

[POSTER] Hybrid Video/Optical See-Through HMD

Fabrizio Cutolo*
Information
Engineering Department,
University of Pisa, Italy

Umberto Fontana
Information
Engineering Department,
University of Pisa, Italy

Marina Carbone
Endocas Center, Dept.
of Translational Research
and New Technologies in
Medicine and Surgery,
University of Pisa, Italy

Renzo D'Amato
Information
Engineering Department,
University of Pisa, Italy

Vincenzo Ferrari
Information
Engineering Department,
University of Pisa, Italy

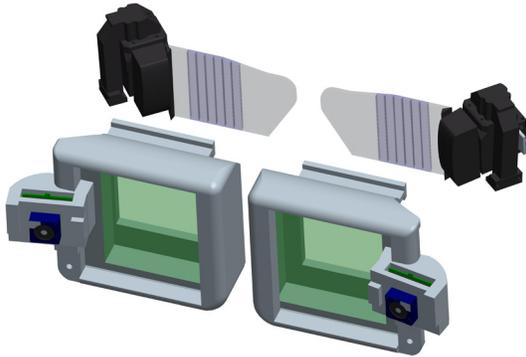


Figure 1: On the left: schematic of the hybrid Video/Optical see-through HMD comprising the supports for the electro-optical shutters and a pair of stereo cameras mounted on the top of a commercial binocular OST display. On the right: embodiment of the hybrid Video/Optical see-through HMD

ABSTRACT

An old but still ongoing subject of debate among augmented reality (AR) experts is about which see-through paradigm is best in wearable AR displays. Video see-through (VST) and optical see-through (OST) paradigms have both their own strengths and shortcomings with respect to technological and human-factor aspects. The major difference between these see-through paradigms is in providing an aided (VST) or unaided (OST) view of the real world. In this work, we present a novel approach for the development of AR stereoscopic head-mounted displays (HMDs) that can provide both the see-through mechanisms. Our idea is to dynamically modify the transparency of the display through a liquid crystal (LC)-based electro-optical shutter applied on the top of a standard OST device opportunely modified for housing a pair of external cameras. A plane-induced homography transformation is used for consistently warping the video images, hence reducing the parallax between cameras and displays. An externally applied drive voltage is used for smoothly controlling the light transmittance of the LC shutters so as to allow an easy transition between the unaided and the camera-mediated view of the real scene. Our tests have proven the efficacy of the proposed solution under worst-case lighting conditions.

Keywords: HMD, Video see-through, Optical see-through.

Index Terms: OST-HMD, VST-HMD, orthoscopy, parallax, image warping, homography induced by a plane, electro-optical shutter, lighting conditions.

* fabrizio.cutolo@endocas.unipi.it

1 INTRODUCTION

Wearable augmented reality (AR) systems based on head-mounted displays (HMDs) intrinsically provide the user with an egocentric viewpoint with a reduced or zero parallax. Standard stereoscopic HMDs provide binocular parallax and motion parallax and smoothly augment the user's perception of the real scene during use. For these reasons AR HMD systems are properly deemed as the most ergonomic solution to those tasks manually performed under user's direct vision since they minimize the extra mental effort required when switching his/her own attention back and forth between the real scene and the augmented view rendered on the display [1].

Most of the AR HMDs fall into two categories according to the see-through paradigm they implement: video see-through (VST) HMDs and optical see-through (OST) HMDs. Typically, in OST HMDs, the user's direct view of the real world is augmented with the projection of virtual information on a beam combiner and then into the user's line of sight [2] (Figure 2). Differently, in VST HMDs the virtual content is merged with the camera images captured by one or two external cameras rigidly fixed on the visor frame (Figure 3).

The industrial pioneers, as well as the early adopters of AR technology, properly considered the camera-mediated view typical of the VST paradigm, as drastically affecting the quality of the visual perception and experience of the real world. By contrast, OST HMDs were/are able to provide the user with a natural view of the real world at full resolution. The fundamental OST paradigm of HMDs is still the same as described by Benton (Google Glass, Microsoft HoloLens, Epson Moverio, Lumus optical, Optinvent ORA-1) [3].

A straightforward implementation of the OST paradigm comprises the employment of a half-silvered mirror as beam combiner to merge real view and virtual content. The user's own view is herein augmented by rendering the virtual content on a two-dimensional (2D) micro display and by sending the image to

the beam combiner. Lenses can be placed between the beam combiner and the display to focus the virtual 2D image so that it appears at a comfortable viewing distance on a semi-transparent surface of projection (SSP) [4, 5]. As an alternative, the use of thin and high-precision light guide technologies allows the removal of the traditional and bulky optical engine placed in front of the user's eyes [6]. The OST paradigm is particularly suitable for augmenting the reality by means of simple virtual elements (models, icons or text labels) but shortcomings remain both from a technical and a perceptual standpoint, especially in case of virtual contents of greater complexity and at close working distances.



Figure 2: Commercial optical see-through head-mounted displays.



Figure 3: Commercial video see-through head-mounted displays

From a technological standpoint, even if automated and not user-dependent eye-to-display calibration methods have been recently proposed [7-10], the problem of ensuring a robust and accurate AR image registration remains an open question.

From a perceptual standpoint, in OST HMDs the difference in the user's perception of the real 3D world and of the 2D projection of the virtual content on the SSP creates conflicts due to mismatched accommodations. These perceptual conflicts are often reflected in reducing the benefits brought by the optical paradigm of leaving unaltered the user's view of the real world. In OST HMDs the user is indeed forced to accommodate his/her eyes for focusing all the virtual objects on the SSP placed at a fixed distance. On the other hand, the focus distance of each physical object in the 3D world depends on its relative distance from the observer and may dynamically vary over time. This means that, even if an accurate geometric registration of virtual objects to the real scene is obtained on the 2D SSP plane, the user may not be able to view both the virtual and real content in focus at the same time. This aspect is particularly significant in complex applications that demand for great hand-eye coordination, since it reduces the user's ability to interact with the real scene whilst maintaining the virtual aid in focus.

Differently, the pixel-wise video mixing technology that underpins the VST paradigm can offer high geometric coherence between virtual and real content. In these systems, a user-specific calibration eye-to-display routine is not necessary, and this is the major advantage of the VST versus the OST approach. Further, in VST systems, real scene and virtual content can be synchronized, whereas in standard OST devices there is an intrinsic lag between the immediate perception of the real scene and the inclusion of the virtual elements. Nevertheless, the user wearing a VST HMD can experience motion sickness if the AR scene is rendered onto the visor internal displays with a significant delay. From a perceptual viewpoint, in VST systems the visual experience of both the real and virtual content is unambiguously controllable by computer graphics, with everything on focus at the same apparent distance from the user. VST systems are much more suited than OST systems, to rendering occlusions between real and virtual

elements or to implementing complex visualization processing modalities that are able to perceptually compensate for the loss of the unobstructed real-world view. On the other hand, VST HMDs have the main drawback of substantially restricting the visibility of the real world depending on the resolution, the field of view (FoV) and the placement of the pair of external cameras.

This paper reports on a novel approach for the development of AR stereoscopic HMDs able to provide both the see-through paradigms. The goal is to offer, on a standard OST HMD, also the benefits of the VST paradigm by using external cameras and offering the possibility of obscuring the view of the real world. In more detail, a pair of liquid crystal (LC)-based optical shutters are mounted on the top of a standard OST HMD to black out its OST capability. Our idea is to dynamically modify the transmittance of the electro-optical shutter so as to allow an easy transition between the unaided and the camera-mediated view of the real scene. The paper is structured as follows: in Section 2, two examples of related works in the field are presented; Section 3 presents the conceptual framework of the proposed approach and describes its technical implementation. Section 4 shows the results of the preliminary experiments carried out for assessing the efficacy of the solution under worst-case lighting conditions. Section 5 discusses few possible use cases in which the hybrid video-optical see-through HMD could be of assistance.

2 RELATED WORKS

In literature, solutions for actively selecting, on a pixel basis, the information coming from standard OST microdisplays have been proposed. In [11] the authors present a prototype that, by means of a masking LCD panel, is able to generate proper occlusions between virtual and real objects in the AR scene. More recently, a less cumbersome wide FoV OST prototype has been proposed [12]; the system, by means of two stacked LCD microdisplays and an RGB backlight, provides spatial light modulators needed for yielding selective per-pixel occlusions and multiple simultaneous focal depths. Our solution is less complex technically and can be easily scaled down: its main feature is to globally controlling the transparency of the see-through displays and it is not intended to address the aforementioned perceptual issues related to the OST paradigm, e.g. forced eye's accommodation due to the SSP generated at a fixed distance.

3 IDEA AND IMPLEMENTATION

Typically, most prototypical embodiments of VST HMDs have been built by mounting stereo camera pairs on top of commercial 3D visors for virtual reality [13-18]. Such systems, if they do not include a system of mirrors for providing rigorous orthostereoscopy [13, 19], suffer from a non-negligible eye-camera parallax and are all quite obtrusive and heavy. Conversely, most OST HMDs feature high ergonomics, high lightness and low bulkiness, being designed as everyday glasses. For this reason, and for achieving a quasi-orthoscopic configuration for the VST modality, we built our novel HMD on top of a commercial OST HMD. By doing so, the pair of external cameras can be mounted closer to the user's exit pupils and the eye-camera parallax is reduced. Obviously, the implementation of the VST modality on our visor requires that the underlying OST HMD works in opaque mode as in [20]. Since our goal was to develop a hybrid VST/OST visor, we had to work on a solution allowing a smooth transition between the VST modality and the OST modality. To this end and with the objective of avoiding any moving parts in the visor, we adopted an electrically-driven shutter in place of an obtrusive mechanical cover. In the following sub-paragraphs we shall briefly outline the technical implementation of our idea and we shall list the components of the hybrid HMD.

3.1 HMD platform with embedded stereo pair

Our hybrid VST/OST is based on a reworked version of a commercial binocular OST HMD (DK-33 by LUMUS, www.lumusvision.com).

The optical engine of the visor is the OE50 which features a 1280x720 resolution, a horizontal FoV of 35.2° and a vertical FoV of 20.3°.

The stereo camera pair, composed by two Sony FCB-MA13 cameras, was mounted laterally at a distance of 98.26 mm. The camera, comprising a 1/2.45" CMOS sensor, is extremely compact in size (16.5 x 10.3 x 18.0 mm) and has the following specifics: horizontal FoV = 53°, vertical FoV = 29° and frame rate of 30 fps at 1920x1080 resolution.

3.2 LC-based electro-optical shutter

For managing the smooth transition between occluded and non-occluded view, therefore with the goal of obtaining a totally hands-free VST/OST switching mechanism, we have adopted an electrically driven LC optical shutter (FOS model by LC-Tec, www.lc-tec.se).

Such LC shutters manage to control the light transmittance by means of an externally applied drive voltage and have the main advantage of allowing analog transitions between fully open and closed states via voltage amplitude modulation (i.e. they can be used as variable filters other than as on/off shutters).

According to the manufacturer's specs, the open state transmittance is $\geq 39.5\%$, whereas the closed state transmittance is close to 0%.

3.3 3D printed supports

The design of the supports was realized using PTC Creo Parametric (www.ptc.com). Their sizes and shapes were established starting from the dimensions and constraints imposed by the outer frame of the OST HMD and considering the size of the electro-optical shutters and of the pair of external cameras.

Overall, the design of the supports was considered so as to be compliant with the targets in terms of system ergonomics, weight and compactness. One support is intended contain the two LC shutters and a second one serves for housing the electronic board of the stereo cameras (Figure 1).

The 3D models of the supports were then printed with a 3D rapid prototyping machine (Stratasys Elite Dimension, www.stratasys.com) (Figure 4). The whole HMD with the external cameras, their electronic boards, and the optical shutters is depicted in Figure 1.

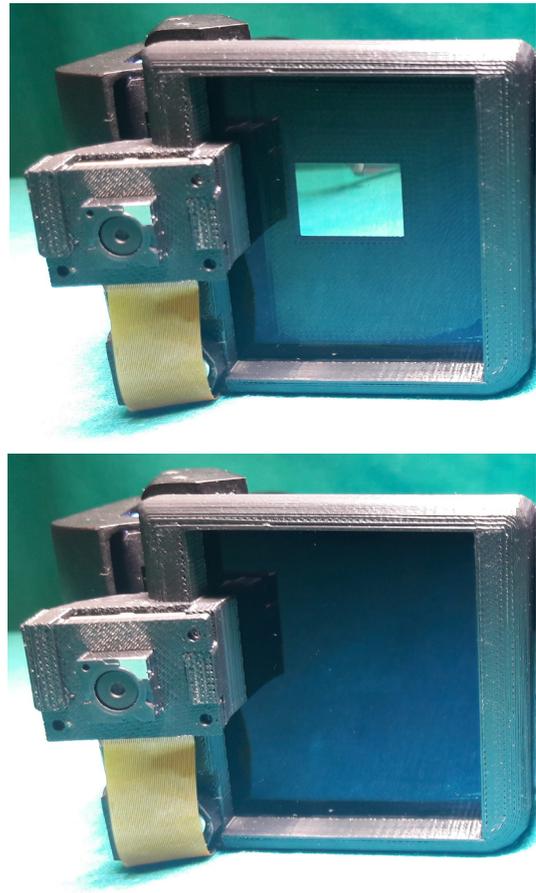


Figure 4: 3D printed external supports for electro-optical shutter and external camera. First row image with the shutter in open-state. Second row image with the shutter in closed-state.

3.4 Camera placement and image warping

The intended use of our hybrid VST/OST HMD is for tasks performed within arm's reach. For this reason cameras are toed in so as to be compliant to a viewing/working distance of 50 cm. The angle of vergence of the stereo pair is of 15.2°, so that the optical axes of the two cameras intersect at the pre-defined working distance [21].

In order to compensate for the different camera placement with respect to the user's eyes position (i.e. eye-camera parallax), camera images are warped before being rendered on the display. Image warping is performed after estimating the homography that maps the image points of an ideal plane, placed at 50 cm from the observer, in the camera's image to those yielded by the same camera placed at the eye's ideal position (i.e. at the center of the display eye box) [22, 23]. The extrinsic and intrinsic parameters of the display's frustum have been retrieved from the specifics of the optical engine. The warped image is subjected to magnification that derives from the mismatch between the FoVs of the camera and the display. In this way, the camera-mediated view of the points placed on the ideal plane should be perceived on the same direction as they would be perceived if seen with the naked eye (i.e. unaided view). Nevertheless, this solution does not ensure a proper visual perception of those points that are far from the ideal plane and if the position of the user's eyes (viewpoints) do not fall exactly at the centers of the displays' eye box. Therefore, under open-state mode and with the OST display

turned on the user may experience different degrees of double vision (from light to severe) depending on the actual viewing distance and on the displacement between his/her own eyes and the ideal centers of the eye-box of the displays.

4 EXPERIMENTS

The goal of our tests was to preliminarily assess the efficacy of our hybrid VST/OST HMD under worst-case lighting conditions.

As previously described, in open-state mode the unaided and the camera-mediated views are perceived as not perfectly overlapped the farther we go from the reference plane considered for computing the homography transformation and since the user's viewpoint hardly matches the center of the eye box of the display (Figure 5). The image was acquired by a third camera placed within the eye box of one of the two displays.

When the electro-optical shutter works in closed-state mode, such double vision effect disappears (Figure 6).



Figure 5: Double vision (severe displacement at close distances) with electro-optical shutter in open-state mode.



Figure 6: No double vision with electro-optical shutter in closed-state mode.

The capacity of the electro-optical shutter to block the environment light depends on the following factors: the user's vision sensitivity, the lighting conditions in the environment, the accuracy in the plane-induced homography estimation in connection with the distance between the reference plane and the area under observation, and the brightness of the OST display. We tested our hybrid HMD on 5 subjects and under different lighting conditions: 1) at a work desk in an office with standard neon lighting on the ceiling; 2) under controlled lighting conditions obtained by means of a photographic spotlight; 3) under worst-case lighting conditions in the operative room under a LED scialytic lamp (Figure 7).

In each lighting conditions we measured the illuminance before the user at approximately a distance of 50 cm from the eyes by using a commercial lux meter by GrandBeing (www.grandbeing.net) (Figure 8). The lux meter has a declared accuracy of $\pm 4\%$. In each one of the selected lighting conditions,

the user looked at his/her own hands and at a 6x4 cm mirror that directly reflected the light source to his/her own eyes.

In each lighting condition, with the electro-optical shutter in closed-state and the OST display turned off, each user was able to perceive the view of the environment through the visor.

With the electro-optical shutter in closed-state and the OST display turned on all the users experienced double vision only in the event of direct reflection of the light coming from the scialytic lamp, as reported in Table 1:

Table 1: Experience of double vision under different lighting conditions

	Desk lighting (330 Lux)		Photographic light (730 Lux)		Scialytic lamp (105800 Lux)	
	Hands	Mirror	Hands	Mirror	Hands	Mirror
Sub.1	No	No	No	No	No	Yes
Sub.2	No	No	No	No	No	Yes
Sub.3	No	No	No	No	No	Yes
Sub.4	No	No	No	No	No	Yes
Sub.5	No	No	No	No	No	Yes



Figure 7: Worst-case lighting conditions in the surgical room under the scialytic lamp.



Figure 8: Illuminance measured in the surgical room with a mirror providing a direct reflection of the light source on the user's eyes.

Even if the electro-optical shutter is not able to completely obscure the light rays coming from the environment, it is enough to avoid double vision when the OST display is turned on. This means that the luminance contribution associated to the light rays coming from the world, after their passage through the electro-optical shutter, is negligible with respect to the luminance generated by the display itself. This effect is aided by the quasi-orthoscopic camera placement that, together with the image warping, allow the quasi-perfect overlapping the real-world light rays with their corresponding pixel-wise reproductions coming from the optical engine of the OST. In this way, directions with high luminance are associated to pixel areas with high brightness and this effect avoids double vision also in presence of direct reflections of the light source to the observer's eyes (which can be considered a worst-case scenario).

All the subjects experienced double vision only in the event of light directly reflecting from the LED scalytic lamp in a surgical room. This condition is, however, rather unrealistic since observing a direct reflection of the scalytic lamp is comparable to staring at the sun with the naked eye (brightest sunlight is about 120.000 Lux).

The electro-optical shutter represents a valid alternative to a mechanical external cover and does not require moving parts (motorized or not) since the switching mechanism between OST and VST is electrically-driven. In future, the optical shutter can be scaled down to be adapted to the optical engine of the OST HMD.

5 CONCLUSION

We presented a new approach for the development of AR stereoscopic head-mounted displays (HMDs) that provides both the see-through mechanisms (OST and VST). Our idea is to dynamically modify the transparency of the display through a liquid crystal (LC)-based electro-optical shutter applied on the top of a standard OST device opportunely modified for housing a pair of external cameras. Our tests have proven the efficacy of the proposed solution under worst-case lighting conditions.

VST and OST paradigms have both their own strengths and shortcomings, which are almost all complementary for a defined use case. A hybrid system can therefore combine the benefits of

both the AR modalities since it allows to switching to the most suitable modality during use.

The idea of integrating the potentialities of both the see-through paradigms is of utmost importance towards the definition of AR visualization modalities able to adapt to highly specific and complex tasks. This is particularly true in the context of image-guided surgery, where the quality of the AR experience and the ability to smoothly integrate the surgeon's perceptive efficiency, represent the key requisites for a successful result.

The availability of a hybrid OST and VST display paves the way to new perceptual studies in the context of AR (e.g. aimed at comparing different registration or rendering methods while switching between both paradigms on the same device).

ACKNOWLEDGEMENTS

FUNDED BY THE HORIZON2020 PROJECT VOSTARS, PROJECT ID: 731974. CALL: ICT-29-2016 - PHOTONICS KET 2016.

REFERENCES

- [1] F. Cutolo, G. Badiali, and V. Ferrari, "Human-PnP: Ergonomic AR Interaction Paradigm for Manual Placement of Rigid Bodies," in *Augmented Environments for Computer-Assisted Interventions: 10th International Workshop, AE-CAI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 9, 2015. Proceedings*, C. A. Linte, Z. Yaniv, and P. Fallavollita, Eds., ed Cham: Springer International Publishing, 2015, pp. 50-60.
- [2] J. P. Rolland, R. L. Holloway, and H. Fuchs, "Comparison of optical and video see-through, head-mounted displays," in *Photonics for Industrial Applications*, 1995, pp. 293-307.
- [3] S. A. Benton, *Selected Papers on Three-dimensional Displays*: SPIE Optical Engineering Press, 2001.
- [4] J. P. Rolland and O. Cakmakci, "The past, present, and future of head-mounted display designs," 2005, pp. 368-377.
- [5] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-Dimensional Displays: A Review and Applications Analysis," *Ieee Transactions on Broadcasting*, vol. 57, pp. 362-371, Jun 2011.
- [6] H. Mukawa, K. Akutsu, I. Matsumura, S. Nakano, T. Yoshida, M. Kuwahara, *et al.*, "Distinguished paper: A full color eyewear display using holographic planar waveguides," *2008 Sid International Symposium, Digest of Technical Papers, Vol Xcix, Books I-Iii*, vol. 39, pp. 89-92, 2008.
- [7] Y. Itoh and G. Klinker, "Interaction-Free Calibration for Optical See-Through Head-Mounted Displays based on 3D Eye Localization," *2014 Ieee Symposium on 3d User Interfaces (3dUI)*, pp. 75-82, 2014.
- [8] Y. Itoh and G. Klinker, "Performance and Sensitivity Analysis of INDICA : Interaction- Free Display Calibration for Optical See-Through Head-Mounted Displays," *2014 Ieee International Symposium on Mixed and Augmented Reality (Ismar) - Science and Technology*, pp. 171-176, 2014.
- [9] A. Plopski, Y. Itoh, C. Nitschke, K. Kiyokawa, G. Klinker, and H. Takemura, "Corneal-Imaging Calibration for Optical See-Through Head-Mounted Displays," *Ieee Transactions on Visualization and Computer Graphics*, vol. 21, pp. 481-490, Apr 2015.

- [10] A. Plopski, J. Orlosky, Y. Itoh, C. Nitschke, K. Kiyokawa, and G. Klinker, "Automated Spatial Calibration of HMD Systems with Unconstrained Eye-cameras," *Proceedings of the 2016 15th Ieee International Symposium on Mixed and Augmented Reality (Ismar)*, pp. 94-99, 2016.
- [11] K. Kiyokawa, M. Billingham, B. Campbell, and E. Woods, "An occlusion-capable optical see-through head mount display for supporting co-located collaboration," *Second Ieee and Acm International Symposium on Mixed and Augmented Reality, Proceedings*, pp. 133-141, 2003.
- [12] A. Maimone and H. Fuchs, "Computational Augmented Reality Eyeglasses," *2013 Ieee International Symposium on Mixed and Augmented Reality (Ismar) - Science and Technology*, pp. 29-38, 2013.
- [13] A. State, K. P. Keller, and H. Fuchs, "Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display," *International Symposium on Mixed and Augmented Reality, Proceedings*, pp. 28-31, 2005.
- [14] A. State, J. Ackerman, G. Hirota, J. Lee, and H. Fuchs, "Dynamic virtual convergence for video see-through head-mounted displays: Maintaining maximum stereo overlap throughout a close-range work space," *Ieee and Acm International Symposium on Augmented Reality, Proceedings*, pp. 137-146, 2001.
- [15] F. Cutolo, C. Freschi, S. Mascioli, P. Parchi, M. Ferrari, and V. Ferrari, "Robust and Accurate Algorithm for Wearable Stereoscopic Augmented Reality with Three Indistinguishable Markers," *Electronics*, vol. 5, p. 59, 2016.
- [16] F. Cutolo, P. D. Parchi, and V. Ferrari, "Video See Through AR Head-Mounted Display for Medical Procedures," *2014 Ieee International Symposium on Mixed and Augmented Reality (Ismar) - Science and Technology*, pp. 393-396, 2014.
- [17] M. Bajura, H. Fuchs, and R. Ohbuchi, "Merging Virtual Objects with the Real-World - Seeing Ultrasound Imagery within the Patient," *Siggraph 92 : Conference Proceedings*, vol. 26, pp. 203-210, 1992.
- [18] F. Sauer, U. J. Schoepf, A. Khamene, S. Vogt, M. Das, and S. G. Silverman, "Augmented reality system for CT-guided interventions: System description and initial phantom trials," *Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display*, vol. 5029, pp. 384-394, 2003.
- [19] A. Takagi, S. Yamazaki, Y. Saito, and N. Taniguchi, "Development of a stereo video see-through HMD for AR systems," *Ieee and Acm International Symposium on Augmented Reality, Proceeding*, pp. 68-77, 2000.
- [20] M. Kanbara, T. Okuma, H. Takemura, and N. Yokoya, "A stereoscopic video see-through augmented reality system based on real-time vision-based registration," in *Proceedings IEEE Virtual Reality 2000 (Cat. No.00CB37048)*, 2000, pp. 255-262.
- [21] V. Ferrari, F. Cutolo, E. M. Calabrò, and M. Ferrari, "[Poster] HMD Video see though AR with unfixed cameras vergence," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2014, pp. 265-266.
- [22] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*: Cambridge university press, 2003.
- [23] M. Tomioka, S. Ikeda, and K. Sato, "Approximated User-Perspective Rendering in Tablet-Based Augmented Reality," *2013 Ieee*