Tangible Tabletop Interface for an Expert to Collaborate with Remote Field Workers

Takeshi Kurata[1] Takahiro Oyabu[12] Nobuchika Sakata[12] Masakatsu Kourogi[1] Hideaki Kuzuoka[2]
[1]AIST, Japan [2]University of Tsukuba, Japan
kurata@ieee.org, {ooyabu, sakata, kuzuoka}@esys.tsukuba.ac.jp, m.kourogi@aist.go.jp

Abstract

This paper presents a new Tangible TableTop (TTT) interface to support remote collaborative works between an expert and multiple field workers in direct and intuitive way. The TTT interface consists of a large touchscreen LCD as a tabletop display and small ultrasonic transmitters that can measure their 3D positions as physical tags on the display. The physical tags represent either each worker or each tool, and are respectively coupled and worked with virtual graphical objects. The TTT interface offers several remarkable features including affordances of the physical tags, Tag Gesture, and bimanual simple manipulation with tags and touchscreen.

1. Introduction

In many practical situations, as in construction work, inspection, and door-to-door repair, it often happens that a small number of the experts and specialists who have the comprehensive understanding and technical expertise have to handle a bunch of inquiries from multiple field workers. Since it is costly and time consuming to develop skilled workers and deploy them to everywhere they are needed, effective remote collaboration technology between experts and field workers is required for alleviating the numerical disproportion.

There have been a variety of efforts to provide field workers with wearable collaborative interfaces (e.g. [3, 4, 8, 11, 12, 13, 15, 18]), because they allow field workers to share views of the real world around them and what they're doing with the remote expert, and also allow them to get visual assistance from the experts. However, most prior works including our previous works have dealt with, at least explicitly, only one-to-one collaboration between an expert and a worker and have not suggested how to expand that into collaboration between an expert and multiple workers.

The goal of this research is to realize a remote collaborative system to make it possible for an expert and multiple workers to communicate with each other in a direct

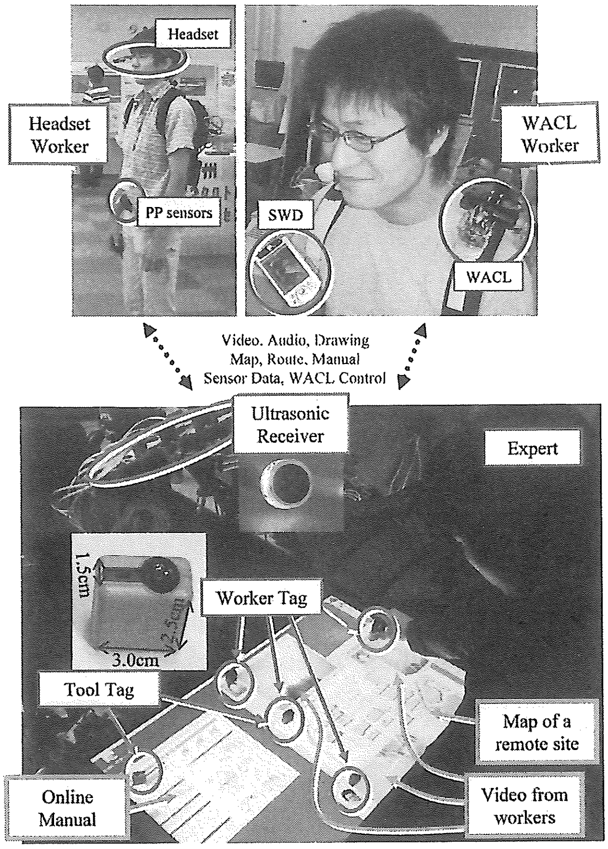

Figure 1. Multiple remote field workers equipped with wearable interfaces (upper) and a Tangible TableTop interface for the expert (lower).

and intuitive way. In contrast to field workers who have to move around the task space, the expert can stay fixed in one place while helping workers by visual/auditory assistance, and thus can employ more enhanced sensor/display devices. On the other hand, the expert has to monitor what each remote worker is doing, prepare appropriate instructions according to their circumstances, and pick up one or more workers to send the instructions promptly and correctly.

In this paper we present a new Tangible TableTop (TTT) interface to support the expert collaborating with remote workers each of who equips with wearable interface (Figure 1). The TTT interface consists of a large touchscreen LCD as a tabletop display and small ultrasonic transmitters that can measure their 3D positions as physical tags on the display. The physical tags represent either each worker or each tool such as a clipboard and map, and are respectively coupled and worked with virtual graphical objects on the display such as a window to communicate with some worker and a window to browse some online manual. The TTT interface offers several remarkable features consisting of including affordances of the physical tags [6], "Tag Gesture" recognized by measuring 3D tag trajectories, and bimanual simple manipulation with tags and touchscreen. The potential benefits of those features are to provide the expert with confidence and comfort in use, and concentration on communication and instruction.

2. Related works

In face-to-face collaboration, a wide variety of communication cues are used for establishing common ground [5], including gaze, facial expression, gesture, speech and non-speech audio. Teleconferencing systems for multiple participants such as the Hydra system [19] can effectively convey many of those cues, especially on gaze information. However, in contrast to such desktop collaboration interfaces with "talking head" video images, wearable collaborative systems are often designed to support users engaged in object manipulation tasks accompanied by moving between different workplaces. In these systems it is most important to provide tools that facilitate effective situational awareness for the remote user and allow them to enhance interaction with the user's surrounding environment.

One of the earliest works on wearable collaborative systems is Kuzuoka's Shared-View system [13] in which a filed worker is equipped with a Head Mounted Display (HMD) and Head Mounted Camera (HMC) and send images of his workspace back to a remote expert. The expert is able to use his finger to indicate regions of interest in the video and the composite image of the finger on the remote video is shown back in the HMD. In this way non-verbal cues can be transmitted in both directions between the expert and the worker. Although neither using a body-worn computer nor dealing with supporting multiple workers, this work demonstrated how an HMD could be used to enhance collaboration on a 3D spatial task.

Tangible User interfaces (TUIs) such as Tangible Bits [9] that couple physical representations with digital representations can be applied for remote collaboration for multiple participants (e.g. [2, 7]). For example, Billinghurst and Kato [2] proposed a collaborative Augmented Reality (AR) system for teleconferencing where a set of small marked user ID cards were used as tangible avatars of remote collaborators, and also another card was used as a tangible tool to share and annotate virtual contents with other participants. One of the limitations of the marker-based collaborative AR system is that the user has to wear an HMD/HMC based headset to see the remote collaborator's face and virtual contents on the cards without spatial displacement between them. However, the idea on tangible avatars is quite appropriate for our requirement in which the expert needs to be able to monitor each worker's circumstance, prepare instructions, and pick up one or more workers and send the instructions as promptly and correctly as possible without any difficulty.

Our work was also inspired by DataTiles [17]. The DataTiles system uses transparent tiles with embedded RFID tags as a modular unit of interaction. By combining a sensor-enhanced tabletop display with the tiles, input and output devices are tightly coupled and it makes the tiles act as tangible windows. In addition, the combination of tiles can activate functions such as Copy-N-Paste. Although the size of each tile (which means the size of each window) is fixed, and possible places to put tiles are discrete due to poor positioning accuracy of the RFID tag/reader as a position sensor, the idea on tangible windows and the combination can be applied for remote collaboration.

3. Wearables for field workers and tangibles for the expert

Figure 1 shows our entire remote collaboration system including a new Tangible TableTop (TTT) interface for the expert that we propose in this paper. Assuming that the system has to support field workers in a variety of situations, we chose two different kinds of wearable visual interfaces:

- Wearable Active Camera/Laser (WACL) with a Shoulder Worn Display (SWD)
- HMD/HMC based Headset

The WACL that involves wearing a steerable camera/laser head around the shoulder is a novel input/output interface device that we developed [18]. The WACL interface allows the remote expert not only to independently set their viewpoint into the wearer's task space, but also to

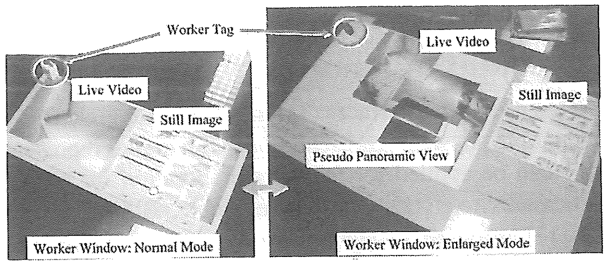

Figure 2. Worker tag and the associated window. Left: Normal mode for both a WACL worker and a headset worker, Right: Enlarged mode for a WACL worker to display the pseudo panoramic views.

point to real objects directly in the task space with the laser spot. In addition, the stabilization function based both on image registration and on a motion sensor (InterSense InterTrax2) attached to the WACL makes the direction of the camera/laser head stable on some level even if the wearer changes his/her posture. However, the visual assistance with the laser spot of the WACL is inferior to the HMD, which has the capability to represent video images.

Results of our user study to compare the head-free WACL interface to the headset interface show that the WACL was more comfortable to wear, was more eye-friendly, and caused less fatigue to the wearer, although there was no significant difference in task completion time. The results also show that experts talked more to workers wearing the WACL when detailed instructions were needed and talked more to workers wearing the headset when view changes were required. The SWD [18] is an additional display device for presenting advanced visual assistance to redress communication asymmetries in the WACL interface which give better impressions to the workers, and impose more burdens on the experts when they need to send detailed instructions.

In addition to the WACL/SWD system, the HMD/HMC based headset is used in our remote collaborative system, since it is also likely to be acceptable to many workers when the workplaces need those wearing headgears such as a helmet and goggle. Either the WACL user or the headset user can get its position and orientation and send the data to the expert by being equipped with a Personal Positioning (PP) system [10] based on data fusion of walking locomotion analysis with self-contained sensors and measurement of absolute position and orientation with a wearable camera and position reference systems such as GPS in a Kalman filtering framework.

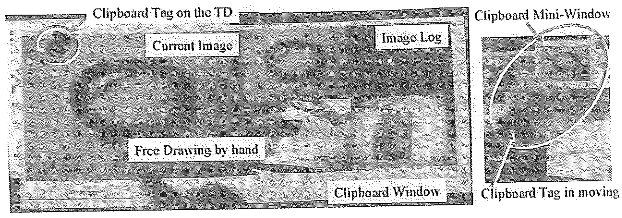

Figure 3. Tool tag and the associated window. Left: Clipboard tag and the window, Right: Every window minifies the size in moving so that the user can easily see the other windows.

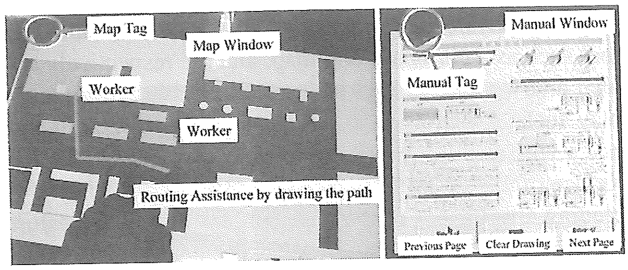

Figure 4. Tool tag and the associated window. Left: Map tag, Right: Manual tag.

4. System architecture of the TTT Interface

The TTT interface currently consists of a 40-inch touch-screen LCD as a tabletop display and small ultrasonic transmitters that can measure their 3D positions as physical tags on the display. The physical tags represent either each worker or each tool such as a clipboard and map, and are respectively coupled and worked with virtual graphical objects on the display such as a window to communicate with some worker and a window to browse some online manual.

We employ a 3D ultrasonic tagging system developed by Nishida et al. [16] as physical tags since we can easily get not only 3D position but ID of each tag and the cost is much lower than other 3D sensors such as magneto-electric sensors. The ultrasonic receiving section receives ultrasonic pulses emitted from each ultrasonic transmitter. The time-of-flight measuring section records the signal travel time from transmission to reception. The 3D position of each tag is obtained by using a random-sample-consensus (RANSAC) based multilateration. The sampling frequency of the entire system is 20 - 50 Hz and each tag shares the frequency on a time-sharing basis. By adaptively controlling the sharing rate of each tag depending on the movement, we can get the trajectory in detail and can save power consumption of tags.

For a touchscreen on the tabletop display, we chose the surface acoustic wave (SAW) technology. When the expert touches the glass surface with his/her finger or soft materials, it absorbs the energy of the acoustic wave on the surface. By comparing the changed wave to a stored reference wave, the touch location can be measured. Since the surface of physical tags is made of plastic, the tags do not affect the touchscreen even if they are put on the touchscreen.

5. Interaction techniques

Every physical tag falls into two types; Worker Tag and Tool Tag. Each worker tag represents either a WACL worker or a headset worker (Figure 2), and has the associated windows that can be used to communicate with the worker. By moving the tag position, the associated window is also moved so that the upper-left corner of the window is located at the tag's position. The expert can draw annotations on both of a live image sent from the worker and a still image captured from the live image or pasted from a clipboard described later. By using the WACL, pseudo panoramic views can be made from live images and pan/tilt angles corresponding to the each image so as to give the remote expert better situational awareness. The expert can see the panorama by switching worker windows from the normal mode to the enlarged mode with touchscreen.

We have three kinds of tool tags available at this moment; Clipboard, Map, and Manual. A clipboard tag and the associated window (Figure 3) can store multiple contents (currently images) and the expert can copy and paste the contents from/to almost all windows. As is the case with worker tags, the expert can draw annotations on the copied contents.

A map tag and the associated window (Figure 4-left) can be used to display the location of each worker equipped with the PP system, and also the expert can draw routing paths and send them to workers who need to move to the next workplace. A manual tag and the associated window (Figure 4-right) stores online manuals. The expert browses them to find the appropriate content, and send it with hand-drawn annotations on the content to the clipboard or directly to some workers.

It should be noticed that bimanual (two-handed) operation is one of great features inherent in TUIs, and a natural asymmetric bimanual operation [14] like drawing annotation with the dominant hand while grasping the tag with the non-dominant hand is induced spontaneously by this TTT interface.

5.1. Tag Gesture

Along with touchscreen operation such as pressing buttons and drawing annotations, tag gestures play an important role in interacting with workers through the TTT interface. Each tag gesture is recognized by a simple finite state

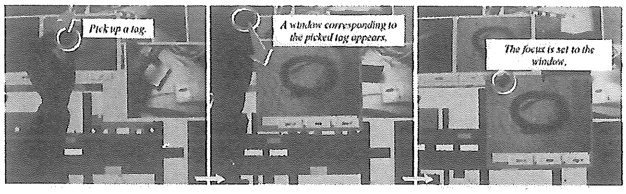

Figure 5. Pick2Focus gesture.

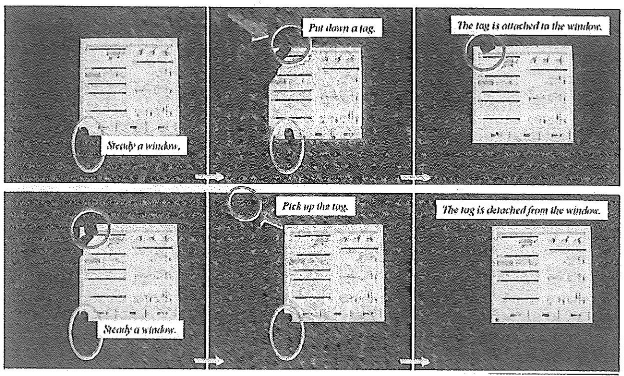

Figure 6. Gestures for Steady2Attach (upper) and Steady2Detach (lower).

machine (FSM) with the 3D trajectory of each tag. We designed tag gestures so that the expert can use them as simply and intuitively as possible.

Pick2Focus This is one of the most basic and simple tag gestures (Figure 5). Pick2Focus is used to literally pick up the window associated with the tag and set the so-called focus to it. Even if the window is completely occluded by other windows, we can easily find and display it on top just by picking up the tag.

Steady2Attach/Detach These gestures are used to associate a tag with a window or to dissociate it from the associated window (Figure 6). This is one example of the combination of using touchscreen and tag gesture. As if we steady a paper to stick something on it and to unstick something else from it, we attach tags to windows and detach tags from windows. These gestures are useful when we change windows that can be manipulated along with physical tags since the number of physical tags should be limited to 5 to 10 depending on the size of the tabletop display in spite of an unlimited number of virtual windows.

Copy-N-Paste Figure 7 shows how to copy an live image from a worker window to the clipboard. This gesture is done just by moving the tag horizontally on the display and by bring it close to a worker or manual tag within about 8 cm. As for Paste gesture, we have to once pick up the clipboard tag by about 8 cm high and put it down on the display while bring it close to a worker tag (Figure 8).

We chose more simple gesture as Copy gesture compared with Paste gesture, since the clipboard window stores

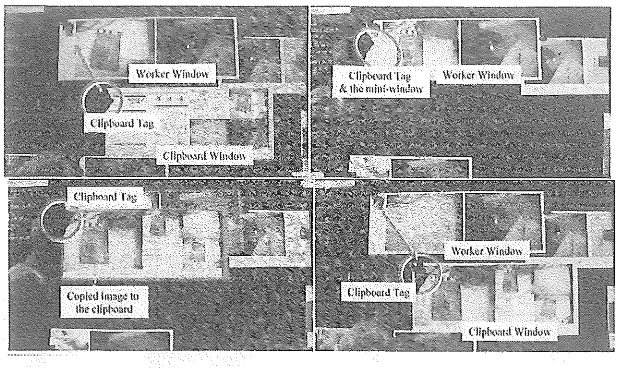

Figure 7. Copy gesture (From upper-left to lower-right).

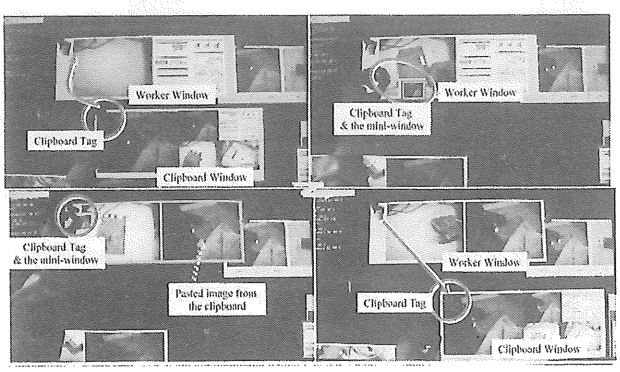

Figure 8. Paste gesture.

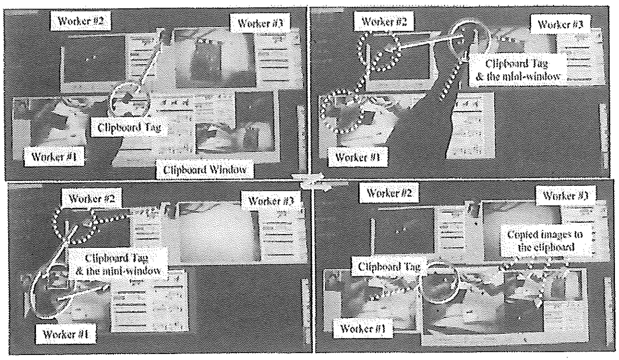

Figure 9. Sequential Copy gesture.

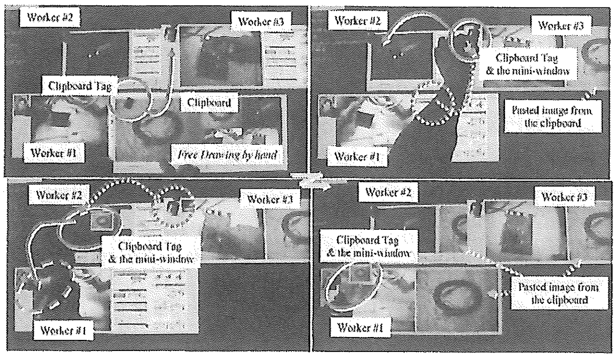

Figure 10. Sequential Paste gesture.

multiple images in the image logs and we can select one out of them even if extra images are copied by mistake. On the other hand, if the expert sends some contents such as annotated images and manuals incorrectly by pasting, it should confuse the worker. To prevent such wrong operations, Paste gesture is defined as a relatively complicated gesture (It is still simple, though). According to the same reason, we chose the same gesture as Paste gesture when sending hand-written routing assistance with map tags and when sending manuals with manual tags directly to workers.

As shown in Figure 3-right and others, every window associated with tags minifies the size in moving so that the expert can easily see the other windows. In addition, if the tag reached 8 cm high, the frame of the mini-window changes in color. Such visual feedback is necessary to indicate state transition that should make the expert have great confidence when using the system.

Sequential Copy-N-Paste and Group Copy-N-Paste Sequential Copy-N-Paste and Group Copy-N-Paste are very useful to store images of multiple places at the same time and to send same instructions to multiple workers at once. Sequential Copy and Paste (Figures 9, 10) are realized just by doing either gesture of Copy or Paste one after another. For Group Copy-N-Paste (Figures 11, 12), all worker tags that we want to copy data from or paste data to are gathered and put together in one place so that we can use Copy or Paste gesture for multiple targets in the same manner as we use it for a single tag.

6. Conclusion

This paper presented a new Tangible TableTop interface to support remote collaborative works between an expert and multiple field workers by offering remarkable features including affordances of the physical tags, Tag Gesture, and bimanual simple manipulation with tags and touchscreen. Since this work is still in very early stage, further work is necessary to identify which of these features that the TTT interface has would alleviate the complexity involved with giving instructions to multiple workers.

In the future, we plan to introduce the idea of "View Management" [1] into the tabletop display. For example, when doing Group Copy-N-Paste gesture, it is very difficult for the expert to see every window brought together at once. We will be able to prevent the windows from occluding each other by maintaining visual constraints on virtual windows associated with physical tags on the display.

ACKNOWLEDGEMENTS : This work is supported in part by Special Coordination Funds for Promoting Science and Technology from MEXT.

References

[1] B. Bell, S. Feiner, and T. Hullerer. View Management for virtual and augmented reality. In *Proc. the 4th Annual ACM Symposium on User Interface Software and Technology (UIST 2001)*, pages 101–110, 2001.

[2] M. Billinghurst and H. Kato. Out and About: Real world teleconferencing. *BT Technology Journal*, 18(1):80–82, 2000.

[3] BT Development. Camnet videotape. Suffolk, Great Britain, 1993.

[4] L. Cheng and J. Robinson. Dealing with speed and robustness issues for video-based registration on a wearable computing platform. In *Proc. ISWC98*, pages 84–91, 1998.

[5] H. H. Clark and S. E. Brennan. *Perspectives on socially shared cognition*, chapter Grounding in communication, pages 127–149. APA Books, Washington, DC, 1991.

[6] J. J. Gibson. *The ecological approach to visual perception*. Erlbaum, NJ, 1979.

[7] J. Hauber, M. Billinghurst, and H. Regenbrecht. Tangible teleconferencing. In *Proc. the Sixth Asia Pacific Conference on Human Computer Interaction (APCHI 2004)*, pages 143–152, 2004.

[8] B. Hestnes, S. Heiestad, P. Brooks, and L. Drageset. Real situations of wearable computers used for video conferencing - and implications for terminal and network design. In *Proc. ISWC2001*, pages 85–93, 2001.

[9] H. Ishii and B. Ullmer. Tangible Bits: Towards seamless interfaces between people, bits and atoms. In *Proc. ACM CHI'97*, pages 234–241, 1997.

[10] M. Kourogi and T. Kurata. Personal positioning based on walking locomotion analysis with self-contained sensors and a wearable camera. In *Proc. ISMAR03*, pages 107–114, 2003.

[11] T. Kurata, N. Sakata, M. Kourogi, H. Kuzuoka, and M. Billinghurst. The advantages and limitations of a wearable active camera/laser in remote collaboration. In *Conference Supplement (Interactive Poster) of the ACM Conference on Computer Supported Cooperative Work (CSCW 2004)*, 2004.

[12] T. Kurata, N. Sakata, M. Kourogi, H. Kuzuoka, and M. Billinghurst. Remote collaboration using a shoulder-worn active camera/laser. In *Proc. 8th IEEE International Symposium on Wearable Computers (ISWC2004)*, pages 62–69, 2004.

[13] H. Kuzuoka. Spatial workpace collaboration: A sharedview video support system for remote collaboration capability. In *Proc. CHI 92*, pages 533–540, 1992.

[14] I. S. MacKenzie and Y. Guiard. The two-handed desktop interface: Are we there yet? In *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI 2001*, pages 351–352, 2001.

[15] S. Mann. Telepointer: Hands-free completely self contained wearable visual augmented reality without headwear and without any infrastructure reliance. In *Proc. ISWC00*, pages 177–178, 2000.

[16] Y. Nishida, H. Aizawa, T. Hori, N. Hoffman, T. Kanade, and M. Kakikura. 3D ultrasonic tagging system for observing human activity. In *Proc. IEEE International Conference on Intelligent Robots and Systems (IROS2003)*, pages 785–701, 2003.

[17] J. Rekimoto, B. Ullmer, and H. Oba. DataTiles: A modular platform for mixed physical and graphical interactions. In *Proc. ACM CHI2001*, pages 269–276, 2001.

[18] N. Sakata, T. Kurata, T. Kato, M. Kourogi, and H. Kuzuoka. WACL: Supporting telecommunications using wearable active camera with laser pointer. In *ISWC 2003*, pages 53–56, 2003.

[19] A. Sellen, W. Buxton, and J. Arnott. Using spatial cues to improve videoconferencing. In *Proc. ACM CHI'92*, pages 651–652, 1992.

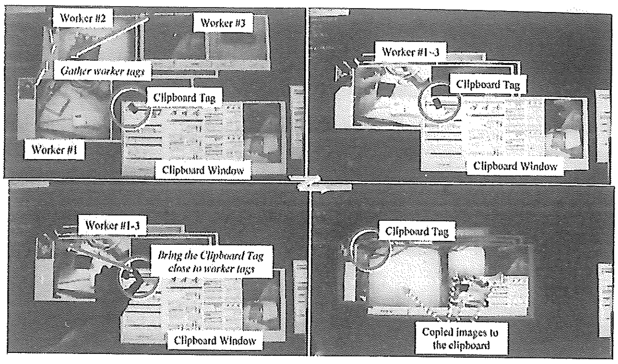

Figure 11. Group Copy gesture.

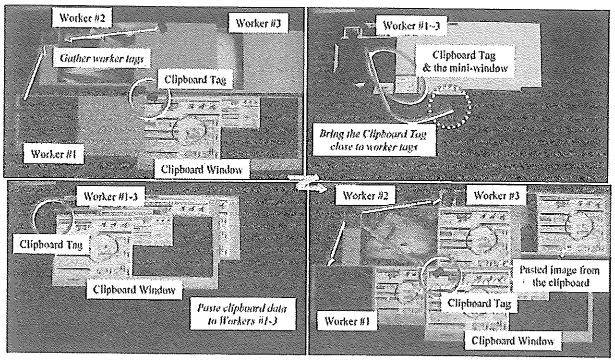

Figure 12. Group Paste gesture.