Touchless Interaction in Surgery

Kenton O'Hara[1], Abigail Sellen[1], Gerardo Gonzalez[1,3] Tom Carrell[2], Helena Mentis[4], Graeme Penney[2], Antonio Criminisi[1], Robert Corish[1], Mark Rouncefield[3], Neville Dastur[2], Andreas Varnavas[2]

[1]Microsoft Research
Cambridge, UK

[2]Kings College London
London, UK

[3]University of Lancaster
Lancaster, UK

[4] University of Maryland
Baltimore, USA

ABSTRACT

With advances in medical imaging technologies in recent decades, we have seen their widespread adoption in the context of surgical procedures. While surgeons are increasingly reliant on these technologies, their ability to interact with them during surgery is restricted by traditional touch-based input mechanisms due to the need to maintain sterility. In response to the need to provide surgeons with control over medical images while maintaining sterility we are seeing a number of research initiatives exploring ways of interacting with these imaging technologies without touching, in particular through the use of gesture and voice control. Given the growing interest in the area, it is an opportune time to take a reflective look at the corpus of initiatives to highlight key lessons learned as well as some of the issues and challenges relevant to the development of these systems. As well as the key technical challenges to be faced, we also highlight how key socio-technical concerns play an important role in the ways we approach the design of these systems and illustrate this through some of our own development experiences in this area. In light of discussion we offer some directions for the future progress of the field.

Author Keywords

Gestural interaction, touchless interaction, space, proxemics, health, surgery, imaging.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Human Factors, Performance.

INTRODUCTION

As medical imaging technologies have advanced in recent decades, we have seen their widespread adoption for surgical procedures. A glance around any operating theatre reveals a large number of visual displays for accessing pre- and intra-operative images such as Computer Tomography (CT), Magnetic Resonance Imagery (MRI), fluoroscopy and various procedure-specific imaging applications. Such images support diagnosis and planning, and offer a virtual "line of sight" into the body during surgery. While surgeons are reliant on the capture, browsing and manipulation of these images, they are constrained by typical interaction mechanisms available, such as keyboard and mouse.

At the heart of these difficulties is the need to maintain a strict boundary between that which is sterile and that which is not. When surgeons are scrubbed up and gloved, they cannot touch these input devices without breaking asepsis. To get around this, a number of strategies are available to them for interacting with images but often these are not ideal. For example, surgeons commonly have other members of the surgical team (such as radiographers or nurses) manipulate images under their instruction [11, 14]. While this can work successfully, it is not without problems. Team members are not always available to help out resulting in frustration and time delays. Issuing instructions, while fine for relatively discrete and simple image interaction requests, can be cumbersome and time consuming. More significantly, though, such indirect manipulation is not conducive to the more analytic and interpretive tasks performed by the surgeon with the medical images. The way that images are interacted with, browsed or selectively manipulated is closely bound up with clinical knowledge and clinical interpretation. Research has shown that surgeons need direct control of image data in order to mentally "get to grips" with what is going on in a procedure [11]; something not achieved by proxy. To achieve this, some clinicians will flick their surgical gown over their hands and manipulate a mouse through the gown [11]. The rear of the gown, which is non-sterile, touches the mouse (also non sterile) while the front of the gown and the hands, which are sterile, remain separated from those surfaces (see Figure 1).

Such practices are not entirely risk free. For non-invasive procedures, these practices are considered justified by the clinical benefits they bring in terms of time saving and direct control of the images. For more invasive procedures, though, such practices are less appropriate. In these circumstances, when the surgeon needs hands-on control of the images, they must resort to the removal of gloves and rescrubbing which can be time consuming. For long procedures, where there may be multiple occasions to interact with images, this can cause significant delays to the procedure, increasing both financial costs and clinical risks.

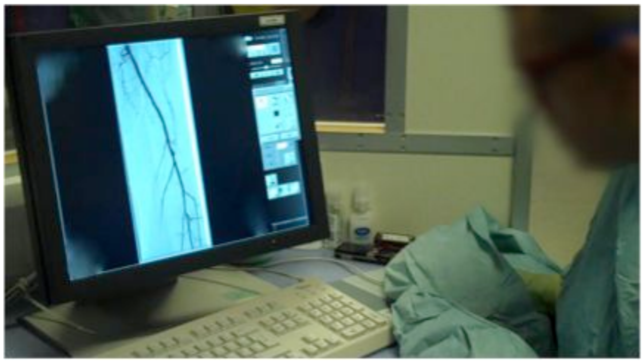


Figure 1. Using the surgical gown to avoid touching non-sterile mouse with sterile gloved hand

Providing surgeons with direct control over image manipulation and navigation while maintaining sterility within the operating theatre, then, holds itself up as an important problem to address [22]; one that has captured the imagination of research groups and commercial entities around the world. For some, the approach is to insert a barrier between the sterile gloves of the surgeon and a non-sterile interaction device, e.g. IDEO's optical mouse-in-a-bag solution [9]. While such solutions have a certain elegance in a their simplicity, their remain certain practical concerns in use at the patient bedside. In addition, barrier-based solutions have certain inherent risks owing to the potential for damage to the barrier. Other approaches then have looked to enable interaction techniques in the Operating Theatre that avoid the need for contact with an input device altogether. The seeds of this interest are in evidence in the middle part of the last decade where we start to see the use of computer vision techniques for controlling medical imaging systems by tracking the in-air gestures of the surgeon. Graetzel and colleagues [4], in one of the earliest examples of these touchless medical imaging systems, allowed the standard mouse functions such as cursor movement and clicking to be controlled by a surgeon using camera-tracked hand gestures. Shortly afterwards, we begin to see a more sophisticated use of air-based gestures for surgical imaging technology in the form of the Wach's et al's Gestix system [21]. Rather than just emulate mouse functionality, the Gestix system introduced possibilities for more bespoke gesture-based control for functionality such as navigation, zooming and rotation.

These initial systems paved an important path in this area and in the last few years we have begun to witness a significant growth in the number of systems and research efforts looking at the touchless control of medical images for surgical settings [e.g. 1, 2, 3, 6, 7, 8, 11, 14, 17, 18, 19, 20]. One of the enablers of this growth has been the emergence of the Kinect sensor and Software Development Kit [5], which has lowered barriers to entry such as financial costs, development complexity and the need to wear trackable markers. The Kinect sensor is based on a laser and a horizontally-displaced infrared camera. The laser projects a known pattern onto the scene. The depth of each point in the scene is estimated by looking at the way the pattern deforms when looked at from the IR camera. Once the scene depth has been estimated, a machine learning-based algorithm automatically interprets each pixel as belonging to the background or one of the 31 parts in which people's body has been subdivided. Finally this information is used to compute the position of the "skeleton" (a stick-man representation of the human user). Kinect has helped overcome some of the inherent challenges of full depth skeleton capture from purely camera-based systems. With this range of systems, then, we are starting to see both common themes relating to this work as well as the opportunity to explore a more diverse set of approaches to this particular problem space. The concerns now are no longer simply to demonstrate the technical feasibility of such solutions. Rather, some of the key challenges now are how best to design and implement such touchless systems to work within the particular demands and circumstances that characterise practices in the operating theatre. With this in mind, and with the growing interest in these issues, now is an opportune time to take a reflective look at the area to highlight some of the lessons learned as well as some of the issues and challenges relevant to the development of these systems. We begin with a brief review of key projects in the area to reveal some of these issues. Such a discussion, though will only take us so far. What will become apparent is that there are a deeper set of socio-technical concerns at play in understanding the design, development and use of these systems. These are the issues that our own research group is currently grappling with.

One of the most high profile of these recent systems is the work that has been deployed and actively used for different kinds of surgery at Sunnybrook Hospital in Toronto [6]. In this system, Kinect is used to navigate through a pre-defined stack of MRI or CT images. A simple and constrained gesture set is used to move either forwards or backwards through the images and to *engage* and *disengage* from the system (an important issue we will revisit later). Any transformations of the image such as rotating, or zooming or other image parameter adjustments are not available in the system unless these manipulations are bound up in the predefined image stack. There is a genuine elegance in the simplicity of the system. The limited number of gestures has benefits in terms of ease of use and system learnability. Such a constrained gesture set too can offer certain reliability benefits enabling use of reliably distinctive gestures - avoiding problems of *gesture bleed* where gestures in a vocabulary share some common kinaesthetic components leading to occasions of system misinterpretation. Given that this system is one of only a few that has been actively deployed and used, such reliability concerns are paramount in the particular design choices made here. Also of note is the adoption of *2-handed gestures* in the design of the gestural vocabulary. Such bimanual techniques can bring certain benefits as well as constrain the way that such systems can come to be used in surgical contexts – this is a key theme in the design of these systems that we will discuss later in further detail.

While there is elegance in the simplicity of the Sunnybrook system, such as approach also brings with it limitations. Interactions with medical images in surgical settings often extend beyond simple navigation possibilities, requiring a much richer set of image manipulation possibilities beyond the need to rotate/pan/zoom, to potentially include the adjustment of various image parameters such as contrast, density functions (to reveal different features such as bone, tissue or blood vessels), opacity and so on. They may even extend to the capability of marking up or annotating images during procedures. Furthermore, such manipulations may apply to whole images or more specific regions of interest defined by the clinician. With these richer possibilities in mind, several recent Kinect-based projects have developed a much larger gesture set to accommodate this increased functionality, as well as to interface with various standardised open source DICOM image viewers and PACS (Picture Archive and Communication System) systems, such as MITO and OsiriX. Notable examples here are the systems by Gallo and colleagues [3], Ebert and his colleagues [1, 2], Ruppert et al [17] and Tan et al [19]. Incorporating these richer functional sets is impressive but also brings with it certain challenges.

One such challenge concerns the notion of *expressive richness*, namely how to map an increasingly large set of functionalities (often involving the continuous adjustment of levels of a parameter) naturally onto a reliably distinctive gesture vocabulary. A number of approaches have been adopted in these systems (e.g. use of modes to distinguish gestures, different input modalities such as speech and the use of composite multi-handed gestures). Using one and two-handed tracking for example not only brings the benefits of bimanual interaction, but also enables a richer set of expressive possibilities. In both Gallo et al's and Ebert et al's systems, their gesture set employ both one and two-handed gestures. Different combinations such as single hand, 2 hands together, 2 hands apart, can then be used to denote particular image parameters that can then be adjusted according to their respective positioning in the x, y and z planes. In more recent versions of their system, Ebert et al have been able to add even further expressive capabilities by developing algorithms capable of more finger level tracking in which spread hands are distinguishable from open palmed hands.

With the larger gesture sets enabled by this expressive richness, there is an accompanying concern with the learnability of such systems [16], in particular as new system functionalities may need to be accommodated. Again we can see some attempts to deal with these issues in the systems of Tan et al [19] and Ruppert et al [17]. These systems build up compound gestures combining dominant and non-dominant hands in consistent and extendable ways. So the non-dominant hand is used for selecting particular functions or modes while the dominant hand moves within the XY and Z planes for the continuous adjustment of image parameters. In this way common gestures can be applied across a range of different functionalities making such systems both more learnable and extendable.

What emerges from this brief discussion is the use of *one and two-handed gestures* as an important theme in the design and understanding of these touchless medical systems that we pick up again later in our discussion. In particular, while the varied approaches appear to be motivated by certain control pragmatics (e.g. need for expressive richness or learnability), what is not apparent is how particular design decisions are motivated by principles of bimanual interaction design [13] or more significantly the broader set of socio-technical issues that arise when considering how these systems might be used in the actual context of a surgical procedure.

Of further interest in these touchless imaging systems is the use of gesture as the predominant means through which interaction is achieved. As we see in the work of Ebart et al [1] and in our own work [7], another intriguing possibility for touchless interaction is the use of voice recognition. Of course there are some acknowledged challenges of voice recognition software in noisy environments such as the operating theatre, and, when used in isolation, it is not really suitable for the manipulation of continuous parameters. But what is significant in the use of voice in these systems is how it is combined with the gestural modality to achieve control. So for discrete actions and functions (e.g. changing a mode or functionality) voice control may offer some important benefits.

GESTURE CONTROL IN SURGERY AS A SOCIO-TECHNICAL CONCERN

A central concern for these systems goes beyond simply developing touchless control mechanisms to overcome issues of sterility. These systems have to be situated in the context of working practices performed by the surgical team and in the setting of the operating theatre. These settings and practices shape and constrain the system design choices we make in terms of things such as the specifics of tracking, design of the gesture vocabulary and distribution of interaction across different input modalities (e.g. voice or gesture tracking). While many of the systems we have discussed have been developed in collaboration with and successfully used by clinical partners, the rationale behind the particular design choices of these systems often remains implicit with respect to the settings and work practices within which these systems need to be deployed. As this field continues to grow and develop, it is worth reflecting on these issues and making them more explicit. To do this we draw on some of our experiences in developing a system and how design choices relate to particular socio-technical concerns arising out of observations in the operating theatre. The focus on our own experiences is for illustrative purposes but our intention is to highlight issues and lessons that are of interest to the broader set of technologies we have been discussing above.

The particular system we describe is being developed for image-guided Vascular Surgery. During such procedures, the surgeon is continuously guided by live fluoroscopy and x-ray images that are presented on a bank of monitors above the patient table. On one of the monitors, a volumetric rendering of the aorta (from preoperative CT data) is overlaid on top of continuously updated x-ray images to help the surgeon visualise where the inserted wires and stents are with respect to the actual structure of the aorta. It is this combined x-ray and volumetric overlay that is manipulated using Kinect-based gesture and voice recognition (see Figure 2 and Figure 3).

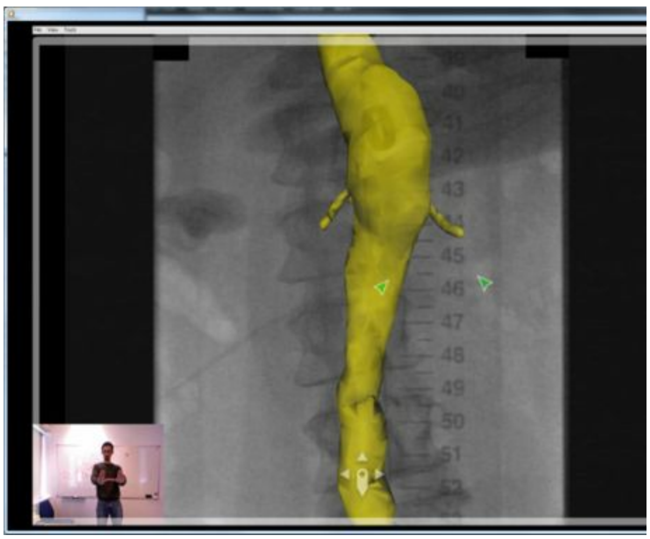


Figure 2. Gesture system for manipulating 3D overlay in vascular surgery

Figure 3. Gesture system for vascular surgery in Theatre.

In the design of our system we have oriented to some notable socio-technical concerns that have broader significance for how we think about the development of these systems. These relate to supporting more *collaboration and control* of these systems, issues of *engagement and disengagement,* and working with *one hand, two hands and hands-free*.

Supporting collaboration and control

In many of the systems discussed so far, the focus has been on providing a single point of control for the surgeon in the operating theatre. While this remains an important goal to support, there are significant collaborative aspects of imaging practices in surgery [e.g. 11, 14]. In the settings we have examined, it is not so much that more than one person wants to control images simultaneously, but rather that it is sometimes important to be able to fluidly hand over control from one person to another. For example, if the surgeon is busy with aspects of the procedure and patient management, other clinical support may need to assume control of the images. There are also times when the clinician leading the procedure may hand over certain responsibilities to another specialist or trainee. A second significant collaborative aspect of imaging practices in the operating theatre concerns collaborative clinical interpretation and discussion in which different members of the surgical team together point and gesticulate around the displayed images.

In our own system design we have attempted to address these issues by tracking the skeletons of multiple team members, using colour-coding to give each team member a distinct pair of cursors corresponding to their two hands. This allows collaborators to point and gesticulate at different parts of the image as they discuss, interpret and plan an appropriate course of action. At any point in time, a team member can raise their hand, and issue a spoken command to request control of the system, so that, as with the other systems we have discussed, there is a notion of a single dominant controller of the images. However, even when in this mode, other team members can point and gesture using visible cursors, and can assume control at any time through issuing a voice command, if required by the demands of the procedure.

Dealing with system engagement and disengagement

As we have we have alluded to in the context of collaborative discussion, gestures and actions in front of the screen are not always for the purpose of system control. As well as gesture in support of conversation, movements in front of the screen may arise through other actions being performed in the context of the procedure or may arise as the surgeon attempts to transition from one gesture to another. These actions all raise the real possibility of the system inadvertently recognising these as system control gestures. What is key in the design of these systems, then, is the need for mechanisms to move between states of system and engagement and disengagement, reinforced with appropriate feedback to signal the system state.

A number of different approaches are seen in the various systems to date that attempt to deal with these issues, each bringing their own set of pros and cons. For example, in the Sunnybrook system [6], they use a deliberately unusual gesture above the head to engage/disengage the system. Such a gesture is not likely to occur in the course of other activity and so can be considered useful in terms of avoiding inadvertent triggering. In the development of our system we have tried a number of different approaches with varying success. For example, to engage the system to recognise gestures, we initially utilised a right handed "waving" gesture to engage and disengage the system. But this suffered from the problem of *gesture transition* whereby the movement necessary to initiate the hand wave gesture was sometimes recognised as a gesture in itself. This relates to the notion of gesture *spotting* which concerns itself with the detection the start and end points of a gesture using low-level kinaesthetic features such as acceleration [12]. While techniques are improving in this area, it remains an inherently difficult challenge for the community. One way to mitigate this issue can be to employ non-classification based techniques whereby continuous image parameters simply correspond to continuous positioning of the hands. But such approaches nevertheless remain prone to the gesture transition problem for a variety of reasons, such as, for example, if parameter adjustment extends beyond the reach of natural arm movements in either plane or if particular areas of the screen are used for additional feature access. To overcome such concerns, we incorporate a clutching mechanism in which arms are withdrawn close to the body to declutch the system and allow movement transition without any corresponding image manipulation by the system.

Another approach we adopted was to use a time-based lock in which holding the hands in position for a period of time. While such an approach has be successful in other domains of gestural interaction, in evaluations with the surgeons, there was a natural tendency for the surgeon to pause and inspect the image or holding a pose to point at a specific feature of the image. These naturally occurring imaging behaviours clashed with the pause-based lock gesture. In our most recent version of the system, then, engaging and disengaging control is achieved through a simple voice command which we have found complements the gesture vocabulary and works well when some discrete change of state is needed.

Others have also explored the possibilities of automatically determining intention to engage and disengage from the system. A good example of this can be seen in the work of Mithun et al [15] which discusses contextual cues, such as gaze, hand position, head orientation and torso orientation to judge whether a surgeon is intending to perform a system readable gesture or not. Such approaches show some promise in avoiding unintentional gestures but it remains an inherently difficult challenge to fully determine human intent on the basis of such cues. For example, such contextual cues are likely to be similar when talking and gesticulating around the image during collaborative discussion as they are when actually intending to interact with the system. As such the design of explicit interaction mechanisms for engagement and disengagement remains an important concern for the field.

One hands, two hands, hands-free

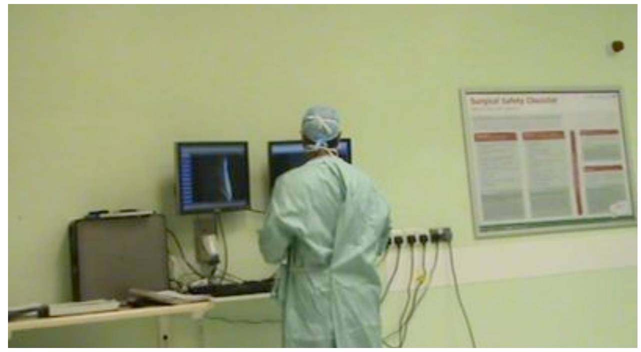
A number of the systems we have discussed have made use of both one and two-handed gestures. As well as increasing the richness of gesture vocabulary and exploiting important properties of bimanual action during interaction, there are important clinical considerations at play in the ways we design our gestural systems with one or two hands. For example sometimes image interaction is needed when the surgeon is holding various medical instruments. This raises questions as to how many hands may be available to perform certain gestural operations at particular moments. In this respect, the design of the gestural vocabulary is not simply a question of having the right number of commands to match the functionality but also determined by the clinical context of use.

In our own system, we have used a range of one and two-handed gestures. For functionalities such as panning and zooming the image, observations and interviews suggested that these manipulations are typically done at points when instruments and catheter wires can be put down. For functionality such as fading the opacity of the overlay or annotating the overlay with markers (to highlight a point of correspondence on the underlying fluoroscopy image), the surgeon may be holding onto the catheter, thereby only having one hand free. For these clinical reasons our system uses two-handed gestures for panning and zooming, but for opacity fading, this can be done with the hand that is free. For marking the overlay we combine one handed tracking with a voice command, allowing it to be carried out while holding the catheter.

What we are arguing here is not that touchless control should be available at all times when the clinicians are using other instruments. Indeed there will be many points in a procedure where image manipulation could be a distraction to the main task at hand. The point is, rather that there are some opportunities for combination to be considered here and that as a consequence, the specification of gesture vocabulary across both hands needs to be done with clinical significance. Clearly, in different kinds of surgical procedures there will be different constraints in terms of how and when image manipulation opportunities can be combined surgical instrument use. This will call for careful consideration of how to accomplish input, especially in cases where both hands may be holding instruments. In such cases it might be possible to exploit voice commands for various hands free manipulations (providing these are suitable to the discrete properties of voice commands), or combine voice with other kinds of input such as foot pedals, gaze input or head movement. The important point here then is that when designing these systems it is important to take a principled approach based not simply on the technical but also the clinical demands as to whether one, two or no hands are free for image interaction.

At the operating table, away from the operating table

The above discussion points to another important consideration around image interaction in surgical settings, namely, where the surgeon is when they need to interact with different imaging systems. For example, we can see in Figure 4, instances of surgeons interacting with images (a) away from the table and (b) at the operating table. Aside from the use of tools at the operating table, the images highlight a number of significant features. First of all, the operating table is noticeably a much more crowded environment with the surgeon often being in close proximity to other members of the surgical team. Not only can this affect our approaches to tracking but also imposes certain constraints on the kinds of movements that are available for gesture design. This is due both to the physical restrictions of working in close proximity to others as well as particular movement constraints imposed by strict sterile practices. In sterile practice, hand movements should be restricted to the area extending forwards from the torso between the hips and nipple level – shoulders upper chest, thighs and back are considered to be more risky and so movements (and thereby gestures) in this space should be avoided. Second, the operating table itself hides the lower half of the surgeon's body, while the surgeon away from the table reveals more of the whole body to the tracking system. The Kinect system (SDK v1.5 [5]) provides two tracking modes, a default mode optimised for full body skeleton tracking and a seating mode optimised for upper torso tracking (head shoulders and arms). While the full body tracking suits the kind of situation illustrated in figure 4(a) the upper torso-tracking mode is better suited for the situation at the operating table shown in figure 4(b). Thirdly, at the operating table, the surgeon's position is defined by the clinical demands of the procedure and one cannot always guarantee an ideal position in front of the gesture sensing equipment both in terms of distance and orientation. As such it can be important to account for and accommodate such variations in the design and development of gestures and tracking capabilities. The examples presented here are intended to illustrate the broader issues at play here and there may be other clinically dependent and theatre dependent configurations that we might want to consider in the development of these systems (such as the surgeon sitting down in front of a PACS system shown in figure 1).



(a)

(b)

Figure 4. Interacting with medical images: (a) away from the operating table (b) at the operating table.

CONCLUSIONS AND MOVING FORWARD

In realising the possibilities for touchless interaction, we are at a very exciting time in the development of these technologies. The numerous initiatives around the world are all testament to the remarkable progress that has been made in recent years. At the same time, this work also highlights the rich and varied ways that we might approach this problem space. Given the momentum being gathered in this space, it is an opportune time to reflect on these different approaches and some of the lessons that can be derived from the corpus of work as a whole. At this point, the goal is not simply about showing the feasibility of touchless control in clinical settings. There remain important design challenges here about the particular ways that we achieve this, from the design of the gesture vocabulary, the appropriate combination of different input modalities and the specific technical concerns of the sensing mechanisms.

Key to these design challenges is the need to attend to the broader socio-technical concerns of these settings. We have shown a number of ways that these can play out in the development of such systems, but as we move forward it will become increasingly important to understand and articulate these issues further especially as these systems become increasingly used in real world settings and across a wider range of different clinical settings. This is not a straightforward issue of asking clinicians to specify what they want in the way of a gesture vocabulary. While the participation of clinicians in the design process is essential this is not a simple question of offloading gesture design to them. As we have discussed, the design of these gestures extends beyond having sufficient expressive richness to meet the functional control needs specified by clinicians, to encompass concerns with reliability and gestural bleed. Rather, it is about understanding how the work of clinical teams is organised with respect to the demands of the procedure and particular properties of the physical setting – how is positioning and movement of the clinical team constrained by the patient, colleagues and artefact use. Further, we need to consider these systems not simply as sterile ways of performing the same imaging as before. Rather, we need to understand what it is that the clinicians are trying to achieve through their particular imaging practices and how these practices are currently shaped and constrained by particular features of the procedure with respect to sterility. By combining this with an understanding of the technical properties of these touchless systems, we can then drive the design of the gestures and systems with a view to how they enable richer forms of image interpretation and richer forms of communication and coordination among the clinical teams.

Related to this is a need for further evaluation of these systems as they are used in real world contexts. The concerns here are not so much with the basic usability of these systems but rather with how such systems change the practices of the surgical team, what do they need to do to accommodate these systems and what factors will constrain the ways that these systems are used. In addition, there are further practical concerns about the use of such systems. An important example here is the potential issue of fatigue ("gorilla arm") with the prolonged use of such systems in theatre that may both impact on the use of the system as well as other physical features of surgical practice. Understanding this and other issues arising in the context of use will be of importance in the further refinement of the design of these systems.

As the field moves forward, it is clear there are many interesting opportunities ahead. While the focus of these systems has been around overcoming the constraints of sterility in the operating theatre, there is a much broader issue of infection control within hospital settings at large. There are an enormous range of devices, systems and applications (from large displays to the increasingly pervasive use of mobile devices such as tablet computers) within such settings for which the kind of touchless interaction mechanisms discussed here might play an exciting role not just for medical professionals but for patients too. Within the operating theatre, there is also enormous potential to be had in the touchless operation of other medical equipment. We are already starting to see some interesting examples here such as the GestureNurse system [10] in which a robotic surgical assistant is controlled through gesture-based commands. With increasing use of robotics in cutting edge surgical procedures, this will present the field with many interesting challenges.

It is useful too to think about these touchless systems not simply in terms of their ability to overcome concerns of sterility, but also in terms of some of their other important properties. An interesting example here can be seen in the increasing use of 3D imaging in the operating theatre. The number of images being produced by modern day scanning technologies is becoming evermore cumbersome to interpret with the traditional slice-by-slice visualisation and review techniques. With the volumetric acquisition of scans, the data are increasingly visualised as 3D reconstructions of the relevant anatomical parts. Such 3D visualisations can be better exploited with full 3D interaction techniques. While a number of systems discussed above (including our own) already allow the manipulation of 3D anatomical models they tend to do so using the standard 2 degrees of freedom available with traditional mouse input. What is exciting about the tracking of hands and gestures in 3D space is that it opens up a much richer set of possibilities for how surgeons might be able to manipulate and interact with these 3D images with the full six degrees of freedom. Indeed, with the addition of stereoscopic visualisation of these 3D renderings we can further consider how such 3D interaction techniques will allow clinicians to perform new kinds of interactions such as reaching inside the anatomical models with which they are interacting. In addition we might consider how touchless gestural interaction mechanisms open up possibilities for interacting with things that are at a distance or out of reach. For example, interacting with images on large wall sized displays or displays that can't be reached while at the operating table. There are many new and exciting opportunities not simply in terms of interaction with traditional theatre and display setups but in terms of radical new ways for how we might conceive the entire design and layout of the operating theatres of the future.

REFERENCES

1. Ebert, L., Hatch, G., Ampanozi, G., Thali, M., and Ross, S. (2011) You Can't Touch This: Touch-free Navigation Through Radiological Images. In *Surgical Innovation, 2011*.
2. Ebert, L., Hatch, G., Ampanozi, G., Thali, M., and Ross, S. (2012) Virtopsy - Kinect Gesture Control Plugin for OsiriX. http://www.youtube.com/watch?v=wyeQi-0yC_w. 1
3. Gallo, L. Alessio Pierluigi Placitelli, A.P. Mario Ciampi, M. (2011) Controller-free exploration of medical image data: experiencing the Kinect. In *Proceedings of 24th International Symposium on Computer-Based Medical Systems (CBMS),* Bristol, UK.
4. Graetzel, C., Fong, T., Grange, S., Baur, C. (2004) A Non-Contact Mouse for Surgeon-Computer Interaction. In *Technology and Health Care,12,* IOS Press.

5. http://www.kinectforwindows.com/
6. http://sunnybrook.ca/uploads/N110314.pdf
7. http://research.microsoft.com/en-us/projects/touchlessinteractionmedical/
8. http://www.youtube.com/user/HappyFirefly0314#p/a/u/0/ j0K5iI_T8SY
9. Ionescu, A. (2006) A Mouse in the O.R. In Ambidextrous, Endless Summer, 2006, p30-32.
10. Jacob, M., Li, Y., Akingba, G., Wachs, J. P. (2012) Gestonurse: a robotic surgical nurse for handling surgical instruments in the operating room. In *Journal of Robotic Surgery, 6(1)*, p. 53 - 63.
11. Johnson, R., O'Hara, K., Sellen, A., Cousins, C., & Criminisi, A. (2011). Exploring the potential for touchless interaction in image-guided interventional radiology. In *Proceedings of the Conference on Human Factors in Computing*, Vancouver, Canada (pp. 3323-3332).
12. Kang, H., Woo Lee, C., Jung, K.: Recognition-based gesture spotting in video games. Pattern Recognition Letters 25(15), 1701–1714 (2004)
13. Leganchuk, A., Zhai, S., and Buxton, W. (1998) Manual and Cognitive Benefits of Two-Handed Input: An Experimental Study. In ACM Transactions on Computer-Human Interaction, 5 (4), 326-359.
14. Mentis, H., O'Hara, K., Sellen, A. and Trivedi, R. (2012) Interaction Proxemics and Image Use in Neurosurgery. In *Proceedings of CHI 2012*, Austin, Texas.
15. Mithun, J. Cange, C., Packer, R. and Juan P. Wachs, J.P. (2012) Intention, Context and Gesture Recognition for Sterile MRI Navigation in the Operating Room. Lecture Notes in Computer Science, 7441, p220-227.
16. Norman, D. (2010) Natural User Interfaces Are Not Natural. In *Interactions (May-June, 2010),* pp6-10.
17. Ruppert, G., Amorim, P., Moares, T., and Silva, J. (2012) Touchless Gesture User Interface for 3D Visualization using Kinect Platform and Open-Source Frameworks. In *Innovative Developments in Virtual and Physical Prototyping - Proceedings of the 5th International Conference on Advanced Research in Virtual and Rapid Prototyping*, Leiria, Portugal, 28 September - 1 October, 2011, pp 215–219.
18. Stern, H., Wachs, J. and Edan. Y. (2008). Optimal Consensus Intuitive Hand Gesture Vocabulary Design. In *Proceedings of the 2008 IEEE International Conference on Semantic Computing* (ICSC '08). IEEE Computer Society, Washington, DC, USA, 96-103.
19. Tan, J., Unal, J., Tucker, T., and Link, K. (2011) Kinect Sensor for Touch-Free Use in Virtual Medicine. http://www.youtube.com/watch?v=id7OZAbFaVI&feature=player_embedded.
20. Kipshagen, T., Graw, M., Tronnier, V.,.Bonsanto, M. and Hofmann, U. (2009) Touch- and marker-free interaction with medical software. In R. Magjarevic et al., editors, *WC '09,* pages 75–78. Springer Berlin Heidelberg.
21. Wachs, J., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M., Handler, J., (2007) Real-Time Hand Gesture Interface for Browsing Medical Images. In *IC MED, 1(2).*
22. Wachs, J., Kolsch, M., Stern, H. and Edan, Y (2011) Vision Based Hand Gesture Applications. In *Communications of the ACM, 54(2),* p60-71.