., Morris, D., Balakrishnan, R., Turner, J., and Landay, J. A. Enabling always-available input with muscle-computer interfaces. In *Proc. ACM UIST '09* (2009), 167–176.

26. Sato, M., Poupyrev, I., and Harrison, C. Touché;: Enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proc. ACM CHI '12* (2012), 483–492.

27. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. Real-time human pose recognition in parts from single depth images. In *Proc. IEEE CVPR '11* (2011), 1297–1304.

28. Song, J., Sörös, G., Pece, F., Fanello, S. R., Izadi, S., Keskin, C., and Hilliges, O. In-air gestures around unmodified mobile devices. In *Proc. ACM UIST '14* (2014), 319–329.

29. Tamaki, E., Miyaki, T., and Rekimoto, J. Brainy hand: An ear-worn hand gesture interaction device. In *Proc. ACM CHI EA '09* (2009), 4255–4260.

30. Taylor, S., Keskin, C., Hilliges, O., Izadi, S., and Helmes, J. Type-hover-swipe in 96 bytes: A motion sensing mechanical keyboard. In *Proc. ACM CHI '14* (2014), 1695–1704.

31. Viola, P., and Jones, M. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE CVPR '01* (2001), 511–518.

32. Yang, X.-D., Hasan, K., Bruce, N., and Irani, P. Surround-see: Enabling peripheral vision on smartphones during active use. In *Proc. ACM UIST '13* (2013).