Object-Based Geometric Distortion Metric for Viewport Rendering of 360° Images

FALAH JABAR[ID]**, JOÃO ASCENSO**[ID]**, (Member, IEEE), AND MARIA PAULA QUELUZ**[ID]
Instituto de Telecomunicações/Instituto Superior Técnico, University of Lisbon, 1049-001 Lisbon, Portugal

Corresponding author: Falah Jabar (falah.jabar@lx.it.pt)

ABSTRACT To visualize omnidirectional (or 360°) visual content, a sphere to plane projection is employed, that maps pixels from the observed sphere region to a 2D image, called as *viewport*. However, this projection introduces geometrical distortions on the rendered image, such as object shape stretching, or shearing, and bending of straight lines, which may affect the user's quality of experience (QoE). This paper proposes an object-based quality metric to assess the subjective impact of the objects shape deformation. The metric uses semantic segmentation to identify the relevant objects in the viewport, where the stretching distortion has a higher perceptual impact, and computes the stretching distortion for each object. Two distinct approaches were exploited and evaluated: the first one, directly computes and compares object shape measures on the sphere and on the viewport; the second one is based on Tissot indicatrices, which are computed for individual objects in the viewport. The experimental results show that while the Tissot based method performs slightly better than the direct shape measurement, both approaches outperform benchmark solutions; furthermore, they are able to classify the viewport quality, with respect to quality scores obtained in a subjective crowdsourcing study, with a correct decision percentage close to 90%. Additionally, the Tissot based approach was used in a global quality metric that finds out the Pannini projection parameters that result in the least perceivable geometric distortion. It is shown that the automatically tuned Pannini projection results in viewports with a more pleasant visual quality than the considered benchmark projections.

INDEX TERMS Omnidirectional images, viewport rendering, geometric distortions, objective quality assessment, Pannini projection, content-aware projection.

I. INTRODUCTION

In recent years, the popularity of omnidirectional visual content and applications is increasing rapidly, notably in virtual reality (VR) and augmented reality (AR) fields. Omnidirectional visual content can already be found in a large set of applications that users can enjoy, including immersive gaming, remote education, virtual shopping, virtual sports, virtual tours, and even broadcasting of live content. However, to have a successful application, the user's quality of experience (QoE) should be high, and thus techniques to assess and improve the quality of experience are an important research topic [1]–[4].

Omnidirectional visual content contains the information of the scene around the camera, covering the whole 360°

The associate editor coordinating the review of this manuscript and approving it for publication was Chua Chin Heng Matthew[ID].

(horizontal) × 180° (vertical) viewing range, referred to as viewing sphere. To store or transmit the omnidirectional visual content, the viewing sphere is mapped on a 2D image using, typically, the equirectangular or cubic projections. When this type of content is played, the user can observe any part of the visual scene by changing the viewing direction ("look around") which creates, to him, the feeling of being physically present; this allows a visual experience more immersive than what is offered by traditional 2D visual content.

There are several ways to display omnidirectional visual content, including HMDs (e.g., HTC Vive and Oculus Rift), smartphones (or tablets), or standard computer monitors. Typically, the HMD provides a better immersive experience, although it is somewhat uncomfortable, expensive and not accessible to all users. Therefore, watching omnidirectional visual content on smartphones or computer monitors
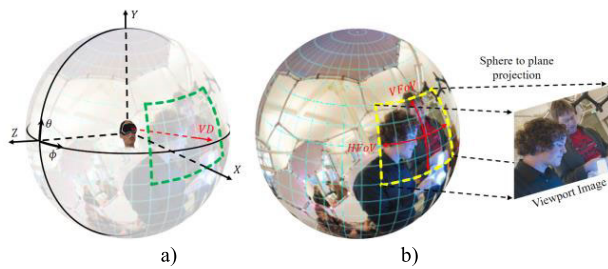
FIGURE 1. Viewport rendering: a) Coordinate system; b) Sphere to plane projection.
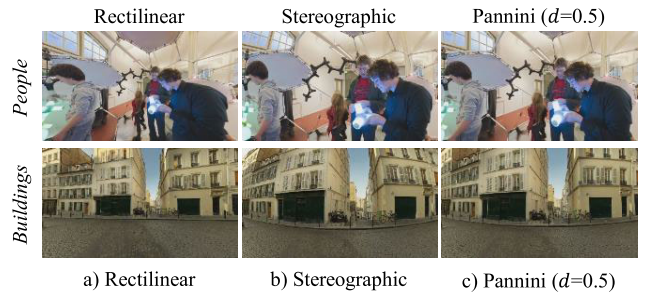

FIGURE 2. Viewport rendered from two omnidirectional images, using a) Rectilinear, b) Stereographic and c) Pannini projections, and a HFoV of 130°.

is rather common. Recently, the large technological industry, such as Google and Facebook, provided several applications and services for smartphones and personal computers (thus, using 2D screens), including Google street view, Facebook 360 photos, and YouTube VR, among others; this is the target application scenario of this paper.

Regardless of the used display, the users only see a fraction (called as *viewport*) of the entire sphere, at a time. The viewport content is defined by the viewing direction (VD, in Fig. 1-a)) and by the horizontal and vertical fields of view (HFoV and VFoV, in Fig. 1-b)); a large field of view (FoV) includes more information in the viewport, making the visual experience more pleasant [5]. To render the viewport in a 2D image, a sphere to plane projection is required; the rectilinear, stereographic, and Pannini projections are often considered for this procedure [6]–[8]. However, any sphere to plane projection introduces geometric distortions, such as stretching of objects and bending of straight lines.

The projection type, the considered FoV, and the image content characteristics are the three main factors that influence the amount of perceived distortion [9]. Concerning the latter, although the geometric distortion at infinitesimal scale (i.e., locally) is independent on the image content, it has a minimum value at the viewport center and increases towards the viewport borders. Also, for nonconformal projections this distortion is not isotropic. At a larger scale, the geometric distortion suffered by an object is thus dependent on its position, shape, and area. The rectilinear projection keeps the straightness of the lines, but stretches the objects, mainly those close to the viewport borders; this effect is clearly visible in the *People* viewport, depicted on the top of Fig. 2. The stereographic projection preserves the shapes locally, but it severely bends the straight lines (fisheye effect), as can be seen on the *Buildings* viewport, depicted on the bottom of Fig. 2. In the Pannini projection, the projection center, *d*, is an adjustable parameter, allowing the projection to vary its main characteristics, from rectilinear to quasi-stereographic. In this projection, the vertical and radial lines are projected without bending, but the horizontal lines are bent, as shown on the *Buildings* viewport. To reduce the bending of horizontal lines, vertical compression can be applied, at the expense of increasing the objects stretching. A good balance between stretching and bending may be achieved by adjusting the two parameters, *d* and *vc* (vertical compression factor). Thus, the

Pannini projection has the advantage of preserving object shapes and the straightness of vertical lines better than the typical rectilinear and stereographic projections, being more suitable for viewport rendering with large FoV; this justifies why it was the selected projection for the study described in this paper.

Since the geometric distortion, and its subjective impact, is dependent on the image content, selecting a proper viewport projection and its parameters may have an important role on the user's QoE. This requires the availability of a content-aware objective quality metric to automatically measure the perceived geometric distortion, after viewport rendering. In Earth cartography, the Tissot's indicatrices [10] have been used to assess the geometric distortions introduced by map projection. However, this metric is content independent, resulting in the same distortion measure, for the same sphere to plane projection, regardless of the image content.

As an example, the same distortion measure will be obtained for the *People* and *Buildings* rectilinear viewports in Fig. 2, although the stretching distortion in *People* is much more annoying than in *Buildings*. Moreover, the line bending, visible in Fig. 2 for *Buildings* with stereographic and Pannini projections, cannot be estimated with Tissot's indicatrices. In the authors previous work [11], several line bending and stretching measures were proposed; the former directly measures the resulting line curvature, and the latter is based on Tissot's indicatrices, improved with the integration of saliency weights that give more importance to the parts of the viewport image that attract the user attention. Both bending and stretching measures were evaluated by correlating their values with perceptual scores, obtained from subjective tests. However, while the bending metric showed to be well correlated with perceptual scores, the stretching measures achieved a poor performance. In fact, the Tissot's indicatrices are local measures and, even if weighted by saliency scores, they are not able to accurately capture the global object distortion.

In this paper, object-based stretching distortion metrics are proposed to automatically assess the subjective impact of the viewport stretching distortion, in 360° image rendering. The new metrics overcome the shortcoming of the stretching distortion measures proposed in [11] by measuring the distortion of individual objects, having semantic meaning (such

as people, cars, or furniture), using semantic segmentation. A procedure, that integrates one of the new metrics with a line bending measure, is also proposed to automatically tune the Pannini projection parameters, d and vc, according to the viewport content. In this context, the main contributions of this paper can be summarized as follows:

- A web-based crowdsourcing subjective evaluation of viewport images, rendered with the Pannini projection, was conducted, to assess the perceptual impact of the stretching distortion. The resulting viewport images and perceptual scores were made available in [12].
- Several object-based stretching distortion metrics are proposed, which were evaluated using the subjective test results. These metrics measure the stretching distortion of viewport objects with semantic meaning, where the stretching distortion has a higher perceptual impact.
- A procedure to automatically tune the Pannini projection parameters, d and vc, according to the image content, is proposed. It integrates one of the new object-based stretching distortion metrics, showing that it can be exploited to minimize, in a perceptual way, the geometric distortions resulting from the rendering process.

The rest of this paper is organized as follows: Section II reviews related works; Section III presents the results of the subjective test campaign to assess the stretching distortion perceptual impact, and the corresponding analysis; Section IV details the proposed object-based stretching distortion metrics. Section V presents the metrics evaluation results. Section VI describes and assesses a potential application for the proposed metric: the automatic tuning of the Pannini projection parameters for a given viewport. Finally, Section VII concludes the paper.

II. RELATED WORK

In cartography, several map projections have been used to flatten the earth surface into a plane. Any sphere to plane projection distorts the spherical surface in different ways, changing its geometric properties such as distance, direction, shape, or area; no projection can preserve, simultaneously, all these properties. The Tissot's indicatrix (or ellipse of distortion) [10], is the geometric distortion metric most used in cartography. This ellipse is obtained after projecting, in the map, an infinitely small circle defined on the sphere; the scale, area, and angular deformations are defined using the relationship between the semi major and semi minor ellipse axis. However, this metric has some drawbacks, notably: *i)* it is content independent; *ii)* it is a local metric, thus global distortion, e.g., bending of straight lines, cannot be directly measured with it.

More recently, a few geometric distortion measures were proposed in the context of content-aware projections for wide-angle images, seeking to minimize the resulting geometric distortions. In [13], the projection was locally, and iteratively, adapted to the image content based on local conformality and line straightness measures. Conformality

measure was computed based on Cauchy-Riemann equations [14], and line straightness measure was computed based on the geometry of straight lines. However, these measures were not validated with respect to perceived geometric distortion. Moreover, the straight lines present in the scene need to be identified manually by the users. In [8], the same line straightness and conformality measures proposed in [13] were used to optimize the Pannini projection parameters. The optimized Pannini and a set of benchmark projections were subjectively evaluated in a crowdsourcing subjective test. However, the test results showed that, in average, the Pannini projection with fixed parameters achieved higher quality scores than the optimized Pannini. Also, as in [13], there was no procedure defined to validate the distortion measures with respect to perceived distortion. In [15], a content-aware projection was proposed to minimize the stretching distortion of human faces in wide-angle photos taken from a mobile device, and with FoV up to 120°. To preserve the human faces, a face stretching measure was proposed. Moreover, a line constraint measure was defined over the regions between the faces and the background to preserve the straightness of the lines. The proposed projection was locally adapted to the stereographic projection on facial regions and evolved to the rectilinear projection over the background. Similar to [8], [13], the distortion measures were not validated with respect to perceived distortion. Also, the stretching measure was designed specifically for faces, and not for general objects. However, correcting the human faces without the rest of the body may create additional artifacts. Furthermore, the distortions of other objects in the scene were not considered.

In the authors previous work [11], several stretching and line bending measures were proposed to measure the geometric distortion in viewport rendering of 360° images. The proposed stretching distortion measures were based on the Tissot's indicatrices but, to make it content dependent, saliency weights were integrated to give more importance to the parts of the viewport image that attract more the user attention. Both bending and stretching measures were validated with respect to perceived geometric distortion, by correlating their values with the perceptual scores obtained from a subjective test campaign; while the bending metrics showed to be well correlated with those scores, the stretching measures achieved a low performance.

Since the human visual system is highly sensitive to the stretching distortion of the objects presented on the image, in this paper several object-based stretching distortion metrics are proposed, seeking to overcome the shortcomings of previously proposed metrics. This allows a better optimization (in a perceptual sense) of the sphere to plan projection, seeking to improve the perceived quality of the rendered viewport.

III. SUBJECTIVE ASSESSMENT OF THE STRETCHING DISTORTION EFFECT

This section describes a crowdsourcing-based subjective evaluation of viewport images, aiming to assess the
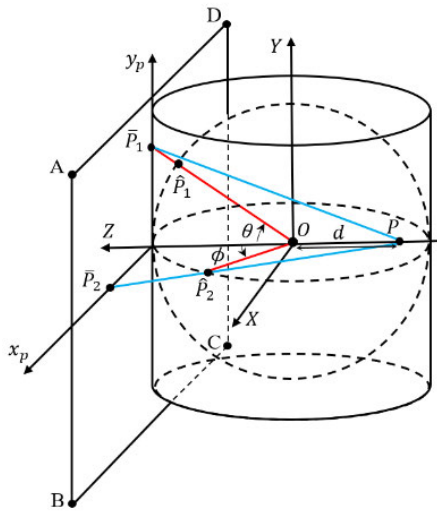
FIGURE 3. Pannini projection of two points, $\hat{P}_1$ and $\hat{P}_2$. The red lines project the points from the sphere surface to the cylinder surface; the blue lines project the points from the cylinder surface to the plane.

perceptual impact of the stretching distortion; the viewport images were rendered using the Pannini projection which, for a particular choice of its projection parameters, results in a pure rectilinear projection, for which the stretching effect (also known as perspective distortion) is most evident. For the sake of completeness, the viewport rendering process using the Pannini projection is firstly described; the used 360° images dataset, the subjective evaluation methodology, and the final subjective test results and analysis are then successively presented.

A. VIEWPORT RENDERING WITH PANNINI PROJECTION
The Pannini projection (PP) [16] uses a cylindrical surface as an intermediary surface to project a point from the viewing sphere to a plane, tangent to the sphere, as depicted in Fig. 3. This procedure consists on two steps: *i)* rectilinear projection from the spherical surface to the cylindrical surface (red lines in Fig. 3); *ii)* perspective projection from the cylindrical surface to the plane (blue lines in Fig. 3), with the projection center at a distance d from the cylinder center. For $d = 0$, the PP becomes the rectilinear projection; increasing d, gradually expands the middle of the resulting image – making it closer to the viewer and reducing the perspective distortion – but the horizontal lines are progressively bent. To reduce this effect, vertical compression can be applied, at the expense of bending the radial lines and/or increasing the stretching of the objects. The vertical compression strength can be adjusted manually, seeking the best compromise between the bending of horizontal and radial lines, and the objects stretching. The forward projection equations for the Pannini projection are given by [16]:

$$x_p = S \sin(\phi), \tag{1}$$

$$y_p = (1 - vc)(S \tan(\theta)) + vc\left(\frac{\tan(\theta)}{\cos(\phi)}\right), \tag{2}$$

with

$$S = \frac{d + 1}{d + \cos(\phi)}, \tag{3}$$

where (ϕ, θ) denote, respectively, the longitude and latitude coordinates of a point on the viewing sphere, vc is the vertical compression factor, and (x_p, y_p) are the Cartesian coordinates of the projected point; these coordinates have their origin at the center of the plane, as depicted in Fig. 3. Inversely, for any point on the plane with Cartesian coordinates (x_p, y_p), the backward projection equations are given by [16]:

$$\phi = \tan^{-1}\left(\frac{x_p}{S\cos(\check{\phi})}\right), \tag{4}$$

$$\theta = \tan^{-1}\left(y_p \Big/ \left[S(1 - vc) + \frac{vc}{\cos(\check{\phi})}\right]\right), \tag{5}$$

where

$$\cos(\check{\phi}) = \frac{-\gamma d + \sqrt{\gamma^2 d^2 - (\gamma + 1)(\gamma d^2 - 1)}}{\gamma + 1}, \tag{6}$$

$$\gamma = \frac{x_p^2}{(d + 1)^2}. \tag{7}$$

For the front viewport, corresponding to a viewing direction $(\phi, \theta) = (0^o, 0^o)$, a pixel with coordinates (m, n) (and with this coordinate system centered on the top-left corner of the viewport plane) is rendered by applying (8) and (9):

$$x_p = V_{hs}\left(\frac{m + 0.5}{W_{vp}} - \frac{1}{2}\right), \qquad 0 \le m < W_{vp} \tag{8}$$

$$y_p = V_{vs}\left(\frac{1}{2} - \frac{n + 0.5}{H_{vp}}\right), \qquad 0 \le n < H_{vp} \tag{9}$$

followed by (4) to (7); in (8) and (9), W_{vp} and H_{vp} are, respectively, the viewport width and height, in pixels; V_{hs}, V_{vs}, are, respectively, the horizontal and vertical viewport sizes, in length units:

$$V_{hs} = 2(d + 1)\frac{\sin\left(\frac{F_h}{2}\right)}{d + \cos\left(\frac{F_h}{2}\right)}, \quad V_{vs} = 2\tan\left(\frac{F_v}{2}\right), \tag{10}$$

where F_h and F_v are, respectively, the horizontal and vertical FoV. For a given F_h and viewport aspect ratio, $AR = V_{hs}/V_{vs}$, F_v can be obtained by:

$$F_v = 2\tan^{-1}\left(\frac{(d + 1)\sin\left(\frac{F_h}{2}\right)}{AR\left(d + \cos\left(\frac{F_h}{2}\right)\right)}\right). \tag{11}$$

The rendering of a viewport corresponding to a generic viewing direction, (ϕ_{VD}, θ_{VD}), requires an additional step, relating the corresponding sphere coordinates, $(\acute{X}, \acute{Y}, \acute{Z})$, with the ones that are obtained for the front viewport, (X, Y, Z):

$$\left(\acute{X}, \acute{Y}, \acute{Z}\right)^T = R(X, Y, Z)^T, \tag{12}$$

where R is the viewport rotation matrix, defined according to (ϕ_{VD}, θ_{VD}).

a) *Ind1* (4100 × 2050) b) *Ind2* (4100 × 2050) c) *Ind3* (6000 × 3000) d) *Ind4* (4000 × 2000) e) *Ind5* (5000 × 2500)

f) *Ind6* (3840 × 1920) g) *Ind7* (5376 × 2688) h) *Outd1* (7500 × 3750) i) *Outd2* (7500 × 3750) j) *Outd3* (5376 × 2688)
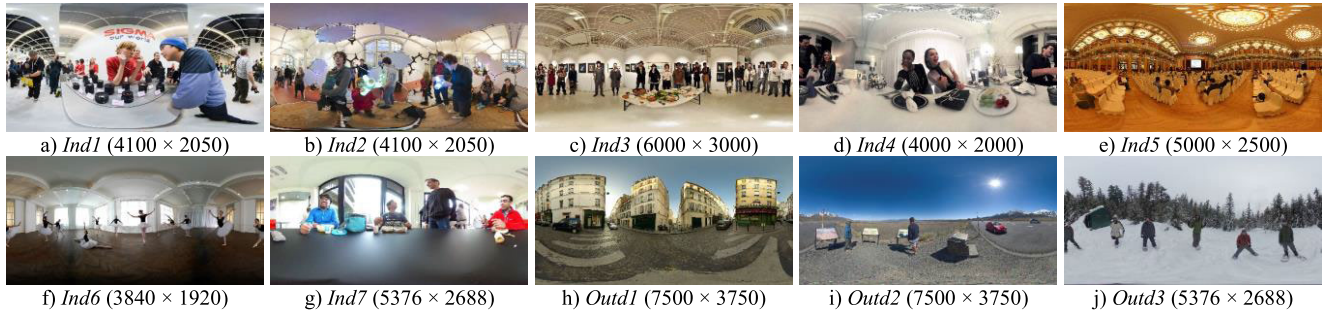
FIGURE 4. ERI images used in the subjective test, and their spatial resolution.

B. DATASET

Ten omnidirectional images in equirectangular format (ERI), extracted from the datasets available in [8] and [17], were used in the subjective assessment. The images, and their spatial resolutions, are depicted in Fig. 4. The selected images contain indoor and outdoor scenes, and include objects near and far away from the camera. For each image, three viewports were rendered, corresponding to three different viewing directions, with 70% overlapping between successive directions. This allows to compare different levels of stretching distortion of the same objects, when these objects appear in different positions on the viewports. The viewports were rendered for the region of the omnidirectional image where the users attention is often attracted for, using true saliency maps available for the images taken from [17], and the attention-related model proposed in [18] for the images taken from [8].

For each viewing direction, two viewports were rendered with the Pannini projection (PP), using different parameter values: $PP_1(d = 0, vc = 0)$; $PP_2(d = 0.5, vc = 0)$. Thus, for each image, six viewports were produced, denoted as VP_i, $i = 1, 2, \ldots, 6$, where VP_1, VP_2, VP_3 correspond to PP_1, and VP_4, VP_5, VP_6 are correspond to PP_2. PP_1, which is a rectilinear projection, was selected since it is often used for viewport rendering of omnidirectional images, and results on strong objects stretching. PP_2 was included to get, for the same viewing directions, different levels of stretching distortion of the objects. The viewports were rendered with a F_h of either 110° or 115° (presented for each ERI on the bottom of Table 1), and with a spatial resolution of 856 × 856 pixels ($AR = 1$); besides being recommended in [19] for subjective tests, this resolution allows the simultaneously display of two viewports, side-by-side, in typical monitors. The F_h of 110° was selected based on the study in [5]; moreover, it is a moderate size often used in HMD displays. As shown in Table 1, for some of the images a F_h of 115° was used to guarantee that the main objects were not cut by the image borders.

C. SUBJECTIVE EVALUATION METHOD

In this work, the pairwise comparison (PC) was chosen as the subjective evaluation method. Nowadays, PC became very popular for image and video quality assessment [20]–[22],
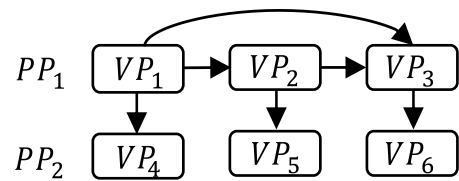

FIGURE 5. Considered comparisons per omnidirectional image.

TABLE 1. Selected pairs (indicated in ✓) and F_h for each 360° images.

Pair	Ind1	Ind2	Ind3	Ind4	Ind5	Ind6	Ind7	Outd1	Outd2	Outd3
(VP_1, VP_2)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(VP_2, VP_3)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(VP_1, VP_3)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(VP_1, VP_4)	✓	✓	✗	✗	✗	✓	✗	✗	✓	✓
(VP_2, VP_5)	✓	✓	✗	✗	✗	✗	✓	✗	✓	✓
(VP_3, VP_6)	✗	✓	✗	✓	✗	✗	✓	✗	✓	✓
F_h	110°	110°	110°	115°	115°	115°	115°	115°	110°	115°

particularly for the evaluation of rendering methods; in this method, the viewer is asked to observe a pair of rendered viewports, shown side by side, and to select the one that has higher quality in his opinion.

As depicted in Fig. 5, six comparisons were made per 360° image: a complete set of comparisons between the viewports rendered with PP_1, and three additional comparisons between the viewports rendered with PP_1 and PP_2, and having the same viewing direction. The comparisons between viewports rendered with PP_2 were not considered to limit the test duration; furthermore, these viewports have similar level of stretching distortion. Also, since for some of the 360° images and viewing directions the bending distortion was visible for viewports rendered with PP_2, these were excluded from the subjective test. Table 1 presents the selected pairs and horizontal field of view, F_h, for each 360° images.

In total, 45 viewport pairs were considered. A web-based crowdsourcing interface was designed to perform the visualization of the stimuli and to collect the PC scores; it presents two viewport images, 'A' and 'B', side by side as shown in Fig. 6, and requires a monitor with a minimum resolution of 1920×1080 pixels, with a minimum diagonal size of 13-inch. To participate in the test, an invitation email with detailed instructions was sent to several observers; the observers were
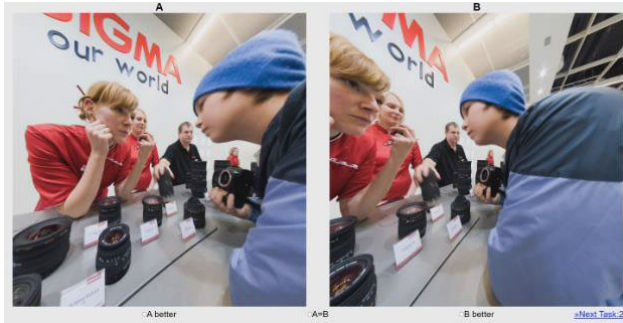
FIGURE 6. PC web-based crowdsourcing interface.

FIGURE 7. CDF with $O = 28$ and $p = 0.5$.

asked to not perform the test if they do not have a monitor with the aforementioned characteristics. Before starting the subjective test, the observers were asked to: i) open the subjective test interface in their web browser and put it on full-screen mode; ii) type their name and age in the top of the page; iii) select their monitor size in inches. The instructions about the subjective test procedure were shown on the same page. Subsequently, to familiarize the observer with the stretching distortion's characteristics and the evaluation interface, a short training video was shown.

The viewports used in the training video were not used for the actual test. During the test, the viewport pairs were shown in random order and position, and the observers were asked to judge which viewport image ('A' or 'B') had the best quality. To avoid random preference selections, the option 'A = B' was also included. A total of 32 subjects, aged between 21 and 58 years, from Instituto Superior Técnico (IST), performed the online subjective evaluation. The omnidirectional images, the rendered viewports, and the resulting PC subjective scores were made publicly available in [12].

D. SUBJECTIVE TESTS RESULTS AND ANALYSIS

Outliers were first detected by computing the transitivity satisfaction rate, R, [23]; this rule is violated when a circular triad is formed among the three stimuli, VP_1, VP_2, and VP_3 (e.g., $VP_1 > VP_2$, $VP_2 > VP_3$, $VP_3 > VP_1$, where > means "better than"). The score reliability, R_o, of observer o, was computed as:

$$R_o = 1 - \frac{d_o}{h_o}, \qquad (13)$$

where d_o is number of detected circular triads and h_o is the total number of possible circular triads for an observer. If $R_o < 0.8$, the observer o is considered to be an outlier. Four outliers were detected, and their subjective scores were not further considered. Next, for each compared viewports pair, (VP_i, VP_j), the winning frequency, w_{ij}, which represents the number of times VP_i was preferred over VP_j, was computed. To solve the tie cases, a score of 0.5 was given to each viewport whenever the observer had chosen the option 'A = B'. Note that $w_{ij} + w_{ji} = O$, where O is the number of observers, and $w_{ii} = 0$. The probability of selecting
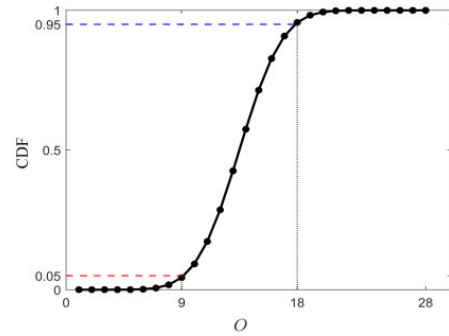
VP_i against VP_j, is given by:

$$P_{ij} = P\left(VP_i > VP_j\right) = \frac{w_{ij}}{O}. \qquad (14)$$

To determine whether the difference on the number of times VP_i was preferred over VP_j (and vice-versa) is statistically significant, a statistical hypothesis test was performed according to the procedure suggested in [20]. After solving the tie cases, the PC scores roughly follow a Bernoulli process $B(O, p)$, where O is the number of subjects and p is the probability of success in a Bernoulli trial. Fig. 7 depicts the cumulative distribution function (CDF) for a Binomial distribution with $O = 28$ (the final number of observers, after outliers removal) and $p = 0.5$, as suggested in [20], meaning that when comparing VP_i and VP_j both have the same chance of being selected. The CDF of the Bernoulli distribution can be expressed as [24]:

$$CDF(k; O, p) = \sum_{i=1}^{\lfloor k \rfloor} \binom{O}{i} p^i (1-p)^{(O-i)}, \qquad (15)$$

where k is the number of times that VP_i was selected over VP_j, p is the probability of selecting VP_i over the VP_j in the Bernoulli trial, and $\lfloor \cdot \rfloor$ is the floor operator. The resulting CDF value is the probability that the observers select, k times, VP_i over VP_j. The critical region for the statistical test is obtained from the CDF. To find out whether the number of times VP_i was preferred over VP_j is statistically significant, thus allowing to conclude that "VP_i is better than VP_j", a one-tailed binomial test was performed with a significant level of 0.05, with the following hypothesizes: H0 (VP_i is equal or worse than VP_j); H1 (VP_i is better than VP_j). In Fig. 7, the CDF has values above the probability of 0.95 for $O \geq 18(F(18, 28, 0.5) = 0.9564)$. Therefore, if $k \geq 18$, the null hypothesis (H0) can be rejected. A similar statistical test was applied to find out if the number of times that VP_j was preferred over VP_i is statistically significant, thus allowing to conclude that "VP_i is worse than VP_j", with the following hypothesizes: H0 (VP_i is equal or better than VP_j); H1 (VP_i is worse than VP_j). In Fig. 7, the CDF has values below the probability of 0.05 for $O \leq 9(F(9, 28, 0.5) = 0.0436)$. Therefore, if $k \leq 9$, the null hypothesis (H0) can be rejected. Note that the Bernoulli process is defined only for integer
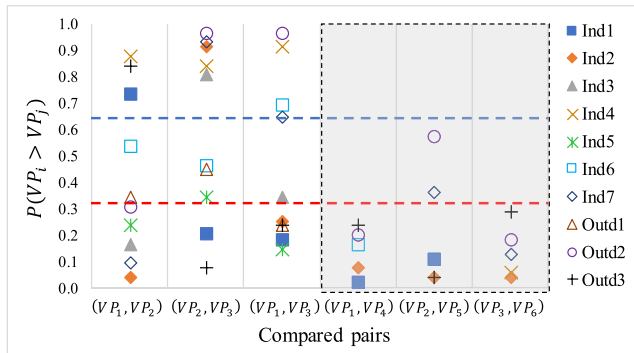
FIGURE 8. Preference probability of selecting VP_i over VP_j for compared pairs.

values, and non-integer values need to be rounded; the floor function is used for this purpose.

Fig. 8 presents the probability of selecting VP_i over VP_j, computed by (14), for each compared pair, and for all considered 360° images. The horizontal blue dashed line corresponds to the case where the vote count for VP_i is equal to, or greater than, 18, i.e., $P\left(VP_i > VP_j\right) = 18/28 = 0.643$. The horizontal red dashed line corresponds to the case where the vote count for VP_i is equal or less than 9, i.e. $P\left(VP_i > VP_j\right) = 9/28 = 0.321$. Values on or above the horizontal blue dashed line, and on or below the horizontal red dashed line, correspond to the cases where the difference in the votes between VP_i and VP_j is statistically significant; in the first case, VP_i has a higher perceived quality than VP_j; on the second case, VP_i has a lower perceived quality than VP_j. The values between the two horizontal dashed lines correspond to the cases where the difference in the votes between VP_i and VP_j is not statistically significant.

From the experimental results, the following conclusions can be obtained:

- As can be observed in Fig. 8, the viewports rendered with $PP_2(d = 0.5, vc = 0)$ were selected over those rendered with $PP_1(d = 0, vc = 0)$, for most of the considered images (values inside the shaded area in Fig. 8). This was expected since the rectilinear projection has a strong stretching effect, and this stretching decreases as the value of d increases.
- For the viewports rendered with PP_1, the preferred ones (by the subjects) are strongly dependent on the position of the main objects, because the stretching distortion has a higher perceptual impact when the objects are close to the image borders, and/or close to the camera. As an example, Fig. 9 shows three pairs of viewports rendered with PP_1, together with the subject selections; due to the difference on viewing direction, the same objects are rendered in different positions of a viewport pair, having different perspective distortion.
- The human perception is very sensitive to the stretching distortion of the human body, which may justify why, as shown in Fig. 9, most of the observers have selected VP_2 over VP_3 for *Outd2*, and VP_2 over VP_1 for *Ind2*,
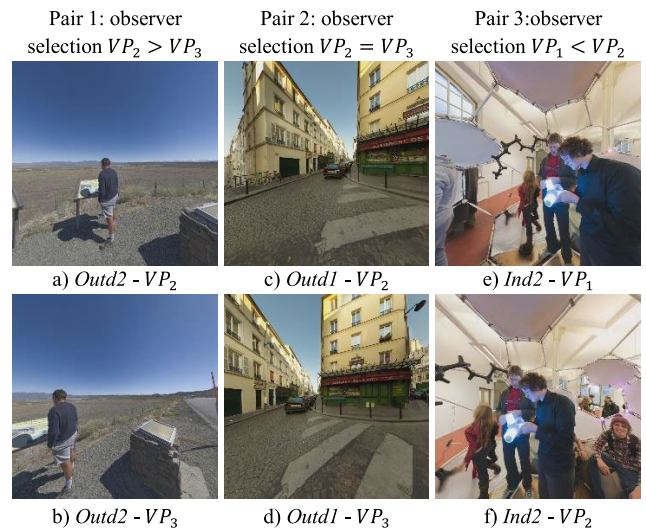

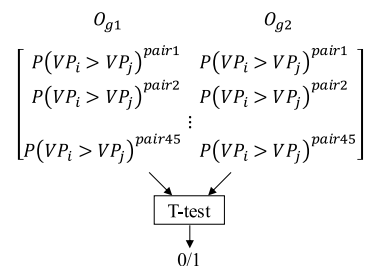**FIGURE 9.** Rectilinear viewports, in pairs, evaluated by observers.

FIGURE 10. T-test procedure for evaluating if there is a significant impact of screen size on the preference probability.

while for *Outd1* there is no clear choice between VP_2 and VP_3.

- The stretching distortion of the background, e.g., sky, floor, building walls, has a lower perceptual impact than the stretching of foreground objects.

Finally, and since the subjects were used different monitor sizes, the impact of it on the subjective results was assessed. For that, the observes were divided into two groups: O_{g1}, containing the observers that have a monitor size in the range [13,16] inch, and O_{g2}, contains the observers that have a monitor size in the range [22,27] inch. Then, a paired sample T-test with a significant level of 0.05 was applied, to compare the preference probability, $P(VP_i > VP_j)$, between groups. This procedure is illustrated in Fig. 10. The T-test results indicate that the null-hypothesis, i.e., that the two groups have similar means, cannot be rejected since the resulting *p*-value, 0.57, is much higher than the significant level of 0.05; this confirms that the difference on the subjective results from the two groups is not statistically significant.

IV. OBJECT-BASED STRETCHING DISTORTION MEASUREMENT

This section describes two new approaches for measuring the object shape distortion in viewport rendering of 360° images.
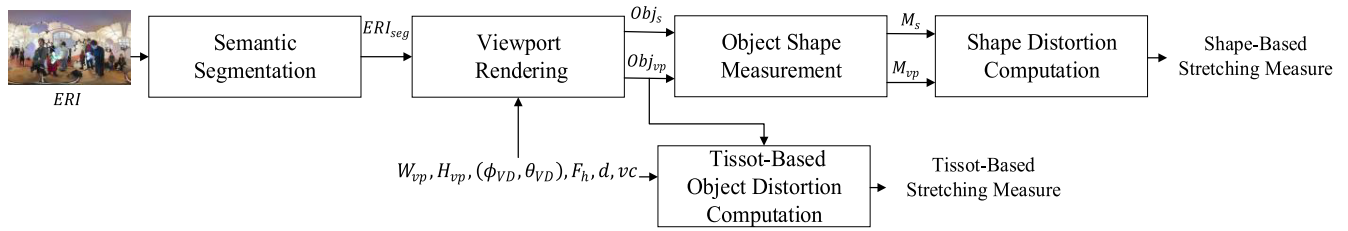
FIGURE 11. The main architecture of the proposed object-based stretching metrics.

The first one directly computes and compares object shape measures on the sphere and on the viewport, thus before and after rendering, while the second is based on the Tissot indicatrices [10], which are computed for individual objects in the rendered viewport. As depicted in Fig. 11, the process starts with the semantic segmentation of the 360° image, in equirectangular format (ERI), producing a segmentation map denoted as ERI_{seg}. Afterwards, for a required viewport horizontal field-of-view, F_h, spatial resolution (W_{vp}, H_{vp}), and viewing direction (ϕ_{VD}, θ_{VD}), the viewport rendering process is applied to ERI_{seg}, resulting in the viewport segmentation map. For the rendering, the Pannini projection, with parameters (d, vc), is used. The objects distortion is then computed, using the two approaches aforementioned. The main steps are described in the following sections.

A. SEMANTIC SEGMENTATION
Semantic segmentation is a process of assigning a label (e.g., person, car, bicycle, and so on) to objects in the image; in this process, multiple objects of the same class have the same label; it has been used in many computer vision tasks, using 2D images. Although some semantic segmentation models have been developed for 360° images (e.g., [25]–[28]), they were designed for the purpose of autonomous driving, with outdoor images. In this paper, to obtain the semantic segmentation of both indoor and outdoor 360° images, the input equirectangular image (ERI) is transformed to cubic format, which results in six 2D, rectilinear projected images (the cube faces), with horizontal and vertical FoVs of 90°. The Auto-DeepLab semantic segmentation model, proposed in [29], is then applied to each cube face. This model is a deep learning-based approach designed for semantic segmentation of 2D images, that was trained, validated, and tested on several datasets that include indoor and outdoor scenes, and has high accuracy. In this work, it was used the Auto-DeepLab with multi-scale inference and the network backbone *Xception-65*, pre-trained on ImageNet [30], and MS-COCO [31] datasets. The training was performed on the PASCAL VOC 2012 dataset [32], which contains 20 foreground object classes and one background class. As described in [29], the training was performed with a polynomial learning rate with an initial value of 0.05, and a crop size of 513×513 pixels. Batch normalization parameters were fine-tuned during training. After obtaining the semantic segmentation for all six cube face images, it is
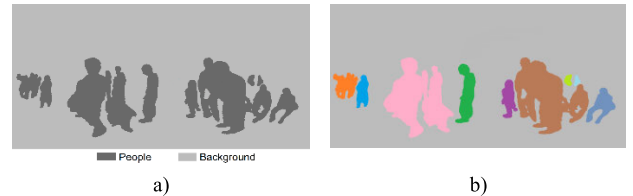

FIGURE 12. a) Semantic segmentation of *Ind2*; b) Disconnected objects.

transformed back to equirectangular format. As an example, Fig. 12-a) depicts the semantic segmentation of *Ind2*, using Auto-DeepLab. As already mentioned, multiple objects of the same class have the same label. To obtain different labels for disconnected objects, the connected component analysis (CCA) [33], with 4-connectivity, is applied to the segmented ERI. Fig. 12-b) depicts the resulting ERI segmentation map after CCA, where each object is represented with a different color.

B. OBJECT SHAPE MEASUREMENT
The object shape distortion measure can be obtained by relating the object shape on the viewing sphere and on the viewport. Several object shape measures have been proposed on the literature [34], [35]. Since the sphere to plan projections typically alter the area of the objects, or objects are stretched in the horizontal and/or vertical directions towards the viewport borders (*cf.* Figs. 2 and 6), three shape measures were considered: area, average width and average height. In cartography, these measures showed good performance when used to characterize the distortion of continents and countries for different map projections [36]; this justifies why they were chosen for the study described in this paper.

After semantic segmentation of the ERI image, it is possible to obtain the semantic segmentation map for any viewport by projecting ERI_{seg}; this allows to obtain the objects in the viewport, Obj_{vp}, linked to the same objects on the viewing sphere, Obj_s. Fig. 13-a) depicts an example of a viewport from *Ind2* and its segmentation map, with three objects (Fig. 13-b)), obtained by projecting ERI_{seg}; the same objects are also identified in ERI_{seg} (Fig. 13-c)). The following object shape measures were considered:

- **Object area**: On the sphere, the object area can be computed by summing up the area covered by parallel lines (defined as a sequence of pixels) within the object. At latitude θ, the parallel line area, $PLA_s(\theta)$, contained
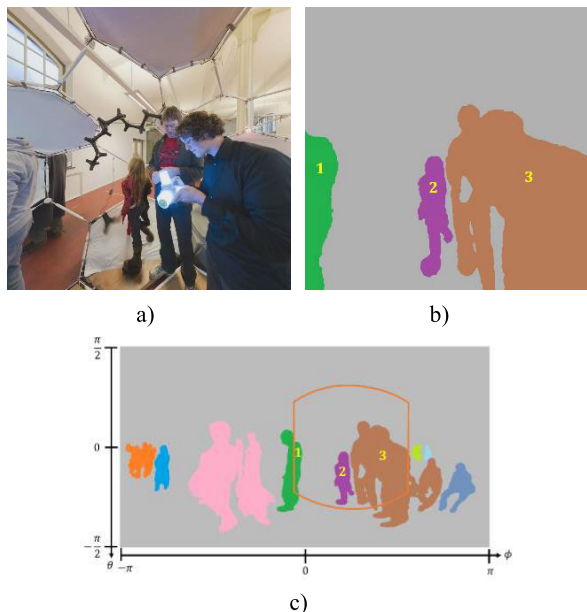
FIGURE 13. a) A viewport of *Ind2*; b) Three identified objects in the viewport; c) Corresponding objects in the ERI.

in an object is given by:

$$PLA_s(\theta) = PA_s(\theta) N_{ERI}^{PL}(\theta), \quad (16)$$

where $PA_s(\theta)$ is the area covered by a pixel at latitude θ, and $N_{ERI}^{PL}(\theta)$ is the total number of pixels within the object at latitude θ. $PA_s(\theta)$ can be approximated by:

$$PA_s(\theta) = \Delta\phi\Delta\theta cos(\theta) = \frac{2\pi}{W_{ERI}} \frac{\pi}{H_{ERI}} \cos(\theta), \quad (17)$$

where W_{ERI} and H_{ERI} are, respectively, the width and height of the ERI image, in pixels, $\Delta\phi\Delta\theta$ is the area covered by a pixel in the ERI image, and $\cos(\theta)$ reflects the decrease in the area (on the sphere) comprised by $\Delta\phi, \Delta\theta$, as θ varies from 0 to ± 90 degrees. The object area, on the sphere, is computed by:

$$OA_s = \sum_{k=1}^{K_{ERI}^{PL}} PLA_s(\theta_k), \quad (18)$$

where K_{ERI}^{PL} is the total number of parallel lines covered by the object, $k = 1 \ldots K_{ERI}^{PL}$ is the index of those lines, and θ_k is the latitude of the k-th parallel line.

The area covered by a pixel on the viewport, PA_{vp}, is given by:

$$PA_{vp} = \frac{V_{hs}}{W_{vp}} \frac{V_{vs}}{H_{vp}}, \quad (19)$$

where V_{hs} and V_{vs} are, respectively, the viewport width and height, in length unit, given by (10). The object area in the viewport is computed by:

$$OA_{vp} = PA_{vp} N_{vp}^{obj}, \quad (20)$$

where N_{vp}^{obj} is the total number of pixels within the object.

- **Object average width**: On the sphere, and at latitude θ, the width of the object, $OW_s(\theta)$, is the length of the parallel line at θ, covered by the object. Since, in discrete domain, each parallel corresponds to a line on the ERI image, $OW_s(\theta)$ can be computed as:

$$OW_s(\theta) = \frac{2\pi}{W_{ERI}} N_{ERI}^{PL}(\theta) \cos(\theta), \quad (21)$$

where $N_{ERI}^{PL}(\theta)$ is the total number of pixels within the object at latitude θ, and $2\pi/W_{ERI}$ is the width covered by a pixel in the ERI image. The object average width, on the sphere, is computed by:

$$OW_s = \frac{1}{K_{ERI}^{PL}} \sum_{k=1}^{K_{ERI}^{PL}} OW_s(\theta_k). \quad (22)$$

On the viewport, the width of the object at line i can be computed as:

$$OW_{vp}(i) = \frac{V_{hs}}{W_{vp}} N_{vp}^l(i), \quad (23)$$

where $N_{vp}^l(i)$ is the total number of pixels covered by the object at line i, and V_{hs}/W_{vp} is the width covered by a pixel in the viewport image. The object average width, on the viewport, is given by:

$$OW_{vp} = \frac{1}{K_{vp}^l} \sum_{i\in\text{Obj}} OW_{vp}(i), \quad (24)$$

with the summation applied to the viewport lines covered by the object, and K_{vp}^l being the total number of those lines.

- **Object average height**: On the sphere, at longitude ϕ, the object height is the length of the meridian line (ML) at ϕ - which corresponds to a column of the ERI image - covered by the object:

$$OH_s(\phi) = \frac{\pi}{H_{ERI}} N_{ERI}^{ML}(\phi), \quad (25)$$

where $N_{ERI}^{ML}(\phi)$ is the total number of pixels within the object at longitude ϕ, and π/H_{ERI} is the height covered by a pixel in the ERI image. The object average height, on the sphere, is given by:

$$OH_s = \frac{1}{K_{ERI}^{ML}} \sum_{k=1}^{K_{ERI}^{ML}} OH_s(\phi_k), \quad (26)$$

where K_{ERI}^{ML} is the total number of meridian lines covered by the object, $k = 1 \ldots K_{ERI}^{ML}$ is the index of those lines, and ϕ_k is the longitude of the k-th meridian line.

On the viewport, the height of the object at viewport column j, can be computed as:

$$OH_{vp}(j) = \frac{V_{vs}}{H_{vp}} N_{vp}^c(j), \quad (27)$$

where $N_{vp}^c(j)$ is the total number of pixels covered by the object at column j, and V_{vs}/H_{vp} is the height covered by

TABLE 2. Resulting *OA*, *OW*, *OH* values for three objects on the sphere, before projection, and in the viewport, after projection.

Object \\ Measure	1	2	3
OA_s	0.067	0.132	0.807
OA_{vp}	0.336	0.220	1.990
OW_s	0.073	0.161	0.605
OW_{vp}	0.215	0.183	1.066
OH_s	0.564	0.470	0.913
OH_{vp}	1.084	0.652	1.386

TABLE 3. Object distortion pooling functions.

$PF_1 = \text{Sum}(\boldsymbol{D})$	(33)	$PF_2 = \text{Sum}(\boldsymbol{D_p})$	(34)
$PF_3 = \text{Average}(\boldsymbol{D})$	(35)	$PF_4 = \text{Average}(\boldsymbol{D_p})$	(36)
$PF_5 = \text{Max}(\boldsymbol{D})$	(37)	$PF_6 = \dfrac{\text{Sum}(\boldsymbol{OA_{vp}} \odot \boldsymbol{D})}{\text{Sum}(\boldsymbol{OA_{vp}})}$	(38)

a pixel in the viewport image. The object average height, on the viewport, is given by:

$$OH_{vp} = \frac{1}{K_{vp}^c} \sum_{j \in \text{Obj}} OH_{vp}(j), \qquad (28)$$

with the summation applied to the viewport columns covered by the object, and K_{vp}^c being the total number of those lines.

It is important to note that all the shape measurements are in length units, and are obtained only for the objects (or parts of the objects) that are rendered on the viewport. As an example, only the parts of objects 1 and 3 that can be seen in Fig. 13-b), were used for the shape measures. Table 2 presents the resulting *OA*, *OW*, and *OH* values, on the sphere and on the viewport, for the three objects of Fig. 13-b). All the measures increase after projection, especially for the objects closer to the viewport borders.

C. SHAPE DISTORTION COMPUTATION
Based on the object shape measures previously presented, the following object shape distortion metrics are defined:

- **Area distortion:** For each object in the viewport, the area distortion is expressed by:

$$OAD = \left| OA_s - OA_{vp} \right|, \qquad (29)$$

where the OA_s and OA_{vp} are computed by (18) and (20), respectively.

- **Width distortion**: The object width distortion is given by:

$$OWD = \left| OW_s - OW_{vp} \right|, \qquad (30)$$

where OW_s and OW_{vp} are given by (22) and (24), respectively. This measure characterizes the horizontal stretching of the object.

- **High distortion**: The object high distortion is computed by:

$$OHD = \left| OH_s - OH_{vp} \right|, \qquad (31)$$

where OH_s and OH_{vp} are computed by (26) and (28), respectively. This measure characterizes the vertical stretching of the object.

- **Total length distortion**: The total length distortion of an object is defined as:
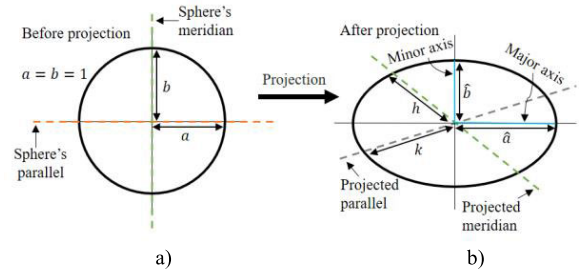
$$OTD = OWD + OHD, \qquad (32)$$

FIGURE 14. a) Infinitesimal unit circle defined on the sphere; b) Its corresponding Tissot's indicatrix after projection on the plane.

where *OWD* and *OHD* are computed by (30) and (31), respectively.

It is important to mention that, besides the absolute difference expressed by (29) to (31), the relative difference was also considered, but did not improve the performance of the metric. To obtain a global viewport stretching distortion measure, several pooling functions were considered to aggregate the shape distortion measure of all detected objects in the viewport. The considered pooling functions are listed in Table 3, where, $\boldsymbol{D}$ is a vector containing one of the distortion measures for all objects in the viewport, and $\boldsymbol{D_p}$ is a vector containing the $p\%$ highest elements of $\boldsymbol{D}$; $\boldsymbol{OA_{vp}}$ is a vector containing the object area on the viewport, and $\odot$ denotes element-wise product. Poolings PF_1 and PF_2 assume that the subjective impact of the distortion increases with the number of objects, while pooling PF_3 and PF_4 consider that the impact varies with the average objects distortion; pooling PF_5 presume that the perceptual impact is mainly influenced by the most distorted object, while pooling PF_6 considers the object area in the viewport, giving more emphasis to the distortion of large objects. The reason for the percentile ($p\%$) is to exclude the objects with low distortion values (e.g., the distortion for objects at the viewport center is low and may not be visible); as $p\%$ approaches 100%, PF_2 and PF_4 will be closer to PF_5; if $p\%$ approaches 0%, PF_2 will be closer to PF_1 and PF_4 will be closer to PF_3. In summary, considering four shape distortion measures with six pooling functions, results in 24 potential shape-based stretching measures.

D. TISSOT-BASED OBJECT DISTORTION COMPUTATION
The Tissot indicatrix [10] has been used for years by cartographers to evaluate and compare distortion on different Earth map projections. This indicatrix is obtained after projection, in the map, an infinitely small circle defined on the sphere; the relationship between the major and minor axis of the resulting ellipse, after projection, enables to compute the local scale, area, and angle distortions, at the
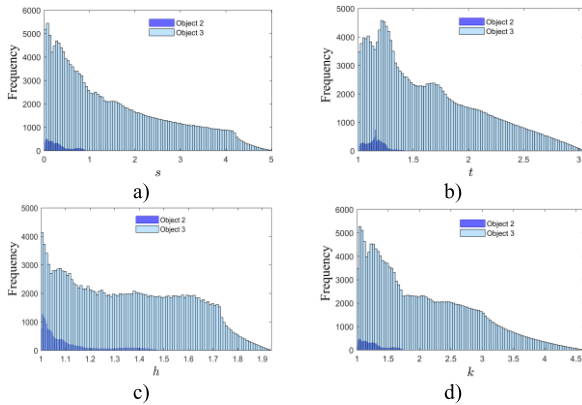
FIGURE 15. Histogram plots of s, t, h, k for two identified viewport objects, Object 2 and Object 3 of Fig. 13-b): a) Local area distortion, s; b) Local shape distortion, t; c) Scale factor, h; d) Scale factors, k.
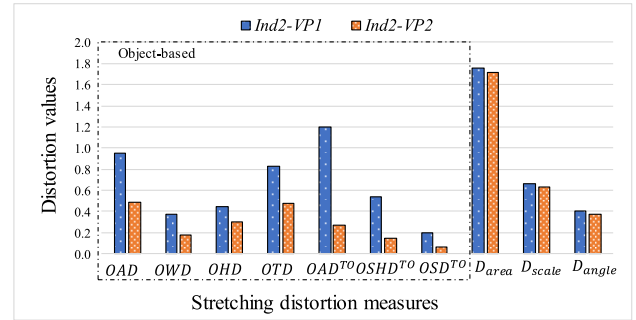

FIGURE 16. Stretching distortion values for the proposed object-based measures, and for the stretching measures proposed in [11], computed for a pair of viewports, $Ind2 - VP_1$ and $Ind2 - VP_2$, presented respectively in Fig. 9-e) and Fig. 9-f).

projected point. Fig. 14 depicts an infinitesimal unit circle defined on the sphere, and its corresponding Tissot's indicatrix after projection on the plane; $\hat{a}$ and $\hat{b}$ are, respectively, the Tissot's indicatrix semi major and minor axis; h and k correspond, respectively, to the scale factor along the projected meridian and parallel. In the authors' previous work [11], three viewport Tissot distortion measures - namely, *area distortion* (D_{area}) *scale distortion* (D_{scale}), and *angle distortion* (D_{angle}) - were defined to measure the stretching distortion in the viewport rendering of 360° images, under the general perspective projection; to make these measures content dependent, saliency weights were used. In this paper, object based Tissot distortion measures are proposed, and obtained according to the two following steps:

1) **Compute local Tissot distortion metrics**: For a given horizontal and vertical field-of-view, F_h and F_v, the viewing area on the sphere is defined by $\phi \in [-F_h/2, F_h/2]$ and $\theta \in [-F_v/2, F_v/2]$; this region is then uniformly sampled with a fixed interval $\Delta\phi$, $\Delta\theta$ (set to 0.05 degrees in this paper). For each sampled point, indexed by i, with spherical coordinates (ϕ_i, θ_i), the corresponding Tissot scale factors h_i and k_i, semi-major, $\hat{a}_i$, and semi-minor, $\hat{b}_i$, axis of the Tissot ellipse are obtained. The details about the computation of these parameters are presented in the appendix. Afterwards, the local area distortion, s_i, and local shape distortion, t_i, are computed as proposed in [11]:

$$s_i = \left(\hat{a}_i\hat{b}_i - 1\right)\cos\theta_i, \qquad (39)$$

$$t_i = \frac{\hat{a}_i}{\hat{b}_i}. \qquad (40)$$

Although the local angle distortion was also initially considered, it did not improve the results, and was not retained for further assessment. Fig. 15 presents histogram plots of s, t, h, k for two objects of Fig. 13-b), namely *Object* 2 (close to the viewport center and with low distortion), and *Object* 3 (close to the border and

with high distortion). As can be seen, s, t, h, k have a wider range of values (and with higher variance) for *Object* 3, than for *Object* 2.

2) **Compute Tissot-based object distortion metrics**: For each object in the viewport, the following object-based Tissot distortion metrics are obtained:

$$OAD^{TO} = Variance\left(s\right), \qquad (41)$$

$$OSHD^{TO} = Variance\left(t\right), \qquad (42)$$

$$OSD^{TO} = \text{Max}\left(Variance\left(h\right), Variance\left(k\right)\right), \quad (43)$$

where OAD^{TO}, $OSHD^{TO}$, OSD^{TO}, are, respectively, the object-based Tissot area, shape, and scale distortion metrics. The superscript *TO* denotes object-based Tissot measure; s and t are vectors containing, respectively, the local area and shape distortions for all points within an object; h and k are vectors containing the scale factors for all points within an object. To obtain a single measure per object, the *Variance* and *Average* functions were considered in (41) to (43); however, the *Variance* function was selected as it showed the best performance.

To obtain a global viewport stretching distortion measure, the pooling functions presented in Table 3 of Section IV-C were used. In this case, using three Tissot based object distortion measures, with six pooling functions, results in 18 potential Tissot based stretching distortion metrics.

Fig. 16 presents the resulting stretching distortion values for a sub-set of the proposed object based stretching measures with pooling function PF_6, and the three stretching measures - D_{area}, D_{scale}, and D_{angle} - that were previously proposed in [11], computed for a pair of viewports, $Ind2\text{-}VP_1$ and $Ind2\text{-}VP_2$ (depicted in Fig. 9 -e) and -f)); the blue and orange bars correspond, respectively, to $Ind2\text{-}VP_1$ and to $Ind2\text{-}VP_2$. As can be figured out, the proposed object based stretching measures allow a higher discrimination between the quality of the two viewport images, than the metrics proposed in [11], since for the former the difference between blue and orange bars are much more evident.

V. METRICS EVALUATION

In this section, the proposed object-based stretching distortion metrics are evaluated and compared to benchmark solutions. The usual measure to evaluate an objective quality metric is the correlation (Pearson and/or Spearman) between objective scores and ground truth opinion scores (typically, MOS or DMOS). However, since in this work pairwise comparison (PC) was used on the subjective tests, it is not possible to obtain MOS or DMOS values for each individual stimulus. Accordingly, the proposed metrics are assessed versus PC scores using the classification errors approach, as suggested in Rec. ITU-T J.149 [37], and applied in related literature [20], [21].

A. CLASSIFICATION ERRORS

According to Rec. ITU-T J.149 [37], a classification error (CE) occurs when the objective and subjective scores lead to different conclusions about the relative quality of a pair of stimuli, VP_i and VP_j. Three types of errors may happen:

- **False Tie (FT)**: when the subjective score indicates that VP_i and VP_j are different, but the objective score indicates that they are similar.
- **False Differentiation (FD):** when the subjective score indicates that VP_i and VP_j are similar, but the objective score indicates that they are different.
- **False Ranking (FR):** when the subjective score indicates that $VP_i(VP_j)$ is better than $VP_j(VP_i)$, but the objective score indicates the opposite.

Let ΔOM represent the minimum difference, between the objective quality scores of two stimuli, that defines when the two stimuli become perceptually distinguishable. As ΔOM increases, more stimuli pairs are considered similar, increasing the occurrence of FT, but the occurrences of FD and FR will decrease. On the contrary, as ΔOM decreases, the occurrence of FT also decreases, but the occurrence of FD and FR will increase. Following ITU-T J.149, the percentage of each error type and of correct decisions are obtained from the considered stimuli pairs as a function of ΔOM, for individual metrics; this allows to compare the metrics and determine the best one for the application under analysis. The best ΔOM value is the one that maximizes the correct decision percentage [21], [37].

B. EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the proposed distortion metrics, the viewport dataset described in Section III-B, and the processed PC scores described in Section III.D, were used. Moreover, the performance of the metrics were compared to the following benchmark solutions: *area distortion* (D_{area}), *scale distortion* (D_{scale}), and *angle distortion* (D_{angle}), proposed in [11], the *conformality* measure (CM) proposed in [13], and the *content-dependent conformality* (CM^{sal}) proposed in [8]; the latter is a modified CM, by integrating the viewport saliency on it. For poolings PF_2, PF_4, several values of $p\%$ were

TABLE 4. Classification errors and correct decision values, in percentage (%), for the proposed and benchmark metrics.

Correct Decision (%)							False Tie (%)						
PF_i / Metric	PF_1	PF_2	PF_3	PF_4	PF_5	PF_6	PF_i / Metric	PF_1	PF_2	PF_3	PF_4	PF_5	PF_6
OAD	80.0	84.4	82.2	82.2	84.4	82.2	OAD	2.2	0.0	0.0	0.0	2.2	4.4
OWD	64.4	71.1	82.2	75.6	73.3	80.0	OWD	0.0	2.2	15.6	24.4	4.4	0.0
OHD	71.1	68.9	82.2	80.0	71.1	77.8	OHD	2.2	2.2	0.0	2.2	0.0	6.7
OTD	68.9	68.9	84.4	86.7	71.1	82.2	OTD	0.0	4.4	2.2	2.2	2.2	8.9
OSD^{TO}	84.4	84.4	86.7	86.7	84.4	84.4	OSD^{TO}	0.0	0.0	0.0	2.2	0.0	2.2
OAD^{TO}	84.4	82.2	86.7	86.7	86.7	88.9	OAD^{TO}	0.0	0.0	0.0	0.0	0.0	0.0
$OSHD^{TO}$	82.2	84.4	84.4	84.4	84.4	82.2	$OSHD^{TO}$	0.0	0.0	0.0	0.0	0.0	2.2

False Differentiation (%)							False Ranking (%)						
PF_i / Metric	PF_1	PF_2	PF_3	PF_4	PF_5	PF_6	PF_i / Metric	PF_1	PF_2	PF_3	PF_4	PF_5	PF_6
OAD	11.1	8.9	13.3	15.6	8.9	8.9	OAD	6.7	6.7	4.4	2.2	4.4	4.4
OWD	17.8	15.6	2.2	0.0	13.3	15.6	OWD	17.8	11.1	0.0	0.0	8.9	4.4
OHD	13.3	13.3	13.3	11.1	13.3	6.7	OHD	13.3	15.6	4.4	6.7	15.6	8.9
OTD	17.8	13.3	8.9	6.7	13.3	6.7	OTD	13.3	13.3	4.4	4.4	13.3	2.2
OSD^{TO}	11.1	8.9	8.9	8.9	13.3	11.1	OSD^{TO}	4.4	6.7	4.4	2.2	2.2	2.2
OAD^{TO}	13.3	13.3	13.3	13.3	13.3	11.1	OAD^{TO}	2.2	4.4	0.0	0.0	0.0	0.0
$OSHD^{TO}$	11.1	11.1	8.9	11.1	11.1	11.1	$OSHD^{TO}$	6.7	4.4	6.7	4.4	4.4	4.4

Benchmark Metrics

Metric	Correct Decision	False Tie	False Differentiation	False Ranking
D_{area}	68.9	0.0	17.8	13.3
D_{scale}	68.9	0.0	15.6	15.6
D_{angle}	66.7	17.8	6.7	8.9
CM^{sal}	66.7	0.0	15.5	17.8
CM	42.2	53.3	4.4	0.0

considered, and the resulting classification errors and correct decision were obtained. The best performance was obtained for $p = 50\%$.

Table 4 reports the classification errors and correct decision values for each proposed distortion measure, using the pooling functions described in section IV-C, and for the benchmark solutions; the ΔOM value that maximized the correct decision percentage was used. As can be figured out, there is a significant performance improvement for the object-based metrics, when compared with the benchmark solutions. Among the proposed metrics, the object-based Tissot metrics achieved the highest Correct Decision and the lowest False Tie percentages. Among the benchmark metrics, the CM has the worst performance. This metric is content independent, and the same metric value is obtained for any viewport image. When the conformality integrates the saliency, as in CM^{sal}, the performance increases, confirming that it brings some additional value to the metric. To find out the best solution among the proposed ones, the true positive rate (TPR), defined by (44), was computed:

$$TPR = \frac{CD}{CD + FT + FD + FR}. \tag{44}$$

Table 5 presents the best pooling function for each metric and the resulting *TPR* values. Since the metric OAD^{TO} with PF_6 has the highest *TPR* value, it was the selected one. For the benchmark metrics, the *TPR* have values of 0.69, 0.69, 0.67, 0.67, and 0.42, for D_{area}, D_{scale}, D_{angle}, CM^{sal}, and CM, respectively. These values are much lower than the

TABLE 5. Selected pooling function for each metric and corresponding *TPR* values.

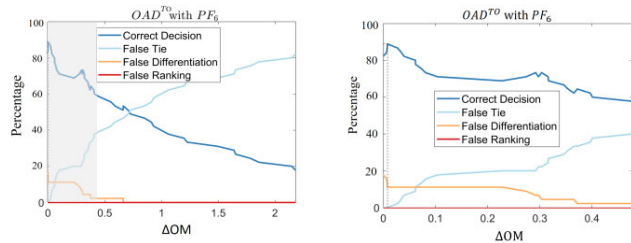| Metric	OAD	OWD	OHD	OTD	OSD^{TO}	OAD^{TO}	$OSHD^{TO}$
Pooling	PF_5	PF_3	PF_3	PF_4	PF_4	PF_6	PF_5
TPR	0.84	0.82	0.82	0.87	0.87	**0.89**	0.84

FIGURE 17. Plots of classification error and correct decision for the best proposed metric. The shaded area on the left side plot is represented in the right side plot.

TPR value of 0.89 obtained for OAD^{TO} with PF_6, being also lower than the *TPR* values obtained for the other proposed metrics.

Fig. 17 depicts the plots of classification errors and correct decision for the selected metric, OAD^{TO} with PF_6, where the dashed line on the right side plot indicates the Δ OM value that maximizes the correct decision. When ΔOM = 0 (all stimuli pairs are considered as perceptually different by the objective metric), the correct decision percentage is 82%, which agrees with the results of the subjective test, where the difference was statistically significant in 82% of the pairs (*cf.* Fig. 8).

To evaluate the impact, on the metric performance, of considering or not the background, the background distortion was computed using the selected metric (OAD^{TO} with PF_6), and included in the metric as an additional measure; after, the classification errors and correct decision values with/without considering the background distortion were compared. Table 6 presents the resulting classification errors and correct decision values, showing that the metric performance decreases when the background distortion is included. This is consistent with fact that the stretching in the background is not as visible as the stretching of foreground objects, and shows the advantage of having an object-based stretching metric. Taking into account the evaluation results of the different metrics, the object-based Tissot area distortion (OAD^{TO}), with pooling function PF_6, is the proposed one to assess the subjective impact of the viewport stretching distortion, in 360° image rendering.

VI. CONTENT-AWARE PANNINI PROJECTION

This section describes a useful application of the proposed stretching distortion metric; it consists in obtaining the optimal - in a perceived quality sense - projection parameters, (d, vc), for the viewport rendering of a 360° image using the general Pannini projection, resulting in a content aware Pannini projection (CAP). As mentioned in Section I, stretching of objects and bending of straight lines are the two main

TABLE 6. Classification errors and correct decision values, in percentage (%), for the best proposed metric, OAD^{TO} with PF_6, considering, or not, the background distortion.

	Not Considering the background distortion	Considering the background distortion
Correct Decision	88.9	71.1
False Tie	0.0	2.2
False Differentiation	11.1	15.6
False Ranking	0.0	11.1

artifacts that condition the perceived geometric distortion of the rendered viewports; furthermore, they have an opposite evolution with the variation of the projection parameters, i.e., stretching decreases and bending increases when d varies from 0 to 1, and/or vc varies from 1 to 0. Thus, the procedure to find the optimal parameters, (d_{opt}, vc_{opt}), seeks the best compromise between these two types of artifacts.

A. METHODOLOGY

For a given input ERI image, viewing direction (ϕ_{VD}, θ_{VD}), and FoV, the resulting viewport stretching and bending metrics are iteratively computed for different combinations of d and vc values, varying d between d_{min} and d_{max}, with a step-size Δd, and vc between vc_{min} and vc_{max}, with a step-size Δvc. For the results presented in this paper, $d_{min} = 0.1, d_{max} = 1, vc_{min} = 0$ and $vc_{max} = 1$; Δd and Δvc were both set to 0.1, resulting in $N = 110$ possible (d, vc) pairs. The optimal parameters, (d_{opt}, vc_{opt}), to be used on the viewport rendering, are obtained by minimizing, over the considered (d, vc) pairs, a cost function, described by (45):

$$
\begin{aligned}
&\left(d_{opt}, vc_{opt}\right) \\
&= \min_{(d,vc)} \left(\alpha \frac{SM(d, vc) - SM_{min}}{SM_{max} - SM_{min}} + \frac{BM(d, vc) - BM_{min}}{BM_{max} - BM_{min}} \right),
\end{aligned}
\tag{45}
$$

where $SM(d, vc)$ and $BM(d, vc)$ are, respectively, the viewport stretching and bending measures for projection parameters (d, vc); α is the stretching to bending ratio; $SM_{min}, SM_{max}, BM_{min},$ and BM_{max} are normalizing constants guaranteeing that the metric values are on the interval [0,1]; they correspond to minimum and maximum SM and BM values that were found for a set of 2200 viewports, rendered from 20 omnidirectional images taken from [8], [17], and using (d, vc) values on the intervals previously specified. For $SM(d, vc)$, the object-based approach showing the best results - OAD^{TO} with PF_6 - was used, while for $BM(d, vc)$ the Line Measure Combination, LMC, proposed in [11], was selected due to its good performance.

In (45), parameter α seeks the best balance between stretching and bending subjective impact and it was learned in a perceptual way using a small data set of Pannini viewports, not contained in the final evaluation dataset. The details about the procedures to obtain the normalization constants and α are provided in the supplementary material.

FIGURE 18. Example of viewports rendered from a different ERI with different projections using horizontal FoV of 150°. The proposed CAP produces viewports more pleasant than the other projections.

B. RESULTS

The proposed content aware Pannini (CAP) projection was compared with several benchmark projections that include rectilinear, stereographic, two Pannini with fixed parameters ($d = 0.5$, $vc = 0$), and ($d = 1$, $vc = 0$), and the optimized Pannini projection (OP) proposed in [8]. For comparison purposes, several viewports were rendered from a set of omnidirectional images available in the datasets of [8], [17]. The viewports were rendered with a horizontal FoV, F_h, of 150° and a spatial resolution of 960×540 pixels (16:9 aspect ratio). Fig. 18 depicts some viewports examples obtained with the proposed CAP and benchmark projections. As can be figured out, the CAP viewports are generally more pleasant than those resulting from the benchmark projections, providing a good compromise between bending and stretching distortions. In particular, the following qualitative comparisons can be made (additional comparison results are provided the supplementary material):

- CAP *vs* rectilinear and stereographic: The viewports resulting from CAP are clearly more pleasant than those resulting from rectilinear and stereographic projections. While the lines are straight in the rectilinear viewports, the perspective effect is very strong and annoying, and the object shapes are too much stretched, notably for the *Office1*, *Office2*, and *Buildings* viewports. Although the objects shape is preserved in the stereographic viewports, the lines are severely bent (fisheye effect).
- CAP *vs* Pannini with fixed parameters: The proposed CAP generates viewports with a good balance between the stretching of objects and bending of lines. This cannot be achieved for Pannini with fixed parameters, as for $vc = 0$ the horizontal lines are rather bent, particularly for $d = 1$.
- CAP *vs* OP: The viewports obtained for CAP have less geometric distortion than the viewports resulting from OP. In particular, for the *Bedroom* viewport the

FIGURE 19. Example of viewports rendered from a different ERI with proposed CAP and CA-GPP using a square FoV of 110°. The proposed CAP produces viewports with less geometric deformations, lines are straighter and the object shapes are less deformed.

horizontal lines on the celling and on the floor are straighter for CAP. In the *Office1* and *Office2* viewports, CAP kept the horizontal lines as straight as OP, but the objects shape (e.g., the chair and monitor on the left side in *Office1* and the chair on the left side in *Office2*) is more conformal for CAP.

The proposed CAP projection was also compared with the content-aware generalized perspective projection (CA-GPP), proposed in [11]. In the CA-GPP, the projection parameter (d) is optimized based on Support Vector Regression (SVR). The SVR model was trained and tested with a viewport dataset rendered with a square FoV of 90° and 110°. For fair comparisons, several viewports were rendered using CAP with a square FoV of 110°, and a spatial resolution of 856 × 856 pixels, as in [11]. Fig. 19 depicts some viewport examples obtained with the proposed CAP and with the CA-GPP. As can be observed, while for *Buildings* the results are similar, for *Friends* and *Dinner* the CAP provides a better tradeoff between stretching and bending; with CA-GPP, for the *Friends* viewport the vertical lines are too much bent and the same happens for the horizontal lines of *Dinner*. Note that while the CA-GPP was optimized for the FoV considered in these results (110°), there was not such optimization for CAP.

VII. CONCLUSION

This paper proposes an object-based quality metric to assess the perceived geometric distortion in viewport images, rendered from 360° images, in planar displays. The metric measures the distortion of individual objects, having semantic meaning (such as people, cars, furniture). The experimental results show that the proposed metric outperforms the

considered benchmark metrics, and is able to assess the viewport quality, with respect to perceptual scores, with a correct decision percentage close to 90%. Also, as a useful application, the proposed metric was integrated in a procedure that optimizes the Pannini projection parameters, according to the viewport content, achieving a good compromise between stretching of objects and bending of straight lines.

As final remarks, it is important to note that for same omnidirectional image, and depending on the viewing direction, the viewport content may vary a lot, and the same will happen to the best projection parameters; however, how to change these parameters during video navigation, requires further investigation, with specific subjective tests where user interaction is allowed. Furthermore, since the proposed content aware Pannini projection is globally adapted to the viewport content (i.e., d and vc have the same values for the whole viewport), stretching and/or bending may be still visible for some image regions and structures; if those parameters are allowed to vary locally, the visibility of geometric distortions could be further reduced. These challenges will be subject of study in future work.

APPENDIX

Tissot Indicatrix for the Pannini projection

The Tissot indicatrices, of a given sphere to plan projection, are characterized by the scaling parameters h, k and by the semi-major, $\hat{a}$, and semi-minor, $\hat{b}$, ellipses axis. To compute these parameters for the Pannini projection (PP), and following the procedure described in the appendix of [11], the partial derivatives of (x_p, y_p) with respect to (ϕ, θ) need to be
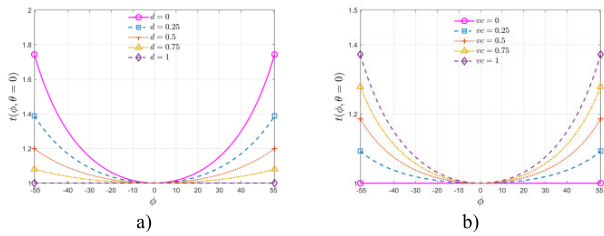
FIGURE 20. The plot of t along the equatorial line ($\theta = 0$) under the PP for: a) varying d and $vc = 0$; b) varying vc and $d = 1$.

computed. From (1) and (2), it follows:

$$\frac{\partial x_p}{\partial \phi} = \frac{(d+1)(d\cos(\phi)+1)}{(d+\cos(\phi))^2}, \tag{46}$$

$$\frac{\partial x_p}{\partial \theta} = 0, \tag{47}$$

$$\frac{\partial y_p}{\partial \phi} = \frac{(1-vc)(d+1)\tan(\theta)\sin(\phi)}{(d+\cos(\phi))^2}$$
$$+\frac{vc\tan(\theta)\sin(\phi)}{\cos^2(\phi)}, \tag{48}$$

$$\frac{\partial y_p}{\partial \theta} = \frac{1}{\cos^2(\theta)}\left[(1-vc)\frac{d+1}{d+\cos(\phi)}+\frac{vc}{\cos(\phi)}\right]. \tag{49}$$

The local area distortion, s, and the local shape distortion, t, can be computed by (39) and (40), respectively. A conformal projection has $t = 1$, and an equal-area projection has $s = 0$. As an example, Fig. 20 depicts the local shape distortion, t, along the equatorial line $\theta(= 0)$ and $\phi \in [-55°, 55°]$, for PP projection with varying parameters d and vc, one at a time. As can be seen in Fig. 20-a), local shape distortion is maximum for the rectilinear projection ($d = 0$). On the other hand, the stereographic PP ($d = 1$, $vc = 0$), is locally conformal ($t = 1$), although horizontal lines are bent, as can be seen in Fig. 18. In the PP, the bending of horizontal lines can be corrected by applying vc, however shape distortion is introduced, as can be concluded from Fig. 20-b).

REFERENCES

[1] M. Domanski, O. Stankiewicz, K. Wegner, and T. Grajek, "Immersive visual media—MPEG-I: 360 video, virtual navigation and beyond," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Poznan, Poland, May 2017, pp. 1–9.

[2] M. Wien, J. M. Boyce, T. Stockhammer, and W.-H. Peng, "Standardization status of immersive video coding," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 5–17, Mar. 2019.

[3] K. Sim, J. Yang, W. Lu, and X. Gao, "MaD-DLS: Mean and deviation of deep and local similarity for image quality assessment," *IEEE Trans. Multimedia*, vol. 23, pp. 4037–4048, 2021.

[4] J. Yang, Y. Zhao, J. Liu, B. Jiang, Q. Meng, W. Lu, and X. Gao, "No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions," *IEEE Trans. Cybern.*, early access, Oct. 15, 2020, doi: 10.1109/TCYB.2020.3024627.

[5] F. Jabar, J. Ascenso, and M. P. Queluz, "Field-of-view effect on the perceived quality of omnidirectional images," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops (ICMEW)*, Jul. 2020, pp. 1–6.

[6] P. Hanhart, Y. He, and Y. Ye, *Viewport-Based Subjective Evaluation of 360-Degree Video Coding*, document ISO/IEC JVET E0071, Geneva, Switzerland, Jan. 2017.

[7] A. Abbas, *GoPro VR Player: A Tool for VR Content Playback*, document ISO/IEC JVET D0177, Chengdu, China, Oct. 2016.

[8] Y. Won Kim, D.-Y. Jo, C.-R. Lee, H.-J. Choi, Y. Hoon Kwon, and K.-J. Yoon, "Automatic content-aware projection for 360° videos," 2017, *arXiv:1704.07528*.

[9] F. Jabar, J. Ascenso, and M. P. Queluz, "Perceptual analysis of perspective projection for viewport rendering in 360° images," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2017, pp. 53–60.

[10] A. Tissot, *Memoire Sur la Representation des Surfaces Projections Cartes Geographiques*. Paris, France: Gauthier Villars, 1881.

[11] F. Jabar, J. Ascenso, and M. P. Queluz, "Objective assessment of perceived geometric distortions in viewport rendering of 360° images," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 49–63, Jan. 2020.

[12] F. Jabar. (Nov. 2021). *IST-Pannini-Dataset*. [Online]. Available: https://ieee-dataport.org/documents/ist-panninidataset

[13] R. Carroll, M. Agrawala, and A. Agarwala, "Optimizing content-preserving projections for wide-angle images," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–43, Aug. 2009.

[14] Wikipedia. *Cauchy-Riemann Equations*. Accessed: Sep. 24, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Cauchy-Riemann _equations

[15] Y. Shih, W.-S. Lai, and C.-K. Liang, "Distortion-free wide-angle portraits on camera phones," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–12, Aug. 2019.

[16] T. K. Sharpless, B. Postle, and D. M. German, "Pannini: A new projection for renderingwide angle perspective images," in *Proc. 6th Int. Conf. Comput. Aesthetics Graph., Vis. Imag.*, London, U.K., Jun. 2010, pp. 1–16.

[17] J. Gutierrez, E. J. David, A. Coutrot, M. P. Da Silva, and P. L. Callet, "Introducing un salient360! benchmark: A platform for evaluating visual attention models for 360° contents," in *Proc. 10th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, Sardinia, Italy, May 2018, pp. 1–3.

[18] R. Monroy, S. Lutz, T. Chalasani, and A. Smolic, "Salnet360: Saliency maps for Omni-directional images with CNN," *Signal Process., Image Commun.*, vol. 69, pp. 26–34, Nov. 2018.

[19] J. Boyce, E. Alshina, A. Abbas, and Y. Ye, *JVET Common Test Conditions and Evaluation Procedures for 360° Video*, document ISO/IEC JVET H1030, Macao, China, Oct. 2017.

[20] P. Hanhart, M. Rerabek, and T. Ebrahimi, "Towards high dynamic range extensions of HEVC: Subjective evaluation of potential coding technologies," in *Proc. SPIE Opt. Eng. Appl.*, San Diego, CA, USA, Sep. 2015, pp. 1–4.

[21] P. Hanhart, L. Krasula, P. Le Callet, and T. Ebrahimi, "How to benchmark objective quality metrics from paired comparison data," in *Proc. 8th Int. Conf. Qual. Multimedia Exper. (QoMEX)*, Lisbon, Portugal, Jun. 2016, pp. 1–6.

[22] P. Hanhart, P. Korshunov, and T. Ebrahimi, "Crowdsourcing evaluation of high dynamic range image compression," in *Proc. SPIE Opt. Eng. Appl.*, San Diego, CA, USA, Sep. 2014, pp. 1–12.

[23] Z. Zhang, J. Zhau, N. Liu, X. Gu, and Y. Zhang, "An improved pairwise comparison scaling method for subjective image quality assessment," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Cagliari, Italy, Jun. 2017, pp. 1–6.

[24] Wikipedia. *Binomial Distribution*. Accessed: Jun. 17, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Binomial_distribution

[25] Y. Xu, K. Wang, K. Yang, D. Sun, and J. Fu, "Semantic segmentation of panoramic images using a synthetic dataset," in *SPIE*, vol. 11169, Strasbourg, France, Sep. 2019, pp. 1–15.

[26] K. Yang, X. Hu, Y. Fang, K. Wang, and R. Stiefelhagen, "Omnisupervised Omnidirectional Semantic Segmentation," *IEEE Trans. Intell. Transp. Syst.*, early access, Sep. 23, 2020, doi: 10.1109/TITS.2020.3023331.

[27] K. Yang, J. Zhang, S. Reiß, X. Hu, and R. Stiefelhagen, "Capturing Omni-range context for omnidirectional segmentation," 2021, *arXiv:2103.05687*.

[28] K. Yang, X. Hu, and R. Stiefelhagen, "Is context-aware CNN ready for the surroundings panoramic semantic segmentation in the wild," *IEEE Trans. IMAGE Process.*, vol. 30, pp. 1866–1881, 2021.

[29] C. Liu, L.-C. Chen, F. Schroff, H. Adam, W. Hua, A. L. Yuille, and L. Fei-Fei, "Auto-DeepLab: Hierarchical neural architecture search for semantic image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 82–92.

[30] O. Russakovsky, J. Deng, H. Su, J. Kraus, S. Satheesh, and S. Ma, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[31] T. Lin, M. Maire, S. Belongie, J. Hays, and P. Perona, "Microsoft COCO: Common Objects in Context," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, Sep. 2014, pp. 740–755.

[32] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, pp. 98–136, Jun. 2014.

[33] Mathworks. *Bwconncomp*. Accessed: Jun. 23, 2021. [Online]. Available: https://www.mathworks.com/help/images/ref/bwconncomp.html

[34] M. A. Wirth, "Shape analysis & measurement," University of Guelph, Guelph, ON, Canada, Tech. Rep. 6320, 2002. [Online]. Available: http://www.cyto.purdue.edu/cdroms/micro2/content/education/wirth0.pdf

[35] L. da F. Costa, *Shape analysis and Classification: Theory and Practice*. Boca Raton, FL, USA: CRC Press, 2000.

[36] M. Basaraner and S. Cetinkaya, "New measures for analysis and comparison of shape distortion in world map projections," *Cartography Geographic Inf. Sci.*, vol. 46, no. 6, pp. 518–531, Nov. 2019.

[37] *Method for Specifying Accuracy and Cross-Calibration of Video Quality Metrics*, document ITU-T Recommendation J.149, Mar. 2004.

JOÃO ASCENSO (Member, IEEE) received the E.E., M.Sc., and Ph.D. degrees from the Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisbon, Portugal, in 1999, 2003, and 2010, respectively, all in electrical and computer engineering. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, and a member of the Multimedia Signal Processing Group, Instituto de Telecomunicações (IT), Lisbon. He has authored over 100 papers in international conference and journals and has more than 3200 citations over 35 papers. His current research interests include visual coding, quality assessment, light-fields, point clouds and holography processing, indexing and searching of audio–visual content, and visual sensor networks. He is also an Elected Member of the IEEE Multimedia Signal Processing Technical Committee and is very active in ISO/IEC JPEG and MPEG standardization activities. He acts as a member of the Organizing Committees of well-known international conferences, such as IEEE ICME 2020, IEEE ISM 2018, QoMEX 2016, and among others. He is also an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA and IEEE TRANSACTIONS ON IMAGE PROCESSING. He was an Associate Editor of IEEE SIGNAL PROCESSING LETTERS.

FALAH JABAR received the B.S. degree in computer engineering from Koya University, Koysinjaq, Kurdistan, Iraq, and the M.S. degree in electrical, electronics and telecommunications from University Technology Malaysia, Johor, Malaysia. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the Universidade de Lisboa, Lisboa, Portugal. In 2016, he joined the Instituto Superior Técnico (IST), Universidade de Lisboa. He is a member of the Multimedia Signal Processing Group, Instituto de Telecomunicações (IT). His current research interests include subjective and objective quality assessment for omnidirectional visual content.

MARIA PAULA QUELUZ received the B.S. and M.S. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), University of Lisbon, Portugal, and the Ph.D. degree from the Catholic University of Louvain, Louvain-la-Neuve, Belgium. She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, IST, and a Senior Research Member of the Instituto de Telecomunicações, Lisbon, Portugal. Her research interests include image/video quality assessment, image/video processing, and wireless communications.

● ● ●