

An Intuitive Multi-Touch Surface and Gesture Based Interaction for Video Surveillance Systems

Ankith Konda, Vikas Reddy, and Prasad K. D. V. Yarlagadda

Abstract—This paper discusses the idea and demonstrates an early prototype of a novel method of interacting with security surveillance footage using natural user interfaces in place of traditional mouse and keyboard interaction. Current surveillance monitoring stations and systems provide the user with a vast array of video feeds from multiple locations on a video wall, relying on the user's ability to distinguish locations of the live feeds from experience or list based key-value pair of location and camera IDs. During an incident, this current method of interaction may cause the user to spend increased amounts time obtaining situational and location awareness, which is counter-productive. The system proposed in this paper demonstrates how a multi-touch screen and natural interaction can enable the surveillance monitoring station users to quickly identify the location of a security camera and efficiently respond to an incident.

Index Terms—Video surveillance stations, multi-touch screen, NUI, human computer interaction.

I. INTRODUCTION

In a recent report, it was identified that the global market for video surveillance systems is expected to grow by more than 80% by 2018 [1]. This is not surprising due to the improvements in computer science and the reduced costs of manufacturing and implementing such systems. As cost of video surveillance systems decreases, the adoption rate increases. Due to the affordability of high quality video surveillance systems, which are capable of identifying many features of an environment than in the recent past, has seen many organisations, big and small, embedding these systems to increase security and reduce loses. Research into video surveillance systems has thus far been focused on automation and 3D visualisations. Automation of video surveillance has been approached from various perspectives, from identifying persons, anomalies, suspicious objects, and suspicious activities [2]. Research into 3D visualisations, has focused on placing the view of a surveillance camera as a 2D plane onto 3D models [3], and more recently blending the view of a camera on the surfaces of 3D models [4].

Automation and 3D visualisations are a novel approach for the future of this market, as artificial intelligence in the future may be developed enough to remove the need for a human input. However this is not true for the near future, although there is a clear increase in the capabilities of computers, it is necessary to have a human input in the process of identifying and taking necessary steps to resolving an incident.

Manuscript received July 15, 2013; revised November 1, 2013.

The authors are with the Queensland University of Technology, India (e-mail: {a.konda, v1.reddy, y.prasad} @qut.edu.au).

The crucial points that have largely been missed within the research of video surveillance systems are the end users. Most researchers have understood the challenges the users of these systems face. However, rather than provide a tool which embeds into the current systems seamlessly, the systems and frameworks proposed have removed the end user, or have failed to understand the usability of such systems from the users perspective.

With advancements in networking technology, we see IP-based video surveillance in heavy use. With IP-based technology, it is possible for us to access the video feeds in real-time without the need for special equipment which was once required. The flexibility of utilising live feeds either over locally networked systems, or Internet-based systems allow us to create applications, which can be adopted quickly by the industry.

With the use of natural user interfaces, we have developed an early prototype of a video surveillance tool that allows the user to quickly gain situational and location awareness during an incident and allow them to plan a response based on information available to them. There are two parts to this tool.

- 1) How the user is able to locate a video surveillance camera based on generic location information passed onto them by an external entity.
- 2) How the user obtains relevant information of an incident, which they have identified on a video wall and the notification process.

The rest of the paper is organized as follows. The next section will discuss related works, which will outline Multi-touch technology, and the benefits of natural user interfaces. In Section III, we will discuss the concept of our proposed system and the build phase followed by Section IV which demonstrates the hardware and the interaction. In Section V, We discuss two scenarios in which our systems can be utilised to increase efficiency, and conclude with our thoughts on how the systems work to improve usability of video surveillance systems.

II. RELATED WORK

A. Multi-Touch

A mouse and keyboard method of interaction is one we are all accustomed to; it is the method of using a mouse and keyboard to interact with what we see on a screen. This type of interaction is widely in use today in our homes and offices. However with advancements in touch screen technology and mobile phones, we are seeing improved software and

interaction methods. This made it possible for us to adopt many devices we use today such as touch and voice enabled mobiles phones and laptops. However within the video surveillance industry, the graphical user interface with a mouse and keyboard remains the main method of interaction for video surveillance systems.

A natural user interface or NUI is an emerging method of the interaction [5], It is similar to that of a graphical user interface where the user still sees a graphical representation of real world objects, otherwise known as the desktop metaphor [6], however it removes the use of mouse and keyboard interaction and replaces this with a more natural, gesture based interaction, such as a multi-touch screen, voice and full body gesture based interaction.

With the 2006 demonstration of the multi-touch technology by Jeff Han [7], the natural user interface was revealed to the public as the future of computer interaction. The 2007 introduction of the apple iPhone which focused entirely on multi-touch interaction, set a new standard for companies such as Google, Microsoft to offer the new interaction methods in their future products, i.e. android based mobile phones, the Xbox kinect, Microsoft surface, Google glass.

With growing adoption of the natural user interface we need to focus our attention of how we can implement natural interaction method with heavily used systems such as video surveillance. In a paper from Kin [8], it was concluded that tasks on a computer were completed faster with a multi-touch screen compared to a mouse and keyboard interaction. In fact, there was an 80% increase of productivity with the use of a natural interaction than with current mouse and keyboard-based interaction.

Sulaiman *et al.* [9], have explored the use of multi-touch surface to manipulate video surveillance footage. In their experiments, they present the user with a multi-touch table as the means to interact with surveillance videos. Their users strongly agreed that natural user interfaces should be implemented to interact with video surveillance systems. The researchers have also observed an increase in efficiency of completing tasks such as moving, scaling videos on multi-touch surface as supposed to the current mouse and keyboard interaction. Although, their research failed to address the contexts in which a natural user interface can become a useful method of interaction.

III. CONCEPT

Typically the current and past research into identifying the location of a video surveillance camera, have focused on visualising the video feed onto 3D models [4], this method of visualisation is very unfamiliar to security personal as they are more used to interacting with maps and floor plans. Hence there is a gap in understanding how 2D visualization coupled with natural user interfaces can be implemented. With the integration of a multi-touch surface to interact with floor plans, we believe there will be an increase in efficiency of the user understanding situational and location awareness.

Fig. 1 demonstrates how the system presents a user with floor plans of their facility, with information of video surveillance camera locations, along with their field of view.

The user is able to rotate, scale and move floor plans within the interface, giving the user a sense of control of the system rather than the current systems in which the user has to conform.



Fig. 1. Floor plan of a facility with camera locations.

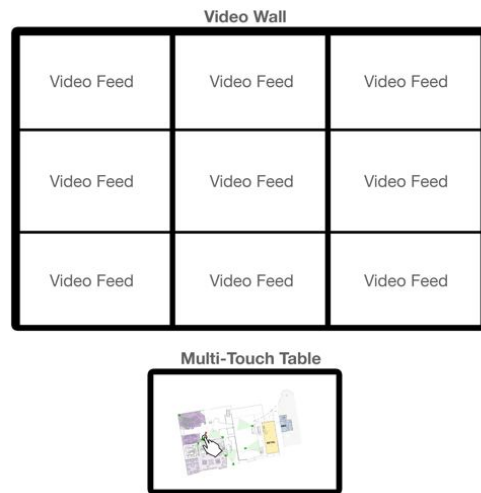


Fig. 2. Concept for the video wall.

Our system can be implemented in very much the same way the current systems are setup. In Fig. 2 we present the user a video wall of all the surveillance footage. This will give the user a sense of familiarity rather than a completely new environment such as a 3D model seen in [4]. The video wall coupled with a gesture-based interaction allows the user to initiate hand gestures to “grab” feeds and place them on their touch surface.

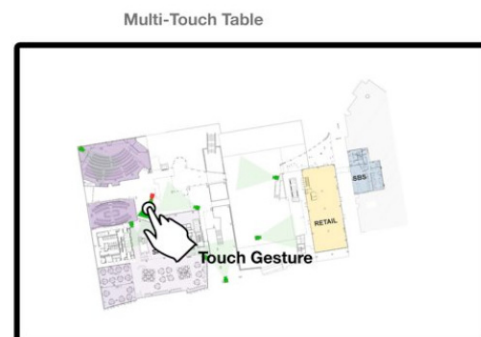


Fig. 3. User selecting a camera on the multi-touch table.

Fig. 3 shows how the user is able to select a video surveillance camera icon, which in turn shows the feed of the selected camera, rather than the current method of selecting a camera from a list.

Fig. 4 demonstrates how the user is able to peruse through

the video wall and if they detect an anomaly or an incident which requires their attention, they will be able use hand gestures to select the video feed rather than the user locating the feed based on the camera ID number.

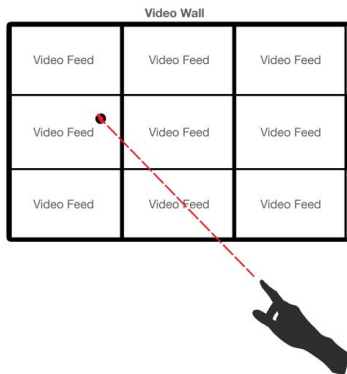


Fig. 4. User has the ability to point to a video feed of interest.

The interaction proposed above showcases how simple implementations of natural user interfaces can allow the user to be in a familiar setting at the same time improve efficiency. We have built an early prototype of this proposed system. In this paper, we demonstrate how we constructed a multi-touch table along with a demonstration of the interface and features of the early prototype software of the proposed system

IV. MULTI-TOUCH TABLE BUILD PHASE

In this section of the paper we discuss and demonstrate the prototype interface.



Fig. 5. Multi-touch table using rear diffused illumination method.



Fig. 6. Interior components of the multi-touch table.

We have developed a multi-touch table using the rear diffused surface illumination method as shows in Fig. 5. The prototype multi-touch table consists of a glass screen with a rear projection material, an ultra short throw projector, infrared led modules and a specialised camera lens which

only captures a certain wavelength of infrared light are used. The top surface is glass with a rear projection diffuser material laminated on the bottom. The projector is positioned vertically, which allows us to present the computer screen, in a horizontal position. The infrared camera is positioned at the bottom-center of the frame allowing the field of view to observe the entirety of the screen. The infrared LED modules have a wavelength value of 780nm, which only the infrared camera can see and are positioned throughout the bottom of the table frame as seen in Fig. 6.

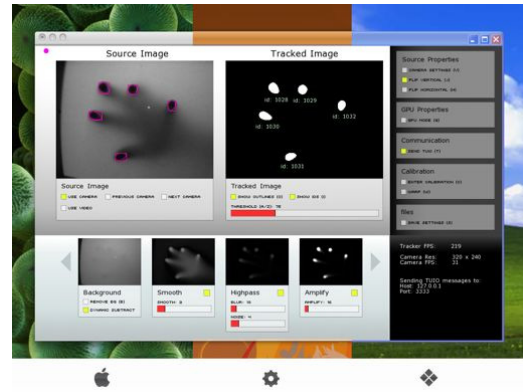


Fig. 7. Community core vision.

When the user touches the glass, the infrared light reflects down towards the camera, even though the camera can see all of the infrared light outputted by the led modules, the fingers on the surface of the glass create bright spots. Through calibration of the screen we can map the bright points to the display so that the position of the touch is accurate. Fig. 7 demonstrates how we can filter bright points, which the camera sees to bright blobs, which are recognized as touch points using the open source software called community core vision or CCV [10]. The flexibility of designing and building the multi-touch table using this method allows for a greater flexibility of how we can detect touch points and how we can translate them into customized software.

Fig. 8–Fig. 12, showcases the working prototype of the proposed system. The user is presented with a video wall which allows them to get a quick overview off the facility. The user is able to interact with floor plans of the facility and tap on the camera of interest and see the video feed on the wall. Finally if the user wishes to find the location of a particular camera on the video wall, they can point to a feed and the multi-touch table highlights the camera of interest.

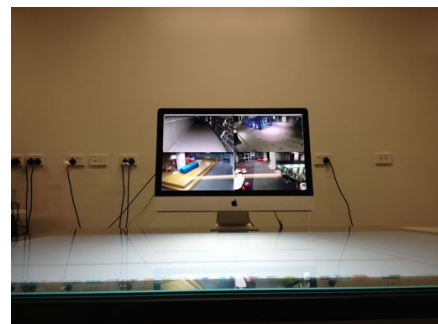


Fig. 8. Video wall showing 4 feeds from within the facility.



Fig. 9. Floor plans of the facility displayed on the multi-touch table.

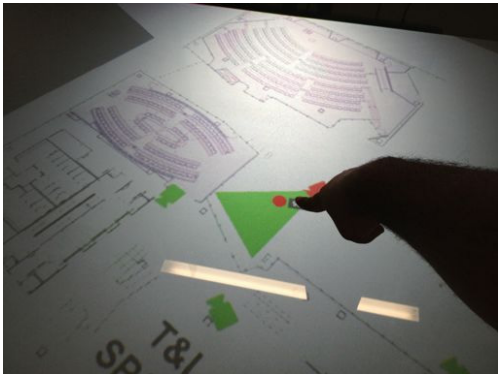


Fig. 10. User is able to touch camera icons to see video feed on the video wall.



Fig. 11. User can point to a feed on the video wall.



Fig. 12. Camera of interest (colored in red) highlights when the user points to a feed on the video wall.

V. SCENARIOS AND DISCUSSION

In this section we briefly discuss two scenarios in which our system can be utilised with efficiency.

Scenario One: During an incident in a populated facility, the user who is represented as a security personal who monitors video surveillance footage receives a message

regarding the location of the incident, for example. “Incident in progress near a Japanese restaurant”. Using the multi-touch table, the user is able to quickly open the floor plan of the facility and quickly locate the only Japanese restaurant within the facility and visualize the locations and orientations of all the security camera feeds available. The user now has the ability to select one or multiple icons and easily visualize the feeds on their video wall. With the additional feature of locating the nearest security officers, whom have locaters built into their communication devices. The user is able to select the person icon on the floor plan to contact the officers and dispatch them to the incident zone.

Scenario Two: During a similar incident, the user observes an incident on their video wall, in order for the user to respond to the incident the user needs to clearly gain situational and location awareness. The user is able to use hand gesture such as “point” or “grab” as shown in Fig. 11, to select the video feed from their video wall. The system automatically locates the camera and presents the user with a floor plan of the camera and the location of the incident and the nearest security officers. This allows the user to quickly communicate with the personal near the incident zone allowing them to respond faster.

VI. CONCLUSION

We presented a novel method of interacting with video surveillance systems to quickly identify locations of cameras, incidents and security personal. The ability to quickly respond to an incident is vital for emergency response team. However, in most surveillance systems, finding the location of cameras could be time-consuming since it requires the operator to have extensive knowledge of the locations of all the cameras analytics [11]. The proposed system addresses this critical problem.

In scenario one, we noticed how the user was only presented information that is crucial to responding to the incident. Furthermore we only presented information that the user is able to perceive from experience. We assume security personal are experienced in understanding floor plan views of a facility rather than 3d visualisations, which are rarely used. Floor plans, which identify locations of assets (cameras or security personal) required responding to an incident, allowing the users to take advantage of prior experiences and the new method of natural interaction to respond quickly to an incident.

In scenario two, the user is able to quickly identify the location of an incident which they have observed on the video wall by using natural gestures rather than identifying the location by entering in the identification number of the camera into a search feature of the system or manually locating the camera by perusing through a list of cameras and locations associated with them.

We demonstrated that the user is presented only with relevant information to remove confusion and loss in productivity. We believe further research can be undertaken to better understand how we can use natural user interfaces for video surveillance systems.

ACKNOWLEDGMENT

This research was supported by the Australian Research Council's Linkage Project "Airports of the Future" (LP0990135). The authors also acknowledge the contribution made by the many aviation industry stakeholders also involved in this project.

REFERENCES

[1] S. E. Staff. Global Video Surveillance Market To Be Worth More Than \$23B. [Online]. Available: <http://www.Securityinfowatch.com/News/10964143/Its-Projects-Global-Video-Surveillance-Market-To-Be-Worth-More-Than-23-Billion-In-2017>

[2] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size And texture," in *Proc. 2011 IEEE Computer Society Conference On Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.51-56, 2011.

[3] W. Yi *et al.*, "Contextualized videos: combining videos with environment models to support situational understanding," *IEEE Transactions On Visualization And Computer Graphics*, 2007. vol. 13, no. 6, pp. 1568-1575.

[4] C. Y. Yuan *Et Al.*, "A 3-D surveillance system using multiple integrated cameras," in *Proc. 2010 IEEE International Conference On Information And Automation (ICIA)*, pp.1930-1935, 2010.

[5] W. Y. Liu, "Natural user interface next mainstream product user interface," in *Proc. 2010 IEEE 11th International Conference On Computer-Aided Industrial Design & Conceptual Design (CAIDCD)*, pp.1-9, 2012.

[6] K. A. Olsen and R. R. Korfhage, "Desktop visualization," in *Proc. IEEE Symposium On Visual Languages*, pp. 2-4, 1994.

[7] J. Han, "Talks jeff han: Unveiling the genius of multi-touch interface design," *Ted Ideas Worth Spreading*, vol. 6, pp. 1, Aug, 2006.

[8] K. Kin, M. Agrawala, and T. Derosé, "Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation," in *Proc. Graphics Interface 2009*, Canadian Information Processing Society: Kelowna, Canada, British Columbia, pp. 119-124, 2009.

[9] S. Sulaiman *et al.*, "Manipulating Surveillance Videos Using A Multi-Touch Interface System," in *Proc. 2012 International Conference On Computer & Information Science (ICCIS)*, pp. 1036-1040, vol. 2, 2012.

[10] Community Core Vision. [Online]. Available: <http://www.Ccv.Nuigroup.Com>.

[11] S. Velastin, "CCTV Video Analytics: Recent Advances And Limitations," in *Visual Informatics: Bridging Research And Practice*, H. Badioze Zaman *et al.*, Ed. 2009, Berlin Heidelberg, Springer pp. 22-34.



Ankith Konda graduated from the University of Queensland in 2011 with a degree in multimedia design. Currently enrolled as a research master student at the Queensland University of Technology. He worked at as a research assistant and casual instructor for four courses within the field of interaction design, at the University of Queensland. Ankith Konda received an award for the "Best User-centered Design project" in his final year of undergraduate program. His work on the virtual interior design project received national attention in Australia due to the innovative idea and concept.



Vikas Reddy received his bachelor of engineering degree in electronics and communication from Visvesvaraya Technological University in 2003 and PhD from the University of Queensland in 2012. He has worked in the signal processing industry for about 5 years where he was involved in the implementation of multimedia compression standards, such as MPEG4, MP3, H263 on Texas Instrument's OMAP/DSP platforms. The IP has been licensed and shipped in numerous mobile handsets around the world including N93, N95, and N82. While doing his PhD, he also worked as a graduate researcher at Queensland Research Laboratory, NICTA. His research interests include computer vision, image understanding, intelligent video surveillance, video compression and embedded systems.



Prasad K. D. V. Yarlagadda obtained his Ph.D. from the Indian institute of technology, Mumbai. broad research interests includes but not limited to; Control systems and automation for process control, knowledge management, simulation and computational modeling of adaptive systems, security and resilience of air transportation systems, Titanium applications for biomedical scaffolding and Tissue engineering and Rapid Prototype Manufacturing and Rapid Tooling. He has worked in industry and universities for over 29 years in India, Hong Kong, Singapore, Papua New Guinea, and Australia. At Present he is Project Director, Airport of the Future project which is multi-disciplinary research project in the field of airport security, facilitation, risk and continuity planning. Prof. Yarlagadda has published more than 325 quality papers in high quality international Journals and conference proceedings. Professor Prasad KDV Yarlagadda received number of awards from various national and international agencies for his outstanding contribution to engineering field in particular to discipline of manufacturing. At present he is Editor-in-Chief GSTF Journal of Engineering Technology, Deputy Editor-In Chief of International Journal of Advances in Manufacturing and Materials Engineering and was also guest editor to number of international journals. He received significant amount of research funding from various government and industrial organizations. He recently received Fryderyk Staub Golden Owl Award from World Academy of Manufacturing and Materials, Poland, for his outstanding contribution to the discipline of materials and manufacturing engineering in the international arena.