IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 June 2024, accepted 30 June 2024, date of publication 8 July 2024, date of current version 24 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3424412

RESEARCH ARTICLE

A Novel Multiple Camera RGB-D Calibration Approach Using Simulated Annealing

MEHRAN TAGHIPOUR-GORJIKOLAIE[1,2], MARCO VOLINO[1], CLARE RUSBRIDGE[3], AND KEVIN WELLS[1]

[1]Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, GU2 7XH Guildford, U.K.
[2]School of Engineering, London South Bank University (LSBU), SE1 0AA London, U.K.
[3]School of Veterinary Medicine, University of Surrey, GU2 7XH Guildford, U.K.

Corresponding author: Mehran Taghipour-Gorjikolaie (mehran.taghipour-gorjikolaie@lsbu.ac.uk)

This work was supported in part by the Dogs Trust Canine Welfare Grant, and in part by the Hannah Hasty Memorial Fund.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the NASPA Sub-Committee (NASPA) under Application No. NASPA-1819-008.

ABSTRACT The development of a cost-effective surface scanning system tailored for live animal image capture can play an important role in biomedical research. The primary aim was to introduce a low-cost system, achieving a surface reconstruction error of less than 2mm, and enabling rapid acquisition speeds of approximately 1 second for a complete 360-degree surface capture. Leveraging a five RGB-D camera configuration, our approach offers a simple, low-cost alternative to conventional lab-based 3D scanning setups. Key to our methodology is a novel calibration strategy aimed at refining intrinsic and extrinsic camera parameters simultaneously for improved accuracy. We introduce a novel 3D calibration object, extending existing techniques employing ArUco markers, and implement a depth correction matrix to enhance depth accuracy. By utilizing Simulated Annealing optimization alongside our custom calibration object, we achieve superior results compared to conventional optimization techniques. Our obtained results show that the proposed depth correction method can reduce the reprojection error from 3.12 to 2.89 pixels. Furthermore, despite the simplicity of our reconstruction method, we observe around a 22% improvement in surface reconstruction compared to factory calibration parameters. Our findings underscore the practicality and efficacy of our proposed system, paving the way for enhanced 3D surface reconstruction for real-world surface capture.

INDEX TERMS Azure Kinect, depth correction, RGB-D calibration, sensors, surface reconstruction.

I. INTRODUCTION

The Azure Kinect DK is the latest generation of Kinect cameras which includes two different types of sensors for 3D scanning: a 12-megapixel RGB sensor that produces 2D images; and a 1-megapixel time-of-flight IR depth sensor [1]. Experimental evaluation demonstrated that the Azure Kinect DK camera can reduce depth error and increase spatial and temporal accuracy in comparison with previous versions of Kinect cameras [1], [2]. These types of cameras are known as RGB-D sensors which are widely used for a range of tasks including robot navigation [3], gait recognition [4], object recognition [5], and 3D model scanning for biomedical applications [6], [7].

In our case, we were motivated to use this technology for developing a 3D surface capture system in order to investigate correlations in canine head surface shape as indicators of inherited neurological disorders (Chiari-like Malformation and Syringomyelia). This system needed to be low cost, portable, but sufficiently accurate to capture the head surface shape for correlation with the internal anatomy obtained from MRI/CT, and thus applied as a triage system for identifying subjects at risk of these clinical conditions.

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhua Sheng.

The first step for preparing a 3D surface scanning system is to calibrate the RGB-D sensors. These sensors are usually factory calibrated but the accuracy is insufficient for sensitive applications such as in a medical and veterinary setting. Updated intrinsic and extrinsic camera parameters can enhance camera performance in such cases [8]. However, the choice of calibration object also plays a key role in the quality of the final 3D reconstructed model. One of the most common calibration objects used in surface reconstruction consists of a planar checkerboard, in which corners of each square are considered as reference points for calibration [8], [9], [10], [11]. Although such a checkerboard pattern could be sufficient for some applications, to increase the accuracy of the detected calibration points ArUco markers may be combined with a checkerboard, referred to as a ChArUco board [12]. However, using planar objects for calibration requires variation in object-camera distance to generate depth information, which has motivated the move to 3D calibration objects that do not require such motion. For example, in [13], [14] two checkerboards coupled at a 90-degree angle have been proposed. This helps to represent depth information that may be perfect for single view 3D reconstruction. For Multiview 3D reconstruction, cubic and spherical versions of such 3D calibration objects have been proposed in [15] and [16], respectively. Although, the 3D scan space could be covered using such objects, these do not offer the advantages of the 90-degree coupled checkerboards, and moreover, require several refinement steps to obtain the final 3D reconstruction. Therefore, in this paper we propose an alternative ArUco marker-based phantom that combines the advantages of these approaches with a fast and easy-to-use approach for calibration.

A further challenge for RGB-D sensor calibration is the accuracy of any depth measurement. Prior works show that most RGB-D sensors suffer significant depth measurement errors, and therefore, in most bespoke camera calibration methods depth correction is also proposed, some of which are based on correction of depth image data directly and some are based on using combinations of RGB and IR images [17], [18], [19], [20], [21], [22], [23]. The main idea for [17] and [18] is to correct depth error based on a range measurement correction, because depth data are produced by range-to-3D transformation. In [17], a single range bias was considered for depth error correction, and in [18] range measurement errors are classified into five clusters using K-means. However, these approaches ignore intrinsic depth sensor parameters for depth correction and treat depth correction and intrinsic parameter estimation as separate steps. This may result in suboptimal intrinsic parameter estimates. In contrast, more accurate depth measurement can lead to better intrinsic parameter estimates, making it important to simultaneously estimate depth correction and intrinsic parameters. Depth information used in [11] and [19], utilizes a spatially varying exponential offset to correct depth error. Such techniques may be useful for capturing complete indoor 3D scenes rather than single isolated target object. Such approaches have been previously based around scanning 2D planar calibration objects requiring multiple different depth views over a large area, in contrast to the relatively small ($<$1m3) targeted for capture in our application. However, such an approach is not possible here as we have no access to the factory-based disparity values. In [20], intrinsic parameters for both RGB and depth sensor data were estimated at the same time by using IR images and a generalization of normal distributions as a Gaussian Processes (GP) proposed for depth correction. The main challenge of the method is the requirement for large amounts of training data, which rendered this approach impossible given the relatively small data set available to us.

It is worth noting that most of the proposed methodologies for camera calibration and depth correction are based on an iterative optimization approach [19]. One of the most common optimization techniques is Damped Least-Squares (DLS), otherwise known as the Levenberg Marquardt (LM) algorithm. LM approach is very sensitive to being trapped in local minima, dependent on initialization of parameter estimates, especially for complex cost functions.

The last step for surface capture is to use an efficient registration technique with low computational cost. A variety of different approaches may be employed for such 3D registration. The most popular approaches are based on extracted features from an RGB image such as SIFT [24], SURF [25], and FAST [26]. Other methods are based on 3D features on overlapped areas between pairs of point clouds generated for two cameras, such as ICP [27], CPD [28], and deep learning-based methods such as Deep VCP [29], and deep closest point (DCP) [30]. These approaches are useful and efficient only when the overlap of the views between cameras are high, and sufficient pairs of match points can be extracted from the overlap area. Additionally, Deep learning-based approaches require substantial training data which in many applications is not viable.

In this paper, the main goal was to develop a low cost ($<$ \$4k) surface scanning system with average reconstruction error of $<$2mm and acquisition speeds of $\sim$1s for a full 360-degree surface capture to allow for unavoidable subject motion of live subjects. We opted for a five-camera approach based on the well-established Azure Kinect technology, offering a cost-effective approach that can be used in uncontrolled settings outside of the laboratory as an alternative to permanent lab-based 3D scanning systems. The full 360-degree field of view for capture using five cameras meant that overlap in camera views was minimal, rendering several of the previously published approaches non-viable. Moreover, early experiments with the factory-supplied intrinsic camera parameters demonstrated that significant further improvement would be needed via a bespoke calibration approach. We therefore sought to develop a novel calibration approach using a novel phantom that extends prior work in this area using ArUco markers. Our approach is based on a novel

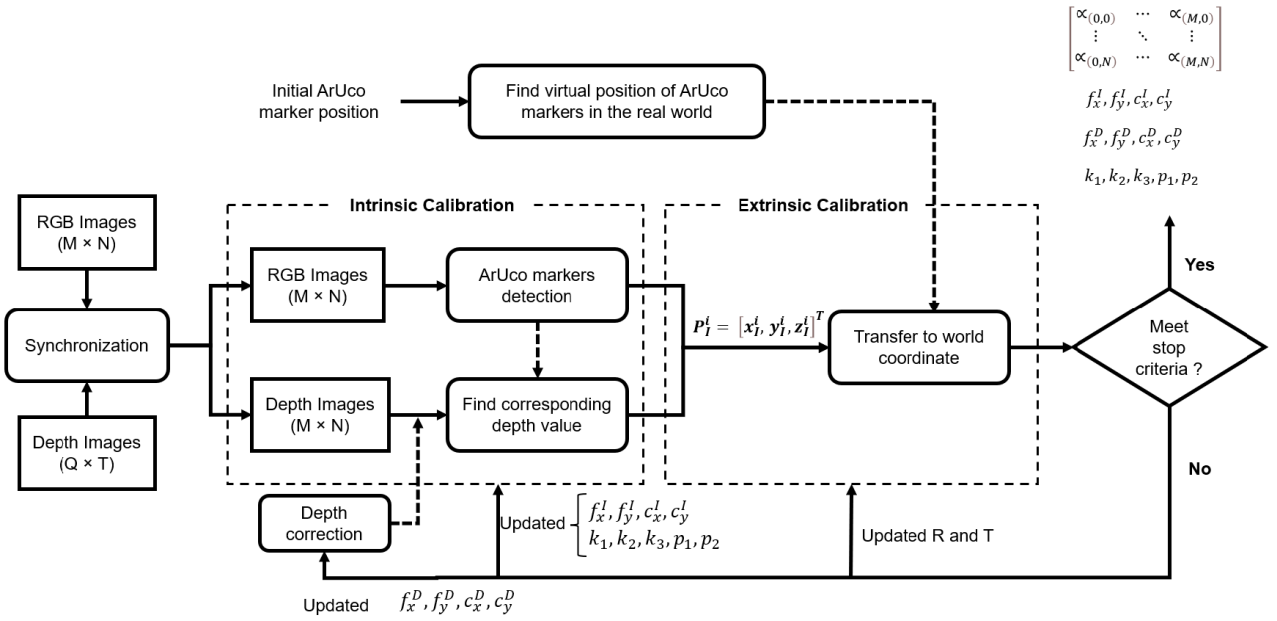

FIGURE 1. Azure Kinect camera calibration includes optimizing RGB intrinsic, and distortion parameters, depth correction and extrinsic calibration.

depth correction matrix, such that each element of the matrix represents a depth correction coefficient for a corresponding pixel in the RGB image which also uses virtual depth calibration parameters. A Simulated Annealing optimization-based approach in conjunction with a novel calibration object is then used for calibration which presents better results than LM or other optimization techniques. This is followed by a simple registration step based on outlier removal.

In the rest of the paper, the proposed RGB-D calibration approach is outlined in Section II which includes details of the novel 3D calibration object, depth correction method, calibration approach and the proposed 3D headspace surface reconstruction. Finally, experimental results are discussed in Section IV.

II. PROPOSED RGB-D CALIBRATION APPROACH

As mentioned before, two items play an important role in RGB-D calibration results: the calibration object and the depth correction model. Consider point i on the object in the world and camera coordinate system as $P_W^i = \left[x_W^i \; y_W^i \; z_W^i\right]^T$ and $P_I^i = \left[x_I^i \; y_I^i \; z_I^i\right]^T$, respectively, and its corresponding projection in the RGB image is $M_I^i = \left[u_I^i \; v_I^i\right]^T$. Equations (1) and (2), represent the transformation of i^{th} point from world to image plane based on a Pinhole camera model [31].

$$u_I^i = \frac{x_I^i f_x^I}{z_C^i} + c_x^I$$
$$v_I^i = \frac{y_I^i f_y^I}{z_C^i} + c_y^I \tag{1}$$

$$\begin{bmatrix} u_I^i \\ v_I^i \\ 1 \end{bmatrix} = \frac{1}{z_I^i} \begin{bmatrix} f_x^I & 0 & c_x^I \\ 0 & f_y^I & c_y^I \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_W^i \\ y_W^i \\ z_W^i \\ 1 \end{bmatrix} \tag{2}$$

where, f_x^I and f_y^I are focal lengths, and c_x^I and c_x^I are the center of the RGB image plane (principal points), which are referred to as intrinsic parameters. Moreover, the same formulation could be considered for depth images. In addition, radial distortion and tangential distortions are considered for the captured RGB image based on a pinhole camera model.

The Azure Kinect camera provides a set of depth values (DV) representing each pixel of the RGB image, which is an estimate of the distance between the origin of camera coordinate system (depth sensor) and corresponding point of the object in the real word to the pixel. However, z_I^i is the distance between the point and the RGB plane. Based on Euclidian distance there is a relationship between DV and the camera coordinate system for each point. It is worth noting that the resolution of the RGB and depth image are different and therefore an alignment (synchronization) is required. Therefore, there are two sets of intrinsic (calibration) parameters for RGB-D cameras: one for the RGB sensor and the other for the depth sensor (henceforth used for depth correction). Initial results can be obtained by using the RGB-D camera pre-set factory calibration factors, but these are not sufficient for delivering high quality 3D reconstruction needed for our application, motivating a need for re-calibration. As shown in Fig. 1, the proposed approach for Azure Kinect camera calibration has been considered as an optimization problem, during which each iteration of the intrinsic RGB, and depth camera distortion parameters are updated. Moreover,

the system is intended to be used outside a controlled laboratory environment in other indoor field locations where it may not be possible to control environmental factors such as illumination.

A. PROPOSED 3D OBJECT

As shown in Fig. 2, the proposed 3D calibration object is based on a set of 3D columns upon which a set of ArUco markers are attached based on a cross-ratio pattern as calibration points. Cross ratio patterns are projective invariant, which means they are robust to camera orientation. Moreover, Fig. 2(a) and (b) show the proposed calibration object which is made by four 40 × 40 mm aluminum profile struts manufactured by PS PRO [32]. Four corners or the middle point of a ArUco marker can be used as calibration features, with which the location of the marker can be detected by threshold segmentation, straight line fitting, and quadrilateral sum conjecture [12].

There are four unique markers on each column, each repeated four times on each column face, yielding in total 20 markers. This produces 190 Pair-Distances (PDs) (virtual lines) that can be defined and calculated between each pair of points by using AM = 20 ArUco markers. However, it was empirically found that 72 PDs are sufficient for covering the target 3D reconstruction space. Fig. 3(a) to (f) illustrate combinations of PDs considered for the proposed calibration.

B. VIRTUAL 3D POINTS OF ARUCO MARKERS

To be able to utilize all the virtual lines shown in Fig. 3, it is necessary to detect all 20 markers on the test object. Therefore, an optimization technique is used to extract optimum detection parameters. Equation (3) is used as a cost function to minimize the Detection Error (DE) during optimization for each of the 5 views.

$$DE = \min\left(|AM - DAM|\right), (AM = 20) \tag{3}$$

where, AM is the maximum number of attached ArUco markers on the calibration object for each view, and DAM is detected ArUco markers in each iteration of optimization algorithm. This optimization must be completed for each view separately to determine optimum parameters for that view. The main part of this step is to produce virtual 3D ArUco markers points representing the position of markers on the calibration object. A total of 72 reference points (PDs) are established as the ground truth for the object. These reference points are meticulously measured using a high-grade digital laser tape measure with an accuracy of ±2mm. Equation (4) shows the proposed cost function based on Euclidean distance to minimize the Virtual Position of ArUco Markers Error (VPAME) during iterations to extract the optimum position of these virtual markers in the real-world coordinate system. As shown in Fig. 4, the middle of ArUco marker number 20, which is in the bottom part of the middle strut is considered as the origin of the calibration object in the real-world coordinate system. Each strut is 40 × 40 mm^2, and thickness of the ArUco marker is 0.8 mm, therefore, the reference point for the front view is $P_W^{20} = \left[\,0\ 0\ 20.8\,\right]$.

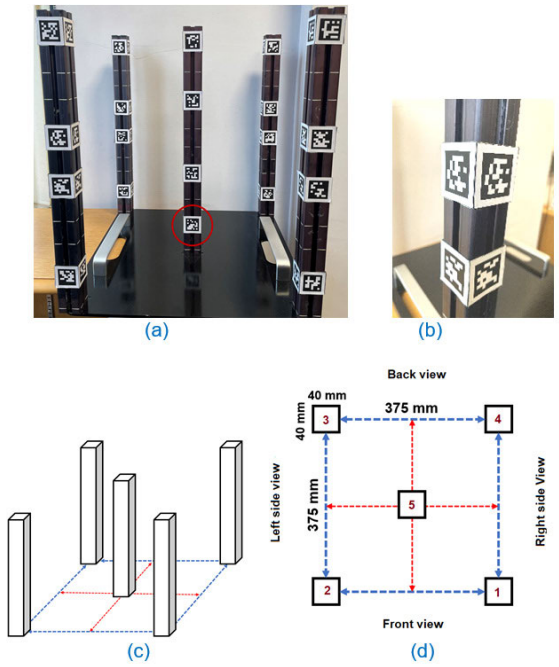

FIGURE 2. Proposed 3D calibration object. (a), and (b) are RGB image of the calibration object and IDs 7 and 12. (c) Right Isometric view. (d) Top view. Red circle which shows the origin of the virtual points is the same position as it can be found in Fig.4.

$$VPAME\left(P_W^1, P_W^2, \ldots, P_W^{19}\right) = \min\left(\frac{1}{L}\sum_{j=1}^{L}\left\|PD_j^R - PD_j^P\right\|\right), (L = 72) \tag{4}$$

where, P represents the virtual position defined by the center of the ArUco markers. PD^R, and PD^P are real (measured) and predicted pair-distances, respectively.

C. DEPTH CORRECTION MODEL

The Azure Kinect camera has four depth camera supported operating modes (Narrow Field Of View (NFOV) unbinned, NFOV 2 × 2 binned, Wide Field Of View (WFOV) 2 × 2 binned, and WFOV unbinned) and two aspect ratios (4:3 and 16:9) with eight RGB camera resolutions. Choosing a suitable depth mode and RGB resolution can impact on the accuracy of calibration and thus the quality of any resulting surface model. As mentioned in [2], NFOV provides reduced systematic error for calculating depth and better pixel overlap is obtained when RGB and depth cameras are in NFOV mode (75°×65°) and 4:3 resolution (90°×74.3°), respectively. However, prior works [1], [2] show that systematic error deteriorates after depth and RGB image synchronization. In the proposed method, depth values have been corrected for each pixel using new virtual depth intrinsic parameters, instead of directly recalibrating the depth sensor.

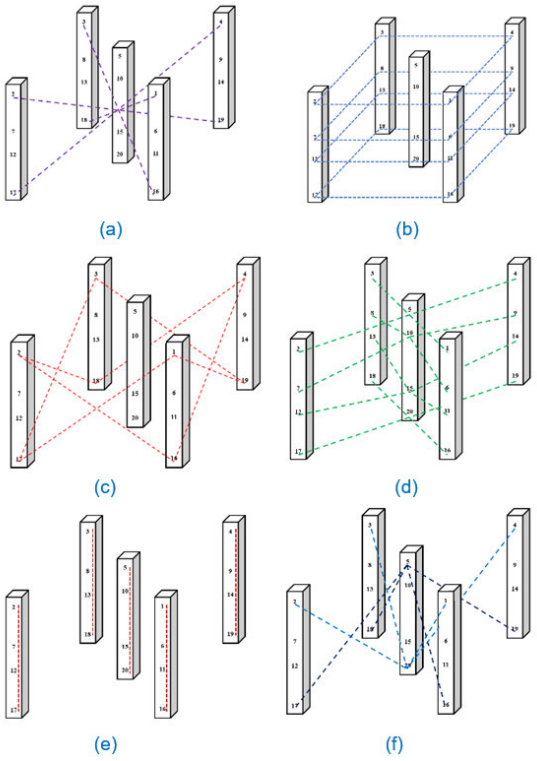

FIGURE 3. Different type of virtual pair-distances (PDs) considered for the proposed calibration. (a) to (f) are referred to as central star PDs, horizontal PDs, side diagonal PDs, central horizontal diagonal PDs, vertical PDs, and central diagonal PDs, respect.

Let D^i is a distance between the origin of the camera and i^{th} point in the image plane.

$$D^i = \sqrt{x_I^{i2} + y_I^{i2} + z_I^{i2}} \tag{5}$$

Equation (5) can be represented by using (1) as

$$D^i = \sqrt{\left(\left(\frac{u_D^i - c_x^D}{f_x^D}\right) z_I^i\right)^2 + \left(\left(\frac{v_D^i - c_y^D}{f_y^D}\right) z_I^i\right)^2 + z_I^{i2}} \tag{6}$$

$$D^i = z_I^i \sqrt{\left(\frac{u_D^i - c_x^D}{f_x^D}\right)^2 + \left(\frac{v_D^i - c_y^D}{f_y^D}\right)^2 + 1} \tag{7}$$

$$z_I^{iC} = \frac{D^i}{\left(\sqrt{\left(\frac{u_D^i - c_x^{Dnew}}{f_x^{Dnew}}\right)^2 + \left(\frac{v_D^i - c_y^{Dnew}}{f_y^{Dnew}}\right)^2 + 1}\right)} \tag{8}$$

$$\alpha = \frac{1}{\left(\sqrt{\left(\frac{u_D^i - c_x^{Dnew}}{f_x^{Dnew}}\right)^2 + \left(\frac{v_D^i - c_y^{Dnew}}{f_y^{Dnew}}\right)^2 + 1}\right)} \tag{9}$$

where, z_I^{iC} is the corrected depth values in depth image, u_D^i and v_D^i are corresponding depth image coordination of i^{th} point which is aligned with the RGB image. $f_x^{D_{new}}$, $f_y^{D_{new}}$, $c_x^{D_{new}}$ and $c_y^{D_{new}}$ are the new intrinsic parameters for the depth sensor. α is a depth correction factor for each point on the object or the corresponding pixel. This means that a Depth Correction Matrix (DCM) can be defined for a M × N 2D image, as

$$DCM = \begin{bmatrix} \alpha_{(1,1)} & \cdots & \alpha_{(M,1)} \\ \vdots & \ddots & \vdots \\ \alpha_{(1,N)} & \cdots & \alpha_{(M,N)} \end{bmatrix} \tag{10}$$

D. AZURE KINECT CAMERA CALIBRATION

The Levenberg Marquardt (LM) algorithm is conventionally used as a deterministic approach for calibration. Although it is reliable and fast, it is highly sensitive to initialisation conditions., To address this issue, Simulated Annealing (SA) which is a metaheuristic optimization algorithm for approximating global optimum in a large search space [33] has been proposed for RGB-D calibration. In the formulation used in this work, 19 parameter estimates must be optimized: 4 parameters for RGB intrinsic parameters, 4 parameters for depth intrinsic parameters, 5 parameters are for RGB distortion, and 6 parameters for rotation and translation. Equation (11) shows the Reprojection Error (RE) which is defined as a cost function of SA optimization,

$$RE = \min\left(\frac{1}{N}\sum_{j=1}^{N}\sqrt{\left(u_I^{jR} - u_I^{jP}\right)^2 + \left(v_I^{jR} - v_I^{jP}\right)^2}\right), \quad (N = 20 \times 4) \tag{11}$$

where, $\left(u_I^{jR}, v_I^{jR}\right)$ and $\left(u_I^{jP}, v_I^{jP}\right)$ represent the real position of corners of the i$^{\text{th}}$ ArUco markers, and reprojected position of corresponded point from world coordinate to image plane.

To start optimization, factory parameters were set to for initializing RGB intrinsic, depth intrinsic, distortion parameters, and transformation matrix (includes rotation and translation with 6 degree of freedom) calculated by ICP method [33]. In each iteration, virtual points of the four corners of each ArUco markers on the calibration object ($4 \times 20 = 80$) were transferred from the real-world coordinates camera coordinates by using updated rotation and translation parameters, then by implementing (1) to (3) with updated distortion, RGB intrinsic parameters, and depth values, new corresponding projected points $\left(u_I^{jP}, v_I^{jP}\right)$ on image plane could be calculated.

E. PROPOSED 3D RECONSTRUCTION

The main challenge for working with animals is that the subject is unlikely to remain stationary for more than a few seconds. Therefore, it is necessary to propose a setup that can capture synchronized RGB-D images from multiple viewpoints, which can then be used to generate the corresponding 3D model.

The main idea of 3D reconstruction is to use the proposed 3D calibration object as a 3D registration object. As shown

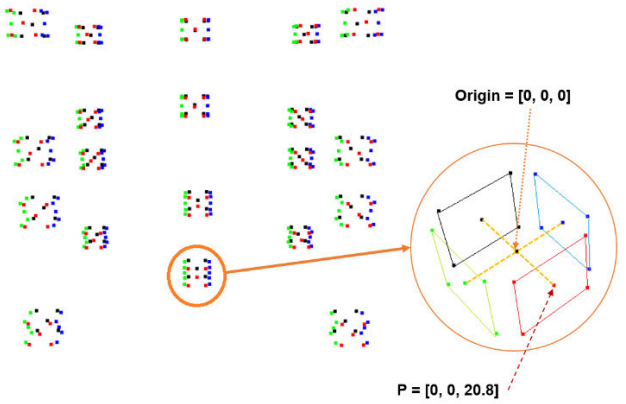

FIGURE 4. Finding common boundaries for pair-registration and final 3D reconstruction. The red circle which shows the origin of the virtual points is the same position as it can be found in Fig.2 a.

in Fig. 5, the proposed structure for 3D scanning includes five Azure Kinect cameras: four cameras positioned at a 90-degree angle relative to each other, with a fifth camera placed beneath the front camera to capture the lower jaw and neck that would otherwise be partially occluded. One camera (facing the subject) is considered as a reference device that leads to a daisy-chain configuration for the remaining four. Moreover, to prevent IR depth illumination interference between cameras, a 160 μs temporal acquisition offset is used for each subordinate camera. All point clouds (PCD) generated from different views are transferred into real world coordinates. The final reconstruction utilizes a sequentially additive approach using four pair-wise registrations, where the first pair-wise registration between PCDs from the top front view and bottom front view are combined to produce an initial partial surface reconstruction. We then add in each PCD from each camera view to extend the partial surface reconstruction until all PCDs from all camera views have been incorporated into a complete surface reconstruction. As shown in Fig. 6, for each pair-registration common boundary between two PCDs are determined, and then outliers are removed to generate reconstructed model.

III. IMPLEMENTATION AND EXPERIMENT

A. OPTIMIZATION ALGORITHMS

To evaluate the performance of all cameras in a variety of different locations in the FoV shown in Fig. 5 (c), 14 samples positions are captured for each view: 10 samples for iterative calibration, and 4 used as test cases.

As previously discussed, experimental results indicate that LM is unsuitable for our strategy due to its susceptibility to becoming trapped in local minima and its failure to achieve acceptable error levels. Consequently, we advocate for the utilization of Meta-heuristic optimization algorithms such as Genetic Algorithm (GA) or Simulated Annealing (SA). Given that the optimization algorithm plays a pivotal role in our proposed method, it is imperative to optimize the controlling parameters effectively. Thus, in this section, we conduct

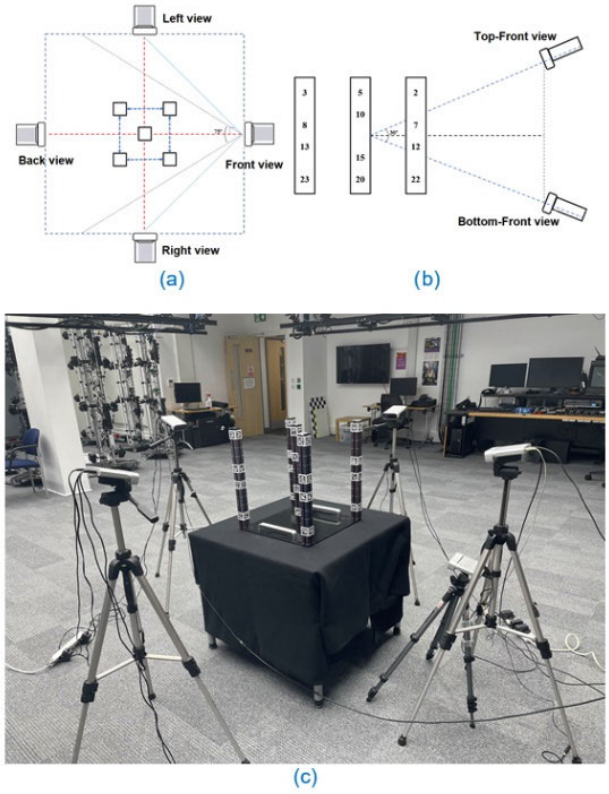

FIGURE 5. A sample point in the camera coordination system Position of the cameras. (a) and (b) are top and right views of the position of the cameras. (c) The position of the cameras in experimental 3D scan setup.

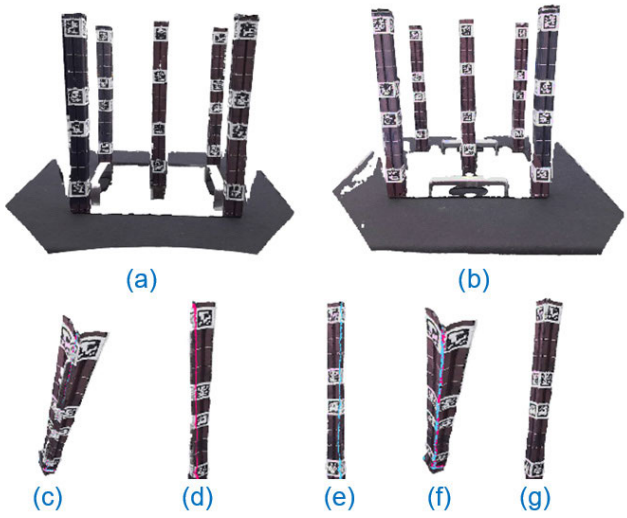

FIGURE 6. Finding common boundaries for pair-registration and final 3D reconstruction. (a) point cloud (PCD) file for front top camera, (b) PCD file for left side view camera, (c) finding overlap area for 4th strut, (d) common boundary on 4th strut on the front.

sensitivity analysis to identify the most suitable parameters for the optimization algorithms.

1) SENSITIVITY ANALYSIS FOR SA

SA has two important controlling parameters: Initial Temperature (T0) and Cooling Schedule (α). TO determines

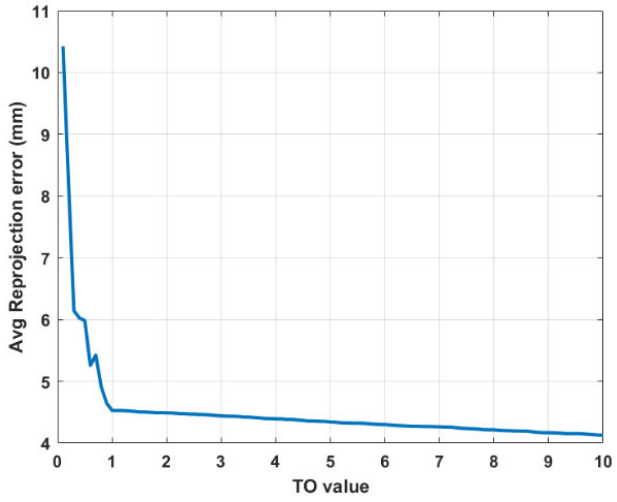

FIGURE 7. Variation of Average of Reprojection error for all 5 cameras by changing TO value and considering fixed value for cooling schedule.

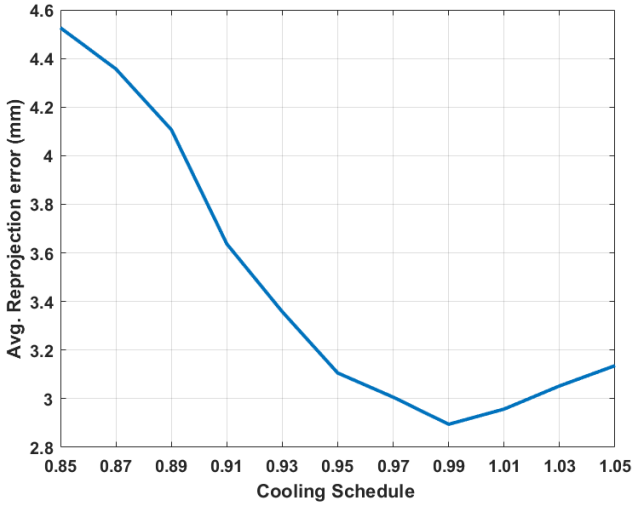

FIGURE 8. Variation of Average of Reprojection error for all 5 cameras by changing the cooling schedule value and considering fixed value for TO.

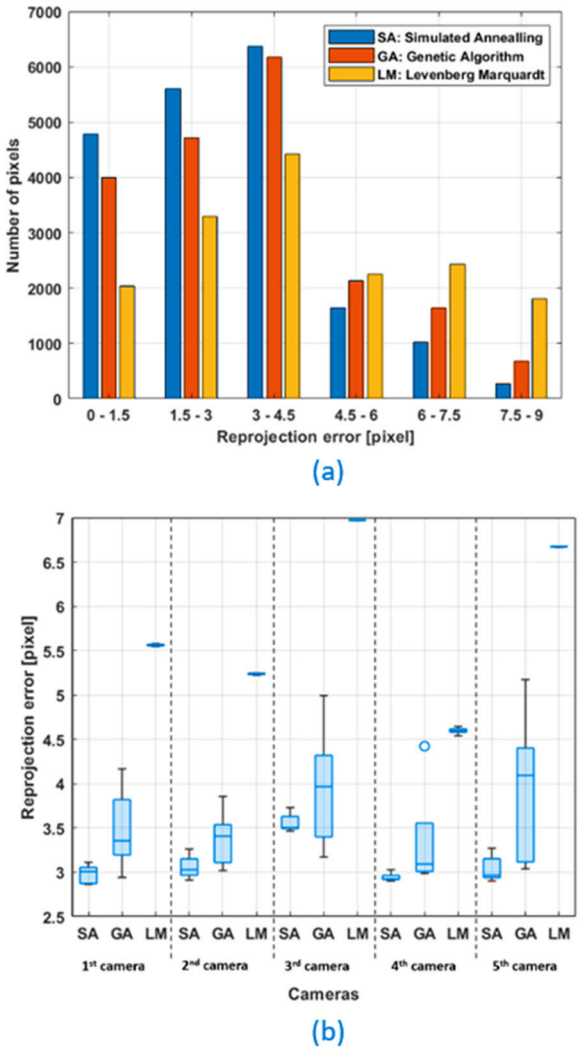

FIGURE 9. Reprojection error of all pixels in all captured images. (a) The distribution of reprojection errors for Factory parameters, SA, GA, and LM based parameters. (b) box-chart of optimization algorithms based on different cameras and 5 separated runs.

the initial "temperature" of the system which is normally between 0.1 to 10. A higher initial temperature allows the algorithm to explore a wider range of solutions initially but may result in a longer time to convergence. Conversely, a lower initial temperature may lead to faster convergence but could risk getting stuck in local optima. The cooling schedule which is in the range of 0.85 to 0.99 defines how the temperature decreases over time. The most common cooling schedules include exponential, linear, and logarithmic. The choice of cooling schedule affects the rate at which the algorithm explores the solution space and balances exploration (at higher temperatures) with exploitation (at lower temperatures). As evident from Fig. 7, the error shows minimal change beyond TO =1. Conversely, higher TO values may escalate completion costs and hinder convergence. Therefore, an optimal TO value appears to be one. By setting TO =1, we can adjust the cooling schedule to identify the most suitable value. Fig. 8 illustrates a descending trend in the average reprojection error until 0.99, beyond which it begins to rise. Hence, 0.99 could be chosen as an optimal choice for alpha.

We applied the same approach to the GA, determining the best parameters as follows: population size = 50, mutation rate = 0.05, crossover rate = 0.75, and selection method = roulette wheel selection. Fig.9 shows reprojection errors of 4 corners of each ArUco marker on the 3D calibration object, with a box and whisker plot for each camera illustrating the reprojection errors. As shown in Fig. 9, SA presents better performance in comparison with the GA and LM methods. For example, more than 52% of reprojection errors are less than 3 pixels, while this value for GA, and LM are 43% and 26%. In addition, the box chart shows that although LM has a very low standard deviation, it has a much lower accuracy, while SA with even lower standard deviation and

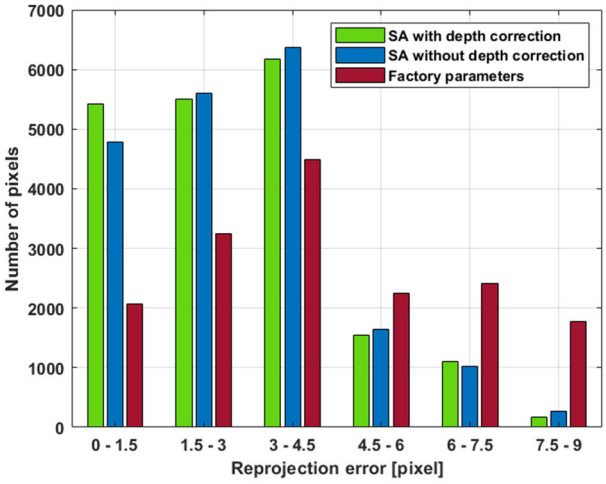

FIGURE 10. Reprojection error of all pixels in all captured images. Evaluating the effect of the proposed depth correction method.

better (lower) mean error in comparison with GA, is more reliable than GA, and LM.

B. DEPTH CORRECTION EVALUATION

Methods reported in [15] and [16], have been used as a depth correction method in our RGB-D calibration approach for comparative purposes. As shown in Fig. 10, our proposed method significantly improves the accuracy in comparison with the factory parameters. Moreover, the average errors with and without depth correction are 2.89 and 3.12 pixels, respectively, which demonstrates that the accuracy improves as a direct result of using the proposed calibration procedure. Furthermore, we conducted comparisons with two well-known 3D calibration objects: with a cubic [15] and a spherical [16] structures, along with their respective calibration strategies, to assess the efficacy of our proposed method. Like Fig. 3, our evaluation encompassed six distinct test scenarios representing horizontal, vertical, diagonal, and depth lines in real-world settings.

As illustrated in Table 1, employing our proposed 3D calibration object results in reduced prediction errors when measuring distances between two points in real-world scenarios. For instance, when measuring central-diagonal lines, the average errors are 2.09 mm, 3.309 mm, and 5.06 mm using the proposed method, the spherical object, and the cubic object, respectively. Notably, across all scenarios, our proposed method consistently outperforms both the spherical and cubic-based approaches, except in cases involving horizontal lines where the cubic-based method exhibits superior performance over the spherical object.

In addition, the comparison reported in Table 1 shows that the proposed method is more accurate than other approaches [15], [16] based on the test samples.

C. 3D RECONSTRUCTION

Fig. 11 shows a 3D scanning setup for the monster and one of real dogs. Moreover, Fig. 9 (c) to (h) present headspace area of RGB images that captured by the proposed setup in different direction.

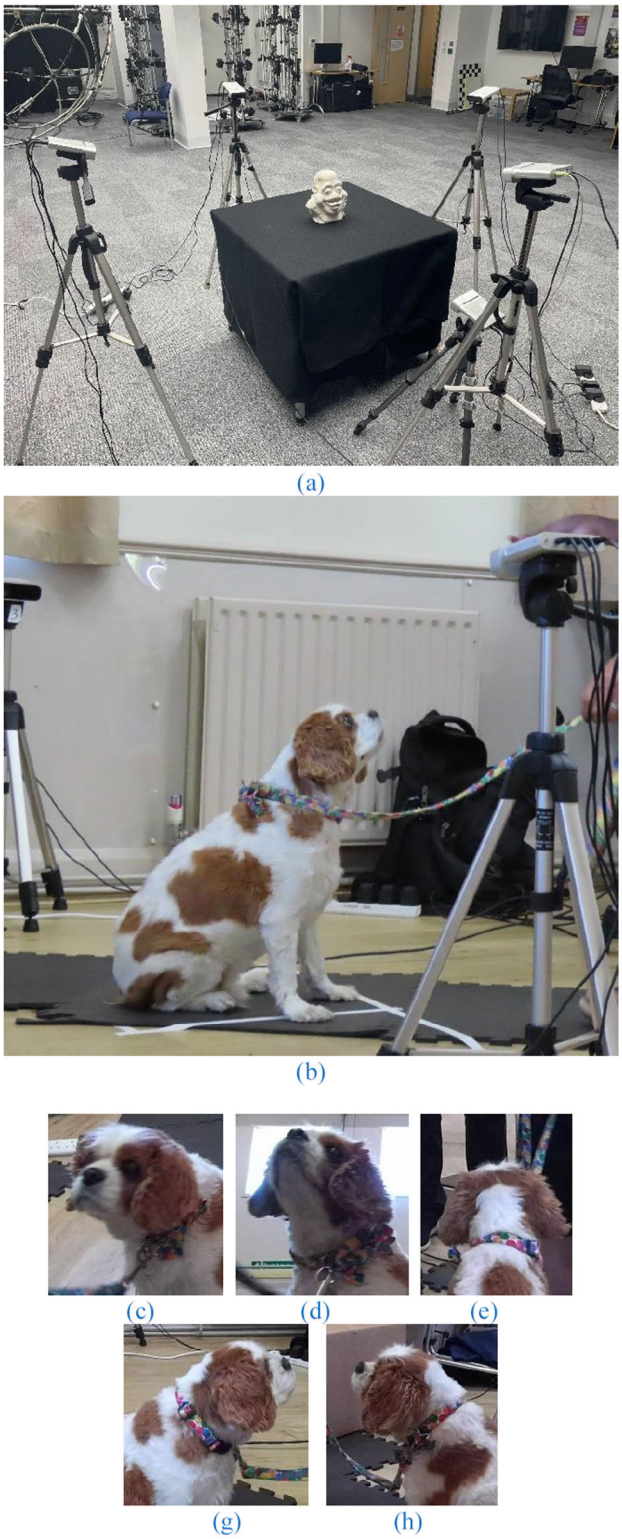

FIGURE 11. Experimental arrangement: (a) Monster head ground truth, (b) live dog during 3D head surface capture, RGB input images used for reconstruction (c) Bottom-front view, (d) Top-front view, (e) Back view, (g) Left view, and (h) Right view.

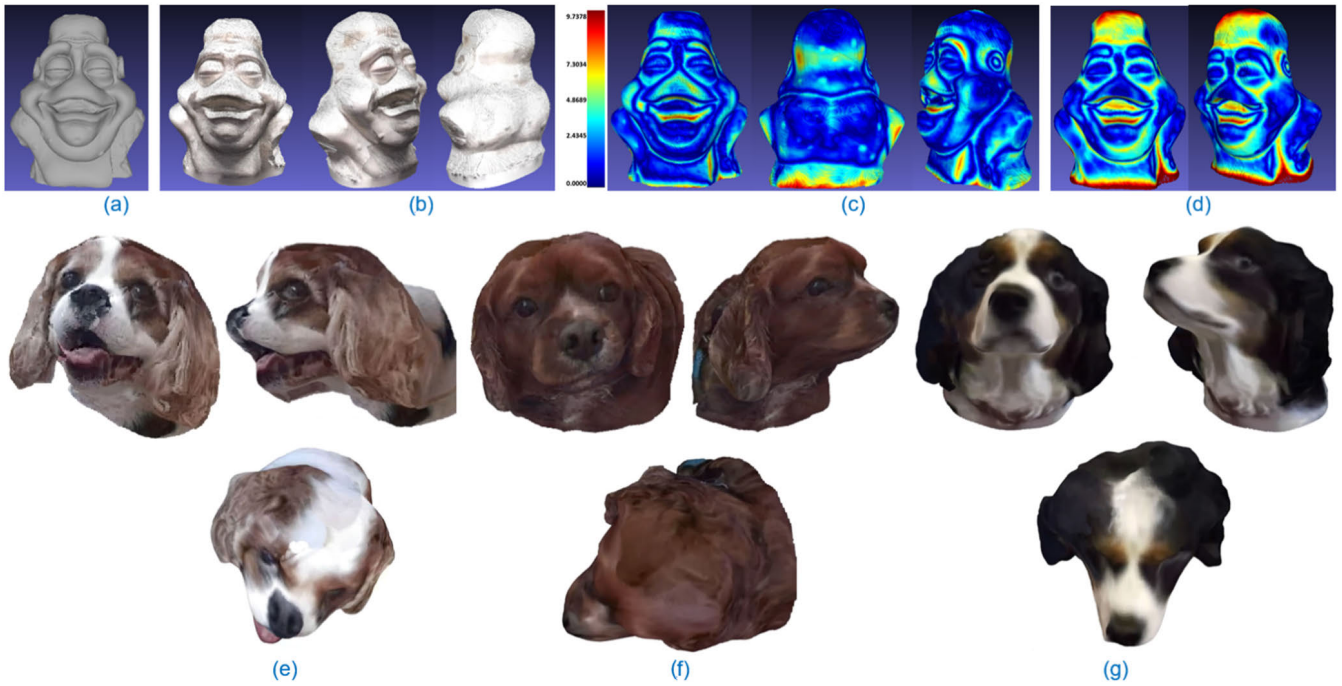

FIGURE 12. Examples of 3D reconstructed models. (a) Monster test object used for quantitative evaluation, (b) 3D reconstructed using proposed method, (c) and (d) Error heatmaps of alignment result from reconstructed model and ground truth using the proposed method and de default calibration parameters (factory parameters), respectively. (e) to (g) PCD of the 3D reconstructed models of three CKCSs. This demonstrates the benefit of the proposed approach compared to using factory settings.

TABLE 1. Comparison between proposed method and exemplar prior work based on the average errors for different virtual lines in calibration object of all cameras for test samples.

Virtual line type	method	Average errors (mm)
central star	Jung et al. [15]	5.0612± 2.2731
	Lindner et al. [16]	3.3090 ± 2.9206
	Proposed method	**2.7076 ± 2.2241**
horizontal	Jung et al. [15]	2.9427 ± 1.9132
	Lindner et al. [16]	3.1640 ± 3.0551
	Proposed method	**1.5449 ± 1.3313**
side diagonal	Jung et al. [15]	4.0268 ± 2.2359
	Lindner et al. [16]	2.6221 ± 2.5086
	Proposed method	**2.1471 ± 1.9343**
central horizontal diagonal	Jung et al. [15]	2.1145 ± 1.2638
	Lindner et al. [16]	2.6615 ± 2.1257
	Proposed method	**2.2267 ± 1.5572**
vertical	Jung et al. [15]	3.8034 ± 1.9169
	Lindner et al. [16]	3.3453 ± 3.3402
	Proposed method	**0.7378 ± 0.6369**
central diagonal	Jung et al. [15]	5.0612 ± 2.2731
	Lindner et al. [16]	3.3090 ± 2.9206
	Proposed method	**2.0963 ± 1.5341**

The main goal of the project is to produce 3D surface models of live dog heads (without sedation or another immobilization). Therefore, a known test object published in [34] has been used as ground truth for evaluating the proposed system. Hausdorff distance was used for measuring the accuracy of the reconstructed model that obtained results are minimum = 0.02 mm, maximum = 9.737849mm, and mean = 1.900918mm with rms = 2.433871. As shown in Fig. 12 (c) most of the difference between the reconstructed model and the ground truth are in prominent part of the Monster face such as eyebrows, Lips, and ears. These are of lesser importance compared to measuring the notch, top of head and neck areas in this application.

Moreover, as shown in Fig. 12 (d) error values for reconstructing by using default calibration parameters (factory parameters) are minimum = 0.1mm, maximum = 20.433292mm, and mean = 2.439069mm with rms = 3.214377. However, as can be seen in Fig. 12 (d) to (f) 3D models of 3D scanned dogs with the proposed system appear qualitatively to be of a high standard, which is attributed to the largely smooth curved surfaces that we wish to reconstruct.

IV. CONCLUSION AND FUTURE WORK

In this paper, we present a 3D scanning setup tailored for animal healthcare monitoring, with potential applications extending to toddler care. Essential criteria for such a system include rapid scanning, high accuracy for medical use, and portability for versatile deployment. Conventional RGB-D scanning systems often fall short on these fronts, with limitations in mobility, depth accuracy, and calibration requirements. Our proposed method aims to address these challenges. Through extensive experiments, we identified the critical role of RGB-D image synchronization in achieving

accurate 3D reconstructions. We observed that oversizing the depth image to match the RGB image reduces its resolution, leading to increased depth prediction errors, particularly at the edges.

To mitigate this issue, we introduce a novel 3D calibration object and propose a depth correction matrix, departing from conventional single or multi-value methods. Furthermore, we employ simulated annealing to minimize the reprojection error during RGB-D calibration, optimizing intrinsic and extrinsic parameters of Azure Kinect cameras dynamically based on their surroundings. The simplicity of our setup facilitates easy installation and usage in diverse uncontrolled environments, requires only a swift (∼10 mins?) calibration of all cameras before scanning begins. This user-friendly approach enhances the applicability of our system to both veterinary and human healthcare scenarios where both CM and SM may be present.

According to the obtained results our method represents a significant advancement in 3D scanning technology, offering enhanced accuracy, speed, and convenience for medical monitoring applications. With its potential to revolutionize healthcare monitoring for animals and toddlers alike, our system holds promise for widespread adoption in clinical and research settings. And finally obtained results show that our bespoke depth correction method can increase the accuracy of the calibration especially in comparison with factory parameters, and finally the proposed 3D calibration object is best choice for calibrating Azure Kinect cameras using for 3D scanning setup. Using a simple, low-cost, 3D-surface scanning and reconstruction system based on five RGB-D sensors combined with a dedicated calibration scheme can provide useful surface capture for dogs at risk of breeding disorders manifest in their physical head presentation. Such an approach may be a useful adjunct or triage approach prior to MRI and CT based investigation. We were successful in obtaining data for our study on CKCS head morphology supporting our claim. Moreover, our future work is to make the 3D reconstruction step currently relies on manual intervention, which we aim to automate. This can be achieved by employing surface-based features to identify common boundaries, thereby removing outliers and refining the final 3D image automatically.

ACKNOWLEDGMENT

The authors wish to thank Dogs Trust Canine Welfare and the Hannah Hasty Memorial Fund for their financial support. We also extend our gratitude to the University of Surrey Veterinary School Clinical Study laboratory technicians, the volunteer dogs, and their owners. In addition, we offer our sincere thanks to the Companion Cavalier Club, Cavalier Matters Charity, and Bliss Cavalier Rescue.

REFERENCES

[1] G. Kurillo, E. Hemingway, M.-L. Cheng, and L. Cheng, "Evaluating the accuracy of the Azure Kinect and Kinect v2," *Sensors*, vol. 22, no. 7, p. 2469, Mar. 2022.

[2] M. Tölgyessy, M. Dekan, L. Chovanec, and P. Hubinský, "Evaluation of the Azure Kinect and its comparison to Kinect V1 and Kinect V2," *Sensors*, vol. 21, no. 2, p. 413, Jan. 2021.

[3] A. Gunatilake, L. Piyathilaka, A. Tran, V. K. Vishwanathan, K. Thiyagarajan, and S. Kodagoda, "Stereo vision combined with laser profiling for mapping of pipeline internal defects," *IEEE Sensors J.*, vol. 21, no. 10, pp. 11926–11934, May 2021.

[4] Y. Guo, F. Deligianni, X. Gu, and G.-Z. Yang, "3-D canonical pose estimation and abnormal gait recognition with a single RGB-D camera," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3617–3624, Oct. 2019.

[5] C.-C. Sun, Y.-H. Wang, and M.-H. Sheu, "Fast motion object detection algorithm using complementary depth image on an RGB-D camera," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5728–5734, Sep. 2017.

[6] S. Shuai, Y. Ling, L. Shihao, Z. Haojie, T. Xuhong, L. Caixing, S. Aidong, and L. Hanxing, "Research on 3D surface reconstruction and body size measurement of pigs based on multi-view RGB-D cameras," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105543.

[7] W. C. Ramos, K. H. E. Beange, and R. B. Graham, "Concurrent validity of a custom computer vision algorithm for measuring lumbar spine motion from RGB-D camera depth data," *Med. Eng. Phys.*, vol. 96, pp. 22–28, Oct. 2021.

[8] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[9] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of Kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, Feb. 2012.

[10] J. Jung, Y. Jeong, J. Park, H. Ha, J. D. Kim, and I.-S. Kweon, "A novel 2.5D pattern for extrinsic calibration of tof and camera fusion system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 3290–3296.

[11] D. Herrera C., J. Kannala, and J. Heikkilä, "Joint depth and color camera calibration with distortion correction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2058–2064, Oct. 2012.

[12] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, Jun. 2014.

[13] F. Rameau, J. Park, O. Bailo, and I. S. Kweon, "MC-calib: A generic and robust calibration toolbox for multi-camera systems," *Comput. Vis. Image Understand.*, vol. 217, Mar. 2022, Art. no. 103353.

[14] W. Darwish, W. Li, S. Tang, B. Wu, and W. Chen, "A robust calibration method for consumer grade RGB-D sensors for precise indoor reconstruction," *IEEE Access*, vol. 7, pp. 8824–8833, 2019.

[15] B.-S. Park, W. Kim, J.-K. Kim, D.-W. Kim, and Y.-H. Seo, "Iterative extrinsic calibration using virtual viewpoint for 3D reconstruction," *Signal Process.*, vol. 197, Aug. 2022, Art. no. 108535.

[16] A. N. Staranowicz, G. R. Brown, F. Morbidi, and G.-L. Mariottini, "Practical and accurate calibration of RGB-D cameras using spheres," *Comput. Vis. Image Understand.*, vol. 137, pp. 102–114, Aug. 2015.

[17] J. Jung, J.-Y. Lee, Y. Jeong, and I. S. Kweon, "Time-of-flight sensor calibration for a color and depth camera pair," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1501–1513, Jul. 2015.

[18] M. Lindner, I. Schiller, A. Kolb, and R. Koch, "Time-of-flight sensor calibration for accurate range sensing," *Comput. Vis. Image Understand.*, vol. 114, no. 12, pp. 1318–1328, Dec. 2010.

[19] F. Basso, E. Menegatti, and A. Pretto, "Robust intrinsic and extrinsic calibration of RGB-D cameras," *IEEE Trans. Robot.*, vol. 34, no. 5, pp. 1315–1332, Oct. 2018.

[20] G. Chen, G. Cui, Z. Jin, F. Wu, and X. Chen, "Accurate intrinsic and extrinsic calibration of RGB-D cameras with GP-based depth correction," *IEEE Sensors J.*, vol. 19, no. 7, pp. 2685–2694, Apr. 2019.

[21] S. Jacob, V. G. Menon, and S. Joseph, "Depth information enhancement using block matching and image pyramiding stereo vision enabled RGB-D sensor," *IEEE Sensors J.*, vol. 20, no. 10, pp. 5406–5414, May 2020.

[22] M. Cao, L. Zheng, and X. Liu, "Single view 3D reconstruction based on improved RGB-D image," *IEEE Sensors J.*, vol. 20, no. 20, pp. 12049–12056, Oct. 2020.

[23] I. El Bouazzaoui, S. A. R. Florez, and A. El Ouardi, "Enhancing RGB-D SLAM performances considering sensor specifications for indoor localization," *IEEE Sensors J.*, vol. 22, no. 6, pp. 4970–4977, Mar. 2022.

[24] L. Ruotsalainen, A. Morrison, M. Mäkelä, J. Rantanen, and N. Sokolova, "Improving computer vision-based perception for collaborative indoor navigation," *IEEE Sensors J.*, vol. 22, no. 6, pp. 4816–4826, Mar. 2022.

[25] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.* Graz, Austria: Springer, 2006, pp. 404–417.
[26] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.* Graz, Austria: Springer, 2006, pp. 430–443.
[27] P. J. Besl and N. D. Mckay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[28] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
[29] W. Lu, G. Wan, Y. Zhou, X. Fu, P. Yuan, and S. Song, "DeepVCP: An end-to-end deep neural network for point cloud registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 12–21.
[30] Y. Wang and J. Solomon, "Deep closest point: Learning representations for point cloud registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 3522–3531.
[31] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[32] *RS PRO Silver Aluminum Profile Strut,* 40 × 40 *Mm, 8mm Groove, 1000 mm Length, RS PRO Silver Aluminum Profile Strut,* 40 × 40 *mm, 8 mm Groove, 1000 mm Length | RS*. Accessed: 2023. [Online]. Available: https://www.rs-online.com
[33] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
[34] N. Roubtsova and J.-Y. Guillemaut, "Colour Helmholtz stereopsis for reconstruction of dynamic scenes with arbitrary unknown reflectance," *Int. J. Comput. Vis.*, vol. 124, no. 1, pp. 18–48, Aug. 2017.

MEHRAN TAGHIPOUR-GORJIKOLAIE received the B.Sc. degree in electrical engineering from the University of Mazandaran, Iran, in 2008, and the M.Sc. (Hons.) and Ph.D. degrees in electronic engineering from the University of Birjand, Iran, in 2011 and 2016, respectively. He was a Lecturer (Assistant Professor) with the University of Birjand, from 2016 to 2021, then he continues his research as a Research Fellow of application AI, ML, and CV in healthcare and wellbeing with CVSSP, University of Surrey, up to 2023. He is currently a Postdoctoral Research Fellow with London South Bank University, involved in application of AI in human healthcare. His research interests include the application of machine learning, artificial intelligence, computer vision, and optimization algorithms.

MARCO VOLINO received the M.Eng. degree in electronic engineering from the University of Surrey, in 2011, and the Ph.D. degree in computer vision and graphics from the Centre for Vision, Speech and Signal Processing, in 2016. He is currently a Lecturer of computer vision and graphics with the University of Surrey. His research interests include the intersection of computer vision, computer graphics, and machine learning to enable digital content production in the creative industries.

CLARE RUSBRIDGE received Bachelor of Veterinary Medicine degree in surgery from Glasgow University, in 1991, the dual Diplomate degree in veterinary neurology and neurosurgery from European College of Veterinary Neurology and the Royal Veterinary College, and the Ph.D. degree from Utrecht University, in 2007. She became a RCVS Specialist of veterinary neurology, in 1997. She is currently a Professor of veterinary neurology with the University of Surrey, and a Senior Neurologist with Wear Referrals. She was made a fellow of the Royal College of Veterinary Surgeons (meritorious contribution to knowledge), in 2016. She has spent over 25 years researching Chiari malformation, syringomyelia, and maladaptive pain. As a result of this work, she received the JA Wight Memorial Award in 2011, the RCVS Impact Award in 2022, and the Pet Plan Charitable Trust Scientific Award for 2023.

KEVIN WELLS received the B.S. degree (Hons.) in physics and microelectronics from Kingston University London, and the Ph.D. degree in physics from Brunel University London. He was previously with the Joint Department of Physics, Institute of Cancer Research/Royal Marsden Hospital, University College London and the University of Bath. He is currently a Professor of AI in animal and human healthcare. He is the Team Leader of the Medical Imaging and Healthcare Team, Centre for Vision, Speech and Signal Processing, University of Surrey. He also leads the DataHub Team at Surrey, concerned with data analytics, insights, and AI associated with enhancing the healthcare of companion animals and livestock. His research interests include the application of machine learning, artificial intelligence, computer vision in diagnostic imaging, and healthcare in human and animals.