ON REDIRECTING DOCUMENTS WITH A MOBILE CAMERA

Qiong Liu, Don Kimber, Hanning Zhou, Patrick Chiu

FX Palo Alto Laboratory, 3400 Hillview Ave. Bldg. 4, Palo Alto, CA, U.S.A.

ABSTRACT

This paper presents a method for facilitating document redirection in a physical environment via a mobile camera. With this method, a user is able to move documents among electronic devices, post a paper document to a selected public display, or make a printout of a white board with simple point-and-capture operations. More specifically, the user can move a document from its source to a destination by capturing a source image and a destination image in a consecutive order. The system uses SIFT (Scale Invariant Feature Transform) features of captured images to identify the devices a user is pointing to, and issues corresponding commands associated with identified devices. Unlike RF/IR based remote controls, this method uses object visual features as an all time 'transmitter' for many tasks, and therefore is easy to deploy. We present experiments on identifying three public displays and a document scanner in a conference room for evaluation.

1. INTRODUCTION

In a meeting environment, people frequently redirect documents to proper locations and media formats for better visibility and usability. Tools designed for document redirection include overhead transparency projector, laptop projector, document projector, printer etc. Since these projectors are hard to move, meeting participants have to walk to a specific location for the document redirection task. This practice is not convenient for meeting participants especially during a discussion session.

This paper is about a method for document redirection in a physical environment. A document can be a physical book, printed page(s), scribbled white boards, or visual contents rendered on electronic displays. Since this method is designed to handle documents in the physical world, it has no direct relations with URL redirection or input/output redirection in operating systems. The method uses mobile cameras to identify documents and enable functions associated with the identified documents. By using this method, a user can easily layout documents in a meeting scenario for better content visibility or usability. Moreover, it is easy to add this document redirection functionality to a camera equipped cell phone or PDA (personal digital assistant).

The main contribution of this paper is an intuitive approach to redirect documents with simple point-and-capture operations. Unlike a laser pointer or IR/RF based remote control [4], this approach uses images captured by a mobile camera to guide the document redirection. With this technology, existing camera enhanced mobile devices, such cell phones and PDAs, will not need extra laser pointers, IR transmitters, mini-projectors, etc. for document redirection tasks. Powered by this technology, it will be easy for a person to control document layout with a cell phone. That can save people from the burden of moving back and forth, or tracking various remote controls in a meeting environment. Moreover, the control task does not demand fixed cameras and IR receivers in control environments. Different from controls based on LED (light emitting diode) or visual identity tags, this method does not need users to install LEDs or paste bar code labels on various objects for document redirection tasks [5, 10].

2. RELATED WORK

Overhead transparency projectors, laptop projectors, document projectors, and printers are traditional tools for document redirection. Since these traditional tools are hard to carry, they require a presenter to stay at a certain location for the document redirection task, and therefore are not very suitable for discussion sessions.

To facilitate document redirection at various locations, researchers used video to map the real world to 2D space for mouse control [6]. This technology requires fixed cameras in the meeting environment. Moreover, because of the regular mouse usage in that technology, people have to carry laptops, or PDAs with pen inputs. The regular mouse based system is acceptable for interacting with a remote location. However, it is not very natural for interacting with onsite documents.

Since physical documents exist in 3D space, it is reasonable to consider a 6 DOF (degree of freedom) input device, such as Intersense™'s precision motion trackers or Polhemus's motion trackers [3, 9], for document redirection. Many 6 DOF input devices were proposed for 3D interfaces that require a specifically designed projection system to show a cursor float in air. Moreover, a physical world model has to be constructed and calibrated for this kind of application. These requirements make it too complicated and expensive for using most 6 DOF input devices.

In the physical world, document redirection is much more constrained than that in a virtual world. These constraints enable us to redirect documents with pointing device, such as an IR based remote control or a laser pointer. A directional IR laser pointer controller can reduce user interface complications by using the physical object distribution cue. However, the requirement for the transmitter and receiver pair still limits their deployments.

After cell phones and PDAs gets more popular, products were developed for bar code reading with camera equipped cell phones [8]. However, the trouble for properly tagging bar code on various documents still limits the widely use of this technology. Additionally, if bar code labels are too small on some documents, meeting participants have to walk close to various labels for photos that initiate a redirection. This requirement is not acceptable in many meeting scenarios. Because of the limitations of existing tools, we want to explore other novel techniques for document redirection.

3. THE METHOD AND A SUPPORTING SYSTEM FOR DOCUMENT REDIRECTION

The basic idea of our method is to figure out the user's document redirection intention based on two consecutive images captured by a mobile camera. Figure 1 shows the system structure for supporting this idea. In operation, a document redirection can be initiated by taking two images and send them to a processing unit through a wireless link. After the processing unit receives these two images, it extracts SIFT features [7] from these images, identifies the object the user is pointing to based on an object and document feature repository, and issues corresponding document redirection commands. It is easy to define a set of rules between image pairs and document redirection actions. For example, we can use the following procedure and operation rules to achieve document redirection.

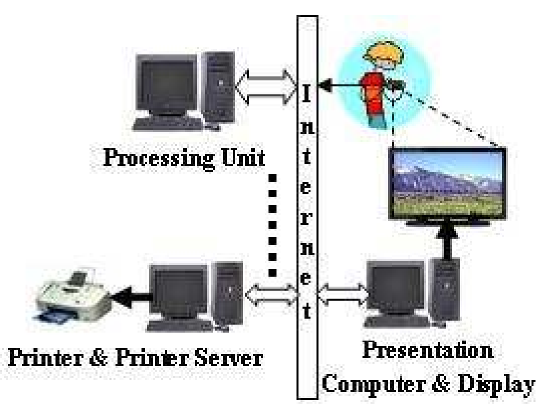

Figure 1. A simple system for document redirection

1. Identify all objects in image 1.
2. Identify the object that is closest to the center of image 1, and name it object 1.
3. Identify all objects in image 2.
4. Identify the object that is closest to the center of image 2, and name it object 2.
5. If object 1 and object 2 are different controllable devices, transfer the document shown on object 1 to object 2. For example, if object 1 is a laptop display and object 2 is a public display, the system should map the laptop display to the public display.
6. If object 1 is not controllable and object 2 has a controllable display, show the image of object 1 on the display of object 2. For example, if object 1 is a paper and object 2 is a public display, the system should post the image of the paper on the public display.
7. If object 1 is controllable and object 2 is not controllable, transfer the content of object 1 to the surface of object 2. For example, if object 1 is a video player and object 2 is a table top, the system should overlay a video saved in the player on the table top.
8. If neither object 1 nor object 2 is controllable, the system may overlay the image of object 1 on object 2 when a proper projection device is available.

The 8 step procedure is an example of one set of rules. Since our control system has a computational unit, it will be easy for users to adjust rules for their specific requirements.

There are various ways to find a screen shot in an image. Chiu et al. [1] found a screen shot in an image based on the matching of DCT coefficients. Even though this approach is not very sensitive to screen segmentation boundaries, it requires that a screen shot occupies a large area in an image, or the screen area in the image is roughly segmented. These requirements are not very proper for finding screen shots in images captured at random locations. When several screen shots are presented in the given image at the same time, this approach will be more problematic for our task unless we can do perfect slide segmentation all the time. Erol et al. [2] tried to find screen shots in an image based on OCR results and line profiles. This approach requires very high resolution image for OCR. Similarly, it is hard to deal with the scenario where several screens are presented in the same image. Moreover, it is hard to identify a screen shot that does not have text in it.

To demonstrate our idea in a practical document redirection scenario, we use the Scale Invariant Feature Transform (SIFT) [7] to find the mobile camera pointed screen in an image submitted by the camera. SIFT computes descriptive local features of an image based on histograms of edge orientation in a window around each point in the image. Each SIFT feature vector has 128 dimensions. The large dimension of this feature can greatly reduce mismatch in various scenarios. SIFT feature can also achieve reliable matching over a wider viewpoint angle. This characteristic is very important for handling images captured at various locations in a conference room. Since the SIFT feature is a local feature, it is reliable even with partial occlusion of a screen. This is also important in various scenarios.

Even though we use the SIFT feature to demonstrate our idea in this paper, it doesn't mean the SIFT feature is the only feature we should use for this document redirection method. We believe other features such as the Gabor feature can help us to achieve similar or even better result. Moreover, features like DCT coefficients, OCR results, or data captured by other sensors on the mobile device may also be considered as additional features for this task in the future.

There are four advantages for using the proposed method. First, cameras are widely installed on cell phones, PDAs, and many other mobile devices. This situation allows us to add our document redirection services on a mobile device easily. Second, this method does not require a control signal transmitter/receiver pair to achieve the control task, and therefore is easy to be deployed. Third, with this method, controls can be easily customized by associating captured photos with control actions. Moreover, it is flexible enough to adjust the DOF up to 6 according to the complexity of control tasks.

Compared with traditional remote controls, the image based controller also has several drawbacks. First, it requires a computational unit that the traditional remote control does not need. Since computational units are becoming cheaper and cheaper and many mobile devices or service centers have computational power, we believe this drawback can be handled easily. Second, if we want to use this system to control traditional devices, the system normally needs an interface unit to interact with various traditional devices. This kind of problem is common for many new designs when people want backward compatibility. So, we don't think it will be a big barrier for this new technology. Additionally, because of the computational speed limit, the system still cannot process images at a high frame rate. Therefore, this approach still cannot be used for time critical object manipulation. Since many document redirection applications, such as posting a paper on a display, are not time critical, our approach is practical for existing technology. On the other hand, we expect the overcome of this problem when more computational resources are provided. Finally, the image based remote control requires texture on object surface. For object without surface texture, we have to use some other techniques to overcome this drawback.

4. EXPERIMENTS AND EVALUATIONS

We deployed the system in a 33' by 24' conference room. Figure 2 shows a back view of the room. In Figure 2, we can see three large displays and a document camera in the room. The experiment is to test document redirection among these four devices with a 640 by 480 mobile camera.

We performed document redirection at 30 random locations uniformly distributed in the room. The documents are 25 slides we prepared for a trip report. In this experiment, around 79% document redirections were performed accurately. We traced the 21% incorrect document redirections and found that all these incorrect redirections were caused by the small image size of the document camera. It seems that the system cannot reliably recognize the document camera anymore when the document-camera's image size is smaller than 100 by 100.

Another problem with the current system is caused by the computation speed. With a 3.4GHz/1GBRam Pentium machine, the system still needs around 6~7 seconds to perform a document redirection. This delay may cause unnecessary confusion to users.

Figure 2. Back view of the conference room

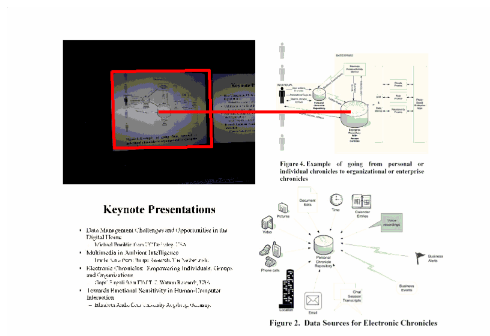

Figure 3. Find the corresponding screen in a typical image.

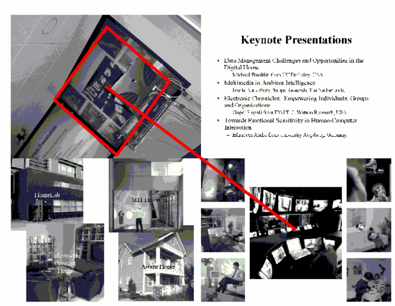

Figure 4. Find the corresponding screen when rotation and occlusion happen in the submitted image.

Image sets in Figure 3-5 illustrate some device identification results. In each set of images, the top left image is an image captured at a random location in the conference room, all other images are object or document images in the system repository. During the document redirection procedure, the current screen shots of three displays are captured by servers running on the support computer of each display and send to the processing unit. When the processing unit receives the mobile camera captured image, it automatically locates the object the user is pointing to based on available object or document images. In each image set, the red rectangle illustrates the boundary of the closest object to the center; the line segment links the bounded region to the identified object. With the formation of this association, it is easy for the processing unit to decide the ID or IP number of the focused object and perform manipulations based on that.

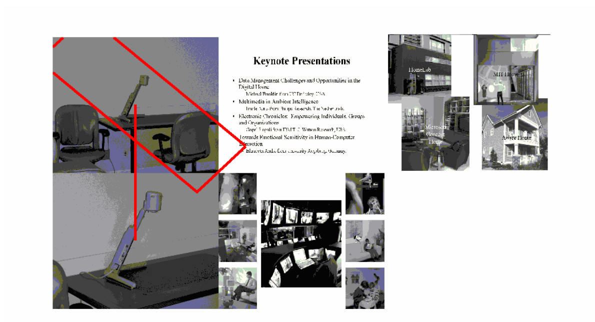

Figure 5. Find the document camera based on the submitted image

From these image sets, we can see that the boundary estimation is not very accurate for the focused display. However, the object identification procedure works reasonably well for document redirection tasks. The last two examples also show experimental results in some extreme conditions. These experiments convinced us that the system can be used for daily meetings.

5. FUTURE WORK

Beyond components for document redirection, we also need to consider UI mechanisms for giving more feedbacks to users. Moreover, users may perform some control actions inaccurately without sufficient interface guidance. To deal with these problems, we are working on four UI components to improve users' experience with this system. These components include a cross in the center of the mobile device video screen for pointing assistance, a bounding box for the identified object during a button click, an indicator on the pointed display, and an audio feedback about the interaction. We are also working on more reliable and fast feature set for object identification. Additionally, we are going to test this system on more meeting support devices.

6. REFERENCES

[1] P. Chiu, A. Kapuskar, S. Reitmeier, and L. Wilcox, "Room with a Rear View: Meeting Capture in a Multimedia Conference Room", IEEE Multimedia Magazine, pp. 48-54, vol. 7, no. 4, Oct-Dec 2000.

[2] B. Erol, and J.J. Hull, Linking Presentation Documents Using Image Analysis, Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA.

[3] Intersense Incorporated, "IS-600 Mark 2 Precision Motion Tracker" http://www.isense.com/products/prec/is600/index.htm

[4] C. Kirstein, and H. Müller, "Interaction with a Projection Screen Using a Camera-Tracked Laser Pointer." Proceedings of The International Conference on Multimedia Modeling. IEEE Computer Society Press, 1998.

[5] N. Kohtake, T. Iwamoto, G. Suzuki , S. Aoki, D. Maruyama, T. Kouda, K. Takashio, H. Tokuda, "u-Photo: A Snapshot-based Interaction Technique for Ubiquitous Embedded Information" Second International Conference on Pervasive Computing (PERVASIVE2004), Advances in Pervasive Computing (ISBN 3-85403-176-9) pp.389 - pp.392, Linz/Wienna Austria, 2004.

[6] C. Liao, Q. Liu, D. Kimber, P. Chiu, J. Foote, and L. Wilcox, "Shared Interactive Video for Teleconferencing." *Proc. ACM Multimedia 2003, pp. 546-55* , November 2, 2003.

[7] D.G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal on Computer Vision, vol. 60, pp. 91-110, 2004.

[8] OP3, "ShotCodes", http://www.op3.com/en/technology, 2005.

[9] Polhemus Incorporated, "ISOTRAK II" http://www.polhemus.com/isotrkds.htm

[10] R. Sharp, "Overview: New Uses for Camera Phones." (Jul. 1, 2004), http://www.deviceforge.com/articles/AT5785815397.html

[11] E. Tokunaga, H. Kimura, N. Kobayashi, and T. Nakajima, "Virtual tangible widgets: seamless universal interaction with personal sensing devices." Proceedings of the 7th international conference on Multimodal interfaces, 2005.

[12] S. Zhai, User Performance in Relation to 3D Input Device Design, *Computer Graphics 32(4)*, November 1998. pp 50-54.