applied optics

Panoramic stereo imaging system for efficient mosaicking: parallax analyses and system design

Junliang Yan,[1,*] Lingsheng Kong,[1] Zhihui Diao,[1] Xiaofeng Liu,[1,2] Lilu Zhu,[1] and Ping Jia[1]

[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Science, Changchun, Jilin 130033, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
*Corresponding author: jlyciomp@163.com

Received 17 October 2017; revised 6 December 2017; accepted 11 December 2017; posted 12 December 2017 (Doc. ID 309365); published 16 January 2018

Panoramic stereo images, captured by distributed devices then mosaicking, are competent contents for virtual reality applications. Mosaicking raw images with different perspectives into satisfying final results is still not efficient enough, even if state-of-the-art algorithms are employed. For improving this efficiency in optical methods, we delve into the potential of the capturing system. Two parallax factors, peak parallax and deviation of parallaxes, are proposed to assess the mosaicking capability. By controlling variables and numerical computation, rules between parallax factors and design parameters have been revealed. Validation by simulations, large capturing distance, more cameras, compact arrangement, and moderate overlaps are suggested as the general design strategy. Benefiting from efficient mosaicking, systems based on our design strategy would have potential for real-time applications.

OCIS codes: (120.4820) Optical systems; (220.4830) Systems design; (000.4430) Numerical approximation and analysis; (110.6880) Three-dimensional image acquisition.

https://doi.org/10.1364/AO.57.000396

1. INTRODUCTION

Benefiting from the captivating experiences and fascinating future application expectations, virtual reality (VR) [1] could claim to be one of the hottest topics in both the technology world and the commercial market. As an essential part that offers the core experience, visual contents matter as much as the head-mounted platforms [2–4]. Compared with the images (including videos) purely rendered by computer graphics, images captured from real scenes seem to be more practical. Images based on real scenes are more competent for some specific VR applications, such as navigation, live telecast, surveillance, and other non-entertainment ones [1,5]. A variety of methods [6], which could be cataloged into the image-based rendering (IBR) [7] technics, have been proposed for producing this kind of contents. Panoramic stereo (PNST) of real scenes is the most common and practical one [8–11].

PNST is a kind of image formation that contains a pair of panoramas and offers stereo of 360° horizontally. In literature, it might be named "omnistereo" [12], "omnidirectional stereo" [13], "stereoscopic panorama" [14], and so on. Even if subtle optical designs have been tried, a single camera is incompetent for capturing PNST. Employing multiple cameras, distributed capturing strategies are widely accepted [15,16]. However, distributed structure brings the problem of mosaicking.

Because of the absence of a common viewing point, mosaicking the raw images from individual cameras to a seamless one usually asks for image processing. For now, the mosaicking process affects the efficiency of the whole work [5,17] and becomes the barrier in promoting VR applications. In this paper, "efficient" and "efficiency" refer to not only the mosaicking speed but also the stereo visual quality of the final product. Certainly, algorithms [14,18] have been proposed to speed up and optimize the mosaicking, but few results have been accepted satisfactorily and widely. The long processing time relative to the capturing indicates that the state-of-the-art algorithms are not versatile for all situations.

If we want to deal with this problem by optical methods, what should we do to lose the burden of the mosaicking and to achieve satisfying PNSTs? By investigating the factors involved in the optical systems and raw images, we find that parallax is the key. For mosaicking, parallax affects not only the speed but also the quality [18].

Parallax indicates the apparent distinction of the same object being viewed from different positions and directions. In photography, it is defined as the included angle of the two principal capturing rays, which emit from the object to each camera. In this paper, we emphasize the importance of the parallax characteristics of the raw images and the optical systems, and then

use them as an assessment of the mosaicking capability. We investigated how design parameters affect the parallaxes and the mosaicking. After that, the capturing system design strategy for efficient PNST mosaicking has been proposed.

This paper starts with the capturing model and the workflow of producing PNST. Parallax factors have been proposed to assess the efficient mosaicking potential. Then, the parallaxes in PNST have been analyzed along with the design parameters by controlling variable and numerical computation. Graphic simulations have been done to verify the validity of the conclusions. Finally, strategies for designing PNST capturing systems have also been proposed.

Before delving into the parallax factors, we introduce and claim the capturing model, common capturing workflow, and some useful conventions.

2. CAPTURING MODEL, CONVENTIONS, AND WORKFLOW

A. Camera and Equivalent Camera

An ideal pinhole camera is employed as the model of a single camera unit. It has only four parameters: nodal point location C, shooting direction $\mathbf{v}$, half-field of view Ω, and focal length f (Fig. 1).

The imaging process is essentially perspective. We use a symbol to indicate the camera with four parameters (Fig. 1). In the context of PNST, two cameras with the same focal length and coincident nodal points could be regarded as equivalent cameras (Fig. 2). When constructing PNSTs, the projection of the images from equivalent cameras could be the same. The operations for these images are just cropping, which is modifying their field of view and direction.

B. Capturing Rig

Researchers have proposed several representative strategies to achieve panoramas and PNSTs. One can create a panorama by rotating a single camera around its nodal point horizontally, recording a series of vertical image strips and mosaicking. By selecting specific sections of these strips, PNST could also be achieved [12,19,20]. These schemes may be suitable for still scenes, but powerless for dynamic situations. Arranging cameras around a common center to form a rig, dynamic scenes could be recorded radially and simultaneously as videos. When changing individual cameras into pairwise ones, PNST could be achieved [10,21,22]. "Project Beyond" [4], "Surround 360" [17], and other capturing devices proposed recently are all based on distributed pairs, though they employ different mosaicking algorithms.

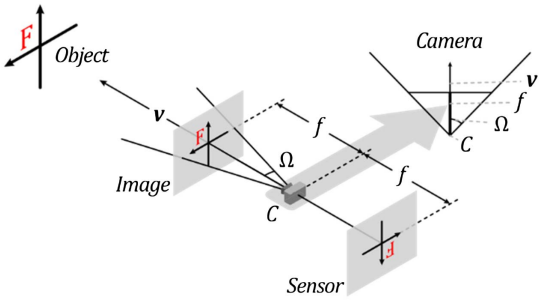

Fig. 1. Ideal camera model with four parameters.

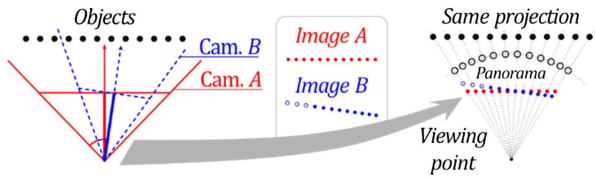

Fig. 2. Equivalent cameras. Cam. A (upright) and Cam. B (oblique) have the same f and C, although different $\mathbf{v}$ and Ω. They capture the same objects and apparently get different images. Essentially, Image A and Image B have the same projection, which means the same for the PNST. We regard Cam. A and Cam. B as equivalent cameras. In fact, Image B is a section of Image A when used in VR.

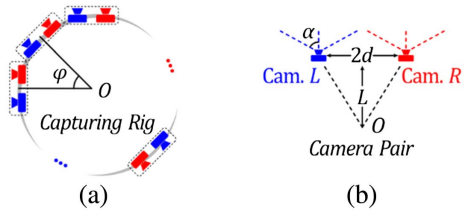

Fig. 3. Capturing rig and camera pair. (a) The overall scheme of a capturing rig (top view). (b) Parameters of a camera pair.

A capturing rig for PNST is an ensemble of several camera pairs with specific parameters and arrangement. A camera pair is the basic unit to form the central symmetrical and radial rig [Fig. 3(a)]. A camera pair is a relatively independent unit and consists of two parallel pairwise cameras. It is usually regarded as a stereo camera and provides two images (stereo) for a certain orientation.

In practicality, all cameras usually have the same specifications and distribute evenly. If n pairs are employed, there are indeed $2n$ cameras. All camera pairs fan out a whole circle to cover a 360° field of view. The interval angle between pairs is φ, then $n \times \varphi = 360°$. Looking inside a single pair [Fig. 3(b)], each camera could be denoted by "Camera Left (Cam. L)" or "Camera Right (Cam. R)" according to the stereo information it offers. Cam. L and Cam. R align with each other with a lateral separation $2d$. The horizontal half-field of view of each camera is α. When installing camera pairs in a rig, the distance from the rig center O to the pair center is L.

When installing the rig to capture a real scene, the related parameters could be described under a uniform coordinate system (Fig. 4). Rectangular coordinates employ the rig center O

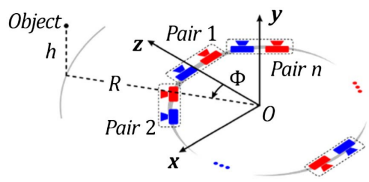

Fig. 4. Parameters related to the capturing rig and the object.

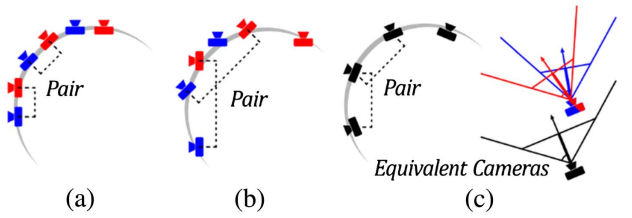

Fig. 5. Three appearances of capturing rigs. (a) Separations of pairwise cameras are relatively small. (b) Separations of pairwise cameras are relatively large. Cameras intrude into adjacent pairs [4]. (c) Cameras apparently distribute symmetrically without pairs or parallel cameras [11]. In PNST capturing, one camera can be equivalently split into two virtual cameras. Cam. L of one pair shares the same position with Cam. R of the adjacent pair.

as their origin. Axis $\boldsymbol{z}$ is parallel to the optical axes of the front camera pair, pointing straight ahead. Axis $\boldsymbol{y}$ is pointing to the zenith, and Axis $\boldsymbol{x}$ is to the left. They create a right-handed system. When looking at the reverse $\boldsymbol{y}$, the direction anticlockwise to Axis $\boldsymbol{z}$ is positive. According to the shooting direction, the camera pair facing ahead is denoted as Pair 1. Pair 2 locates on the left, and Pair n is on the right. Capturing distance from the origin O to the object we focused on is R. The azimuth angle is Φ.

Capturing rigs with different structure parameters may look different in appearance. According to the equivalent camera convention, the capturing model is still applicable for most configurations (Fig. 5).

C. Workflow

"Distributed capturing then mosaicking" [15,16] could be summarized as the general workflow for PNST. If dynamic images are needed, all cameras should be triggered simultaneously. We show the workflow concisely in Fig. 6.

3. PARALLAX IN PNST

For efficient, even real-time, applications, current algorithms used in the image-processing stage still show the incompetence.

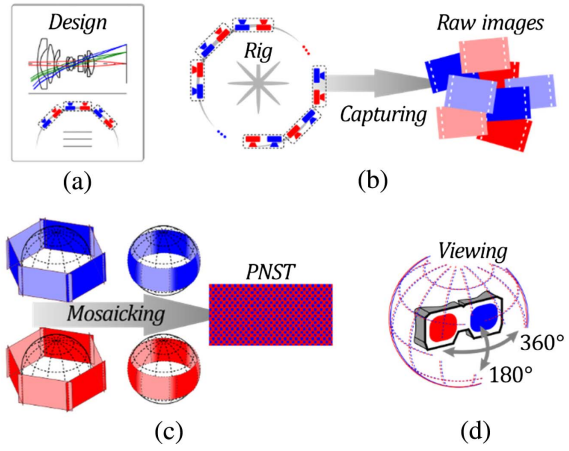

Fig. 6. Workflow of capturing and producing PNST for VR application. (a) Planning and designing. (b) Capturing. (c) Image processing. (d) Viewing.

Therefore, we intend to speed up the workflow in the previous stage of designing by optical methods. Controlling the system design parameters would affect the mosaicking capability of the raw images and the final PNST.

We investigated a quantity of capturing and mosaicking tasks and found that "parallax" is the key. Images with large parallax would be difficult to mosaic [18] and cannot lead to a satisfying final PNST. Parallax affects not only the processing speed but also the viewing quality. Annoyances for mosaicking lurk in the systems that provide raw images with undesirable parallax. However, little attention on parallax has been paid in the irretrievable designing stage for now. If one controls the involved parallax of the system by design parameters, the raw images would be amicable for efficient mosaicking. This leads a quick processing for a satisfying viewing result, time-saving workflow, and efficient applications.

According to the mechanism of PNST, there are two sorts of parallaxes affecting the potential of the efficient mosaicking. One exists between the consecutive camera pairs and has an impact on the mosaicking procedure. The other one exists in the whole PNST and relates to the stereo sense of viewing. We defined them as "mosaicking parallax" (MP) and "stereo parallax" (SP) correspondingly.

A. Mosaicking Parallax

For seamless mosaicking, each camera shoots oversized raw images overlapping with the adjacent ones. In overlaps, one object point would be captured by two or more cameras from different directions. The difference between the shooting directions is the MP, which is always represented by the included angle (Fig. 7). When the MP is too large to be ignored, ambiguities appear, and satisfying mosaicking is not easy to achieve.

Obviously, no MPs exist in the sections where objects are shot by only one camera or equivalent cameras. Because of the symmetry and periodicity, we could just discuss the MPs related to the left cameras in one cycle (from 0° to $360°/n$) for instance. Figure 8 indicates the relevant parameters in the top view.

If there are n left (or right) cameras employed in a rig, the panorama for a single eye contains n overlaps. Each overlap region is limited by the capturing ranges of the adjacent cameras, which begins from the marginal field of view of Cam. $(n+1)L$ and ends at the one of Cam. nL (Fig. 8). We use two azimuth angles, $\Phi_{(n+1)LB}$ and Φ_{nLE}, to indicate the corresponding boundaries of the nth overlap.

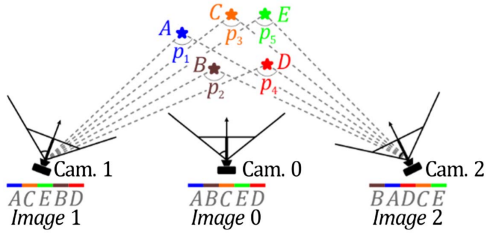

Fig. 7. Different images (Image 1 and Image 2) of the same objects (A to E) are achieved because of the MPs (p_1 to p_5). The existence of ambiguities would hardly lead to satisfying mosaicking, which is supposed to be as "correct" as Image 0.

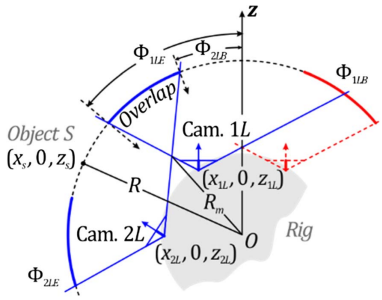

Fig. 8. One overlap and the relevant parameters in one cycle of the whole rig (top view).

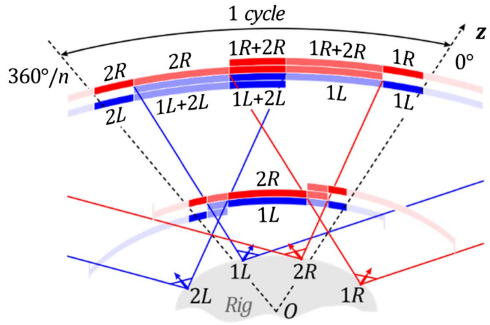

Fig. 9. SPs in PNST would be complicated when the combination of the raw images is provided by more than one camera pair. Narrow (blue or red) annuluses indicate raw images captured by left or right cameras (left or right). Any region of the PNST should consist of at least two pairwise images to create stereo disparities. Some regions may involve more than two raw images.

By geometrics and the coordinates defined in Fig. 8, the boundaries' azimuth angles of the first overlap (Φ_{2LB} and Φ_{1LE}) can be described as

$$\Phi_{2LB} = \arccos \frac{\sqrt{(U^2+1)R^2 - (x_{2L} - Uz_{2L})^2} - U(x_{2L} - Uz_{2L})}{R(U^2+1)}, \tag{1}$$

$$\Phi_{1LE} = \arccos \frac{\sqrt{(V^2+1)R^2 - (x_{1L} - Vz_{1L})^2} - V(x_{1L} - Vz_{1L})}{R(V^2+1)}, \tag{2}$$

where U and V could be defined as

$$\begin{cases} U = \frac{x_s - x_{2L}}{z_s - z_{2L}} = \tan\left(\frac{360^\circ}{N} - \alpha\right) \\ V = \frac{x_s - x_{1L}}{z_s - z_{1L}} = \tan\alpha \end{cases}. \tag{3}$$

Checking boundaries is an effective way of judging whether an object locates in overlaps. $\Phi_{(n+1)LB}$ and Φ_{nLE} are the functions of the capturing distance R. If objects are too close to the rig center, $\Phi_{(n+1)LB}$ would be larger than Φ_{nLE}. It means that no cameras could shoot them; then, the continuous panorama of these areas could not be achieved. This critical distance, minimal capturing distance R_m, could be calculated by setting $\Phi_{2LB} = \Phi_{1LE}$:

$$R_m = \frac{1}{|U - V|}[(x_{1L} - x_{2L} + Uz_{2L} - Vz_{1L})^2 + (Ux_{1L} - Vx_{2L} + UVz_{2L} - UVz_{1L})^2]^{\frac{1}{2}}. \tag{4}$$

For those objects in overlaps, MP could be calculated by two principal rays vector $\mathbf{a}_{nL}$ and $\mathbf{a}_{(n+1)L}$, which emit from the same object to the corresponding camera nodal points:

$$\mathrm{MP}(S) = \langle \mathbf{a}_{nL}, \mathbf{a}_{(n+1)L} \rangle, \tag{5}$$

where

$$\langle \mathbf{a}_1, \mathbf{a}_2 \rangle = \arccos \frac{\mathbf{a}_1 \cdot \mathbf{a}_2}{|\mathbf{a}_1||\mathbf{a}_2|}, \qquad \langle \mathbf{a}_1, \mathbf{a}_2 \rangle \in [0^\circ, 180^\circ]. \tag{6}$$

Notation $\langle \mathbf{a}_{nL}, \mathbf{a}_{(n+1)L} \rangle$ represents the included angle between vector $\mathbf{a}_{nL}$ and $\mathbf{a}_{(n+1)L}$. The subscript "L" could be replaced by "R" for the parameters related to the right cameras.

B. Stereo Parallax

SP is a measurement of the stereo disparity. It is the included angle calculated by the two rays from the same object pointing to the two pairwise cameras. SP can be calculated by

$$\mathrm{SP}(S) = \langle \mathbf{a}_L, \mathbf{a}_R \rangle. \tag{7}$$

The traditional stereo image is constructed by two images with specific parallax. However, images for PNST may be constructed by more than two raw images. The SPs in PNST are more complicated (Fig. 9).

Because of the overlaps, a stereo section of PNST may involve as much as four images provided by different cameras. Focusing on the first cycle of the whole capturing range, SP could be cataloged into five situations (Table 1).

For numerically dealing with these situations according to Eq. (7), we use mean values for calculating viewing angles and representing each overall SP. This could be interpreted as a fusing and even operation. For viewing assessment, it could be regarded as setting a middle viewing direction with blur. For instance, the first three situations could be expressed as

$$\mathrm{SP}_1 = \langle \mathbf{a}_{1L}, \mathbf{a}_{1R} \rangle, \tag{8}$$

$$\mathrm{SP}_2 = \frac{\langle \mathbf{a}_{1L}, \mathbf{a}_{1R} \rangle + \langle \mathbf{a}_{1L}, \mathbf{a}_{2R} \rangle}{2}, \tag{9}$$

Table 1. SP Situations for Constructing the First Cycle of a PNST

	Source Camera(s) for Left Panorama	Source Camera(s) for Right Panorama	Range of the View
1	$1L$	$1R$	$(0, \Phi_{2RB})$
2	$1L$	$1R + 2R$	(Φ_{2RB}, Φ_{2LB})
3	$1L + 2L$	$1R + 2R$	(Φ_{2LB}, Φ_{1RE})
4	$1L + 2L$	$2R$	(Φ_{1RE}, Φ_{1LE})
5	$2L$	$2R$	$(\Phi_{1LE}, 360^\circ/n)$

$$\mathrm{SP}_3 = \frac{\langle \mathbf{a}_{1L}, \mathbf{a}_{1R}\rangle + \langle \mathbf{a}_{1L}, \mathbf{a}_{2R}\rangle + \langle \mathbf{a}_{2L}, \mathbf{a}_{1R}\rangle + \langle \mathbf{a}_{2L}, \mathbf{a}_{2R}\rangle}{4}. \tag{10}$$

If the object is too close to the rig, some stereo information may be offered by the images from Cam. $2R$ and Cam. $1L$, which are not from the same pair (Fig. 9). In reality, the capturing distance is always large enough to get rid of this problem. If the accurate value of this "minimal stereo distance" is needed, one can calculate it by setting $\Phi_{2LB} = \Phi_{1RE}$.

C. Potential of Efficient Mosaicking

The potential of efficient mosaicking is the characteristic of raw images. It indicates their capability and difficulty for mosaicking into PNST. If raw images could lead to time-saving and satisfying results, we can approve the potential. This potential could be translated into the requirements for MP and SP. (1) For mosaicking, MPs should approach zero. (2) For visual quality, SPs should keep constant for a constant capturing distance.

For evaluating this potential numerically, we defined two parallax factors. In statistics, "maximum" and "standard deviation" are two adequate measures for the two requirements above. Basing on the definitions in statistics, we proposed peak parallax (PP) and deviation of parallaxes (DoP) accordingly to evaluate the efficient mosaicking potential numerically. PP is the maximum (peak value) of all the MPs for the entire panorama. This could be calculated from the PNST for single eye and indicates the difficulty of mosaicking. DoP is the standard deviation of all the SPs for the entire PNST. It assesses the rationality and the comfort of stereo vision. Equations (11) and (12) are the corresponding mathematical formulas,

$$\mathrm{PP} = \max\{\mathrm{MP}(N)\}, \tag{11}$$

$$\mathrm{DoP} = \sqrt{\frac{\sum_{N=0}^{N_{\max}}[\mathrm{SP}(N) - \mathrm{SP}_0]}{N_{\max}}}, \tag{12}$$

where N is the index referring to the discrete samples of the investigating range. Operator "max{∗}" indicates the maximum value of ∗. SP_0 represents the mean value of all the SPs. One can extract the discrete parallaxes from the PNST and calculate PP and DoP by these formulas.

4. ANALYSIS AND SIMULATION

The parallax factors have influences on the PNST results and direct the design indeed. Therefore, we connected the parallax factors with the design parameters of the capturing system to discover how design parameters affect the mosaicking efficiency.

For a capturing system, there are five main parameters, which would affect the mosaicking and the final result.

(1) Distance of capturing: R.
(2) Distance between pairwise cameras: $2d$.
(3) Protruding length of camera pair: L.
(4) Quantity of camera pairs: n.
(5) Overlap ratio: η.

Table 2. Reference System Parameters[a]

Parameters	R	n	$2d$	L	α	η
Value	1.5 m	8	150 mm	100 mm	30°	25%

[a] R_m: 388.6 mm, PP: 3.35°, DoP: 0.76.

If more cameras are employed, each camera contributes less field of view. Therefore, a half-field of view α is not an independent parameter. We use overlap ratio η, which could be invariant and easily controlled in different configurations, to describe the instinct property of the capturing range. The relationship between η and α is

$$\eta = \frac{2n \cdot \alpha - 360°}{2n \cdot \alpha}. \tag{13}$$

Figure 4 could be consulted for these definitions. One can hardly derive the analytical relationship between the design parameters and the parallaxes. Therefore, we treat them numerically to find out the rules inside.

The minimal unit of images is a pixel. The analyses should be done at this level. It means that index "N" in Eqs. (11) and (12) should be chosen according to the resolution of the capturing system and then the display system. We employed a reference capturing system (Table 2) and analyzed the variation of PP and DoP for different parameters by controlling the variable method. The reference parameters are selected according to two constraints. One is that the total field of view should cover 360° horizontally. The other is that all of the devices in the system should not overlap with each other physically. Actually, the values in Table 2 are relatively arbitrary without loss of generality.

A. Relationship Between the Design Parameters and the Parallax Factors

We varied the design parameters separately and computed the corresponding PPs and DoPs. All of the result curves are summarized in Figs. 10–14.

Changing the value of R from 1.5 to 10 m while maintaining the others, we achieve the varying curves of PPs and DoPs versus R (Fig. 10). PPs and DoPs decrease along with the increasing of R. One should notice that the objects within 0.7 m could not be captured continuously. In the final PNST for viewing, further objects will maintain a continuous sense of stereo, and closer ones would show some unreasonable information in the overlaps because of the mosaicking. In one word, a large capturing distance brings easier mosaicking. However,

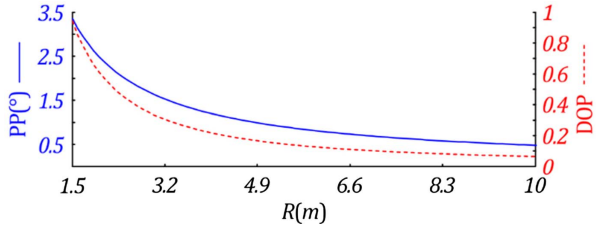

Fig. 10. PP (solid) and DoP (dashed) curves versus different capturing distance R.

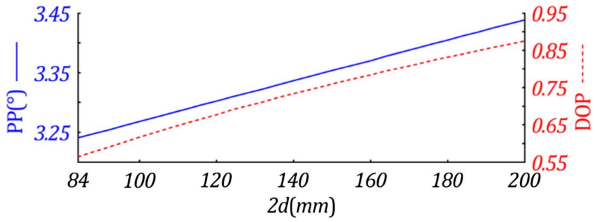

Fig. 11. PP (solid) and DoP (dashed) curves versus different pairwise cameras separating distance $2d$.

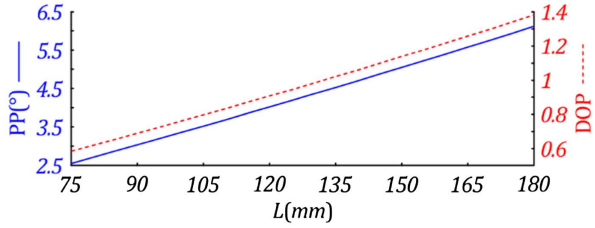

Fig. 12. PP (solid) and DoP (dashed) curves versus the different protruding length L of camera pairs.

the loss of the sense of stereo caused by small disparity (relative short baseline) should be tolerated.

Obviously, the distance ($2d$) between pairwise cameras affects the stereo disparity. When this distance is too small or too large to offer appreciable disparities, the sense of stereo vanishes. We varied the value of $2d$ from 84 to 200 mm and achieved the varying curves of PPs and DoPs (Fig. 11). Changing d will change the appearance of the capturing rig. The camera structure should be limited by the rig model; so, the separating distance between the pairwise cameras should not be too large. In our reference system, the maximum $2d$ is 200 mm. In this range, the PPs and DoPs increase along with the increasing $2d$. Large separation leads to large PP and makes the mosaicking difficult. The uniformity of the disparities between camera pairs is also influenced.

Real cameras have specific volumes. Camera pairs would protrude from the system center to avoid interference (Fig. 4). We varied the protruding distance L from 75 to 180 mm and plotted the corresponding curves in Fig. 12. PPs and DoPs increase with the increasing L. This trend is intuitive. Protruding violates the idealization of mosaicking, which calls for a single viewing point.

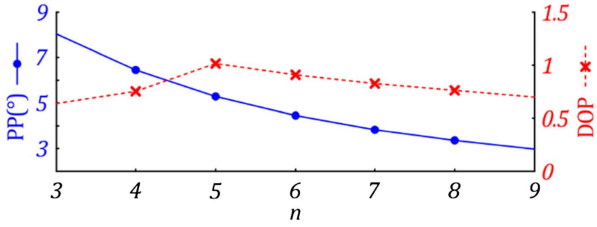

Fig. 13. PP (solid) and DoP (dashed) curves versus different camera pair quantity n.

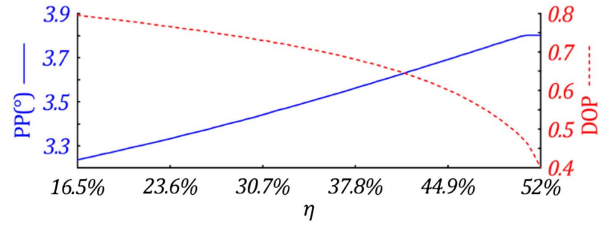

Fig. 14. PP (solid) and DoP (dashed) curves versus different overlap ratio η.

We should pay attention to how the separating distance and the protruding distance affect the appearance of the overall system (Fig. 5). These two parameters also influence the minimal capturing distance R_m [Eq. (4)].

If the field of view of each camera is large enough, the 360° capturing range could be easily covered, no matter how many cameras are employed. Nevertheless, the quantity n of camera pairs affects not only the efficiency of mosaicking but also the overall configuration. We hold the value of the overlap ratio η and varied n from three to nine and compute the corresponding discrete PPs and DoPs (Fig. 13). The PPs will become small when we use more cameras. It shows a similar rule with the system proposed by Peleg *et al.* [12]. A series of small image strips with small overlaps is equal to capturing from a single viewing point. DoPs do not vary monotonically with n. The overall trend is little changed. In a word, more cameras bring smooth overlaps, yet more pieces need to be dealt with.

The overlaps of raw images are necessary for seamless mosaicking. Overlap ratio η is employed to restrict and evaluate the capturing range of each camera. We varied η from 16.5% to 52% (from 27° to 47° for α) and computed the corresponding PPs and DoPs. Varying curves are shown in Fig. 14. When increasing η, PPs ascend, whereas DoPs descend. It means that, the more the overlaps contain, the more difficult the mosaicking is. Large overlaps bring more ambiguous image information. However, large overlaps would make the transitions smooth, not only for the seams but also for the parallaxes. This is caused by the average operation for SP. When considering the overlap ratio, we should balance PP and DoP.

Based on the analyses, PP and DoP always have the coherent varying trends when the structure parameters are changing, except for η. A smooth panorama is good for the experience of interaction and presence. In a PNST, reasonable senses of stereo for some specific sections, not everywhere, are already good enough and acceptable for now. So, we should give priority to a good panorama. We proposed using large R, n and small d, L, η as a general design strategy for PNST capturing system.

B. Verification

For verifying the influences of the parallax factors and the design parameters, we designed two systems according to different design tendencies and simulated their capturing process in the computer. Afterwards, the raw images of each system are mosaicked by only projection and arrangement without any other algorithms. This process of mosaicking could be named "blind" and be regarded as a "fast" mosaicking. If the result, the PNST, of the blind mosaicking has a good quality for viewing, we can

Table 3. Parameters of the Two Systems for Comparison

	R	n	$2d$	L	α	η	PP	DoP
A	3 m	10	100 mm	150 mm	24°	25%	0.29°	1.91
B	1 m	4	250 mm	200 mm	60°	25%	19.86°	5.45

conclude that the corresponding capturing system is suitable for efficient mosaicking. The validity of parallax factors for evaluation could be verified.

We designed System A with small PP and DoP according to the design rules we received. As a comparison, System B, which ignored the considerations for efficient mosaicking, is employed in this simulation (Table 3). We built them in "3ds Max" [23]. We used these virtual cameras to capture the virtual scene of a modeled cathedral, and then got 20 and 8 raw images, respectively. After the blind mosaicking, we achieved two PNSTs, which could be split into four panoramas. We used the corresponding panoramas to check the result of mosaicking (referring to the factor PP) and used the pairwise panoramas to check the quality of stereo (referring to the factor DoP). The assessment in this paper is effective though subjective [24,25].

The panorama in Fig. 15 is mosaicked by 10 raw images offered by System A, and the one in Fig. 16 is mosaicked by four raw images from System B. They are both for the left eye in PNSTs. The visual quality corresponding to System A is obviously much better than the one of System B. Details of the seams are zoomed in for each panorama. Both of them have defects because of the absence of mosaicking algorithms. The zoomed sections in Fig. 16 show some big "mistakes." After all, there are five windows on the front wall of the modeled scene, and System B cannot reserve this information well. System B violated the strategy we proposed and cannot get efficient mosaicking like System A.

For comparing the qualities of stereo, we captured another scene with more details, for example, pillars. Figures 17 and 18 are generated from System A and System B correspondingly.

The sense of stereo comes from the disparity of pairwise images. We choose some object points in Figs. 17 and 18 and measured their SPs in distances. This is more effective than angles for images. DoPs of both PNSTs have been calculated and listed in Table 4.

System B has a small capturing distance, so the SPs are larger than the ones of System A. The DoP of System B is almost 8 times as large as the one of System A. We put both of the PNSTs in a VR device [3] and viewed in 360°. The image of System A has a moderate sense of stereo, although there are some defects in the overlaps. The image of System B has a sensitive sense of stereo in some sections but is confused in other parts. Besides, the image of System B causes some

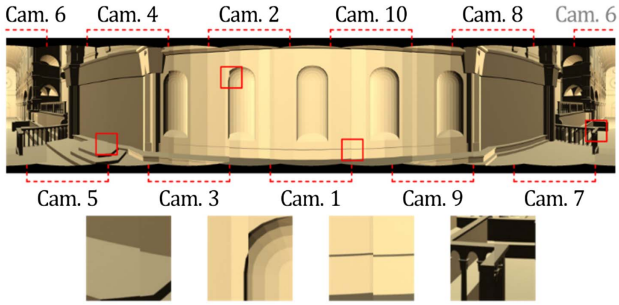

Fig. 15. Ten raw images from System A constructed the panorama by blind mosaicking. Mosaicking defects are zoomed in below.

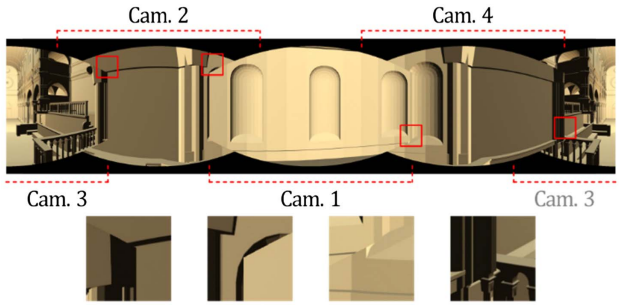

Fig. 16. Four raw images from System B constructed the panorama by blind mosaicking. Mosaicking defects are zoomed in below.

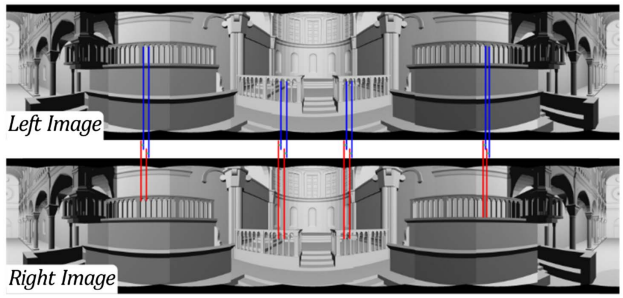

Fig. 17. Pairwise panorama generated from System A.

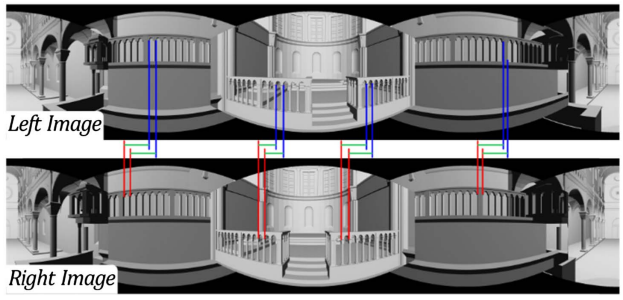

Fig. 18. Pairwise panorama generated from System B.

Table 4. Parallax Factors of the PNSTs by Different Systems

	SP (mm)				DoP
A	0.339	0.337	0.339	0.369	0.017
	0.334	0.320	0.362	0.361	
B	3.558	3.671	3.535	3.756	0.131
	3.536	3.318	3.657	3.544	

uncomfortable feelings when changing viewing directions frequently.

Before capturing, we have predicted the parallax factors of each system. Through the simulation results, the blind mosaicking qualities meet the tendency of the PP and the DoP. Now, we conclude that, when designing a capturing system for PNST, PP and DoP have the capacity to evaluate the potential of efficient mosaicking.

C. Design Strategy

According to the analyses and simulations above, we have found out how design parameters affect the parallax factors, and then the speed and quality of mosaicking.

When designing, one should control the system parameters comprehensively. The general strategy could be summed up here. An efficient mosaicking capturing system calls for large capturing distance, more cameras, compact arrangement, and moderate overlaps. In a catchy form, large R, n, and small d, L, η are good for efficient mosaicking.

For real applications, non-customized devices and other factors limit the design. The optimized design strategy might be surrendered to the quantity, volume, fields of view, and aberrations of the existing cameras. The methods in this work would also help to find a better structure for efficient mosaicking. Nevertheless, our design rules offer a way to a better result instead of the optima, especially when the optimized design is hard to get in practice.

5. SUMMARY AND OUTLOOK

Efficient mosaicking is the key to the applications of PNSTs and VR. We proposed that the potential of raw images should be considered preferentially, and parallax-based factors have been employed to describe this kind of characteristic. A system with large capturing distance, more cameras, compact arrangement, and moderate overlaps could have a good potential for efficient mosaicking. Before the capturing, one can design an efficient system by controlling parameters and following the strategy we proposed.

The research in this issue, as well as the conclusions and results, focus on the promotion of applications. We pursue acceptable PNSTs for applications rather than the best capturing results or perfect depictions of real scenes. Many technologies are involved in the applications of VR, such as computer graphics, optics, controls, and so on. One could also get an efficient capturing system for PNSTs by other methods. What we have researched is for lightening the burden of image processing, which could currently be regarded as a major obstacle. The researchers from other fields could also get enlightenment from this work.

Funding. Youth Innovation Promotion Association of the Chinese Academy of Sciences (2017264); Innovation Foundation of the Changchun Institute of Optics, Fine Mechanics and Physics, CAS (Y586320150).

REFERENCES

1. J. Hecht, "Optical dreams, virtual reality," Opt. Photon. News **27**(6), 24–31 (2016).
2. Y. Wang, W. Liu, X. Meng, H. Fu, D. Zhang, Y. Kang, R. Feng, Z. Wei, X. Zhu, and G. Jiang, "Development of an immersive virtual reality head-mounted display with high performance," Appl. Opt. **55**, 6969–6977 (2016).
3. http://www.samsung.com/global/galaxy/gear-vr/.
4. http://www.thinktankteam.info/beyond/.
5. R. Anderson, D. Gallup, J. T. Barron, J. Kontkanen, N. Snavely, C. Hernández, S. Agarwal, and S. M. Seitz, "Jump: virtual reality video," ACM Trans. Graph. **35**, 198 (2016).
6. J. Thatte, J.-B. Boin, H. Lakshman, G. Wetzstein, and B. Girod, "Depth augmented stereo panorama for cinematic virtual reality with focus cues," in *IEEE International Conference on Image processing* (IEEE, 2016), pp. 1569–1573.
7. S. B. Kang, "Survey of image-based rendering techniques," Proc. SPIE **3641**, 2–16 (1998).
8. R. Bunschoten and B. Kröse, "Robust scene reconstruction from an omnidirectional vision system," IEEE Trans. Robot. Autom **19**, 351–357 (2003).
9. H. Ishiguro, M. Yamamoto, and S. Tsuji, "Omni-directional stereo," IEEE Trans. Pattern Anal. Mach. Intell. **14**, 257–262 (1992).
10. R. O. Reynolds, P. H. Smith, D. G. Crowe, M. Bigler, and M. Pollard, "Design of a stereo multispectral CCD camera for Mars pathfinder," Proc. SPIE **2542**, 197–206 (1995).
11. B. K. Cabral, "Introducing Facebook Surround 360: an open, high-quality 3D-360 video capture system," https://code.facebook.com/posts/1755691291326688/introducing-facebook-surround-360-an-open-high-quality-3d-360-video-capture-system/.
12. S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: panoramic stereo imaging," IEEE Trans. Pattern Anal. Mach. Intell. **23**, 279–290 (2001).
13. F. Zhou, X. Chai, X. Chen, and Y. Song, "Omnidirectional stereo vision sensor based on single camera and catoptric system," Appl. Opt. **55**, 6813–6820 (2016).
14. F. Zhang and F. Liu, "Casual stereoscopic panorama stitching," in *Conference Computer Vision and Pattern Recognition* (IEEE, 2015), pp. 2002–2010.
15. Z. Zhu, "Omnidirectional stereo vision," in *Workshop on Omnidirectional Vision, in the 10th IEEE ICAR* (2001).
16. L. E. Gurrieri and E. Dubois, "Acquisition of omnidirectional stereoscopic images and videos of dynamic scenes: a review," J. Electron. Imaging **22**, 030902 (2013).
17. F. Briggs, "Surround 360 is now open source," http://code.facebook.com/posts/265413023819735/surround-360-is-now-open-source/.
18. F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *Conference Computer Vision and Pattern Recognition* (IEEE, 2014), pp. 3262–3269.
19. C. Richardt, Y. Pritch, H. Zimmer, and A. Sorkine-Hornung, "Megastereo: constructing high-resolution stereo panoramas," in *Conference Computer Vision and Pattern Recognition* (IEEE, 2013), pp. 1256–1263.
20. F. Amjadi and S. Roy, "Comparison of radial and tangential geometries for cylindrical panorama," in *Fourth International Conference on 3D Vision* (2016), pp. 649–657.
21. F. Hongfei, J. Jinyuan, W. Hongkai, and T. luchen, "Immersive roaming of stereoscopic panorama," in *International Conference on Cyberworlds* (IEEE, 2008), pp. 377–382.
22. R. G. Baker, "Immersive imaging system," U.S. patent US7224382 B2 (29 May 2007).
23. http://www.autodesk.com/products/3ds-max/overview.
24. N. Yun, Z. Feng, J. Yang, and J. Lei, "The objective quality assessment of stereo image," Neurocomputing **120**, 121–129 (2013).
25. A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," Signal Process. **28**, 870–883 (2013).