Projector-Based Augmented Reality in Surgery without Calibration

J.-P. Tardif[1], S. Roy[1], J. Meunier[1,2]

[1]Department of Computer Science (DIRO), University of Montréal, Qc, Canada

[2]Biomedical Engineering Institute, University of Montréal and École Polytechnique, Qc, Canada

Abstract—**Augmented reality (AR) is becoming an important tool in surgery to support the surgeon and improve operation quality, safety and duration. However the AR setup with head-mounted display (HMD) and other equipments is often considered cumbersome by surgeons and limits its wide use in the operating room. To reduce this burden, we introduce a new approach to display undistorted image data directly on the patient (skin, bone, surgery linen etc.) without explicit camera and projector calibration. With a single camera used to capture the surgeon's field of view, the calibration is implicitly represented as a mapping establishing the correspondence of each pixel of a camera to a pixel from a projector. After this mapping has been carried out, one can display an image corrected for the surgeon. Results are presented showing the simplicity and potential of the method for an operating room.**

Keywords—Augmented reality, Surgery, Calibration, Camera, Projector

I. INTRODUCTION

Recently, augmented reality (AR) has been undergoing a very significant growth in many fields and particularly in surgery. The improvement and miniaturization of cameras and projectors and the increasing computing power will add even more possibilities in the future. In medicine, the use of computer, 3D numerical images and robot to help in planning and performing surgery is not new. Computer-aided surgery (CAS) has already benefited from AR by allowing the surgeon to look directly at the operating field instead of at a monitor. The most common design commercially available is based on a head-mounted display (HMD), a device embedding a small display which projects a virtual image on semi-transparent glasses allowing the simultaneous viewing of the real and virtual scenery [1]. Unfortunately this system does not provide a common focal plane for the real and virtual images. Video-based solutions exist to circumvent this problem but at the price of lower image quality and bigger weight [2,3]. Semi-transparent monitors placed between the surgeon and the patient offer another possibility, but suffer of parallax problems not easily solved [4]. For all these designs real time tracking of the HMD and its precise calibration (to get the projective transformation between 3D coordinates and the HMD coordinates) are needed.

Finally HMD are cumbersome and certainly limits their acceptance by surgeons in the operating room.

Considering all the aforementioned problems, we propose to display undistorted AR data directly on the patient (skin, bone, surgery linen etc.) thus solving several problems by simply removing the HMD of the usual AR setup (Fig. 1). Systems for projecting over non-flat surfaces already exist. It has been demonstrated that once the projectors are calibrated, texture can be painted over objects whose geometry is known [5,6]. Structured light techniques [7,8] or stereovision (with landmark projection over the surface) are simple ways to get the 3D geometry of the surface [9].

The next sections describe the image correction scheme for projecting undistorted images on the patient from the surgeon's point of view. To bypass calibration, the method exploits structured light to generate a mapping between a projector and a camera used to capture the surgeon's field of view. After this mapping has been carried out, one can display an image corrected for the surgeon.
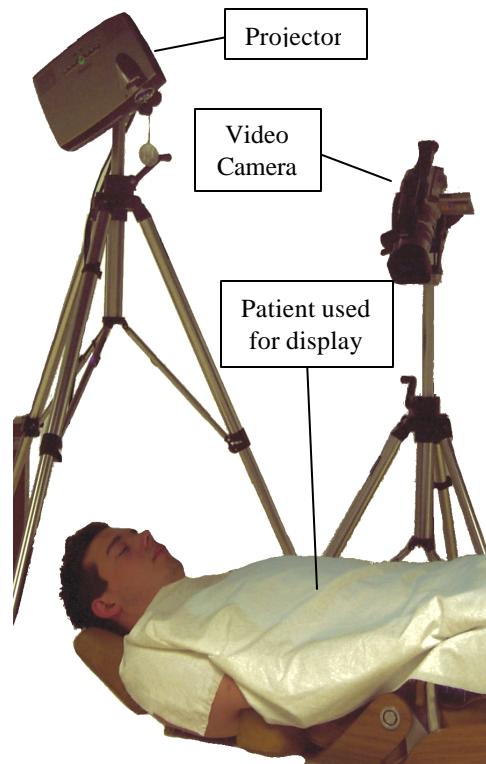

Fig. 1. Experimental setup for the proposed operating room with augmented reality.

II. METHODOLOGY

In order to project an image on a screen, some assumptions must be made. In general, the screen is considered flat and the projector axis perpendicular to it. Thus, minimal distortions appear to an audience in front of the screen. Notice that the assumptions involve the relative position and orientation of the observer, the screen and the projector. Those constraints cannot be met when an arbitrary projection surface is used. In this case, information about the projector-viewer-screen system has to be determined dynamically to correct projected images to avoid distortion. The approach commonly used starts by finding a function between the observer and the screen, and another one between the projector and the screen. These require calibrating the camera and projector [9]. Another approach, limited to a flat projection surface, involves the use of homographies to model the transformations from image and projector planes to the screen. The main advantage of homographies is their representation by 3 x 3 invertible matrices defined as [10]:

$$\begin{bmatrix} s \\ t \\ 1 \end{bmatrix} \cong \mathbf{H_p} \begin{bmatrix} x_{screen} \\ y_{screen} \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cong \mathbf{H_c} \begin{bmatrix} x_{screen} \\ y_{screen} \\ 1 \end{bmatrix}$$

where $\cong$ represents equivalence up to a scale factor. (x_{screen}, y_{screen}), (s,t), (u,v) are respectively an image point in the screen, projector and camera coordinates systems. From the matrices $\mathbf{H_p}$ and $\mathbf{H_c}$, a relation between the projector and the camera can be established:

$$\begin{bmatrix} s \\ t \\ 1 \end{bmatrix} \cong \mathbf{H_p} \mathbf{H_c}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

Unfortunately homographies provide a linear mapping and are not directly useful when the screen is non-flat. Instead, we have previously proposed [11] to replace the relation $\mathbf{H_p} \mathbf{H_c}^{-1}$ by a piecewise linear mapping function $\mathbf{R}$ relating the camera and the projector directly. The next subsections explain each step in more details.

A. Structured light

Structured light is commonly used in the field of 3D surface reconstruction [7]. Simple alternate black and white stripes are used to build a correspondence from a point of the camera to a coordinate in the projector one bit at a time. For instance, for a n bit coordinate encoding, each bit b (b in $\{1,...,n\}$) is processed individually and yields an image of coded stripes, each of width 2^{b-1} pixels. The concatenation of all bits provides the complete coordinate. Fig. 2 gives an example of the coded images we choose for projecting on the patient. Many other coding schemes are possible, we used the simplest possible pattern consisting of two set of horizontal and vertical stripes encoding s and t coordinates.

B. Direct mapping assessment

In order to compute a mapping function $\mathbf{R}$ from (u,v) to (s,t), we first decompose it into partial mapping functions $\mathbf{R}_b^s$ and, $\mathbf{R}_b^t$ mapping the bit b of the s and t coordinates respectively. These mapping functions are built by observing with the camera the projection of the corresponding stripe image and its inverse. Stripe identification is done with pixel by pixel difference between the image and its inverse, yielding Δ^s and Δ^t values between -255 and 255. The bit mapping $\mathbf{R}_b^s$ can now be defined as:

$$\mathbf{R}_b^s(u,v) = \begin{cases} 0 & \text{if } \Delta^s(u,v) < -t \\ 1 & \text{if } \Delta^s(u,v) > t \\ \text{rejected} & \text{otherwise} \end{cases}$$

where t is a threshold set empirically to 50. Exactly the same process using horizontal stripes defines $\mathbf{R}_b^t$ from Δ^t values. When a pixel is rejected, it will not be used anymore for the rest of the algorithm. To obtain a complete mapping $\mathbf{R}(u,v)$ from camera to projector, we simply concatenate the bit functions $\mathbf{R}_b^s$ and $\mathbf{R}_b^t$.

(a)   (b)
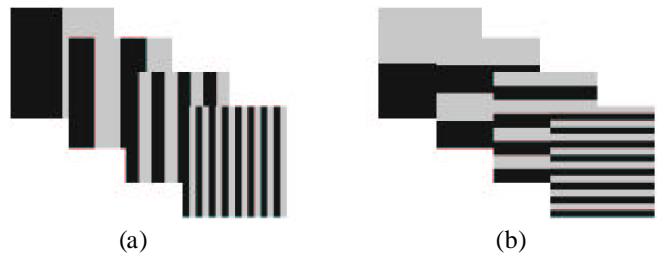
Fig. 2. Projected patterns for bits $b = 4,3,2,1$ for the s (a) and t (b) coordinate respectively.

C. Inverse mapping assessment and correction

For all projector pixels (s, t), we define a set of camera pixels $C(s,t) = \{(u,v) \mid \mathbf{R}(u,v) = (s,t)\}$ which is a contained region of the camera image. To get the inverse mapping $\mathbf{R}^{-1}(s,t)$ we simply select the center of $C(s,t)$. Applying this process for all non-empty $C(s,t)$, this defines an under-sampling of the function $\mathbf{R}^{-1}$. To complete the construction of $\mathbf{R}^{-1}$, an interpolation scheme is used [11]. In practice we

found that "bigger" projector pixels were needed to limit the number of rejected or unusable pixels. Typically, a 1024 x 768 image rebuilt with six bits (big pixel = 16 x 16 pixels), has new dimensions 64 x 48, representing 3072 squares for approximating the function. Once $\mathbf{R}^{-1}$ is found, the appropriate construction of the projector image to get an undistorted image for the camera (surgeon) is easily obtained by determining the color of each point *(s,t)* of the projector by looking at $\mathbf{R}^{-1}$*(s,t)* in the target image.

III. EXPERIMENTAL SETUP

Even if the implementation does not depend on the projector or camera resolution, the quality of the results increases with the resolution of each device. In the following experiments, we used a Sony Digital Handycam DCR-VX2000 (720 x 480 pixel resolution) video camera and a Compaq iPAQ MP4800 XGA DLP projectors (1024 x 768 pixel resolution) as shown in fig. 1 In most cases, acquisition time is proportional to the resolution of the camera. Mapping computation time of the projector using the video camera was below two minutes and could be improved in the future by optimizing the code and reducing the resolution if necessary. After the mapping is carried out, the image correction can be done in less than a second, but can be easily done in real-time on current video hardware technology.

III. RESULTS

Fig. 3 illustrates the process of projector-based AR with a test image projected on a "surgery" linen placed over a volunteer (patient) lying on an inclinable chair (see fig. 1). The projection from the surgeon's (or camera) point of view and before correction is shown in fig. 3(a). As could be expected, the image is severely distorted. However, after correction (and windowing) with our mapping function, the surgeon's view in fig. 3(b) appears mostly undistorted despite the non planar shape of the projection surface (patient). Fig 3(b) is to be compared with fig. 3(c), the original image used for testing. This image is a background grid with a coronary angiogram, an ECG signal, a circle, some text and a small logo of the authors' university. Several types of artifacts are noticeable. First, in some areas the projector does not "see" the whole projection screen (linen) or some stripes of the structured light could not be well acquired producing black regions corresponding to missing data; this is visible in the upper right corner of figure 3(b) for instance. The other artifacts occur due to the interpolation process. Notice that some adjustments were necessary to avoid saturation due to the high reflectivity of the linen and the proximity of the projector (delivering 2000 lumens of light).

IV. DISCUSSION

The results presented in this paper could be improved by using more bits per pixel, however due to subsurface scattering of light on the linen fabric used in our setup it was not possible to get more than 6 bits/pixel to code the coordinates with structured light. With a more appropriate textile we believe that 7 or 8 bits/pixel and thus a better mapping could be achieved.

Automatic determination of the rejection threshold τ and stripe width (number of bits/pixel) and its adaptation across different regions of the screen would result in better reconstruction.

Like every system using structured light, the optical characteristics of each device itself limit the possible screen shape that can be reconstructed. For instance, the depth of field of both camera and projector restricts the geometry and size of the screen. However, for most CAS applications this should not be a major problem.

The assessment of the mapping function $\mathbf{R}$ can be done concurrently with the display of surgical information by using the idea proposed by Raskar et al. [12] where the structured light patterns and its inverse are superimposed to the surgical information image consecutively and rapidly becoming imperceptive to an observer.

Addition of more projectors to cover a larger surface or to cover completely a complex curved surface is also possible [11]. More projectors can also ensure that no data is missing due to possible occlusion of the projector rays by the surgeon, his assistants or any objects. To support an arbitrary number of simultaneous projectors we simply compute a function $\mathbf{R}$ for each projector, one at a time. However, a scheme for intensity blending must be developed to take into account multiple projector sources.

Acquisition time could be decreased using improved patterns. Furthermore, hardware acceleration of video cards could be used to boost the speed of the construction of mapping function as well as the corrected image generation. This would allow real-time use in the operating room

Several other difficult problems were not addressed in this paper such as color and illumination correction when the surface is not white and matt or when illumination from the environment interferes with the projections. We believe that these issues are less crucial for simple graphical or textual display but they will need to be considered and corrected if more complex data (e.g. color images) are considered in the future.

V. CONCLUSION

We have presented a new image projection method that allows displaying undistorted image data directly on the patient in the operating room keeping the surgeon's field of view on the patient instead of alternating form a monitor to

the patient when doing CAS. Relying on a robust structured light approach, the method is simple and accurate and can readily be adapted to multi-projector configurations to eliminate shadows (occlusion). Furthermore the method does not require any direct calibration and could be used with any type of projector lenses.

REFERENCES

[1] Azuma, Ronald T. A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments 6, 4 (August 1997), 355 - 385.

[2] Fuchs, Henry, Mark A. Livingston, Ramesh Raskar, D'nardo Colucci, Kurtis Keller, Andrei State, Jessica R. Crawford, Paul Rademacher, Samuel H. Drake, and Anthony A. Meyer, MD. "Augmented Reality Visualization for Laparoscopic Surgery." Proceedings of First International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '98), 11-13 October 1998, Massachusetts Institute of Technology, Cambridge, MA, USA.

[3] Fuchs, Henry, Andrei State, Etta D. Pisano, William F. Garrett, Gentaro Hirota, Mark A. Livingston, Mary C. Whitton, and Stephen M. Pizer. "(Towards) Performing Ultrasound-Guided Needle Biopsies from within a Head-Mounted Display." Proceedings of Visualization in Biomedical Computing 1996, (Hamburg, Germany, September 22-25, 1996), pgs. 591-600.

[4] S. Nakajima, K.Nakamura, K. Masamune, I. Sakuma, T. Dohi: „Three-dimensional medical imaging display with computer-generated integral photography"; Computerized Medical Imaging and Graphics 25(3), 235-241,2001

[5] Ramesh Raskar, Kok-Lim Low, and Greg Welch. Shader lamps: animating real objects with image-based illumination. Technical Report TR00-027, 06 2000.

[6] R. Raskar, R. Ziegler, and T. Willwacher. Cartoon dioramas in motion. In International Symposium on Non-Photorealistic Animation and Rendering (NPAR), June 2002.

[7] Szymon Rusinkiewicz, Olaf Hall-Holt, and Marc Levoy. Real-time 3d model acquisition. In ACM Transactions on Graphics, volume 21, pages 438–446, 2002.

[8] Li Zhang, Brian Curless, and Steven M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In 1st international symposium on 3D data processing, visualization, and transmission, Padova, Italy, June 2002.

[9] Trucco E., Verri A. – Introductory techniques for 3-D computer vision, Prentice Hall, 1998.

[10] R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge, 2000.

[11] J-P Tardif,. S Roy, M Trudeau Multi-projectors for arbitrary surfaces without calibration nor reconstruction 4th International Conference on 3-D Digital Imaging and Modeling, Banff, Canada,October 6-10, 2003 (submitted)

[12] R. Raskar, M. Cutts, A. Lake, L. Stesin and H. Fuchs, "The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Display," in Proc. SIGGRAPH 98, Orlando, Florida, USA, 1998.
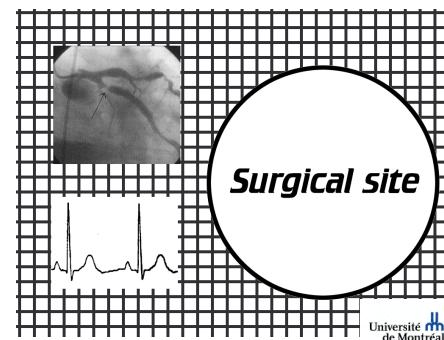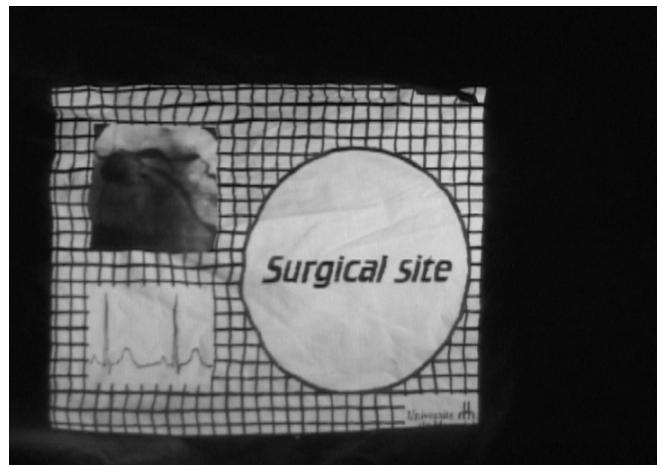
Fig. 3. Experimental results showing (a) the surgeon's (or video camera) view before correction and (b) after correction (and windowing) showing reduced distortion and (c) the original image used for testing