Mobile Camera Supported Document Redirection

Qiong Liu, Paul McEvoy, Cheng-Jia Lai
FX Palo Alto Laboratory
3400 Hillview Avenue, Bldg. 4
Palo Alto, CA 94304
1-650-813-6957
liu@fxpal.com

ABSTRACT
In this demonstration, we are going to illustrate how to use a mobile camera to redirect documents to various devices connected to the same network.

Categories and Subject Descriptors
H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities; H.5.2 [User Interfaces]: Interaction styles, User-centered design, Prototyping; H.5.3 [Group and Organization Interfaces]: Computer-supported cooperative work.

General Terms
Management, Documentation, Design, Experimentation.

Keywords
Mobile camera supported document redirection, device control.

1. INTRODUCTION
This demo is about a system for document redirection in a physical environment. In a meeting environment, people frequently transfer documents to proper locations and media formats for better visibility and usability. We call this document location and format change document redirection. Tools designed for document redirection include overhead transparency projectors, laptop projectors, document projectors, printers etc. Since these devices 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.

The proposed system uses a mobile camera to achieve document drag-and-drop in a physical environment. It uses an image of the document source to locate the starting point of the drag-and-drop and uses an image of the destination to locate the end point of the drag-and-drop. With this system, a user can easily change the location and format of a document 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).

MM'06, October 23–27, 2006, Santa Barbara, California, USA.
ACM 1-59593-447-2/06/0010.

The main contribution of this paper is an intuitive approach to redirecting documents with simple point-and-capture operations. Unlike a laser pointer or IR/RF based remote control [3], this approach uses images captured by a mobile camera to guide the document redirection. With this technology, existing camera enhanced mobile devices, such as 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 display 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 require LEDs or bar code labels on various objects for document redirection tasks [4, 6].

2. DEMONSTRATION

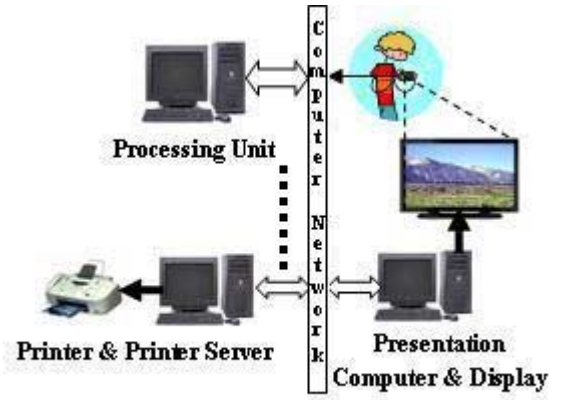

Figure 1. A simple system for document redirection

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 illustrates the system architecture for supporting this idea. In this system, we assume that all controllable devices can communicate with the processing unit through a network such as the Internet, a Virtual Private Network (VPN), or an ad hoc Bluetooth network in the working environment. These controllable devices initially register at the processing unit with their network locations, e.g. their IP addresses along with the TCP port numbers, and periodically send their updated information (e.g. screen shot images) to a repository on the processing unit. During a document redirection, a user captures document source and destination images with a mobile camera (e.g. a cell phone) and sends them to the processing unit through a wireless network such as IEEE 802.11b/g, Bluetooth,

GPRS, etc. The processing unit matches these two images against updated device images in the repository using the Scale Invariant Feature Transform (SIFT) features. After the document source and destination are identified through image matches, the processing unit can start transmitting proper data to the identified device through the network.

With this image-based document source and destination identification mechanism, facility users are released from discovering unknown network addresses and identifying unfamiliar names of those controllable devices during a document redirection activity especially as those devices can be dynamically plugged in or removed from the network. Additionally, if the document source is a paper or book and the processing unit cannot find a match for the document source image, the source image will be rendered on the document destination.

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 a target 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 SIFT features [5] 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 features 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. On the other hand, the SIFT features cannot distinguish devices that have the same visual features. In that scenario, the system either needs a different cue to make the distinction, or makes a random guess.

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 to 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 degree of freedom 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.

We are going to use some physical papers and several networked computer screens to demonstrate the document redirection system. The demonstration will follow the following procedure:

1. Capture image 1 and send it to a server.
2. Identify all objects (i.e. papers and screens) in image 1 on the server.
3. Identify the object that is closest to the center of image 1, and name it object 1.
4. Capture image 2 and send it to the server.
5. Identify all objects in image 2.
6. Identify the object that is closest to the center of image 2, and name it object 2.
7. If object 1 and object 2 are different computer screens, transfer the document shown on screen 1 to screen 2.
8. If object 1 is a page of paper and object 2 is a computer display, show the image of the paper on the display.

3. 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] 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.

[4] 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.

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

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