Robust Head Mounted Wearable Eye Tracking System for Dynamical Calibration

Antonio Lanata
Research Center "E.Piaggio", University of Pisa.

Alberto Greco
Research Center "E.Piaggio", University of Pisa.

Gaetano Valenza
Research Center "E.Piaggio", University of Pisa.

Enzo Pasquale Scilingo
Research Center "E.Piaggio", University of Pisa.

In this work, a new head mounted eye tracking system is presented. Based on computer vision techniques, the system integrates eye images and head movement, in real time, performing a robust gaze point tracking. Nystagmus movements due to vestibulo-ocular reflex are monitored and integrated. The system proposed here is a strongly improved version of a previous platform called HATCAM, which was robust against changes of illumination conditions. The new version, called HAT-Move, is equipped with accurate inertial motion unit to detect the head movement enabling eye gaze even in dynamical conditions. HAT-Move performance is investigated in a group of healthy subjects in both static and dynamic conditions, i.e. when head is kept still or free to move. Evaluation was performed in terms of amplitude of the angular error between the real coordinates of the fixed points and those computed by the system in two experimental setups, specifically, in laboratory settings and in a 3D virtual reality (VR) scenario. The achieved results showed that HAT-Move is able to achieve eye gaze angular error of about 1 degree along both horizontal and vertical directions.

Keywords: Eye tracking, head movement, vestibulo-ocular reflex, gaze point, statistical comparison, dynamical calibration.

Introduction

Eye Gaze Tracking (EGT) system includes a device able to continuously acquire and follow eye position over time and compute gaze point coordinates in the environment around the subject through an analytical relationship. In the current literature, EGTs have been exploited in different fields, ranging from the medical field (detecting the relationship between oculomotor characteristics and cognition and/or mental states) to user interaction (helping people with disabilities or level attention detection), from multimedia to product design (C. Morimoto & Mimica, 2005). Currently, most EGTs are based on the Video-OculoGraphy (VOG) technique, which is a method for tracking eye movements through computer vision techniques used to process eye images (Van der Geest & Frens, 2002). Through VOG, pupil position and iris landmarks are detected by means of image processing algorithms and used to calculate both eye rotation angles and the eye center. VOG-based EGTs can be classified into two main categories identified by the position of the camera, that is dedicated to acquiring eye images with respect to the user. In particular, if the camera is placed on a fixed support in front of the subject the systems are named *remote* EGTs, when placed on the head of the subject they are named $Head - Mounted$ EGTs (HMEGTs), i.e. portable EGTs. Of course, the choice between the two types of systems poses different technical and methodological challenges. Generally, in laboratory environment *remote* EGTs are employed. They allow a quite precise measure, but impose limitations in the kind of information that can be retrieved. Specifically, *remote* EGTs could make use of either the chin support to block the user's head resulting unsuitable for long term acquisitions but with a good accuracy, or intelligent algorithms such as Active Shape Model (ASM) (T. F. Cootes, Taylor, Cooper, & Graham, 1995) to detect the user's eye allowing for limited head movements. They often require really expensive high definition camera that make *remote* EGTs

suitable for investigating oculomotor strategy in neu-rological investigation, (Meyer, Böhme, Martinetz, & Barth, 2006). On the contrary, HMEGTs have opened new scenarios for research and for markets where the user is free to move his head on which the equipment is directly mounted (Zhu & Ji, 2007). HMEGTs make it possible to investigate the eye-movement during nat-ural tasks in uncontrolled environments (M. Land & Lee, 1994) and in real life scenarios. Although in the real life scenario some uncontrollable external factors such as illumination changes could make the experi-mental setting time-consuming and not easily repeat-able, at the same time HMEGTs allow for dynamical monitoring even in case of time-variant experimental paradigms (Hayhoe & Ballard, 2005). One of the most significant approach to prevent luminosity change is-sues is based on infrared (IR) illumination. In partic-ular, spectral (reflective) properties of the pupil, un-der near-IR illumination, are exploited to maximize the image-contrast (C. Morimoto, Amir, & Flickner, 2002; C. Morimoto & Mimica, 2005). Nevertheless, in order to combine the advantage of supervised laboratory set-tings with real-life-like environments one of the most interesting approach is based on virtual reality im-mersive space (Fahrenberg, Myrtek, Pawlik, & Perrez, 2007). Virtual Reality (VR) offers an excellent compro-mise between laboratory and natural world account-ing for a systematic control of the stimuli and the vari-ables involved. Given the advantages of immersive VR spaces, HMEGTs have opened up new exploiting direc-tions of the human interaction over time (Henderson, 2003; Jacob & Karn, 2003) focussing on understand-ing and coding naturalistic behaviour (Hayhoe & Bal-lard, 2005). However, even though many researches exploited HMEGTs for studying user attention, their perception of surrounding objects, and user interest as well as eye pattern in affective stimulation obtain-ing good results, (Lanatà, Valenza, & Scilingo, 2013; de Lemos, Sadeghnia, Ólafsdóttir, & Jensen, 2008; Partala & Surakka, 2003; Lanata, Armato, Valenza, & Scilingo, 2011), the performances of these systems drastically decade when the user head is free to move for example in the investigation of VR space navigation (C. Mori-moto & Mimica, 2005). In this context, this work aims at developing a new HMEGT (named HAT-Move) to be used either in real life or in completely immersive 3D-virtual reality worlds.

Head Movement Issue

The possibility of freely moving the head with HMEGTs requires a robust and reliable identification and tracking of the pupil center and gaze point. Gener-ally, all of the eye tracking systems developed both for market and research purpose make use of an uneasy-to-perform calibration procedure, that should be very accurate. As a matter of fact, the better is the calibration the better is the outcome of the EGT. In particular, given a certain number of points (i.e, calibration points) in the real world and fixed their coordinates on the acquired image, the calibration is an analytical relationship that maps the coordinates of the eye gaze computed by the system into the coordinates of the calibration points (Hartley & Zisserman, 2000).. This procedure can be differentiated by both the number of calibration points and the kind of mathematical models used to generate this relationship, (Ramanauskas, Daunys, & Dervinis, 2008). In the literature many efforts have been spent to improve gaze estimation in terms of increasing toler-ance to head movements, and as a consequence to im-prove and simplify the calibration process (Cerrolaza, Villanueva, & Cabeza, 2008; Johnson, Liu, Thomas, & Spencer, 2007; Evans, Jacobs, Tarduno, & Pelz, 2012). In fact, in spite of several attempts of EGTs to enable head movements which can be found in the literature, the head movement keeps remaining an open issue. Bab-cock et al. introduced a projected grid of 9-points in front of the person, (Babcock, Pelz, & Peak, 2003), as a reference for calibrating the eye position with the cam-era scene image. Even though this system is designed to be used in natural task it requires many calibration steps during the experiments. Moreover, since a mea-sure of the movement of the head is missing the recal-ibration process is based on human-operator expertise. Rothkopf et al. tried to use head movements to dis-ambiguate the different types of eye movements when subjects move head and body (Rothkopf & Pelz, 2004). However, the performance of the algorithm declines for large motions in roll, and sometimes the algorithm fails completely. Data analysis showed that the record-ings suffered from a significant noise level compared to experimental conditions in which the subject does not move neither the body nor the head, highlighting that much more complex patterns of eye and head move-ments had to be considered. Johnson et al. allowed relatively unrestricted head and/or body movements (Johnson et al., 2007) tracking them by a visual motion tracker but large errors were found. Moreover two rel-evant drawbacks have been reported: the eye tracker became less reliable at larger eye rotations and errors arose if there was a shift in the relative position 3D ref-erence on the eye-tracker's visor and the participant's head. As a result of this background it is worthwhile noting that augmenting the eye trackers capacity of dif-ferentiating among the movements of the head and the eyes could improve their ability in performing an ef-fective and robust gaze detection. In this view, HAT-

Move aims at overcoming the state of the art proposing a novel method for integrating head movement contribution in a free to move eye gaze detection.

Eye Movement

One of the crucial point of the proposed system is the ability in detecting and monitoring the eye movement. Specifically, eye movements can be classified as pursuit and smooth pursuit, saccades and nystagmus movement. Pursuit movements or smooth pursuit are eye movements used for tracking an object in movement, therefore a moving image has to remain constrained to the fovea to achieve a stable image seen by the user. Fovea is a small area of the retina with a very high visual acuity, and it covers about 2 degrees of visual angle. Saccades are rapid movements of eyes for scanning a visual scene. They are also present when the subject is fixating one point (Findlay, 2009). Nystagmus indicates involuntary eye movements. More specifically, when the head rotates about any axis, distant visual images are sustained by rotating eyes in the opposite direction on the respective axis. Physiologically, nystagmus is a form of involuntary eye movement that is part of the vestibulo-ocular reflex (VOR), characterized by alternating smooth pursuit in one direction and saccadic movement in the other direction (Westheimer & McKee, 1975). As a matter of fact the brain must turn the eyes so that the image of the fixed object falls on the fovea. Of note, the pursuit system keeps up the moving object in order to allow the brain to move the eyes in opposite direction to head motion, otherwise, image slip on the retina and a blurred image is produced.

In this work, we propose an innovative version of the already partially proposed wearable and wireless eye tracking system (Armato, Lanatà, & Scilingo, 2013). The new system, HAT-Move, is comprised of only one light camera able to capture simultaneously the scene and both eyes of the subject through a mirror. The system calculates the gaze point in real-time both indoor and outdoor. Moreover, a precise and fast Inertial Motion Unit (IMU) was added to render it independent from involuntary head movements, during the calibration phase, or to render it robust to head movements needed to focus on objects out of the visual space. Specifically, in this study we evaluated the contribution, in terms of angular error, of the head movement integration with respect to standard eye tracking. Moreover, HAT-Move system inherits from the previous version the robustness against illumination variation implementing a normalization of the illumination through a Discrete Cosine Transform
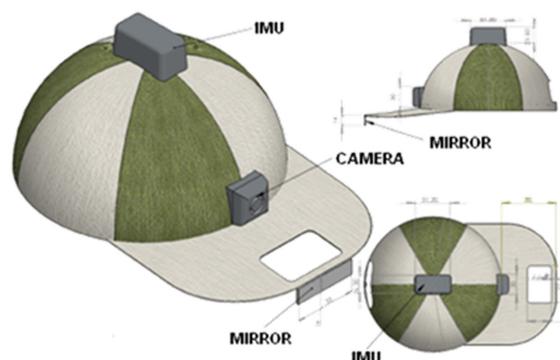
Figure 1. Version I: "Baseball-like" hat configuration.
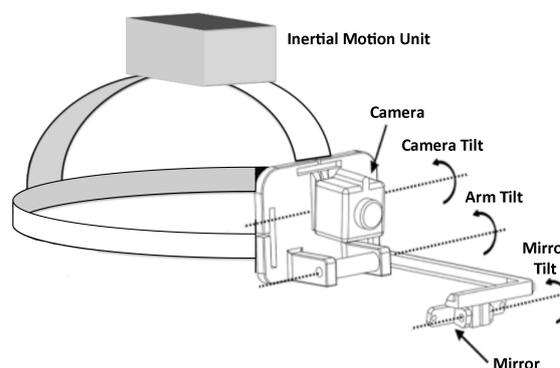
Figure 2. Version II: "Head band" configuration.

(DCT) of the acquired images. Further details are reported in Section .

Materials and Methods

The HAT-Move eye tracking system was developed in two configurations: the "baseball" hat (see fig.1) and the head band (see fig. 2). They are technically and functionally equivalent, although the former can be considered aesthetically more pleasant. It is comprised of a wireless camera that is light and small, with an Audio/Video (A/V) transmitter of up to $30m$ of distance. The camera has a resolution of 628 x 586 pixels with $F2.0$, $D45°$ optic, and 25 frames per second (f.p.s.). In addition, the InfraRed (IR) filter, which is actually present in each camera, was removed and a wide-angle-lens was added allowing to enlarge the view angle and acquire natural infrared components increasing both the image resolution and the contrast between pupil and iris. This system is able to simultaneously record the visual scene in front of the subject and eye position. This is achieved through a mirror(5 x $0.6cm$) placed in front of the user's head (see fig. 2). The system is completely customizable on the user's forehead (see fig.2). In addition, a wireless Inertial Motion Unit is placed atop the head close to the azimuth rota-
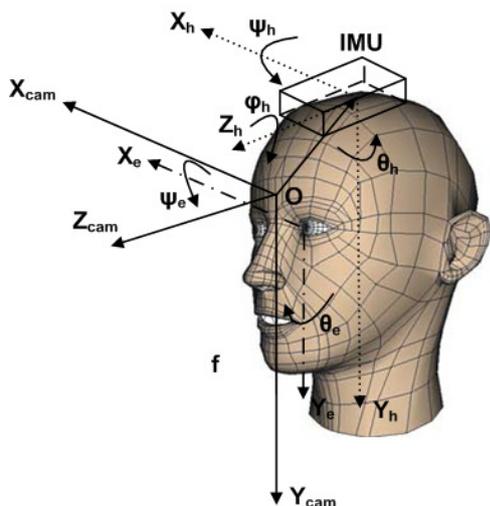
Figure 3. Representation of the head along with IMU. In the figure X_e, Y_e are the axis in the reference frame of the eye; X_{cam}, Y_{cam}, Z_{cam} are the axis in the reference frame of the camera; X_h, Y_h, Z_h are the axis in the reference frame of the center of the head where the IMU is placed; ψ_h, θ_h, ϕ_h are the Euler-angles of the head rotation while ψ_e, θ_e, are the Euler-angles of the eye rotation.

tion center (see fig. 3). The IMU allows for the acquisition of head movements and rotations during natural activities, allowing for the correction of eye gaze estimation taking into account both the movements during the calibration phase and the "Vestibulo-Ocular Reflex" (VOR) contributions. The adopted IMU provides the three rotation Euler-angles of the head (see Fig.3) with a sampling frequency of 100 Hz. The system is intended to be wearable, minimally invasive, capable of eye tracking, estimating pupil size, lightweight, and equipped for wireless communication. Moreover the system is designed to be attractive and aesthetic (in a baseball-like version), and able to process the eye-gaze pattern in real-time. The HAT-Move system uses a passive approach for capturing ambient light reflected by the eye (VOG). The customized image acquisition system is able to acquire both natural light along with its IR components that are already present in the natural light bandwidth. Therefore, the system presents IR lightning advantages increasing the pupil-iris contrast and avoiding any possible eye injury due to artificial IR illuminators. The block diagram in Figure 4 shows the methodology used to process the acquired image, in which both eyes and scene are presented. The whole processing chain is comprised of a series of algorithms for the detection of the eye center and for the integration of the head rotation angles and the correction of involuntary head movements. Specifically, the eye center detection is achieved through the fol-
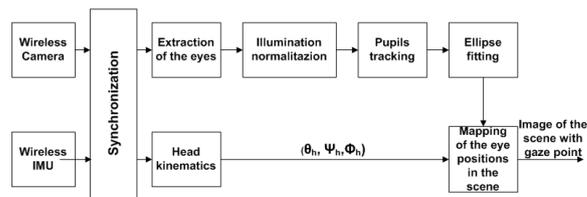
Figure 4. Block diagram showing all the algorithmic stages of the processing of eyes and outside scene.

lowing steps: eye region extraction algorithm, photometric normalization algorithm of illumination, extraction of the pupil contour, and ellipse fitting algorithm as well. Afterwards, the center of the eye is detected, Eulero-head-angles together with the pupil center are integrated into the mapping function to map the eye center and movements into the image plane. The processing chain is fully described in the following sections.

Extraction of the eye region

The region containing the eye must be extracted from the image in which both scene and eyes are simultaneously acquired (see fig. 5). It is obtained through an automatic detection of the rectangular area including the eye (named Region Of Interest (ROI)), (see fig. 5 and 6). For this purpose a modified version of the Active Shape Model was used (ASM). The ASM is an artificial intelligence algorithm generally used for the detection of objects by means their shape (T. F. Cootes et al., 1995), in particular, several modified versions of this algorithm were implemented over time for face detection applications. More specifically, after defining the "landmarks", or distinguishable points present on every image, e.g. the eye corner locations, the shape was represented by an x-y coordinate vector of those landmarks which characterize the specific shape. It is worthwhile noting that a shape of an object does not change when it is moved, rotated or scaled, and the average Euclidean distance between the shape points is minimized through a similar transformation in order to align one shape with an other. Indeed, ASM is a recursive algorithm that starts from an initial tentative shape and adjusts the locations of shape points by means of a template matching the image texture around each point, aimed at adapting the initial shape to a global shape model. The entire search is repeated at each level in an image pyramid, from coarse to fine resolution; specific details can be found in (T. Cootes & Taylor, n.d.). In our study, after the HAT-Move is worn, the first thirty seconds were used to train the ASM and then to detect the ROI. Since the system is mounted on the head, the extracted ROI does not change through-
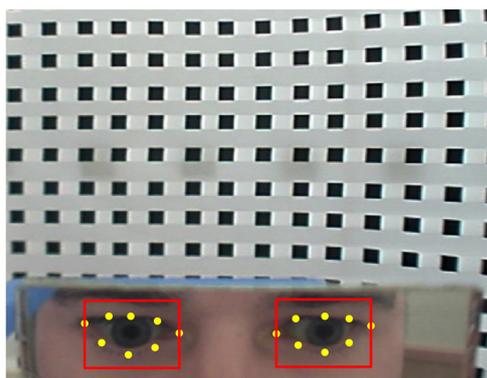
Figure 5. Example of single frame captured by the camera. The rectangular area marked by red represents the ROI.

Figure 6. Red component of the ROI.

out the experiment. In addition, only the red-image-component is converted in gray scale and used as input to the other processing blocks (see fig. 6). This image component, indeed, is especially helpful in enhancing the contrast between pupil and background.

Illumination normalization

Illumination normalization relies on an algorithmic strategy to keep stable illumination conditions throughout the captured images. More specifically, environmental illumination changes are reflected in the acquired images as a variation of the eye representation in terms of intensity thereby strongly reducing the contrast between eyes and landmark. The standard approach is based on the Retinex theory, (E. H. Land & McCann, 1971) whereby the effect of a non-uniform illumination is eliminated and is completely independent of any a-priori knowledge of the surface reflectance and light source composition. According to this theory, the image intensity $I(x,y)$ can be simplified and formulated as follows:

$$I(x,y) = R(x,y)L(x,y) \qquad (1)$$

where $R(x,y)$ is the reflectance and $L(x,y)$ is the illuminance at each point (x,y). The luminance L is assumed to contain low frequency components of the image while the reflectance R mainly includes the high frequency components of the image. The technique

Figure 7. Eye image after the application of illumination normalization algorithm by DCT

used here is the result of a our previous work, in which seven different photonormalization techniques were tested and compared, i.e. Single Scale Retinex (SSR), Multi Scale Retinex (MSR), Single scale self-Quotient Image (SQI), Multi Scale selfQuotient image (MSQ), Discrete Cosine Transform(DCT), WAvelet-based Normalization algorithm (WAN), and Normalization Histogram (NH). In that paper our key approach was to evaluate, in terms of accuracy of eye tracking, pupillometry and execution time under three different brightness conditions: Laboratory, Sunlight, and Darkness. Results showed that the DCT method was the most effective in terms of accuracy of gaze point and pupillometry. More details can be found in (Armato et al., 2013). The DCT technique compensates for illumination variations by truncating the low frequency components of the discrete cosine transform (DCT) in the logarithmic domain. In the logarithmic domain the theory is formulated as follows:

$$logI(x,y) = log(R(x,y) + log(L(x,y) \qquad (2)$$

Figure 7 shows the output of the DCT algorithm applied to gray scale image reported in figure 6.

Pupil tracking and ellipse fitting

This section deals with the method used to extract the pupil contours. The method is comprised of several blocks in which the acquired eye image is first binarized in order to separate the pupil from the background by using a threshold in the image histogram; then a geometrical method was used to reconstruct pupil contours and to remove outliers belonging to the background, details of this algorithm con be found in (Armato et al., 2013). Following the geometrical detection of the points belonging to the pupil, an ellipse fitting algorithm is implemented for pupil contour reconstruction and for detecting the center of the eye. In the literature, the ellipse is considered to be the best geometrical figure representing the eye, being the eye image captured by the camera a projection of the eye in the mirror. Over the last decade many ellipse fitting algorithms have been proposed (Forsyth &
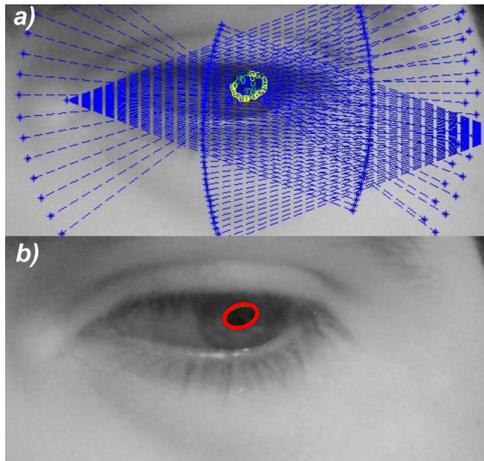
Figure 8. Results of the pupil tracking and ellipse fitting algorithm. a) In blue are represented geometrical construction for pupil contour detection; Contour points are evinced in yellow. b) In red the fitted ellipse is highlighted.

Ponce, 2002; Bennett, Burridge, & Saito, 2002), although most work offline. In our system we used the Least Square (LS) technique, which is based on finding a set of parameters that minimizes the distance between the data points and the ellipse, (Fitzgibbon, Pilu, & Fisher, 2002). According to the literature this technique fulfills the real time requirement (Duchowski, 2007). Specifically, we follow the algorithm proposed by Fitzgibbon et al., which is a direct computational method (i.e. B2AC, it is the exact name of Fitzgibbon's algorithm which is based on the solution of a quadratic polynomial) based on the algebraic distance with a quadratic constraint, in which a gaussian noise is added for algorithm stabilization, (Maini, 2005). Afterwards, the center of the eye is computed as the center of the fitted ellipse. A detailed description of the methods can be found in (Armato et al., 2013). Figure 8 shows the result of pupil tracking and the ellipse fitting algorithm for reconstructing pupil contours.

Mapping of the position of the eye

The mapping procedure aims at associating the instantaneous position of the center of the eye to a point of the scene. This point is named *gazepoint*. This procedure is mainly based on a mathematical function, named *mappingfunction*, which is an equation system constituted of two second order polynomial functions (C. H. Morimoto, Koons, Amir, & Flickner, 2000) defined as:

$$x_{si} = a_{11} + a_{12}x_{ei} + a_{13}y_{ei} + a_{14}x_{ei}y_{ei} + a_{15}x_{ei}^2 + a_{16}y_{ei}^2 \quad (3)$$

$$y_{si} = a_{21} + a_{22}x_{ei} + a_{23}y_{ei} + a_{24}x_{ei}y_{ei} + a_{25}x_{ei}^2 + a_{26}y_{ei}^2 \quad (4)$$
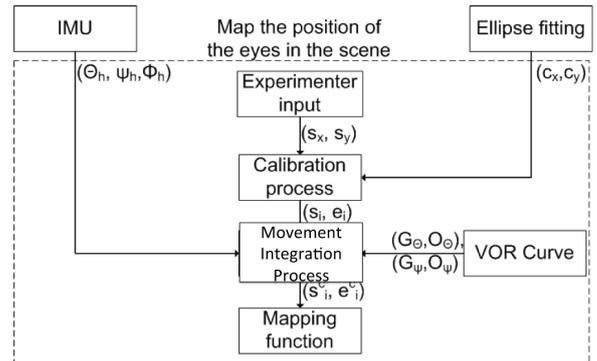
Figure 9. Block Diagram of the mapping function calculation process.

where x_{si}, y_{si} are the coordinates of a point on the image plane (i.e. the coordinates of the point on the screen mapped into the image plane captured by the camera), and x_{ei}, y_{ei} are the coordinates of the center of the eye coming from the ellipse fitting block, referred to the image plane as well. The procedure is intended to solve the equation system by means of a calibration process. Once the system is positioned onto the subject's head in a manner that eyes and scene are simultaneously presented in the image captured by the camera, the user is asked to look at some specific points on the screen (calibration process). These points are identified by coordinates $s_i = (x_{si}, y_{si})$ referred to the image plane (i.e. the image captured by the camera), (see fig.5). Since the coordinates of the calibration points are known to solve the equation system, we have to compute the coefficients $a_{1,1-to-6}$, and $a_{2,1-to-6}$ that are unknowns. The results are achieved because each calibration point defines 2 equations, and, considering a 9-point calibration process, the system is over constrained with 12 unknowns and 18 equations and can be solved using Least Square Method (LSM). Head movements mainly affect the calibration process, resulting in artifact movement that degrades eye estimation and consequentially the point of gaze as well. Two different problems related to these movements arise. The first consists of the modification of the image plane position, which follows the head rotations being attached to the forehead, while the second is due to the involuntary movement. In this work an innovative integration process is implemented, as described in the next paragraph.

Movement Integration Process

This integration process is related to the adjustment of the eye gaze as a consequence of changes in the calibration plane orientation with respect to the camera and the compensating eye rotations against head rota-

6

tions. These issues are mainly due to the user's inability to hold the head still for the whole duration of the calibration process. Consequently, these movements reduce the system accuracy. This process is based on data gathered from the IMU. First of all, according to figure 10, $O_hX_hY_hZ_h$, we define the cartesian system in the reference frame of the IMU; $OX_{cam}Y_{cam}Z_{cam}$ the cartesian system in the reference frame of the camera; OX_eY_e a cartesian system in the reference frame of the center of the eye; $O'x_iy_i$ the cartesian system on the image plane; f the focal distance; $c(x_c,y_c)$ the projection of the central point of the calibration plane onto the image plane; $s_i(x_{si},y_{si})$ the projection of $P(x,y,z)$, which is a generic calibration point, onto the image plane. Moreover, we also define the following rotations: θ_h, ψ_h and ϕ_h head rotation angles around Y_h, X_h, and Z_h axes, respectively; and θ_e and ψ_e the eye rotation angles around Y_e and X_e, respectively. The Movement Integration (MI) is performed during the acquisition of the 9 target points $s_i = (x_{si},y_{si})$, the 9 points related to eye positions $e_i = (x_{ei},y_{ei})$, and synchronously the Euler angles of the head (θ_h, ϕ_h and ψ_h see fig.10). In
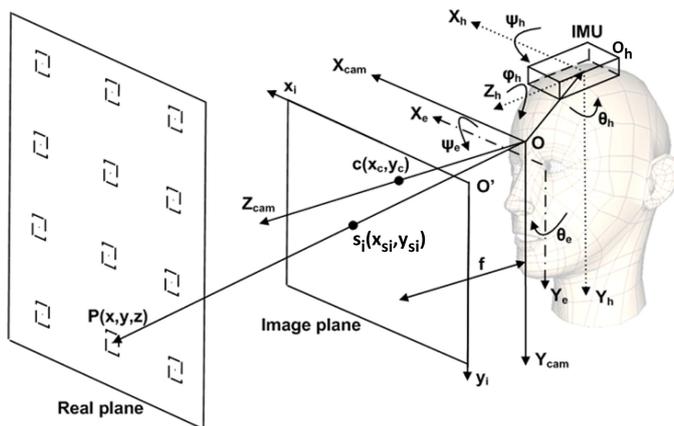
Figure 10. Representation of the head, IMU, image plane and real plane, respectively. In the figure, X_e, Y_e are the axes in the reference frame of the eye; $X_{cam}, Y_{cam}, Z_{cam}$ are the axes in the reference frame of the camera; x_i,y_i are the axes in the reference frame of the image plane; X_h, Y_h, Z_h are the axes in the reference frame of the center of the head where the IMU is placed; ψ_h, θ_h, ϕ_h are the Euler-angles of the head rotation while ψ_e, θ_e, are the Euler-angles of the eye rotation; $c(x_c,y_c)$ is the projection of the central point of the calibration plane on the image plane; $s_i(x_{si},y_{si})$ the projection of $P(x,y,z)$, which is a generic calibration point, on the image plane.

particular, the MI process performs both the realignment of eye center position on the image plane when the VOR occurs and the remapping of the calibrated space onto the image plane when the head of the user is rotating. Hence at the end of the process, the mapping function will compute the adjusted coordinates of

the eye center x_{ei}, y_{ei}, and the corrected coordinates of the calibration point, $s_i = (x_{si},y_{si})$, both referred to the image plane. Referring to the eye rotation angles, they were estimated taking advantage of VOR contributions by means of Vestibulo-Ocular calibration curves. These curves are estimated to quantify eye rotations by a mathematical model for transforming eye rotations, expressed in degrees, into movements of the eye center along the vertical and horizontal axes, expressed in pixels (Crawford & Vilis, 1991). Here, vestibulo-ocular curves are two curves (one is for the rotation ψ around x axis and the other is for the rotation θ around y axis) computed asking the user to rotate the head around the horizontal and vertical axes fixing the gaze on point "C" in front of him while acquiring eye tracking, gaze, head movements and the scene over time. A linear fitting is applied to both rotations for extracting gain and offset, as expressed by the formula:

$$\tilde{\theta}_e = G_{\theta_e}P_x + O_{\theta_e} \tag{5}$$

$$\tilde{\psi}_e = G_{\psi_e}P_y + O_{\psi_e} \tag{6}$$

Specifically, the adjusted eye rotations are computed according to the following equations:

$$\theta'_e \overset{\triangle}{=} \tilde{\theta}_e - \theta_h + \triangle(\theta_h) \tag{7}$$

$$\psi'_e \overset{\triangle}{=} \tilde{\psi}_e + \psi_h + \triangle(\psi_h) \tag{8}$$

where θ'_e and ψ'_e are the corrected eye angles and $\tilde{\theta}_e$ and $\tilde{\psi}_e$ are the eye angles of a specific subject wearing the system computed as explained in eq. 5 and 6. $\triangle(\theta_h)$ and $\triangle(\psi_h)$ are obtained as a decomposition of ϕ_h, which is the head rotation around the z axis. More specifically, when a head rotation around z occurs the IMU provides a value of ϕ different from 0 ($\phi \neq 0$) and taking into account P_x and P_y continuously provided by HAT-Move, which are related to this variation of ϕ, the values of $\triangle\theta_e$ and $\triangle\psi_e$ are obtained by means of the equations 5, 6. Afterwards the corrected angles are calculated by using equations 7 and 8, and then applying the equations 5, 6 the corrected coordinates of the eye center are obtained. At this point starting from the new eye angles by means of the eqs. 5, 6 the corrected coordinates of the eye center are simply computed. The correction of the projection of the calibration point on the rotated image plane is carried out by means of geometrical considerations. Figure 11 shows a transverse view of the subject, the image plane and the calibration plane (or real plane), during the initial conditions. In this case, the subject is aligned with the central point of the calibration plane (point number five, named C, see fig.14). Let us define a generic calibration point P
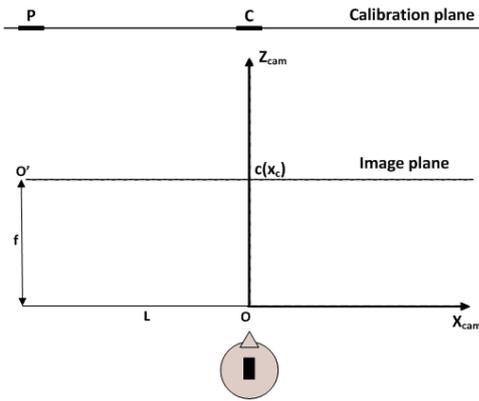
Figure 11. Initial condition of the calibration.
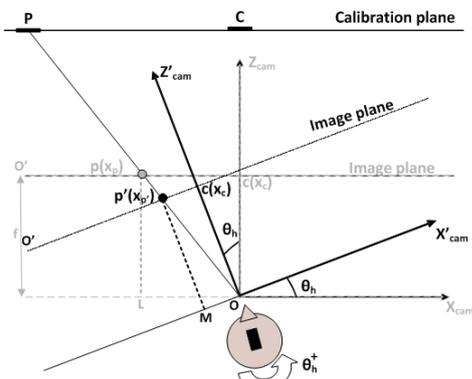
Figure 12. Representation of the positive θ_h rotation.
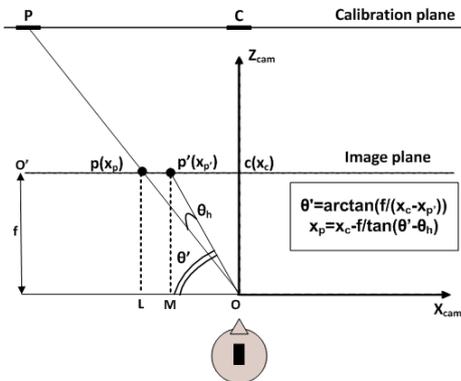
Figure 13. Representation of the error after positive θ_h rotation

and its projection onto image plane p. During a positive head rotation of θ_h (see fig.12) (which is a reasonable head movement during the calibration process) the projection of P in the image plane results p' instead of p (see fig.13). It means that an error is made in the calibration process and consequently it will be propagated in the gaze estimation. Therefore, in taking into account the acquired head rotations, it is possible to correct the projection of P in the exact position by

means of an algorithm based on some geometrical considerations reported in figure 13. In the case shown in fig.13 x'_p results to be less than x_c, but other cases can be identified:

- $x_{p'} > x_c$ and $(\theta' + \theta_h) < 90°$;
- $x_{p'} > x_c$ and $(\theta' + \theta_h) > 90°$;

and for $\theta_h < 0$, we have:

- $x_{p'} > x_c$ and $(\theta' - \theta_h) < 90°$;
- $x_{p'} < x_c$ and $(\theta' - \theta_h) < 90°$;
- $x_{p'} < x_c$ and $(\theta' - \theta_h) > 90°$;

Considering all of the cases mentioned above, the correction along the *x* axis can be reassumed by the following equations:

$$\theta' = arctan\left(\frac{f}{|x_c - x_{p'}|}\right) \quad (9)$$

if $x'_p < x_c$ then

$$x_p = xc - \frac{f}{tan(\theta' - \theta_h)} \quad (10)$$

otherwise

$$x_p = x_c + \frac{f}{tan(\theta' + \theta_h)} \quad (11)$$

The y_p corrections are the same as x_p corrections using the angle ψ_h instead of θ_h; however, the relationships are inverted with respect to the sign of the angle. Therefore:

$$\psi' = arctan\left(\frac{f}{|y_c - y_{p'}|}\right) \quad (12)$$

if $y_{p'} < y_c$ then

$$y_p = y_c - \frac{f}{tan(\psi' - \psi_h)} \quad (13)$$

otherwise

$$y_p = y_c + \frac{f}{tan(\psi' + \psi_h)} \quad (14)$$

The correction by ϕ_h gives a contribution for both *x* and *y* coordinates. All cases for both $\phi_h < 0$ and $\phi_h > 0$ can be summarized by the following equations:

if $(x'_p < x_c$ and $y'_p < y'_c)$ or $(x'_p > x_c$ and $y'_p > y'_c)$ then

$$x_p = x_c - \frac{(x_c - x'_p)}{sin(\phi')}sin(\phi' + \phi_h) \quad (15)$$

$$y_p = y_c - \frac{(y_c - y'_p)}{sin(\phi')}cos(\phi' + \phi_h) \quad (16)$$

otherwise

$$x_p = x_c - \frac{(x_c - x'_p)}{sin(\phi')}sin(\phi' - \phi_h) \quad (17)$$

─────────

[1] the symbol $\triangleq$ indicates that it is a new formulated equation

$$y_p = y_c + \frac{(y_c - y'_p)}{sin(\phi')}cos(\phi' - \phi_h) \qquad (18)$$

The correct coordinates of both the eye center and the calibration points will be used in the mapping function system (eq. 3, 4) to detect the new gaze point.

Experimental setup

This section deals with two experimental setups. The first aims at defining a protocol to validate the accuracy of the HAT-Move and the relevance of correction process (hereinafter named Accuracy Estimation) in laboratory settings, while the second is mainly a proof of concept concerning the use of the HAT-Move system in a 3D VR scenario (hereinafter named VR Estimation). To this extent we evaluated the applicability of the HAT-Move with the correction process in a quasi naturalistic environment when the subject is free to move his head and to walk around in the VR space..

Accuracy Estimation. The first experiment was performed by group of 11 subjects who did not present any ocular pathologies. All of the subjects were asked to sit on a comfortable chair placed in front of a wall (3×3 m^2) at a distance of 2 meters, while wearing the system. The wall surface was constituted of a set of black squares (2 cm per side) immersed into a white background. This controlled environment permitted to verify the system functionality during the whole experimental session. More specifically, experiments were divided into two main blocks; the first was related to VOR curves computation, and the second was related to the estimation of the eye gaze tracking. 11 subjects of both genders were recruited having different eye colors. 8 subjects had dark eyes and 3 had bright eyes. The average age was 27.8 years. For the first session, the VOR curves were computed on the whole group of subjects asking them to rotate their head first around x axis (ψ angle) and then around y axis (θ angle) fixing point C placed in front of them Possible involuntary movements of the head around the azimuth z axis (ϕ angle) have been considered through their contributions along both ψ and θ. During the experiment, the subjects were asked to rotate the head around the horizontal and vertical axes fixing point C placed in front of them.These calibration curves are intended to be used for solving the equation system 5 and 6 for each subject. Specifically, the system can be solved by imposing some constraints. The first constraint regards the initial condition; in fact, when the user is looking at the central calibration point C, before moving the head, his head-angles are forced to be null, and P_x and P_y are considered equal to the starting eye coordinates extracted from the ROI. During the required movements around
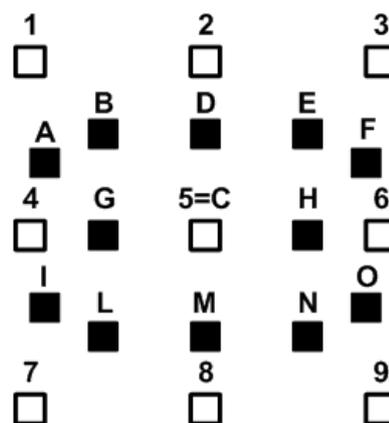
Figure 14. calibration plane

the axes, with a fixed gaze while the head is rotating, the IMU values exactly correspond to the eye rotation angles but in opposite directions (these angles were captured at a sampling frequency of 100 Hz). Therefore, by a linear fitting applied to both rotations, gains were extracted for each subject. Afterwards, by using the specific average gains (G_{θ_e}, G_{ψ_e}) in the equation system, 5 and 6 each specific offset for each subject was computed using the initial condition, where the eye angles were null. At the end of the process, given the G_{θ_e}, G_{ψ_e}, O_{θ_e}, and O_{ψ_e} as well as P_x and P_y in the image plane all corresponding eye angles ($\tilde{\theta}_e$, $\tilde{\psi}_e$) are determined. Moreover, both data with and without VOR contribution were collected and compared. The second session was organized into two phases. During the first phase the subjects were invited to look at the central point of the calibration plane (C point in the fig.14), initial condition, and to the other calibration points with their eyes indicated by the numbers, in an arbitrary order (fig.14). Simultaneously, the experimenter marked the corresponding point seen on the image plane. In the second phase, the subjects were invited to follow the square indicated by letters with their eyes. This second phase was performed in two configurations. The first configuration was carried out with a chin support, where the head was completely still. Afterwards, the second configuration was conducted without any support, so the head was free-to-move. The results of the two configurations were statistically compared.

VR Estimation. Eleven subjects were enrolled for the 3D VR scenario experiment. Our VR application consists of an immersive room equipped with a number of sensors and effectors including three projection screens, a sound system, and a tracking system (fig.15). The subjects were asked to wear the system and, after the calibration phase, to freely walk around the room

looking at the three screens. A total number of 10 circles were shown one by one in a random order and unknown position (see fig.15). The whole experiment duration was of 10 minutes for each subject. During the calibration phase the head movement correction was performed. The accuracy of the system was calculated
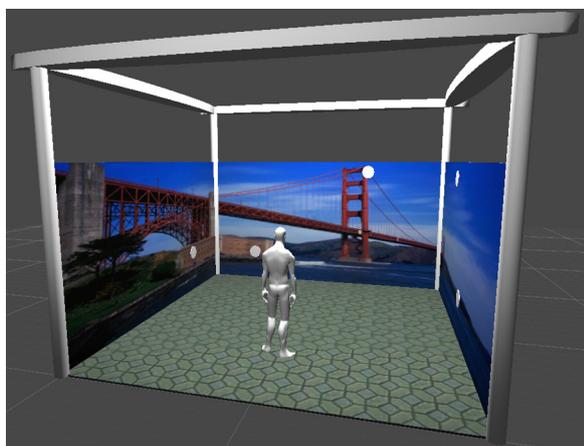
Figure 15. Virtual Reality environment used for the completely free movement scenario.

in terms of median of angular errors between the circle positions and the estimated eye position across all the subjects, Moreover, a statistical comparison was performed between the accuracy results in the laboratory and in the VR conditions.

Experimental results

In this section we report on the achieved results for both experimental setups.

Accuracy estimation

Here, the computation of the VOR curves as well as the computation of the accuracy of the system along *x*, and *y* axes were computed in terms of angular error. In particular, by means of the angular error, we evaluated the effect of the correction process comparing the gaze point in three different "modalities". Specifically, the first was computed when the head was completely still, making use of a chin support (hereinafter called *Stationary Mode*), the second and third were performed without any support. More specifically, the second was computed applying only the integration of the calibration plane orientation changes (hereinafter called *Screen Point*), and the third was obtained applying both the integration of plane orientation changes and VOR contribution (hereinafter called *Screen Point + VOR*).

Table 1 reports average gain and standard deviation of VOR for both vertical (θ) and horizontal (ψ) axes.

Table 1
Average Gain VOR

G_{θ_e}	G_{ψ_e}
52.44±2.83	61.36±3.46

These values were then used as gain corrections to estimate the specific offset for each subject. The accuracy,

Table 2
X accuracy with head in stationary mode

Subject	Without MI	Screen Point	Screen Point + VOR
1	1.01±0.70	1.08±0.71	0.74±0.86
2	0.69±0.45	0.49±0.40	1.29±0.87
3	1.28±1.12	1.52±1.11	1.32±0.96
4	0.82±0.61	0.84±0.62	0.82±0.65
5	0.48±0.34	0.54±0.36	0.57±0.53
6	0.76±0.55	0.74±0.56	0.73±0.57
7	1.74±1.30	1.81±1.36	1.72±1.11
8	0.85±0.37	0.79±0.42	0.43±0.16
9	0.67±0.42	0.70±0.52	0.67±0.52
10	1.51±0.59	1.54±0.60	1.43±0.60
11	1.16±0.72	1.21±0.70	1.25±1.07
Median	**0.85±0.31**	**0.84±0.37**	**0.82±0.37**

Table 3
Y accuracy with head in stationary mode

Subject	Without MI	Screen Point	Screen Point + VOR
1	1.93±1.01	1.98±1.22	1.93±1.14
2	1.38±0.97	1.29±1.08	2.26±1.38
3	1.62±1.22	1.85±1.50	1.31±0.97
4	0.84±0.82	1.29±1.06	0.97±0.90
5	1.37±0.52	1.37±0.56	1.32±0.46
6	3.20±1.41	3.07±1.44	3.22±1.41
7	2.82±1.38	2.84±1.39	2.83±1.37
8	0.89±0.52	0.98±0.54	0.77±0.54
9	1.65±1.36	2.47±0.80	1.90±1.08
10	0.92±0.60	0.79±0.55	0.85±0.60
11	1.70±0.92	1.40±0.96	1.61±1.05
Median	**1.62±0.54**	**1.40±0.62**	**1.61±0.64**

as mentioned above, was calculated in terms of median and median absolute dispersion of the gaze error, i.e. the error between the real observed position and the estimated value in pixels (e_{pixel}), then transformed in degrees according to the equation 19

$$e_{degree} \stackrel{\triangle}{=} arctan(\frac{e_{pixel}}{d_{pixel}}) \qquad (19)$$

where d_{pixel} represents the distance between the subject and the calibration plane. Tables 2 - 3 show median and median absolute dispersion of errors per subject, expressed in degrees, in stationary head condition for x and y axes (similar information are shown from the figures 16 - 17). In particular, the first column refers to the values without any correction, the column *Screen Point* is referred to the error values, in the image plane, with the correction of the calibration plane orientation only actuated by IMU, and the column *Screen Point + VOR* is referred to the values, in the image plane, with the MI of both calibration plane orientation and VOR together. The corrections are reported for three head rotation angles θ, ϕ, and ψ.

Table 4
X accuracy with the head free to move, corrections are reported for three rotation angles

Subject	Screen Point	Screen Point + VOR
1	11.49±15.35	1.74±1.26
2	5.32±4.54	0.85±0.52
3	5.15±4.28	1.15±1.10
4	3.10±1.93	0.56±0.37
5	3.57±2.40	0.37±0.24
6	9.84±5.09	0.69±0.14
7	5.67±1.51	0.65±0.30
8	11.89±5.76	0.80±0.84
9	4.68±1.82	1.74±1.18
10	1.58±0.85	1.74±1.02
11	7.73±6.27	1.39±1.10
Median	**5.32±2.81**	**0.85±0.44**

Table 5
Y accuracy with the head free to move, corrections are reported for three rotation angles

Subject	Screen Point	Screen Point + VOR
1	10.91±11.99	3.13±1.45
2	4.75±4.87	1.24±1.06
3	1.11±0.84	1.34±0.99
4	3.44±2.40	3.09±1.02
5	2.18±1.90	1.18±0.88
6	15.55±1.92	6.50±0.77
7	10.98±5.02	2.20±1.03
8	8.50±5.11	1.78±1.07
9	3.11±2.58	2.25±1.85
10	1.89±1.50	1.51±1.06
11	4.11±4.40	1.21±1.03
Median	**4.70±3.94**	**1.78±1.05**

Tables 4 - 5 show median and median absolute disper-
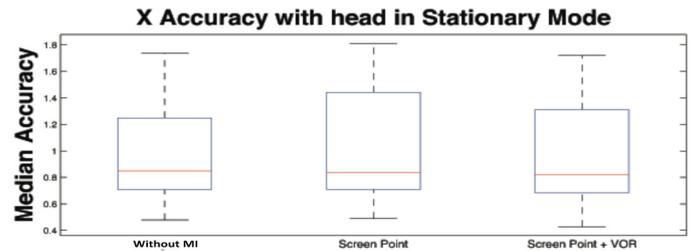
Figure 16. Box plot of data from the statistical comparison of the Median and Median Absolute Dispersion (MAD) along with *x* axis, with and without head movements. In the case of head movements without any VOR correction.
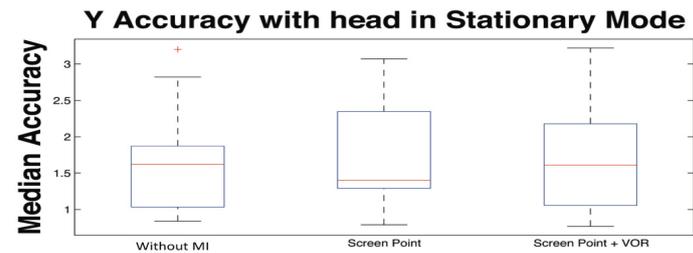
Figure 17. Box plot of data from the statistical comparison of the Median and Median Absolute Dispersion (MAD) along with *y* axis, with and without head movements. In the case of head movements without any VOR correction.
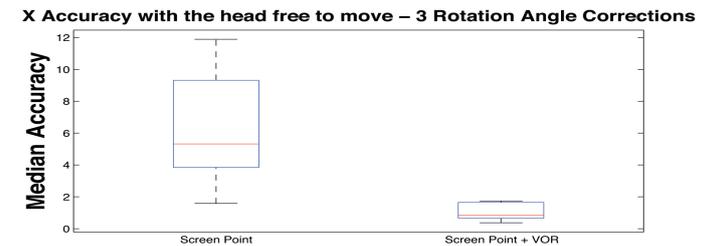
Figure 18. Box plot of data from the statistical comparison of the Median and Median Absolute Dispersion (MAD) along with *x* axis, with and without head movements. In the case of head movements with VOR correction.
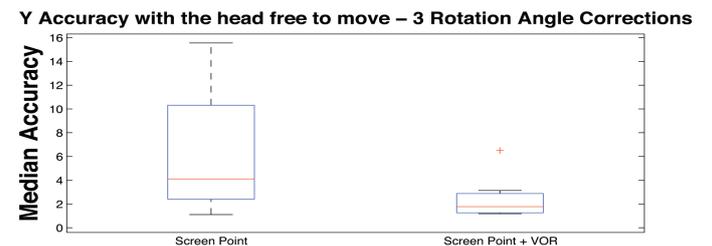
Figure 19. Box plot of data from the statistical comparison of the Median and Median Absolute Dispersion (MAD) along with *y* axis, with and without head movements. In the case of head movements with VOR correction.

sion of errors per subject, expressed in degrees, with free head movement condition for *x* and *y* axes (the same information cane seen in figures 18 - 19). In these

11

Tables the columns report only the values for the column *Screen Point*, which refers to the error values applying the correction of the calibration plane orientation and the column *Screen Point + VOR*, which is referred to the values applying the correction of both calibration plane orientation and VOR together. In these Tables the corrections are reported for 3 head rotation angles θ, ϕ, and ψ. The head movements were related to an average rotation amplitude of 20 degrees around the three axes. Tables 4 - 5 do not report the column "Without MI", with respect to Tables 2 - 3 because the calibration process in case of head movement cannot be computed. More specifically, calibration would produce wrong random values by chance. As a matter of fact, Figure 20 reports an example of calibration during head movement. In this figure, it can be noticed that calibration points are strictly concentrated on a small part of the image plane making the system completely unusable. Table 6 show results from Friedman non para-
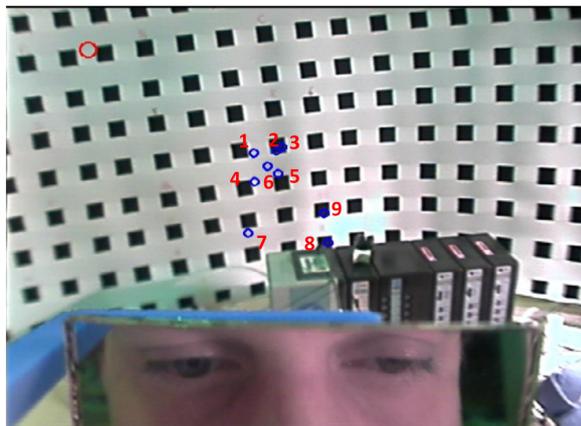
Figure 20. Calibration of Subject 6. The 9 calibration points are marked in blue; the calibration was performed during head movements. It is an example of the errors produced by uncorrected head movement into calibration

Table 6
Results of the Friedman non parametric test with Bonferroni correction applied to Stationary Mode and Head Movement conditions "without MI" and with "Screen point + VOR" for x and y axes, respectively. In the Table p − values are shown.

	Without MI	Screen point + VOR
x Stationary Mode	0.959	0.998
y Stationary Mode	0.957	0.940
x Head Movement	3.27E-005	0.881
y Head Movement	1.5E-3	0.563

metric test applied to stationary head (stationary) and head free to move (movement) conditions, with and without MI. More specifically, the test returns the probability that the different samples were not belonging to the same population. A pairwise comparison was performed for every pair of samples after the rejection of the null hypothesis carrying out a Mann-Whitney test with a Bonferroni adjustment. Results show that in Stationary mode the error between actual and computed coordinates of the gaze point estimated with and without MI are statistically equivalent, while when the head is moving the errors belong to different populations when MI (Screen + VOR) is not used. On the contrary, after the integration of the head movement, the statistical analysis reported that no significant difference is achieved among median in stationary head (stationary) and head free to move (movement) conditions.

VR estimation

Results achieved in the VR scenario are reported in Table 7 and are expressed in terms of angular errors along the *X* and *Y* axes. The Table shows for each subject the median ± the median absolute value of the angular errors computed on the gaze points of the 10 circle targets. The last raw represents the inter-subject median angular error, which resulted to be equal to 1.45 degree for the X coordinate and 1.54 degree for the Y coordinate . Moreover, a statistical comparison

Table 7
Evaluation results of median ± MAD angular error computed for X and Y axes in a 3D VR scenario

Subjects	X	Y
1	1.53 ± 0.70	2.86 ± 0.20
2	1.45 ± 0.38	1.08 ± 0.19
3	1.67 ± 1.33	1.28 ± 1.11
4	0.75 ± 0.51	1.12 ± 0.47
5	1.81 ± 1.253	1.62 ± 0.77
6	3.11 ± 0.54	2.64 ± 1.12
7	0.75 ± 0.51	1.12 ± 0.47
8	0.70 ± 0.53	1.54 ± 0.76
9	0.94 ± 0.47	4.79 ± 0.86
10	3.97 ± 1.46	4.64 ± 1.02
11	0.85 ± 0.57	1.53 ± 0.66
Median	**1.45 ± 0.60**	**1.54 ± 0.42**

between the results achieved in laboratory and virtual reality conditions was performed by means of a Mann-Whitney test. For both coordinates, no significant differences were found between the two experimental setup conditions (X p-value > 0.05 - Y p-value > 0.05) confirming the usability of the HAT-Move even in free-movement conditions.

Real-time estimation

The execution time of the main tasks and the entire software integrated into the system was about of 34.47 ms. The working frequency is about 29 Hz, which is greater than camera sampling frequency (25Hz), therefore the real-time requirement was fulfilled. ¡

Discussion and conclusion

Even though many current HMEGT systems are used without any integration of the movement of the head either for image plane orientation nor for the Vestibulo-Ocular Reflex, they are used with partial head movements. This study pointed out that HMEGT systems are strongly affected by head movements. More specifically, when the chin support is used the angular errors are acceptable. This conclusion is supported by the literature and also confirmed by this study, see figures 16 and 17, and tables 2, 3,. Indeed, no strong divergence between the median values for both errors along *x* and *y* is present, and confirmed by Friedman nonparametric test for paired sample with *Bonferroni* post-hoc correction for multiple comparison reported in Table 6 no significant difference is shown for stationary mode. However, the same statistical tests highlighted that when head slightly moves the errors dramatically increase. It is worthwhile noting that the multiple comparisons reported in figures showed that during head movements the gaze point diverge with errors of $4 - 5$ degrees. This experimental evidence suggests that the proposed corrections are essential to achieve an accurate eye gaze tracking in dynamical conditions. In fact, since eye tracking systems are often used also in medical environments for detecting pathologies which can range from behavioral to neurological fields, these errors could bring to misleading interpretations. The effectiveness of the head movement integration has been proven by the statistical comparison on *x* and *y* directions. In fact, it is possible to reduce the angular errors achieving no statistical difference between stationary mode and head movements, showing that the system keeps the same accuracy in both modalities. The estimated median error with head movement is reduced from $5.32°$ with a standard deviation of $2.81°$ (without VOR correction) to $0.85°$ with a standard deviation of $0.44°$ and from $4.70°$ with a standard deviation of $3.94°$ (without VOR correction) to $1.78°$ with a standard deviation of $1.05°$ for *x* and *y* axes, respectively. The obtained accuracy results confirm the reliability and robustness of the proposed system. Moreover, the difference between the accuracy along *x* and *y* can be due to angular position of the camera (which is above of the eyes) which reduces the accuracy of the vertical positions of the pupil. In addition, this system fulfilled the realtime requirement being the execution time of the algorithm lower than the time interval between two consecutive video frames. In order to test the system even in conditions in which the subject was completely free to move and walk, we have developed an experiment in a virtual reality environment. We asked the subjects to look at random points on the screen of the VR moving into the room. Accuracy resulted to be equal to 1.45 degree for the X coordinate and 1.54 degree for the Y coordinate. No significant differences were found from the accuracy in the laboratory conditions. This result confirm the robustness of the proposed system even in scenarios similar to real environments

The main limitation of the system is the low frame rate of the camera. This limitation does not allow the system to acquire fast saccadic movements, which are known to be in the time range of 30 ms, while it is able to acquire slow saccadic movements around 100 ms. The proposed system is equipped with low cost hardware and it results extremely lightweight, unobtrusive, and aesthetically attractive providing a good acceptability by the end users. Thanks to these properties and its technological specifications the HAT-Move system allows investigating how humans interact with the external environment continuously. This ecological approach (Bronfenbrenner, 1977) could be pursued either at individual or community level with the aim of analyzing and coding both activities and relationships. More specifically, this kind of information can be really useful in studying non-verbal social behavior in both healthy and impaired people (e.g. affected by behavioral pathologies such as autistic disorders) as well as to improve the scientific knowledge on human interpersonal relationships. Furthermore, HAT-Move system has been already shown to be useful for studying eye pattern as a response of emotional stimulation with good and promising results (Lanatà et al., 2013; Lanata, Armato, et al., 2011). As a matter of fact eye feature pattern could provide a new point of view in the study of personal and interpersonal aspects of human feelings. In such a way eye information which we already showed to be informative in emotional response, could be integrated with other sets of physiological data (Betella et al., 2013) such as cardiac information (Lanata, Valenza, Mancuso, & Scilingo, 2011) or heart rate variability (Valenza, Allegrini, Lanata, & Scilingo, 2012) , respiration activity(Valenza, Lanatá, & Scilingo, 2013), as well as electrodermal response (Greco et al., 2012) in a multivariate approach (Valenza et al., 2014) in order to create a very complete descriptive set of data able to explain the non-verbal phe-

nomenology behind implicit (autonomic nervous system response) and explicit human response to external stimulation in real or virtual scenario (Wagner et al., 2013). It could be really helpful also in the investigation of several pathologies where the unobtrusiveness of the instrumentations could allow of monitoring naturally evoked response of participants, (Lanatà, Valenza, & Scilingo, 2012; Valenza, Lanata, Scilingo, & De Rossi, 2010; Lanatà et al., 2010; Armato et al., 2009; Valenza et al., 2010; Lanata, Valenza, et al., 2011). This latter issue could be used either on healthy participants or subject suffering from pathologies such as autism (Mazzei et al., 2012) as well as alterations of mood (Valenza, Gentili, Lanata, & Scilingo, 2013) etc. in which social skills are a strong impact on the lifestyle of it should improve the scientific knowledge on human interpersonal relationships, emotions etc. Moreover, to achieve these aims further efforts will be devoted to integrating a high-speed camera with high resolution in order to capture fast saccadic movements and to providing better accuracy. Moreover, an optimization process will be addressed to develop new multithreading algorithms based on the interocular distance in order to obtain a 3D eye tracking system.

References

Armato, A., Lanatà, A., & Scilingo, E. P. (2013). Comparative study on photometric normalization algorithms for an innovative, robust and real-time eye gaze tracker. *Journal of real-time image processing*, *8*(1), 21–33.

Armato, A., Nardini, E., Lanata, A., Valenza, G., Mancuso, C., Scilingo, E. P., & De Rossi, D. (2009). An fpga based arrhythmia recognition system for wearable applications. In *Intelligent systems design and applications, 2009. isda'09. ninth international conference on* (pp. 660–664).

Babcock, J., Pelz, J., & Peak, J. (2003). *The wearable eyetracker: a tool for the study of high-level visual tasks*. MSS-CCD.

Bennett, N., Burridge, R., & Saito, N. (2002). A method to detect and characterize ellipses using the Hough transform. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *21*(7), 652–657.

Betella, A., Martínez, E., Zucca, R., Arsiwalla, X. D., Omedas, P., Wierenga, S., ... others (2013). Advanced interfaces to stem the data deluge in mixed reality: placing human (un) consciousness in the loop. In *Acm siggraph 2013 posters* (p. 68).

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American psychologist*, *32*(7), 513.

Cerrolaza, J. J., Villanueva, A., & Cabeza, R. (2008). Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. In *Proceedings of the 2008 symposium on eye tracking research & applications* (pp. 259–266).

Cootes, T., & Taylor, C. (n.d.). *Technical report: Statistical models of appearance for computer vision. the university of manchester school of medicine, 2004.*

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. *Computer vision and image understanding*, *61*(1), 38–59.

Crawford, J., & Vilis, T. (1991). Axes of eye rotation and Listing's law during rotations of the head. *Journal of neurophysiology*, *65*(3), 407.

de Lemos, J., Sadeghnia, G. R., Ólafsdóttir, Í., & Jensen, O. (2008). Measuring emotions using eye tracking. In *Proceedings of measuring behavior* (Vol. 226).

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. Springer-Verlag New York Inc.

Evans, K. M., Jacobs, R. A., Tarduno, J. A., & Pelz, J. B. (2012). Collecting and analyzing eye-tracking data in outdoor environments. *Journal of Eye Movement Research*, *5*(2), 1–19.

Fahrenberg, J., Myrtek, M., Pawlik, K., & Perrez, M. (2007). Ambulatory assessment–monitoring behavior in daily life settings: A behavioral-scientific challenge for psychology. *European Journal of Psychological Assessment*, *23*(4), 206.

Findlay, J. M. (2009). Saccadic eye movement programming: Sensory and attentional factors. *Psychological research*, *73*(2), 127–135.

Fitzgibbon, A., Pilu, M., & Fisher, R. (2002). Direct least square fitting of ellipses. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *21*(5), 476–480.

Forsyth, D., & Ponce, J. (2002). *Computer vision: a modern approach*. Prentice Hall Professional Technical Reference.

Greco, A., Lanata, A., Valenza, G., Rota, G., Vanello, N., & Scilingo, E. P. (2012). On the deconvolution analysis of electrodermal activity in bipolar patients. In *Engineering in medicine and biology society (embc), 2012 annual international conference of the ieee* (pp. 6691–6694).

Hartley, R., & Zisserman, A. (2000). *Multiple view geometry in computer vision* (Vol. 2). Cambridge Univ Press.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in cognitive sciences*, *9*(4), 188–194.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences*, *7*(11), 498–504.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, *2*(3), 4.

Johnson, J. S., Liu, L., Thomas, G., & Spencer, J. P. (2007). Calibration algorithm for eyetracking with unrestricted head movement. *Behavior research methods*, *39*(1), 123–132.

Lanata, A., Armato, A., Valenza, G., & Scilingo, E. P. (2011). Eye tracking and pupil size variation as response to affective stimuli: a preliminary study. In *Pervasive computing technologies for healthcare (pervasivehealth), 2011 5th international conference on* (pp. 78–84).

Lanatà, A., Scilingo, E. P., Nardini, E., Loriga, G., Paradiso, R., & De-Rossi, D. (2010). Comparative evaluation of susceptibility to motion artifact in different wearable systems for monitoring respiratory rate. *Information Technology in Biomedicine, IEEE Transactions on*, *14*(2), 378–386.

Lanata, A., Valenza, G., Mancuso, C., & Scilingo, E. P. (2011). Robust multiple cardiac arrhythmia detection through bispectrum analysis. *Expert Systems with Applications*, *38*(6), 6798–6804.

Lanatà, A., Valenza, G., & Scilingo, E. P. (2012). A novel eda glove based on textile-integrated electrodes for affective computing. *Medical & biological engineering & computing*, *50*(11), 1163–1172.

Lanatà, A., Valenza, G., & Scilingo, E. P. (2013). Eye gaze patterns in emotional pictures. *Journal of Ambient Intelligence and Humanized Computing*, *4*(6), 705–715.

Land, E. H., & McCann, J. (1971). Lightness and retinex theory. *JOSA*, *61*(1), 1–11.

Land, M., & Lee, D. (1994). Where we look when we steer. *Nature*, *369*(6483), 742–744.

Maini, E. (2005). Robust ellipse-specific fitting for real-time machine vision. *Brain, Vision, and Artificial Intelligence*, 318–327.

Mazzei, D., Greco, A., Lazzeri, N., Zaraki, A., Lanata, A., Igliozzi, R., ... others (2012). Robotic social therapy on children with autism: preliminary evaluation through multi-parametric analysis. In *Privacy, security, risk and trust (passat), 2012 international conference on and 2012 international confernece on social computing (socialcom)* (pp. 955–960).

Meyer, A., Böhme, M., Martinetz, T., & Barth, E. (2006). A single-camera remote eye tracker. In *Perception and interactive technologies* (pp. 208–211). Springer.

Morimoto, C., Amir, A., & Flickner, M. (2002). Detecting eye position and gaze from a single camera and 2 light sources. *Pattern Recognition*, *4*, 40314.

Morimoto, C., & Mimica, M. (2005). Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, *98*(1), 4–24.

Morimoto, C. H., Koons, D., Amir, A., & Flickner, M. (2000). Pupil detection and tracking using multiple light sources. *Image and vision computing*, *18*(4), 331–335.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies*, *59*(1), 185–198.

Ramanauskas, N., Daunys, G., & Dervinis, D. (2008). Investigation of calibration techniques in video based eye tracking system. In *Computers helping people with special needs* (pp. 1208–1215). Springer.

Rothkopf, C. A., & Pelz, J. B. (2004). Head movement estimation for wearable eye tracker. In *Proceedings of the 2004 symposium on eye tracking research & applications* (pp. 123–130).

Valenza, G., Allegrini, P., Lanata, A., & Scilingo, E. P. (2012). Dominant lyapunov exponent and approximate entropy in heart rate variability during emotional visual elicitation. *Frontiers in neuroengineering*, *5*.

Valenza, G., Gentili, C., Lanata, A., & Scilingo, E. P. (2013). Mood recognition in bipolar patients through the psyche platform: preliminary evaluations and perspectives. *Artificial intelligence in medicine*, *57*(1), 49–58.

Valenza, G., Lanatá, A., & Scilingo, E. P. (2013). Improving emotion recognition systems by embedding cardiorespiratory coupling. *Physiological measurement*, *34*(4), 449.

Valenza, G., Lanata, A., Scilingo, E. P., & De Rossi, D. (2010). Towards a smart glove: Arousal recognition based on textile electrodermal response. In *Engineering in medicine and biology society (embc), international conference of the ieee* (pp. 3598–3601).

Valenza, G., Nardelli, M., Lanata, A., Gentili, C., Bertschy, G., Paradiso, R., & Scilingo, E. P. (2014). Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis. *Biomedical and Health Informatics, IEEE Journal of*, *18*(5), 1625–1635.

Van der Geest, J., & Frens, M. (2002). Recording eye movements with video-oculography and scleral search coils: a direct comparison of two methods. *Journal of neuroscience methods*, *114*(2), 185–195.

Wagner, J., Lingenfelser, F., André, E., Mazzei, D., Tognetti, A., Lanata, A., ... others (2013). A sensing architecture for empathetic data systems. In *Proceedings of the 4th augmented human international conference* (pp. 96–99).

Westheimer, G., & McKee, S. P. (1975). Visual acuity in the presence of retinal-image motion. *JOSA*, *65*(7), 847–850.

Zhu, Z., & Ji, Q. (2007). Novel eye gaze tracking techniques under natural head movement. *Biomedical Engineering, IEEE Transactions on*, *54*(12), 2246–2260.