Abstract

Medical Mixed Reality helps surgeons to contextualize intraoperative data with video of the surgical scene. Nonetheless, the surgical scene and anatomical target are often occluded by surgical instruments and surgeon hands. In this paper and to our knowledge, we propose a multi-layer visualization in Medical Mixed Reality solution which subtly improves a surgeon’s visualization by making transparent the occluding objects. As an example scenario, we use an augmented reality C-arm fluoroscope device. A video image is created using a volumetric-based image synthesis technique and stereo-RGBD cameras mounted on the C-arm. From this synthesized view, the background which is occluded by the surgical instruments and surgeon hands is recovered by modifying the volumetric-based image synthesis technique. The occluding objects can therefore become transparent over the surgical scene. Experimentation with different augmented reality scenarios yield results demonstrating that the background of the surgical scenes can be recovered with accuracy between 45%-99%. In conclusion, we present a solution that a Mixed Reality solution for medicine, providing transparency to objects occluding the surgical scene. This work is also the first application of volumetric field for Diminished Reality/Mixed Reality.

Keywords: Diminished Reality, Mixed Reality, Multi-Layer, Medicine, Surgery, Visualization

1 Introduction

The term “Surgery” comes from the Greek “Kheirourgia” which means handiwork. Despite numerous technological improvements in the last centuries, surgery remains a manual work where surgeons perform complex tasks using their hands and surgical instrumentation. As it is yet not possible to retrieve the view as seen directly by the surgeon, numerous works are using video cameras to record the entire surgical scene. Such a solution is applicable for training medical students using “first-person” view cameras [1], or more commonly for Medical Augmented Reality where another modality (intraoperative or preoperative) is overlaid over the video to give context to the medical data. Having the hands and instrumentation positioned in the field of action inherently signifies the occlusion of the surgical scene and the anatomy being treated. This is true both from the surgeon viewpoint or any imaging modality viewpoint. It would be advantageous if there was a solution to display to the surgeon any occluded region of interest without losing the information about the action that is given by the hands and surgical instrument positions. Introducing transparency links the problem to the Diminished Reality field of study. Such an application would then combine Diminished Reality with Augmented Reality, providing a Mixed Reality Visualization.

1.1 Related Works

Medical Augmented Reality can be classified into 2 main categories: preoperative data (CT, MRI) overlaid over intraoperative data (video, X-ray images), or intraoperative data overlaid over intraoperative data stemming from another modality. The first category uses preoperative data to segment 3D models of organs [15, 11] or plan paths/entry points [3] that can then be rendered during surgeries using video coming from an external camera or an endoscope. The second category uses intraoperative data acquired during surgery to display over another type of intraoperative data, most of the time video. The overlaid intraoperative data can be 3D such as 3D Freehand SPECT images [17], 2D such as X-ray images [15], OCT images [20] or ultrasound [22].

The Camera Augmented Mobile C-arm by Navab et al. [15] has been the first Augmented Reality device to enter an Operating Room and has been used on over 40 patients [14]. A video camera is placed next to the C-arm source and a mirror construction fixed under the X-ray source allows the alignment of the optical axis and centers of both modalities such that an exact overlay of X-ray and video is possible. The main drawback of this work is its mirror construction, which restricts the surgical workspace available for the surgeon and requires invasive engineering on the C-arm. Habert et al. [6] proposed to augment a C-arm with 2 RGBD cameras placed on the side of the X-ray source. Using the RGBD data, the video image from the X-ray source viewpoint can be synthesized and the X-ray image can be overlaid in a similar fashion to Navab et al. [15]. A volumetric reconstruction of the scene is computed using the RGBD data from the 2 cameras, following the principle of Truncated Signed Distance Field (TSDF), used for example by Kinect Fusion [16]. Then, the image is synthesized using raytracing from the X-ray source viewpoint. Knowing that the reconstruction is volumetric and that the 2 RGBD cameras are positioned on the sides of the X-ray source, the cameras provide more information than is actually used during raytracing. Indeed, the raytracing will stop at the first voxel representing the surface (where the field is equal to zero). If, instead of stopping at this voxel, the raytracing would go further and search for the second voxel where the field is zero along the ray, a second layer could be synthesized beyond the first layer. Thus, using a depth augmented C-arm technology, this method would allow visualization of several layers. These include front and back layers, which are equivalent to any instrument and clinician hand above the patient anatomy, and the X-ray image plane respectively.

Making the front layer transparent or even disappear in order to visualize what is beyond has been studied in Diminished Reality (DR). In contrast to Augmented Reality where graphics are overlaid to a real-scene, DR withdraws or attenuates real elements from a scene. The works in DR can be divided into 3 categories according to the background recovering method: multi-viewpoint, temporal, and inpainting. The temporal methods [23] [3] [2] suppose that the camera have seen the scene without the occluder (or at another position) and use this previous information to recover...
the current occluded pixels. The inpainting methods recover the occluded part of an image with information from its non-occluded part using patch-based methods \[7\][10] or combined pixels methods \[8\]. The multi-viewpoint techniques use additional cameras that can observe the occluded background totally, or partially in order to recover it from the occluded viewpoint. Jarusirisawad and Saitoo \[9\] use perspective wrapping from the non-occluded cameras to the occluded camera to recover background pixels. More recently, using RGBD cameras, several works \[13, 21\] have generated surface mesh models of the background from one or multiple side cameras. Observing the mesh from the occluded viewpoint requires only a rigid transformation, avoiding distortions due to wrapping. Sugi-imoto et al. \[24\] use the 3D geometry to backproject to the side views the occluded pixels and therefore recover it. By design, the multi-viewpoint recovery can be used for the stereo-RGBD augmented C-arm which contains 2 RGBD cameras are placed on the side of the X-ray source viewpoint. Instead of using a mesh, the volumetric field can be used. However, no work in literature has used volumetric field such as TSDF to recover background information to the best of our knowledge. Concerning the visualization of the foreground of the front layer in combination with the back layer, the most used technique is transparency \[4, 24\]. As explained by Livingston et al. in their review of depth cues for “X-ray” vision augmented reality \[12\], transparency is indeed the most natural depth cues as it can be experienced in the real world with transparent objects.

1.2 Contribution

In this paper, we propose a mixed reality multi-layer visualization of the surgeon hands and surgical instruments using a stereo-RGBD augmented C-arm fluoroscope. This visualization consists of multiple layers which can be blended into one single view along the line of sight of the surgeon while offering different output as the blending values are chosen differently. The front layer synthesized from the X-ray source viewpoint by the stereo-RGBD augmented C-arm contains the surgeon hands and surgical instruments, the second layer is the background containing the surgical target (i.e. also synthesized by our algorithm), while the last layer is the X-ray image displaying the anatomy. As any layer can be blended to the others, our visualization proposes, for example, to display transparent hands on the background, on which the X-ray can also be blended. The blending parameters can be chosen on the fly and according to preferences or workflow steps.

In summary, this work presents the potential in positively impacting the following areas:

- User-adjustable multiple layer visualization for Medical Mixed Reality
- Improved training medical students and residents by visualizing multiple layers to better understand surgical instrument positioning and alignment, as opposed to visualizing the global scene using traditional augmented reality methods.
- First work in Medical Mixed Reality combining Diminished and Augmented Reality
- First use of volumetric field using TSDF for Diminished/Mixed Reality

2 Methodology

The setup, calibration methods, and image synthesisization used in this paper have been previously published by \[6\]. In the interest of brevity, we will not describe the calibration steps but we will thoroughly describe the synthesisization process since it is vital to our Mixed Reality multi-layer visualization contribution.

2.1 Setup

The setup comprises 2 RGBD cameras (Kinect v2) placed on the side of a X-ray source \[1\]. Each RGBD camera outputs a depth image, an infrared image and a wide-angle video image. Their fields-of-view are overlapping over the C-arm detector. Kinect v2 has been chosen because its depth information does not interfere with a similar sensor. The depth and video images are recorded using the libfreenect2 library \[5\]. The mapping from depth to video image is provided by the library.

The synchronization between images from the two cameras has been performed manually because two Kinect v2 can not used on a single standard computer and are therefore run on two separate computers. As a consequence, every sequence is recorded at a lower framerate than a standard 30fps video.

2.2 Image synthesisization

Once the system has been calibrated following the steps from \[6\], the video image from the X-ray viewpoint can be synthesized. First, the origin of the 3D world coordinate space \(\Omega_R \subset \mathbb{R}^3\) is positioned at the center of the volumetric grid, around the C-arm intensifier. Knowing the poses of the two RGBD cameras relative to the X-ray source, the projection matrices \(\Pi_1^c\) and \(\Pi_2^c\) for the 2 RGBD sensors can be computed. The notations relative to the cameras are defined as follows: optical center of the first camera \(C_1^c\), its depth image \(I_1^d\) and color image \(I_1^c\) (respectively, in the second camera \(C_2^c\), \(I_2^d\) and \(I_2^c\)).

To render the color image from the X-ray source viewpoint, a volumetric TSDF field \(f_v : \Omega_R \rightarrow [-1, 1] \) is created which maps a 3D point \(x \in \Omega_R\) to a truncated signed distance value. This value is the weighted mean of the truncated signed distance values \(v^1(x)\) and \(v^2(x)\) computed respectively in the 2 RGBD sensor cameras. Therefore, the field \(f_v\) follows Equation \(1\).

\[
f_v(x) = \frac{w^1(x)v^1(x) + w^2(x)v^2(x)}{w^1(x) + w^2(x)}
\]

where \(w^1\) and \(w^2\) are the weights for each camera. The weights are used to reject truncated signed values according to specific conditions (described in Equation \(2\)). For each camera \(i \in \{1, 2\}\), the weights \(w^i(x)\) for each truncated signed value are computed as:

\[
w^i(x) = \begin{cases} 1 & \text{if } |\Pi_i^c(x)| - ||x - C^i|| < -\eta \\ 0 & \text{else} \end{cases}
\]
where $\eta$ is a tolerance on the visibility of $x$ (we use $\eta = 6 mm$). For each view $i \in \{1, 2\}$, $v^i(x)$ represents geometrically the difference in between the distance from $x$ to the optical center of the camera $i C^i$ and the depth value obtained by projecting $x$ into camera $i$, on which a scaled truncation to the interval $[-1,1]$ is applied. The truncated signed distances $v^i(x)$ are computed according to Equation 3:

$$
v^i(x) = \phi(I^i_\delta(|\Pi^i_\delta(x)|) - |x - C^i|) \text{ with } \phi(s) = \begin{cases} \text{sgn}(s) & \text{if } |s| > \frac{\delta}{2} \\ \delta & \text{else} \end{cases}
$$

with $\delta$ being a tolerance parameter to handle noise in depth measurements ($\delta = 2 mm$ in our method). Alongside with the TSDF $f_c$, we also create a volumetric color field $f_c : \Omega \rightarrow [0..255]^3$ following Equation 4:

$$
f_c(x) = \frac{w^1(x) I^1_\delta(|\Pi^1_\delta(x)|) + w^2(x) I^2_\delta(|\Pi^2_\delta(x)|)}{w^1(x) + w^2(x)}
$$

The scene to synthesize is represented in the volumetric grid by the voxels whose TSDF values is equal to 0. The color image $I_c$ from the X-ray viewpoint is therefore generated by performing raytracing from the X-ray viewpoint on the TSDF field $f_c$. For every pixel in the image to be synthesized, a ray is traced passing through the X-ray source and the pixel. Raytracing consists at searching the closest to the X-ray source voxel $y$ respecting the condition $f_c(y) = 0$ along this ray. To speed up this step, the search for the 0-value is performed by binary search. Once the $y$ has been found, the color $f_c(y)$ is applied to the pixel in the synthesized image $I_c$. A depth image $I_d$ can be synthesized by calculating the distance between $y$ and the X-ray source.

### 2.3 Multi-Layer Image Generation

After the first raytracing step, the video image $I_v$, as seen by the X-ray source viewpoint, as well as its corresponding depth image $I_d$ are generated. The volumetric TSDF field is a dense representation which contains information about the full 3D space around the C-arm detector whereas the raytracing stops only at the first found 0-value voxel. Therefore, the TSDF field contains more information than is actually used until now. Beyond the hands synthesized by the first raytracing, more 0-value can be present along the ray. This is true especially since the 2 RGBD cameras are placed on the side of the C-arm, giving additional information from another viewpoint. This situation is illustrated in Figure 2 where the background occluded by a hand from the X-ray source viewpoint (the blue point) can be seen by at least one of the 2 cameras. In a TSDF representation, this means those occluded background voxels also have a 0-value. To find those additional 0-value, a modified “second run” raytracing must be performed on the foreground (e.g. surgeon hands or surgical tools).

#### 2.3.1 Hand segmentation

As a first step, the foreground needs to be segmented from the synthesized video image $I_v$ and depth image $I_d$. A background model is computed from an initialization sequence of depth images where no hands or surgical instruments are introduced yet. An average depth image is created by averaging the depth at every pixel along the initialization sequence. Then, for every new image (with potential hands or surgical instruments present), the depth image $I_d$ is compared to the mean image in order to create a binary mask image $I_m$. For every pixel whose depth is lower than the average depth minus a margin (3 cm), the pixel is classified as foreground and is set as white in $I_m$. If the pixel is classified as background, then it is set as black in $I_m$. The method is rudimentary compared to background subtraction methods, however the margin allows the background to change shape (in the limit of the margin). A noise removal step is added using morphological opening on the mask image. An example of scaled depth image and its corresponding mask are shown on Figure 3.

#### 2.3.2 Second-run raytracing

Once the foreground has been segmented, a second raytracing can be performed on the pixels classified as hands or surgical instruments. Instead of beginning the raytracing from the X-ray source viewpoint, the ray search starts at the voxel $y$ found at the first raytracing run plus a margin of 4 cm. This margin is the insurance to not find a 0-value still related to the foreground. The starting voxel $y$ can be easily retrieved using the depth image $I_d$ resulting from the first raytracing. The raytracing is then performed forward using binary search in a similar fashion to the first run of raytracing. As a result, a color image of the background can be synthesized and combined to the color image from the first raytracing run (excluding the foreground segmented pixels) creating a complete background image $I_b$.

#### 2.3.3 Multi-Layer Visualization

On top of the background image $I_b$, the foreground layer extracted from $I_c$ can be overlaid with transparency as well as the X-ray image $I_{ray}$. A multi-layer image $I_{layers}$ can then be created by blending all the layers according to Equation 5:

$$
I_{layers}(p) = \begin{cases} \alpha I_v(p) + \beta I_b(p) + \gamma I_{ray}(p) & \text{if } p \in \text{foreground} \\ (1 - \delta)I_b(p) + \delta I_{ray}(p) & \text{else} \end{cases}
$$

where $(\alpha, \beta, \gamma, \delta) \in [0,1]^4$ with $\alpha + \beta + \gamma = 1$ are the blending parameters associated with each level. They can also be seen as specific weight values which emphasize a specific layer during the blending process.
The visualization scheme we propose allows us then to observe three layers of structures (displayed in Figure 4) according to those parameters.

Figure 4: Layers in our visualization, all can be observed depending on the chosen blending values $\alpha, \beta, \gamma, \delta$

The furthest layer is the X-ray, which can be observed in its totality in the image $I_{layers}$ with $(\alpha, \beta, \gamma, \delta) = (0, 0, 1, 1)$. As we get closer to the camera, another layer is the background structure recovered using volumetric field. It can be observed with $(\alpha, \beta, \gamma, \delta) = (0, 1, 0, 0)$. Finally the front layer comprising the hands and instruments can be observed in the image $I_{layers}$ using $(\alpha, \beta, \gamma, \delta) = (1, 0, 0, 0)$. Our visualization scheme allows to see in transparency the different layers (anatomy by X-ray, background, front layer) by choosing blending parameters $(\alpha, \beta, \gamma, \delta)$ non equal to 0 and 1. The choice of blending values depends on multiple parameters such as surgeon preferences, step in the surgical workflow, type of instrument used. It can be changed on the fly during surgery according to such parameters. For example, once an instrument has already penetrated the skin, the background is not necessary to visualize. The transparent hands can be overlaid directly on the X-ray image, skipping the background layer. This scenario corresponds to blending parameters $(\beta, \delta) = (0, 1)$, $\alpha = 1 - \gamma$ with $0 < \gamma < 1$. With the configuration $(\alpha, \beta, \gamma, \delta) = (1, 0, 0, 1)$, the visualization consists of fully opaque hands or surgical tools on the X-ray image, giving a similar output as [19] which aimed at obtaining a natural ordering of hands over X-ray image. As every layer is known at any point in a sequence, the multi-layer visualization can be replayed to medical students and residents for example with other blending parameters than the one used in surgery. They can have full control for the observation of the layers having the choice to emphasize particular layers of interest for their learning.

3 Results

3.1 Experimental protocol

Six sequences have been recorded depicting example scenarios which include both surgeon hands and surgical tools. Both a realistic hand model phantom and a real patient hand are used and positioned on a surgical table. A clinician wearing purple examination gloves introduces partial occlusions randomly to the scene. Sequences 1 and 3 contain the motion of the clinicians hand above the hand model phantom at 20 cm and 30 cm respectively. Sequences 2 and 4 contain the motion of a clinician hand closed and above the hand model phantom at 20 cm and 30 cm respectively. Sequences 3 and 4 also contain incision lines drawn using a marker on the hand model phantom. Finally, Sequences 5 and 6 are recorded with surgical tools above a real patient hand. Sequence 5 includes actions using a surgical hammer aiming for a cross target drawn on the patient hand. Sequence 6 includes a scalpel targeting the same cross. The heights of the surgical instruments to the patient hand vary up from 5 cm to 30 cm.

3.2 Background recovery

For every sequence, the mean value for the percentage of recovered pixels is calculated and indicated in Table 1. The natural observation in Table 1 is that the closer the surgeon hand and surgical tools are to the anatomy the larger the occlusion in both side cameras will be. This signifies a lower percentage of recovered pixels by our algorithm which is demonstrated.

| Sequences | 1   | 2   | 3   | 4   | 5   | 6   |
|-----------|-----|-----|-----|-----|-----|-----|
| Pixels recovered (in %) | 69.3 | 65.2 | 88.2 | 97.4 | 84.1 | 45.2 |

Table 1: Background recovery results

Sequences 1 and 2 were recorded with surgeon hand open (69.3%) and closed (65.2%). Less pixels are recovered for the close hand scenario as mainly the fist is present in the scene. The fist is also not recovered in the other scenario but the fingers are also occluding which are easier to recover from (due to their thin shape), in percentage, the open hand scenario recovers more, even if occluding more. Sequences 3 and 4 resulted in larger recovery percentages (88.2% and 97.4% respectively) because the surgeon hand was farther away from the hand model. This implies that there is a greater probability for the background voxels to be seen by the RGBD sensors. Sequence 6 with a scalpel confirms that the height strongly influences the recovery. The scalpel scenario which includes numerous images with hands and instruments close to the background (less than 10 cm) shows a low recovery result as expected. Due to the hammer’s shape, the sequence 5 shows however a higher recovery percentage.

3.3 Visualization results

In Figure 5, for each scenario, one selected image $I_{i, layers}$ in the sequence can observed with different values of $\alpha, \beta, \gamma$ and $\delta$. Each row $i$ corresponds to the sequence $i$. From left to right, the layer visualized in $I_{i, layer}$ is getting closer to the X-ray source viewpoint. In the column (a), the furthest layer (the X-ray image) is displayed. In the column (b), the second layer (the background), in the column (c), the blending of the front layer with the background, in the column (d), the blending of the three layers and finally, in the column (e), the closest layer is shown. Additional images from the sequences can be visualized in the supplementary video where interaction between the layers by changing the blending values can be observed.

Despite the fact that the background cannot be ideally recovered, a manual post processing step involving inpainting is applied and displayed in the column (f) of Figure 5. We believe that the multi-layer visualization concept is an interesting and profound solution offering numerous possibilities in the surgical areas, as well as, the mixed reality communities.

Similar to results from Habert et al. [6], the images resulting from synthesisation are not as sharp as a real video image. The area synthesized by our algorithm is approximately 20 cm x 20 cm (C-arm detector size), which is small compared to the wide-angle field of view from the Kinect v2. Reduced to the area of synthesisation, the video and depth from Kinect is not of high resolution enough for sharper results. More specialized hardware with smaller field of view and higher resolution RGBD data would solve this problem. Moreover, several artifacts can be seen around the hand and surgical instruments in the synthesized image due to high difference and noise in depth in the RGBD data from the 2 cameras. However, our results demonstrate that our method is working well, since the incision line and cross drawn on the hand model and patient hand are perfectly visible in the recovered background image and can be seen in transparency through the hands and surgical tools in the images of Figure 5 column (c) and (d). In the scalpel sequence.
Figure 5: Per row $i$, multi-layer image $I_{layers}$ of one selected frame in the sequence $i$ with different blending parameters $(\alpha, \beta, \gamma, \delta)$.
(sequence 6) in Figure 5 column (b), it can be seen that the tip of the scalp is considered as background. This is due to the margin of few centimeters used for background segmentation. In this image, the scalp is actually touching the skin.

4 Discussion

Involving temporal priors can help alleviate occlusion. Methods involving volumetric fields [16] use temporal information as the field is sequentially updating with new information, instead of fully being reinitialized as per our method. The percentage of pixels recovered is also dependent of the side cameras configuration. In our clinical case, the camera setup is constrained by the C-arm design and the disparity between the X-ray source and the two RGB cameras is low. A higher disparity would lead to less occlusion in at least one of the cameras. Even with our constrained and difficult clinical setup, the results are extremely promising and we are convinced the work could also be easily extended to less restrictive settings. A potential application is Industrial Diminished/Mediative Reality where workers wearing a HMD with two cameras placed on its side (with a higher disparity than our setup) could see their viewpoint synthesized with their hands in transparency.

5 Conclusion

In this paper, we have presented the first work combining Diminished and Augmented Reality in medical domain. Our visualization scheme proposes a user-adjustable multiple layer visualization where each layer can be blended with others. The multiple layers comprise the anatomy with the X-ray image, the patient background, and the surgeon hand and surgical instruments. The result of our visualization scheme offers the clinician to choose which layer(s) are to become transparent depending on the surgical scenario or workflow step. Beyond the medical domain, this work is the first use of volumetric field for background recovery in Diminished Reality and Mixed Reality. Future works should involve adding additional layers, by disassociating the surgeon hand layer from the surgical instruments layer, in order to adjust further the visualization to the user preferences.

References

[1] N. Bizzotto, A. Sandri, F. Lavini, C. Dalloca, and D. Regis. Video in operating room gopro hero3 camera on surgeons head to film operations test. Surgical innovation, page 1553350613513514, 2013.

[2] V. Buchmann, T. Nilsen, and M. Billinghurst. Interaction with partially transparent hands and objects. In Proceedings of the Sixth Australasian conference on User interface-Volume 40, pages 17–20. Australian Computer Society, Inc., 2005.

[3] F. I. Cosco, C. Garre, F. Bruno, M. Muzzupappa, and M. A. Otaduy. Augmented touch without visual obstruction. In Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on, pages 99–102. IEEE, 2009.

[4] L. T. De Paolis and G. Aloisio. Augmented reality in minimally invasive surgery. In Advances in Biomedical Sensing, Measurements, Instrumentation and Systems, pages 305–320. Springer, 2010.

[5] F. Echtler and al. libfreenect2: Release 0.1.1. Jan. 2016.

[6] S. Habert, M. Meng, W. Kehl, X. Wang, F. Tombari, P. Fallavollita, and N. Navab. Augmenting mobile c-arm fluoroscopes via stereo-rbgd sensors for multimodal visualization. In Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on, pages 72–75. Sept 2015.

[7] J. Herling and W. Broll. Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments. In Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on, pages 207–212. IEEE, 2010.

[8] J. Herling and W. Broll. Pixmix: A real-time approach to high-quality diminished reality. In Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on, pages 141–150. IEEE, 2012.

[9] S. Jarusirisawad and H. Saito. Diminished reality via multiple handheld cameras. In Distributed Smart Cameras, 2007. ICDCS ’07. First ACM/IEEE International Conference on, pages 251–258, Sept 2007.

[10] N. Kawai, T. Sato, and N. Yokoya. Diminished reality considering background structures. In Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on, pages 259–260. IEEE, 2013.

[11] H. Liao, T. Inomata, I. Sakuma, and T. Doi. 3-d augmented reality for mini-guided surgery using integral videography autostereoscopic image overlay. Biomedical Engineering, IEEE Transactions on, 57(6):1476–1486, 2010.

[12] M. A. Livingston, A. Dey, C. Sandor, and B. H. Thomas. Pursuit of X-ray vision for augmented reality. Springer, 2013.

[13] S. Meerits and H. Saito. Real-time diminished reality for dynamic scenes. In Mixed and Augmented Reality Workshops (ISMARW), 2015 IEEE International Symposium on, pages 53–59. IEEE, 2015.

[14] N. Navab, T. Blumi, L. Wang, A. Okur, and T. Wendler. First deployments of augmented reality in operating rooms. Computer, (7):48–55, 2012.

[15] N. Navab, S.-M. Heining, and J. Traub. Camera augmented mobile c-arm (camc): calibration, accuracy study, and clinical applications. Medical Imaging, IEEE Transactions on, 29(7):1412–1423, 2010.

[16] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In Mixed and augmented reality(ISMAR), 2011 10th IEEE international symposium on, pages 127–136. IEEE, 2011.

[17] A. Okur, S.-A. Ahmadi, A. Bigdelou, T. Wendler, and N. Navab. Mr in or: First analysis of arvt visualization in 100 intra-operative freehand spect acquisitions. In Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on, pages 211–218. IEEE, 2011.

[18] S. Onda, T. Okamoto, M. Kanehira, S. Fujikoa, N. Suzuki, A. Hat tori, and K. Yanaga. Short rigid scope and stereo-scope designed specifically for open abdominal navigation surgery: clinical application for hepatobiliary and pancreatic surgery. Journal of hepatobiliary-pancreatic sciences, 20(4):448–453, 2013.

[19] O. Pauly, B. Diotte, S. Habert, S. Weidert, E. Euler, P. Fallavollita, and N. Navab. Relevance-based visualization to improve surgeon perception in. Information Processing in Computer-Assisted Interventions, pages 178–185. Springer, 2014.

[20] H. Roodaki, K. Filipatos, A. Eslami, and N. Navab. Introducing augmented reality to optical coherence tomography in ophthalmic microsurgery. In Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on, pages 1–6. IEEE, 2015.

[21] H. Saito, T. Honda, Y. Nakayama, and F. De Sobrier. Camera pose estimation for mixed and diminished reality in ftv. In 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2014, pages 1–4. IEEE, 2014.

[22] F. Sauer, A. Khamene, B. Bascle, L. Schinunang, F. Wenzel, and S. Vogt. Augmented reality visualization of ultrasound images: system description, calibration, and features. In Augmented Reality, 2001. Proceedings. IEEE and ACM International Symposium on, pages 30–39. IEEE, 2001.

[23] Y. Shen, F. Lu, X. Cao, and H. Foroosh. Video completion for perspective camera under constrained motion. In Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, volume 7, pages 63–66. IEEE, 2006.

[24] K. Sugimoto, H. Fujii, A. Yamashita, and H. Asama. Half-diminished reality image using three rgb-d sensors for remote control robots. In Safety, Security, and Rescue Robotics (SSRR), 2014 IEEE International Symposium on, pages 1–6. IEEE, 2014.