# Towards the Development of a Quasi-Orthoscopic Hybrid Video/Optical See-Through HMD for Manual Tasks

Fabrizio Cutolo[1,2(✉)], Nadia Cattari[2], Umberto Fontana[1,2], Vincenzo Ferrari[1,2]

[1] Information Engineering Department, University of Pisa, Italy.
fabrizio.cutolo@endocas.unipi.it

[2] EndoCAS Center, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Italy.

**Abstract.** Augmented Reality (AR) based on head-mounted displays (HMDs) represent the most ergonomic and efficient solution for aiding complex manual tasks performed under direct vision and it is considered an enabling technology of the fourth industrial revolution (Industry 4.0). This work intends to give a harmonized and consistent overview of the workflow carried out so far in the framework of the European Project VOSTARS towards the development of a novel AR HMD specifically designed to aiding complex manual tasks within arm's reach. First results and future developments are being presented.

**Keywords:** Augmented Reality, Head-mounted Displays, Orthoscopy, Visual Perception, Industry 4.0.

## 1    Introduction

Augmented Reality (AR) allows merging natural perception of the real environment with computer-generated information and it is considered an enabling technology of the fourth industrial revolution (Industry 4.0); it is indeed aimed at aiding complex tasks as a tool in the manufacturing industry such as in assembly tasks, maintenance and repair, training, quality control, and commissioning [1-4]. In recent years, AR has already proven to be a promising technology also in the education field [5-7] and in the healthcare industry, as demonstrated by the increasing number of publications in the medical training field [8, 9], surgical field [10-15] and rehabilitation [16].

In the ideal AR system, especially if designed for aiding manual tasks in the peripersonal space, there should not be any perceivable difference between the user's natural experience of the world and his/her augmented experience through the AR interface [17]. For this goal, the conditions to be satisfied are twofold: accurate spatial registration between real world scene and computer-generated content and ergonomic interaction with the AR scene [18]. According to the words of one of the pioneer researchers in the field of AR [19], "*The basic goal of an AR system is to enhance the user's*

*perception of and interaction with the real world through supplementing the real world with 3D virtual objects that appear to coexist in the same space as the real world".*

For this reason, AR systems based on head-mounted displays (HMDs), are intrinsically the most ergonomic and efficient solutions for aiding manual tasks, thanks to their ability to preserve a natural and egocentric viewpoint thus allowing the user to freely interact with the augmented scene with his/her own hands [20, 21]. Since 2016, the authors have been coordinating the European project VOSTARS (Video and Optical See-Through Augmented Reality Surgical Systems, Project ID: 731974) [22], whose aim is to develop and validate an innovative and immersive HMD to act as surgical navigator. The aim of the project is to bring 'surgical navigation' directly in front of the surgeon's eyes and so to aid him/her during complex manipulative tasks. More generally the ambition is to overcome the drawbacks of state-of-the-art AR HMDs and so massively revolutionize the paradigms through which visual AR is delivered. In line with this, this paper provides a detailed insight into the achievements of the early stage of development of the ongoing European project. The paper explicitly refers to some previously published works where early results were individually unveiled. Therefore, it intends to give a harmonized and consistent overview of the workflow carried out so far towards the development of a new AR visor specifically devoted to aiding complex manual tasks within arm's reach.

## 2 Materials and Methods

In AR HMDs, the see-through capability can be accomplished either through the video see-through (VST) mechanism or through the optical see-through (OST) mechanism. In binocular VST HMDs, the view of the real world is captured by a pair of stereo cameras rigidly anchored to the visor with an anthropometric interaxial distance (Fig. 1). The stereo views of the world are presented onto the binocular micro displays of the HMD after being digitally combined with the virtual content. AR image registration is here easily achieved but the real view of the world is mediated by the cameras. By contrast, in OST HMDs, the user's direct view of the world is mostly preserved albeit at expenses of a reduced AR image registration. Here, the direct view of the real world is optically merged, through a beam combiner, with the computer-generated content (Fig. 2). This aspect confers a clear advantage over VST solutions, particularly when used to interact with objects at close distances, since it allows the user to maintain almost unaltered his/her own natural visual experience. The computer-generated content is projected onto a microdisplay (i.e., image source) and then optically redirected to the user's retina with a bit of delay.
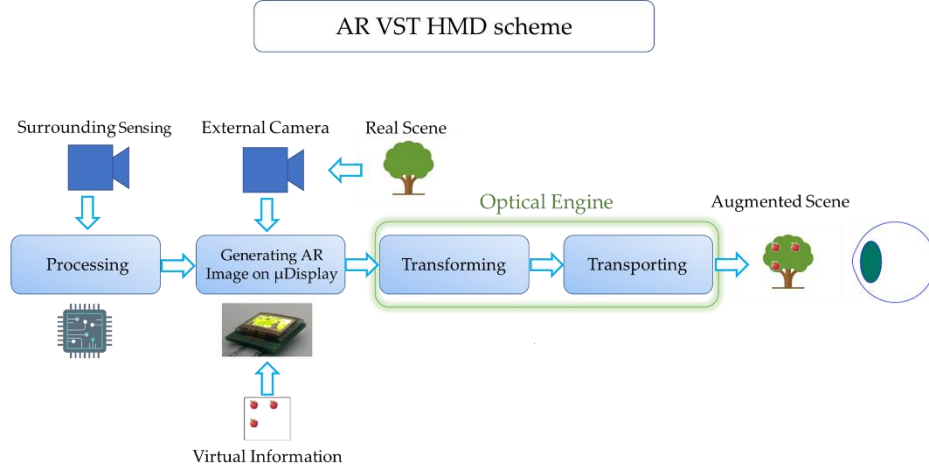
**Fig. 1.** Schematic representation of the Video See-Through Augmented Reality paradigm
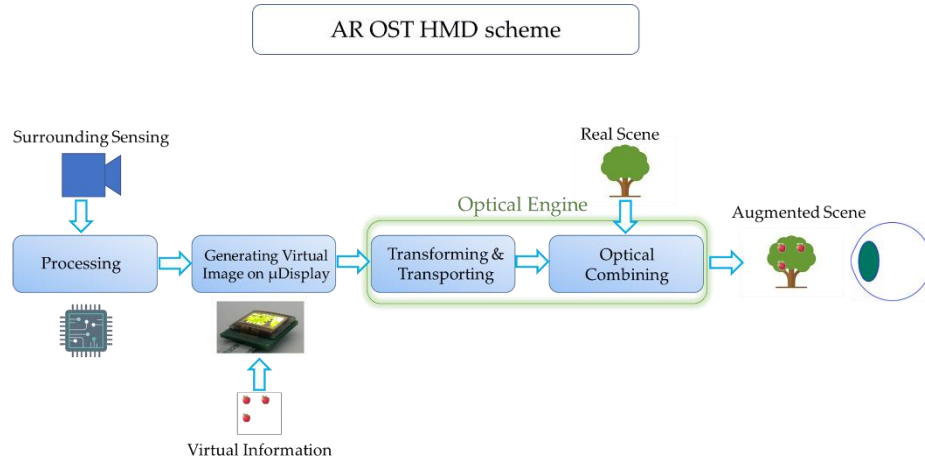


**Fig. 2.** Schematic representation of the Optical See-Through Augmented Reality paradigm

Our goal is to design and validate a hybrid OST/VST HMD able to provide the benefits of both see-through paradigms and to ensure a perceptually consistent AR experience to the user at close distances. To do so, we are presenting, in a coherent fashion, the technical solutions implemented so far towards the realization of a novel AR HMD devoted to aid complex manipulative tasks. Each technical solution was tested on a different embodiment of hybrid OST/VST.

This section is structured as follows: Section 2.1 reports on a novel solution that allows the development of stereoscopic AR HMDs able to provide both the see-through mechanisms (optical see-through and video see-through); Section 2.2 describes a software solution that is able to partially recover natural stereo fusion of the scene in quasi-

orthoscopic video see-through HMDs; finally, Section 2.3 describes a closed-loop method for the automatic calibration of optical see-through HMDs with infinity focus.

## 2.1    Hybrid Video/Optical See-Through HMD

In 2017, we presented a novel approach for the development of an HMD able to provide both the see-through modalities [23]. The proposed solution is based on the use of a pair of liquid-crystal (LC) shutters that can be electronically controlled so to modify the transparency of a generic see-through display. This allows us to switch from the augmented and camera-mediated view, with the optical shutter on closed-state (i.e., VST mode), to the see-through (i.e., OST mode) view, with the optical shutter on open-state. Under OST mode, only the computer-generated elements are rendered onto the microdisplay whereas under VST mode the real 2D scene and virtual content are digitally blended and then rendered onto the display. A first custom-made prototype implementing such mechanism was realized by assembling the two optical shutters on a commercial waveguide-based OST display opportunely modified for housing also the stereo cameras needed for the VST modality (Fig. 3).

The proposed solution is therefore able to combine the benefits of both the AR modalities since it allows switching to the most suitable modality during use and without relying on any moving part (i.e. as with a physical shutter).
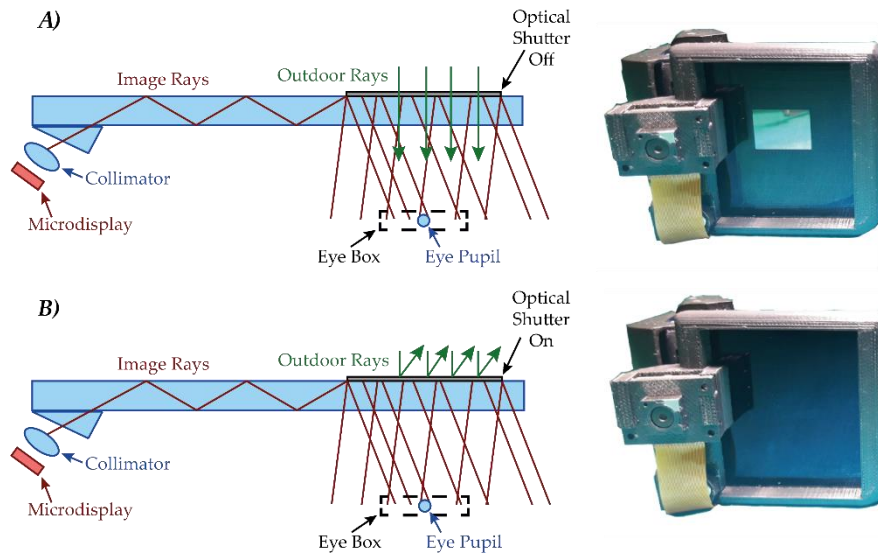


**Fig. 3.** Switch between Optical and Video See-Through modality by means of the optical shutter. On the top (A), the shutter in open-state, on the bottom (B) the shutter in closed-state

## 2.2 Perspective Preserving Solution for Quasi-Orthoscopic Binocular Video See-Through HMDs

From a human-factor standpoint, non-rigorously stereo-orthoscopic VST HMDs, as the one presented in the previous section, raise issues related to the user's interaction with the augmented content. In such systems, depth perception through stereopsis is adversely affected by sources of spatial perception errors. The issues affecting depth perception are mostly bound to the non-orthoscopic placement of the stereo cameras on the HMD.

In a study published in 2018 we presented a software solution aimed at mitigating the effect of the eye-to-camera parallax [24]. The idea is to partially resolve eye-camera parallax by opportunely warping the camera images through a perspective-preserving homographic transformation that accounts for: the intrinsic parameters of the cameras ($K_C$), the intrinsic parameters of the see-through displays ($K_D$), the geometry of the binocular VST HMD ($R_C^D | \vec{t}_C^D$), and the distance ($d^{C \to \pi}$) and orientation ($\vec{n}^T$) of the reference plane with respect to the user (Fig. 4) [25, 26]. The plane-induced homography is:

$$H_C^D = \left( K_D \left( R_C^D + \frac{\vec{t}_C^D \cdot \vec{n}^T}{d^{C \to \pi}} \right) K_C^{-1} \right)$$

(1)

While the pixel-to-pixel relation between camera points and display points is:

$$\lambda x_D = H_C^D (R_C^D, t_C^D, K_C, K_D) x_C$$

(2)
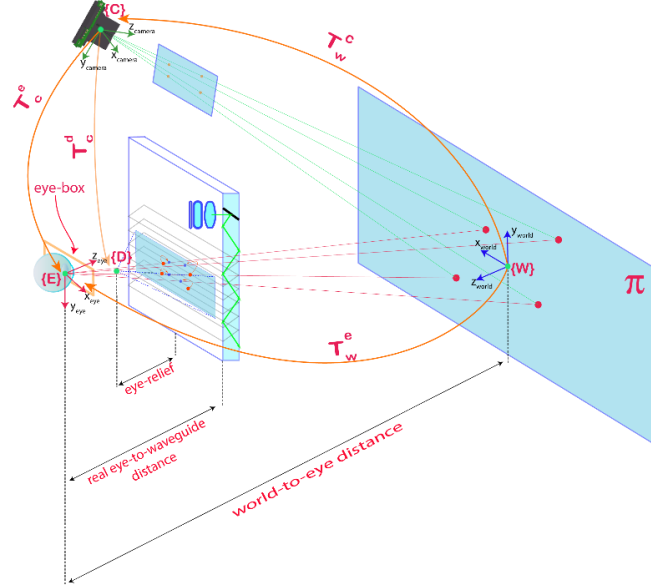


**Fig. 4.** Rigid transformation involved in the evaluation of the perspective preserving homography induced by a reference plane.

By means of this plane-induced homography applied over the camera frames, the perceived distortion of space around the pre-defined reference plane can be dramatically reduced and therefore the user is able to interact with the augmented scene with reduced discomfort.

In Fig. 5 a new custom-made prototype of the hybrid OST/VST HMD used for testing the proposed solution is shown. As in the first prototype, the visor is based on a reworked version of a commercial binocular OST HMD (DK-33 by LUMUS). The binocular HMD features a quasi-orthoscopic placement of the external cameras with an anthropometric interaxial distance.
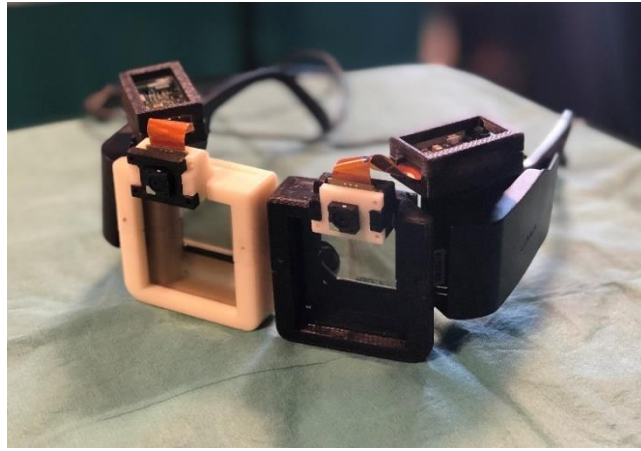


**Fig. 5.** Embodiment of the hybrid Video/Optical See-Through HMD used for assessing the perspective preserving solution that mitigates the parallax due to a non-orthoscopic setting of the.

### 2.3 Closed – Loop Calibration for Optical See-Through HMDs with Infinity Focus

To achieve an accurate alignment between the user's view of the real 3D world and the computer-generated 2D graphics projected onto the see-through display, in any OST HMD a dedicated calibration routine is needed to estimate the projective parameters of the combined eye-display pinhole model. State-of-the-art OST calibration methods models the eye-NED system as an off-axis pinhole camera model, and therefore include the contribution of the eyes' positions also into the modelling of the intrinsic matrix of the eye-NED. In a study recently presented at ISMAR 2018 [27], we presented a method for robustly calibrating OST NEDs that explicitly ignores this assumption and that is specifically suited for OST displays having the imaging plane at infinity as the one used in our second prototype of hybrid OST/VST HMD.

Our assumption is that, being the imaging plane at infinity, the eye-NED can be modelled as an on-axis pinhole camera and therefore the eye-to-display parallax only affects the extrinsic component of the homographic transformation between eye and display in its centered position (i.e., at the center of the display eye-box). The overall

transformation matrix that properly maps the world points onto the image points of the display is:

$$P_{def} = H_E^D \cdot K_D \cdot T_W^E = \lambda K_{eq} \cdot [R_{eq} \,|\, t_{eq}] \qquad (3)$$

Where $K_D \cdot T_W^E$ is the projection matrix associated to the display. In order to determine the correct points onto the display (with no parallax), a perspective preserving homography ($H_C^D$) is to be estimated to correlate the two different pinhole models, in our case the eye/display model in its actual position and the eye/display model in its centered position. This transformation compensates the eye to display parallax. By opportunely decomposing the resulting $P_{def}$, we obtain an equivalent intrinsic matrix and an equivalent extrinsic matrix. These two matrices are then used for modelling the parameters of the virtual camera correctly. By using $K_{eq}$ and $R_{eq} \,|\, t_{eq}$, in the rendering engine, the OST AR system can project pixels onto the display so to be properly aligned with their corresponding 3D point in the world.

The calibration procedure was tested on the same prototype presented in Section 2.2, here used under OST modality. The promising results of the calibration are shown in Fig. 6.



**Fig. 6.** Example of OST view after calibration.

## 3    Discussion and Future works

In this work, we have unveiled an overview of the workflow carried out so far towards the development of a novel AR visor specifically devoted to aiding complex manual tasks within arm's reach. Overall, the achievements of the development phase carried out so far are:

- Implementation of a hybrid HMD which can work both under OST and VST modality and switch between each modality upon request. The switching

mechanism is made possible by means of a pair of opto-electronic LC-shutters placed ahead of the optical combiners of the OST HMDs.

- A quasi-orthoscopic setting of the visor under VST modality, with a reduced eye-to-camera parallax and an anthropometric interaxial distance that reduces distortions in depth perception.
- A software solution that further mitigates the effect of the eye-to-camera parallax by means of a perspective-preserving image warping applied to the camera frames under VST modality. This transformation allows the recovering of a correct perception of the relative depths in the peripersonal space around a pre-defined reference plane.
- A closed-loop calibration routine designed for an OST HMD with infinity focus. The calibration routine yields accurate results in terms of real-to-virtual alignment also under OST modality.

As future developments, a third hybrid visor is currently under development that will comprise all the above described key features:

- A pair of OST near-eye-displays with the associated optical engines.
- A pair of opto-electronic shutters.
- A pair of external cameras needed for capturing real-world views under VST modality.
- A dedicated software framework that manages the OST/VST switching and all the needed image transformations needed for mitigating the perceptual issues related to the eye-to-camera parallax and for ensuring a perceptually consistent AR experience to the user at close distances.

This visor will undergo specific usability tests in order to assess the overall efficacy of all the proposed solutions.

## References

1. Ong, S.K., Yuan, M.L., Nee, A.Y.C.: Augmented reality applications in manufacturing: a survey. International Journal of Production Research 46, 2707-2742 (2008)
2. Henderson, S., Feiner, S.: Exploring the Benefits of Augmented Reality Documentation for Maintenance and Repair. Ieee T Vis Comput Gr 17, 1355-1368 (2011)
3. Nee, A.Y.C., Ong, S.K., Chryssolouris, G., Mourtzis, D.: Augmented reality applications in design and manufacturing. Cirp Ann-Manuf Techn 61, 657-679 (2012)
4. Wang, X., Ong, S.K., Nee, A.Y.C.: A comprehensive survey of augmented reality assembly research. Adv Manuf 4, 1-22 (2016)
5. Bower, M., Howe, C., McCredie, N., Robinson, A., Grover, D.: Augmented Reality in Education - Cases, Places, and Potentials. Ieee Ann Conf Int (2013)
6. Radu, I.: Augmented reality in education: a meta-review and cross-media analysis. Personal and Ubiquitous Computing 18, 1533-1543 (2014)

7. Bacca, J., Baldiris, S., Fabregat, R., Graf, S., Kinshuk: Augmented Reality Trends in Education: A Systematic Review of Research and Applications. Educ Technol Soc 17, 133-149 (2014)

8. Qu, M., Hou, Y.K., Xu, Y.R., Shen, C.C., Zhu, M., Xie, L., Wang, H., Zhang, Y., Chai, G.: Precise positioning of an intraoral distractor using augmented reality in patients with hemifacial microsomia. J Cranio Maxill Surg 43, 106-112 (2015)

9. Voinea, A., Moldoveanu, A., Moldoveanu, F.: Efficient Learning Technique in Medical Education Based on Virtual and Augmented Reality. Iceri Proc 8757-8764 (2016)

10. Meola, A., Cutolo, F., Carbone, M., Cagnazzo, F., Ferrari, M., Ferrari, V.: Augmented reality in neurosurgery: a systematic review. Neurosurg Rev 40, 537-548 (2017)

11. Kersten-Oertel, M., Jannin, P., Collins, D.L.: The state of the art of visualization in mixed reality image guided surgery. Computerized Medical Imaging and Graphics 37, 98-112 (2013)

12. Fida, B., Cutolo, F., di Franco, G., Ferrari, M., Ferrari, V.: Augmented reality in open surgery. Updates in surgery 70, 389-400 (2018)

13. Eckert, M., Volmerg, J.S., Friedrich, C.M.: Augmented Reality in Medicine: Systematic and Bibliographic Review. JMIR mHealth and uHealth 7, e10967 (2019)

14. Ferreira Reis, A., Wirth, G.J., Iselin, C.E.: [Augmented reality in urology : present and future]. Revue medicale suisse 14, 2154-2157 (2018)

15. Yoon, J.W., Chen, R.E., Kim, E.J., Akinduro, O.O., Kerezoudis, P., Han, P.K., Si, P., Freeman, W.D., Diaz, R.J., Komotar, R.J., Pirris, S.M., Brown, B.L., Bydon, M., Wang, M.Y., Wharen, R.E., Jr., Quinones-Hinojosa, A.: Augmented reality for the surgeon: Systematic review. The international journal of medical robotics + computer assisted surgery : MRCAS 14, e1914 (2018)

16. Dunn, J., Yeo, E., Moghaddampour, P., Chau, B., Humbert, S.: Virtual and augmented reality in the treatment of phantom limb pain: A literature review. Neurorehabilitation 40, 595-601 (2017)

17. Cutolo, F., Meola, A., Carbone, M., Sinceri, S., Cagnazzo, F., Denaro, E., Esposito, N., Ferrari, M., Ferrari, V.: A new head-mounted display-based augmented reality system in neurosurgical oncology: a study on phantom. Comput Assist Surg 22, 39-53 (2017)

18. Cutolo, F., Freschi, C., Mascioli, S., Parchi, P., Ferrari, M., Ferrari, V.: Robust and Accurate Algorithm for Wearable Stereoscopic Augmented Reality with Three Indistinguishable Markers. Electronics 5, 59 (2016)

19. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. Ieee Computer Graphics and Applications 21, 34-47 (2001)

20. Cutolo, F., Badiali, G., Ferrari, V.: Human-PnP: Ergonomic AR Interaction Paradigm for Manual Placement of Rigid Bodies. Augmented Environments for Computer-Assisted Interventions, Ae-Cai 2015 9365, 50-60 (2015)

21. Cutolo, F., Carbone, M., Parchi, P.D., Ferrari, V., Lisanti, M., Ferrari, M.: Application of a New Wearable Augmented Reality Video See-Through Display to Aid Percutaneous Procedures in Spine Surgery. Lect Notes Comput Sc 9769, 43-54 (2016)

22. http://www.vostars.eu/

23. Cutolo, F., Fontana, U., Carbone, M., D'Amato, R., Ferrari, V.: Hybrid Video/Optical See-Through HMD. Adjunct Proceedings of the 2017 Ieee International Symposium on Mixed and Augmented Reality (Ismar-Adjunct) 52-57 (2017)

24. Cutolo, F., Fontana, U., Ferrari, V.: Perspective Preserving Solution for Quasi-Orthoscopic Video See-Through HMDs. Technologies 6, (2018)

25. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2003)

26. Tomioka, M., Ikeda, S., Sato, K.: Approximated User-Perspective Rendering in Tablet-Based Augmented Reality. Int Sym Mix Augment 21-28 (2013)

27. Fontana, U., Cutolo, F., Cattari, N., Ferrari, V.: Closed – Loop Calibration for Optical See-Through Near Eye Display with Infinity Focus. Adjunct Proceedings of the 2018 Ieee International Symposium on Mixed and Augmented Reality (2018)