Pose Estimation for Sensors Which Capture Cylindric Panoramas

Fay Huang[1], Reinhard Klette[2], and Yun-Hao Xie[1]

[1] Institute of Computer Science and Information Engineering,
National Ilan University, Taiwan, R.O.C.
[2] Department of Computer Science,
The University of Auckland, New Zealand

Abstract. This paper shows that there exist linear models for sensor pose estimation for multi-view panoramas defined by a symmetric or leveled pair of cylindric images. It assumes that pairs of corresponding points have been detected already in those pairs of panoramas. For the first time a cost function is formulated whose minimization solves the pose estimation problem for these two general cases of multi-view panoramas, specified by unconstrained sensor parameter values but only minor constraints on sensor poses. (Note that due to the non-linearity of the panoramic projection geometry, the modeling of sensor pose estimation typically results into non-linear forms which incur numerical instability.)

1 Review and Basic Notions

A panoramic image is recorded by a panoramic sensor, such as a rotating camera. Sensor pose estimation deals with recovering the relative pose of two (calibrated) sensors. Compared to planar images or catadioptric images, there is very few literature on sensor pose estimation from cylindric panoramas.

Ishikuro el at. [1] dealt with a very restricted case of the sensor pose estimation problem, in which the given panoramas are acquired at the same altitude and with parallel rotation axes. Kang and Szeliski [2] discussed the sensor pose estimation problem only for single-center panoramas. Neither generalized to the multi-view case (i.e., different intrinsic sensor parameters and arbitrary sensor poses) nor practically relevant cases (e.g., multi-view panoramas as in [3, 4]) of sensor pose estimation have been studied or discussed in the literatures before. This paper provides (for the first time) a cost function whose minimization solves the pose estimation problem for two general cases of cylindric panoramas.

A 360° cylindric panoramic image can be acquired by various means, such as a rotating video or matrix-sensor camera, a catadioptric sensor (with a subsequent mapping onto a cylinder), or a rotating sensor-line camera, as commercially available from various producers since the late 1990s. For simplifying our discussion, we assume a sensor model close to the latter one which has a fixed rotation axis and takes images consecutively at equidistant angles. (Rotating

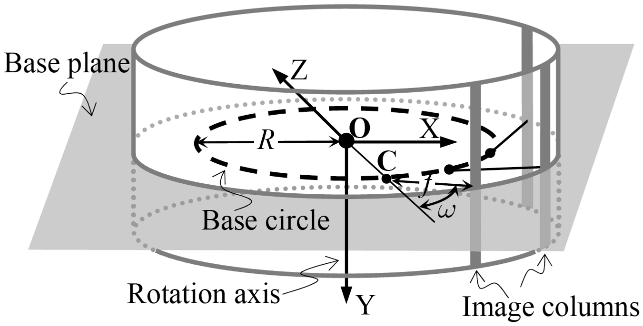

Fig. 1. Sensor model for cylindric panoramas, showing three image columns with their projection centers.

sensor-line cameras allow maximum accuracy, and have been used, e.g., in major architectural photogrammetric projects; see [5]). The projection center of the camera does not have to be on the rotation axis. In the case of an off-axis distance $R > 0$, the resulting panoramic images are refereed to as *multi-projection center panoramas.*

A *multi-view panorama* [6] is a set of multi- or single-projection center cylindric panoramas which were recorded at different locations or with different capturing parameters. In particular, they might be acquired with respect to different rotation axes. In comparison to a single axial panorama [7, 8], the advantages of multi-view panoramas are known to include enlarged visibility and improved stereo reconstruction opportunities; in short, they define multi-view image analysis for the cylindric panoramic case.

2 Cylindric Panoramas - General Case

We generalize from various panoramic imaging models [1, 7, 9]. The model consists of multiple projection centers and a cylindric image surface; see Fig. 1. $\mathbf{C}$ denotes a projection center. Projection centers are uniformly distributed on the *base circle* which is in the *base plane* and has center $\mathbf{O}$, which is also the origin of the sensor coordinate system. The *off-axis distance* R (radius of the base circle) describes the distance between any projection center and the rotation axis.

A cylindric panorama is partitioned into *image columns* of equal width which are parallel to the rotation axis. W is the number of image columns. There is a one-to-one ordered mapping between image columns and projection centers. The distance between a projection center and its associated image column is the *effective focal length* f. The *principal angle* ω is between a projection ray in the base plane, emitting from $\mathbf{C}$, and the normal vector of the base circle at point $\mathbf{C}$. R, f, ω, and W are the four intrinsic sensor parameters, characterizing a panoramic image $E_{\mathcal{P}}(R,\, f,\, \omega,\, W)$. The affine transform between two sensor coordinate systems is described by a 3×3 rotation matrix $\mathbf{R} = [\mathbf{r}_1^{\mathrm{T}}\mathbf{r}_2^{\mathrm{T}}\mathbf{r}_3^{\mathrm{T}}]^{\mathrm{T}}$ and a 3×1 translation vector $\mathbf{T} = (t_x, t_y, t_z)^{\mathrm{T}}$.

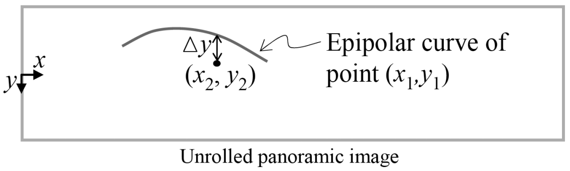

Fig. 2. Row difference $\triangle y$ between the actual corresponding image point (x_2, y_2) and the point where epipolar curve and column x_2 intersect.

Let (x_1, y_1) and (x_2, y_2) denote the image coordinates of the projection of a 3D point in two panoramas $E_{\mathcal{P}_1}(R_1,\ f_1,\ \omega_1,\ W_1)$ and $E_{\mathcal{P}_2}(R_2,\ f_2,\ \omega_2,\ W_2)$, respectively. If multiple pairs of corresponding image points are provided, say (x_{1i}, y_{1i}) and (x_{2i}, y_{2i}), for $i = 1, 2, \ldots, n$, then we are able to estimate sensor poses by minimizing the following cost function,

$$min\sum_{i=1}^{n}\left(y_{2i}-\frac{f_2\mathbf{r}_2^{\mathrm{T}}\cdot\mathbf{V}}{\sin\delta_{2i}\mathbf{r}_1^{\mathrm{T}}\cdot\mathbf{V}+\cos\delta_{2i}\mathbf{r}_3^{\mathrm{T}}\cdot\mathbf{V}-R_2\cos\omega_2}\right)^2 \qquad (1)$$

where $\alpha_{ki} = \frac{2\pi x_{ki}}{W_k}$, $\delta_{ki} = (\alpha_{ki}+\omega_k)$, $\beta_{ki} = \tan^{-1}(\frac{y_{ki}}{f_k})$, and $k = 1$ or 2. Moreover,

$$\mathbf{V} = \mathbf{A} + \frac{R_2\sin\omega_2 + \cos\delta_{2i}\mathbf{r}_1^{\mathrm{T}}\cdot\mathbf{A} - \sin\delta_{2i}\mathbf{r}_3^{\mathrm{T}}\cdot\mathbf{A}}{\sin\delta_{2i}\mathbf{r}_3^{\mathrm{T}}\cdot\mathbf{B} - \cos\delta_{2i}\mathbf{r}_1^{\mathrm{T}}\cdot\mathbf{B}}\mathbf{B} \qquad (2)$$

$$\mathbf{A} = \begin{pmatrix} R_1\sin\alpha_{1i} - t_x \\ -t_y \\ R_1\cos\alpha_{1i} - t_z \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} \sin\delta_{1i}\cos\beta_{1i} \\ \sin\beta_{1i} \\ \cos\delta_{1i}\cos\beta_{1i} \end{pmatrix}$$

The cost function is defined to be the image row difference $\triangle y$, see Fig. 2. Epipolar curves are calculated based on point coordinates x_{1i}, y_{1i} and sensor parameters; see [6]. The estimation of sensor poses appears to be rather difficult, if not impossible for the unrestricted case.

3 Two Standard Multi-view Cases

However, when using panoramic sensors, such as rotating sensor-line systems, then it is actually standard to aim for a set of leveled panoramas, and for symmetric panoramas if stereo-viewing is also intended; see [5].

Two Symmetric Pairs. $E_{\mathcal{P}_1}(R, f,\ \omega,\ W)$ and $E_{\mathcal{P}_2}(R, f,\ \text{-}\omega,\ W)$ define a symmetric pair of panoramas, both defined for the same sensor coordinate system. Epipolar curves are in this case lines which may be identified with image rows (see proofs in [4, 6, 10]). Dense image correspondences can be calculated by using stereo matching algorithms as developed for stereo pairs of planar images. If 3D data are collected from multiple pairs of symmetric panoramas, acquired at different locations, then data fusion becomes a challenge, and the registration step requires that the sensor pose estimation problem has been solved. In our sensor pose estimation approach, first, for each symmetric pair, transform pairs of corresponding image points into directional unit vectors pointing to the reconstructed 3D points; second, establish a geometric relation between these two bunches of unit vectors that are respectively defined in two sensor coordinate systems.

The theorem below shows how the directional unit vector of a 3D point (with respect to **O**) is derived from a pair of corresponding points (x_1, y) and (x_2, y) on symmetric panoramas $E_{\mathcal{P}_1}(R,\, f,\, \omega,\, W)$ and $E_{\mathcal{P}_2}(R,\, f,\, \text{-}\omega,\, W)$. Let **u** be the directional unit vector of that 3D point which projects into (x_1, y) and (x_2, y).

Theorem 1. *Let* $\alpha = \frac{(x_1+x_2)\pi}{W}$ *and* $\beta = \frac{(x_2-x_1)\pi}{W}$. *We have that*

$$\boldsymbol{u} = \frac{\left(\sin\omega \sin\alpha,\, \frac{y}{f}\sin\beta, \sin\omega\cos\alpha\right)^T}{\sqrt{\sin^2\omega + \frac{y^2}{f^2}\sin^2\beta}} \tag{3}$$

This applies to approaches in [6, 3] where angle ω can take any value. There is no dependence of the off-axis distance R in those two formulas. – For sensor pose estimation from two symmetric pairs (see Fig. 3), we may apply reasoning and results as in [11].

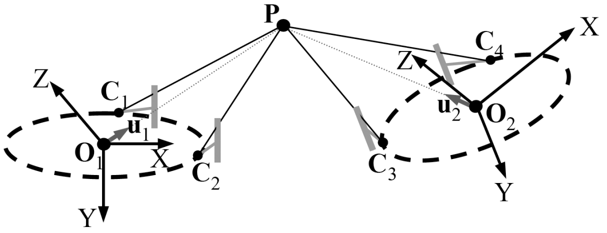

Fig. 3. Four corresponding points in two pairs of symmetric panoramas and its preimage **P** in 3D space. Vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are two corresponding directional unit vectors of **P** with respect to the sensor coordinate systems $\mathbf{O}_1$ and $\mathbf{O}_2$, respectively.

Theorem 2. *Given are at least eight pairs of corresponding points in two pairs of symmetric panoramas, where the associated sensor parameters are known except for R. The relative sensor poses can then be recovered by the normalized 8-point algorithm up to a scale factor.*

If all the sensor parameter values (including R) are pre-calibrated and given, then the exact sensor pose associated to $\mathbf{O}_2$ can be recovered with respect to the world coordinate system at $\mathbf{O}_1$. Then, these distances can be used as a reference to recover the unknown scale factor in Theorem 2.

One Leveled Pair. In this second common approach for capturing multi-view panoramas, the only constraint is that all associated base planes have to be parallel (say, to the sea level), to be guaranteed by a lever. See Figure 4.

Leveled panoramas allow large “overlapping” fields of views. The larger the common field of view, the higher the probability that object surfaces are visible in more than one panorama. This supports reliable stereo reconstruction and smooth view-transitions between multiple panoramas.

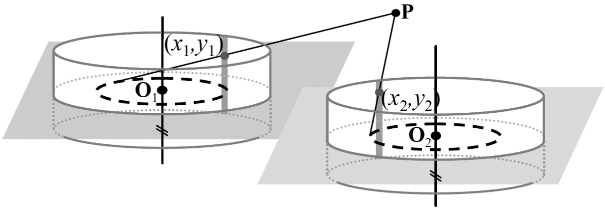

Fig. 4. A pair of leveled panoramas and a pair of corresponding image points.

The sensor pose estimation criteria of a leveled pair are specified in the following theorem. Both leveled panoramas are acquired by two sensors with identical intrinsic parameters, and the sensor poses are related by a single rotation angle ϕ with respect to the rotation axis and a translation vector $(t_x, t_y, t_z)^{\mathrm{T}}$. Let $X_1 = \cos\phi$, $X_2 = \sin\phi$, $X_3 = t_x$, $X_4 = t_z$, $X_5 = t_y$ and

$$\begin{array}{rcll}
c_{1i} & = & y_{2i} R \sin(\delta_{1i} - \alpha_{2i}) + y_{1i} R \sin(\delta_{2i} - \alpha_{1i}) & \\
c_{2i} & = & y_{1i} R \cos(\delta_{2i} - \alpha_{1i}) - y_{2i} R \cos(\delta_{1i} - \alpha_{2i}) & \\
c_{3i} & = & -y_{2i} \cos\delta_{1i} & c_{4i} = y_{2i} \sin\delta_{1i} \\
c_{5i} & = & y_{1i} \cos\delta_{2i} & c_{6i} = -y_{1i} \sin\delta_{2i} \\
c_{7i} & = & f \sin(\alpha_{2i} - \alpha_{1i}) & c_{8i} = f \cos(\alpha_{2i} - \alpha_{1i}) \\
c_{9i} & = & -(y_{1i} + y_{2i}) R \sin\omega &
\end{array}$$

where $\alpha_{ki} = \frac{2\pi x_{ki}}{W}$, $\delta_{ki} = (\alpha_{ki} + \omega)$, and $k = 1$ or 2.

Theorem 3. *Given a set of corresponding pairs of points (x_{1i}, y_{1i}) and (x_{2i}, y_{2i}), where $i = 1, 2, \ldots, n$, the values of ϕ, t_x, t_y, and t_z can be estimated by minimizing the following sum,*

$$\sum_{i=1}^{n} (c_{1i}X_1 + c_{2i}X_2 + c_{3i}X_3 + c_{4i}X_4 + c_{5i}X_1X_3 + c_{6i}X_1X_4$$
$$+ c_{7i}X_1X_5 + c_{6i}X_2X_3 - c_{5i}X_2X_4 + c_{8i}X_2X_5 + c_{9i})^2$$

subjected to the constraints $X_1^2 + X_2^2 = 1$, $X_1^2 \leq 1$, and $X_2^2 \leq 1$.

4 Experiments

We carried out real-world experiments on estimating sensor poses at different indoor or outdoor locations, using different cameras. Camera and sensor were calibrated separately in advance; camera's intrinsic parameters are thus known and kept unaltered during image acquisition. Figure 5 just illustrates one example of a leveled pair. In our experiments, the estimated sensor pose was denoted as $\hat{\mathbf{R}}$ and $\hat{\mathbf{T}}$. The error measurement for rotation was defined as $\arccos\left(\left(tr(\mathbf{R}\hat{\mathbf{R}}^{\mathrm{T}}) - 1\right)/2\right)$ and the error measurement for translation was defined as $\arccos\left(\mathbf{T} \cdot \hat{\mathbf{T}} / \|\mathbf{T}\| \|\hat{\mathbf{T}}\|\right)$, i.e., the angle between $\mathbf{T}$ and $\hat{\mathbf{T}}$, both in degrees.

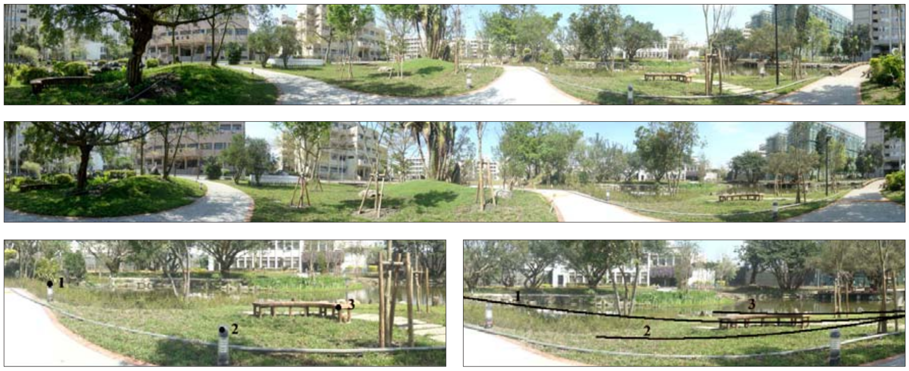

Fig. 5. A symmetric pair of panoramas, with enlarged areas also showing used points.

We used the SVD method for estimating $\hat{\mathbf{R}}$ and $\hat{\mathbf{T}}$ when symmetric pairs were used. When only a leveled pair was used, the sequential quadratic programming method was used for optimization. We achieved in average 1.24° error in the rotation estimation and 4.65° error in the translation. Figure 6 and Fig. 5 bottom show three particular epipolar curves calculated based on erroneous estimations from the leveled case. The mean y-difference between identified corresponding points and the calculated epipolar curves is 1.2 pixel. Less than three pixel error in vertical direction was typical.

We also conducted an error sensitivity analysis with simulated image data, in analogy to the real-world experiment, for both estimation approaches. Figure 7 plots how errors in detecting corresponding points impact the estimation result. The horizontal axes show various error sizes up to ten pixel. In the analysis, for example, a five-pixel input error means that each pair of corresponding image points was corrupted by errors of max/min five pixel in both x- and y-values, and the errors are modeled as Gaussian-distributed random numbers.

In the case of symmetric panoramas, the curves of the estimation errors for rotation and translation show both a monotonic increase (measured for 500 runs). For up to ten-pixel input error, the estimation errors of rotation matrix or translation vector are less than one or three degrees, respectively. This analysis suggests that we had input errors of about six to eight pixel in our real-world

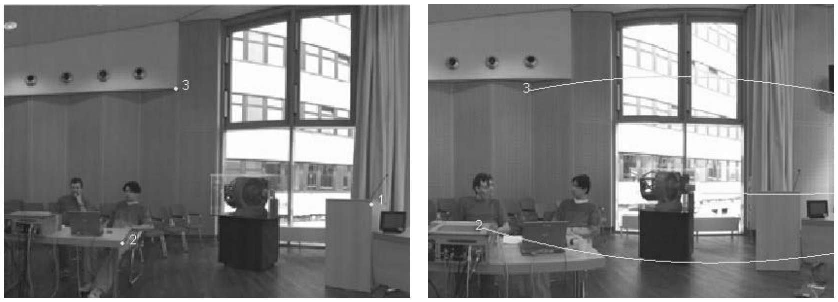

Fig. 6. Three epipolar curves calculated based on the pose estimation results.

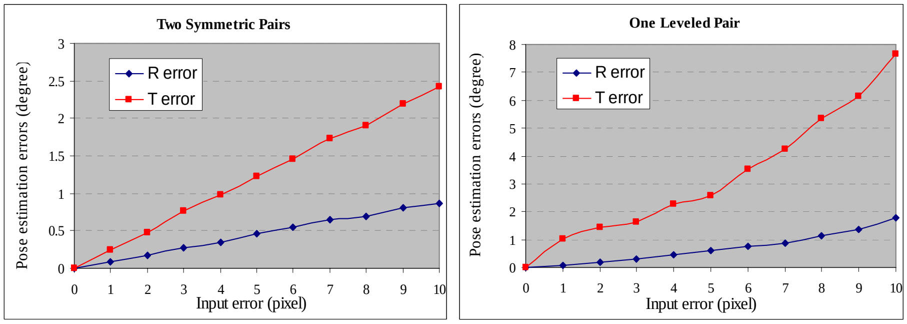

Fig. 7. Error sensitivity analysis for the symmetric or leveled case (synthetic images).

experiment. However, this conclusion did not match our expectations. Hence, a further error analysis was conducted to test how the sensor calibration errors of R and ω affect the pose estimation results. In the symmetric case, R is an independent variable; and if ω has a one-degree error, then it would produce a four-pixel error in the input data. Therefore, the accuracy of sensor calibration, especially for ω, is crucial to the pose estimation result.

For the case of leveled panoramas, the errors for both $\hat{\mathbf{R}}$ and $\hat{\mathbf{T}}$ are about two point five times the errors in the symmetric-panorama case. It suggests that the quadratic programming approach is more sensitive to input errors than the SVD method. Also, the assignment of initial values has significant impact onto the estimation result. According to our experiments, the estimation result was mostly sensitive to the 'sign' of the initial values but not to their quantities nor inter-ratios. In particular, zeros were not good for an initial guess in our case. The plots in this case indicate that we had input errors of about eight to nine pixel in our real-world experiment, which are close but slightly bigger than the conclusion drawn in the symmetric-panorama case. Error analysis on R and ω was carried out as well. It concludes that the error of R has a very minor impact on pose estimation results. Moreover, a k-degree error of ω would cause about a k-degree error in the estimated $\hat{\mathbf{T}}$, for any real number k, but an error in ω has very little impact on the estimation of $\mathbf{R}$. The conclusion drawn here is coherent to the symmetric-panorama case.

Finally, more synthetic experiments lead to conclusions that the resolution of the input panoramic images, and the distribution of the selected corresponding points are also two critical factors for pose estimation. The panoramic image resolution, especially the width, should be as large as possible. The corresponding points should be distributed uniformly and sparsely on the entire panoramic images. A larger set of corresponding points, say greater than 100, would not guarantee a better estimation result. A much better result can be achieved if image resolution of $1,000 \times 10,000$ is used instead, and the nearest scene point is no less than four meters from both sensors. The estimation errors can be less than 0.5 degrees for both $\mathbf{R}$ and $\mathbf{T}$, allowing for both cases even up to ten-pixel input error.

5 Conclusions

For the case of two symmetric pairs, we showed that the (common) normalized 8-point algorithm can be utilized in this case. Experimentally, we found that the normalization step of the normalized 8-point algorithm for improving the accuracy and satiability was ignorable in our case,[3] and this makes a difference to the planar image case. The results for the leveled-panorama case were greatly improved and reasonably stable. The proposed approaches are able to achieve high accuracy of less than 0.5 degree error in general, if high-resolution panoramic images are used and corresponding image points are carefully selected.

According to our error sensitivity analysis, the estimation of $\mathbf{T}$ is generally more sensitive to noise than the estimation of $\mathbf{R}$, and both estimation errors have approximately a linear relation to the input errors (as concluded from extensive simulations). We may also conclude that sensor pose estimation from leveled panoramas is more sensitive to errors than from pairs of symmetric panoramas. For future work it is of interest to develop an algorithm, or a framework, that takes care of sensor calibration and pose estimation at once, similar to self-calibration for the planar image case.

References

1. Ishiguro, H., Yamamoto, M., Tsuji, S.: Omni-directional stereo. PAMI **14** (1992) 257–262
2. Kang, S.B., Szeliski, R.: 3-d scene data recovery using omnidirectional multibaseline stereo. IJCV **25** (1997) 167–183
3. Peleg, S., Ben-Ezra, M.: Stereo panorama with a single camera. In: Proc. CVPR'99, Fort Collins, Colorado, USA (1999) 395–401
4. Shum, H.Y., He, L.W.: Rendering with concentric mosaics. In: Proc. SIGGRAPH'99, Los Angeles, California, USA (1999) 299–306
5. Huang, F., Klette, R., Scheibe, K.: Panoramic Imaging: Sensor-Line Cameras and Laser Range-Finders. Wiley, West Sussex, England (2008)
6. Huang, F., Wei, S.K., Klette, R.: Geometrical fundamentals of polycentric panoramas. In: Proc. ICCV'01, Vancouver, Canada (2001) I:560–565
7. Li, Y., Shum, H.Y., Tang, C.K., Szeliski, R.: Stereo reconstruction from multiperspective panoramas. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 45–62
8. Scheibe, K., Suppa, M., Hirschmäller, H., Strackenbrock, B., Huang, F., Liu, R., Hirzinger, G.: Multi-scale 3d-modeling. In: Proc. PSIVT'06, Hsinchu, Taiwan (2006) 96–107
9. Murray, D.: Recovering range using virtual multicamera stereo. CVIU **61** (1995) 285–291
10. Seitz, S.: The space of all stereo images. In: Proc. ICCV'01, Vancouver, Canada (2001) 26–33
11. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge Uni. Press, United Kingdom (2000)

[3] The reason is that the 3D geometry of corresponding image points on panoramas is likely not as skewed or clustered as in the planar image case, which is mainly attributed to the panoramic 3D representation itself.