

Calibration of an Integrated Robotic Multimodal Range Scanner

Christopher S. Yang, *Member, IEEE*, Phillip Curtis, *Student Member, IEEE*, and Pierre Payeur, *Member, IEEE*

Abstract—Collecting dense range measurements in uncontrolled environments is a challenging problem, as the quality of the measurements is highly dependent on the lighting conditions and the texture of the target surfaces. This dependence affects the registration and data-fusion processes and, consequently, degrades the accuracy of the surface or occupancy models that are computed from the range measurements. Typical approaches to address this issue have concentrated on improving a specific type of range sensor. On the other hand, the overall quality of the sensing can also be enhanced through the development of a mechanism that combines the various range-sensing technologies to form a multimodal range sensor. The resulting problem of the merging datasets can then be solved in two ways: system calibration of the multimodal sensor or data fitting of all the datasets into a single model, of which the latter is more widely implemented. The lack of multimodal-system calibration approaches is due to their complicated and lengthy nature, where individual calibration procedures must be applied to each subsystem and then applied between the subsystems of the multimodal range sensor. This paper proposes a technique to alleviate the problems encountered in a multimodal-system calibration. Straightforward and generic guidelines for the calibration are defined and applied to an in-house integrated multimodal system built from a laser-range-finder system, two structured-lighting systems, and a stereovision system. The system's intracalibration and intercalibration processes are detailed. Reconstructed renderings of the datasets collected with the calibrated multimodal range sensor, without the use of data fitting, are also presented. From these results, the potential benefits of multimodal calibration over the computationally intensive data-fitting methods and the advantages of merging the subsystem's strengths to complement other subsystem's weaknesses are put in evidence.

Index Terms—Active vision, calibration, laser range finder, multimodal scanning, range sensing, robotic sensing systems, stereovision, structured lighting.

I. INTRODUCTION

THE INTRODUCTION of increasingly smaller and portable range-sensing technologies has opened the door for researchers to explore many different environments in new ways, like the Mars Pathfinder, which provided the first stereoscopic views of the Martian landscape in 1997 [1]. However, with scalability of the technology and the reduced manufacturing costs of range-sensing devices, there remains a speculation of their efficiency and reliability. Today's consumers can pur-

chase newly improved high-resolution cameras for a fraction of the cost of their predecessors, but software implementations that handle the disparity techniques have stood at a standstill with little improvement.

The introduction of active sensing techniques such as laser range scanning [2], [3] and structured lighting [4] has provided different approaches to the range sensing, which resolve the conditions where a classical stereovision cannot perceive depth. For example, a light pattern projected on a scene can be used to detect the depth of the objects in obscure and darkened environments, whereas stereovision is dependent upon the illumination of the environment and the texturing of the scene [5], [6]. Laser range sensors, which lately have been decreasing in power consumption and size, became the forefront in depth-sensing technology, providing far more accurate depth estimation than the previous techniques. However, the greatest drawback of most of the current active sensors is their ability to detect depth only over a single plane or a sparse detection grid of a nonreflective object located within close proximity of the scanner. This implies the repetitive process of moving the active scanner to various and strategic poses to complete a full-scene scan, unlike the stereovision where one sampling is sufficient.

With each technology providing advantages and drawbacks in their respective domains, a possible solution consists of the combination of the efforts of various range-sensing techniques to create a multimodal or a joint range-sensing mechanism that would provide the different depth perspectives of a scene from a common viewpoint. Although this approach is very promising in providing optimal-depth data, the question of registration between the range sensors remains a critical issue in ensuring consistency between the measurements.

The iterative and tedious manual processes, inherent in using multiple original-equipment-manufacturer (OEM) systems, lead to errors in the registration and a lack of repeatability that must be avoided. This implies the necessity of a fully integrated sensing system, which merges the various OEM solutions, to ensure that each subsystem operates in tandem with each other subsystem [7]. Such an integrated solution can only be efficient if robust intracalibration and intercalibration procedures are defined.

In this paper, an original approach that achieves the automated calibration within the multimodal scanning systems is proposed. The robotic integration of a prototype of a multimodal range sensor is also discussed. Finally, the quality of the datasets and the validation of the calibration procedure are analyzed with the experimental range measurements collected using the prototype.

Manuscript received June 15, 2005; revised April 21, 2006.

The authors are with the Vision, Imaging, Video, and Autonomous Systems Research Laboratory, School of Information Technology and Engineering, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: csyang@site.uottawa.ca; pc Curtis@site.uottawa.ca; ppayeur@site.uottawa.ca).

Digital Object Identifier 10.1109/TIM.2006.876410

II. BACKGROUND AND CONCEPTS

The idea of using two or more range-sensing technologies is not a new concept to the field of robot vision. Although the idealism is to perfect the range-sensing technique of an individual system such that the complexity of the calibration between subsystems and the introduction of additional errors are avoided, the concept of multimodalities is becoming more popular in the field. The term “multimodal” has not been widely used; however, different implementations of multimodal systems have been built to integrate range-sensing datasets in order to improve the modeling of a scanned region.

From a high-level view, multimodal systems can be defined by their multiple and diverse modes of range sensing used to perceive the scenery. Systems that use multiple yet identical modes are not considered multimodal since only a single mode is used. With this in mind, the coined definitions of active and passive range-sensing systems [8], [9] can be extended for the purpose of the multimodal systems into the homogeneous and heterogeneous sensing systems. A homogeneous sensing system is defined as the application of range-sensing technologies that are built from all active or all passive subsystems. Likewise, heterogeneous sensing systems use both active and passive subsystems in tandem.

A successful example of a multimodal homogeneous range-sensing system is the application of two active range-sensing systems: a laser range finder and a sonar/acoustic sensing system, as proposed in [10] and [11]. With these two methods, the laser range finder and the sonar system sample separately without knowledge of each other’s extrinsic locations. Once the individual scans have been completed by each subsystem, the two datasets are merged together to provide a single map of the environment. The success of this multimodal system is dependent upon the environment in which the system operates. For example, if the environment is simply a maze where walls are the only objects, the system performs admirably. However, in a complex environment where there are objects of varying height, there is no mechanism that correlates the measurement of the laser range finder with that of the sonar system.

A common example of a multimodal heterogeneous range-sensing system is that of the omnidirectional stereo and a rotating laser-range-finder system, as proposed by Miura *et al.* in [12]. In these systems, a passive sensor, omnidirectional stereo, is merged with the active laser range finder. The complexity of merging the datasets of both range-sensing technologies is clearly defined by Miura *et al.*, who, unlike [10] and [11], outline the dilemma of the different possible perceptions of an object. As a solution, Miura *et al.* propose the use of probabilistic grids to aid in classifying each subsystem based upon their strengths, weaknesses, and limitations, which are all important factors when merging the datasets.

Inspired by these various approaches, a different heterogeneous sensor is introduced in this paper, which combines three active range-sensing systems (one laser-range-finder system and two structured-lighting systems) and one passive range-sensing system (stereovision system). It is expected that a combination of these three common range-sensing technologies

will provide a high robustness in uncontrolled environments. However, as a result of the multiple separate data sources, the problem of the calibration needs to be solved in order to correctly fuse the data.

Although there are many different approaches to system calibration for both the intrinsic and extrinsic properties, the classical technique of Tsai’s camera-calibration model for stereovision calibration [13], [14], a refined version of Chen and Kak’s structured-lighting-subsystem calibration [15], and the Pless and Zhang’s closed-form solution for the camera to laser-range-finder relationship [16] were selected. These methods are combined in an original way in order to minimize the manipulation and the data collection.

Chen and Kak’s model of a structured-lighting-system calibration could be replaced by other well-known calibration methods such as the one noticed by Trucco and Fisher in [17]. However, Chen and Kak provide the interesting idea of using a simple calibration target, which relies upon the movement of a robot end effector carrying the laser projector and camera. The use of the two-dimensional (2-D) projectivity theorems to derive a transformation matrix that converts the detected structured light emitted on defined edges in the world space provides a simpler calibration approach that can be easily used in an automated calibration.

The closed-form solution defined by Pless and Zhang to relate a laser range finder and camera system [16] is appropriate for the intersubsystem calibration between the laser range finder and the stereo system. But instead of using a planar pattern placed in different poses with respect to the camera and laser range finder, the same calibration target proposed by Chen and Kak is exercised such that the stereo system can detect the defined edges of a known object in the world space.

III. MULTIMODAL-SENSOR DESIGN

The proposed multimodal range-sensing system consists of four subsystems using three range-sensing techniques. The first subsystem is a laser range finder, which provides 2-D data along a scan line marked by a visible red line projected on the scene. The second subsystem is a stereovision system built from two charge-coupled device (CCD) cameras mounted in close proximity to the laser range finder. The third and fourth subsystems are structured-lighting systems that use the left and right stereovision cameras independently to detect the projected structured light emitted from the laser range finder.

A quick assessment of the range-sensing technologies indicates that in structured lighting the cameras must be placed in a noncoplanar position to the laser-range-finder structured light. Structured lighting is highly dependent upon the location of the laser pattern detected in the image. If the cameras were placed coplanar to the projected line stripe, the detected line in the system would remain fixed regardless of the depth of the object in the path of the structured light. This would make it impossible for the structured-lighting system to extract the range data. For the stereovision system, the cameras must have an appropriate baseline, selected such that the objects are visible from both cameras, and the disparity is detectable between the acquired images. Fig. 1 shows the selected configuration where



Fig. 1. VRex stereovision system mounted on a Jupiter laser range finder, which is attached to the CRS F3T robot end effector.

the stereovision cameras are mounted on top of the laser sensor with a bracket that permits the adjustment of the baseline.

Selecting a global frame of reference with respect to which all of the three-dimensional (3-D) datasets are defined is a critical aspect of the design. For the proposed multimodal range sensor, three of the four range sensors can conveniently provide the depth information from the laser-projector optical center. Since structured lighting is highly dependent on the visible light stripe, it is appropriate to directly represent the structured-lighting depth from the structured-light projector frame of reference. This frame of reference is also the laser-range-finder frame of reference, eliminating the need for the intersubsystem laser range finder and structured-lighting calibration.

Stereovision, on the other hand, provides the depth information with respect to a single camera optical center. Therefore, it is necessary to build an additional calibration mechanism to translate the stereovision depth to the laser-range-finder perspectives.

IV. MULTIMODAL CALIBRATION

However, the largest difficulty of using a multimodal system is the organization required to calibrate each subsystem and interoperating subsystems. A systematic approach is proposed, which reduces the number of transformations to provide a resilient automated calibration mechanism.

For the selected combination of sensing technologies, advantage can be taken from the widely known calibration procedures such as the Tsai's camera-calibration technique [13] or the internal self-calibration functionalities integrated in the device controllers. In the present case, Tsai's approach is used to calibrate individual CCD cameras and to determine intrinsic and extrinsic parameters between the pairs of camera reference frames. Likewise, the Jupiter laser range finder used in the experimental setup has a fixed calibration provided by the manufacturer [18]. The original part of the proposed calibration procedure is related to the estimation of the registration for the dual structured-lighting system and to the extrinsic calibration between the stereovision pair and the laser range finder. These aspects are detailed in the following sections.

A. Structured-Lighting-Subsystems Calibration

Since the stereovision cameras and the structured lighting from the laser range finder are available, the visible trace of the laser on the scene is exploited to create two structured-lighting subsystems: one for each CCD camera. The proposed calibration mechanism for these subsystems is based upon Chen and Kak's structured-lighting-calibration technique [15], where a recovery conversion matrix can be built from a minimal set of four known points on the surface of a given object. The relationship provided by Chen and Kak is given as follows:

$$\rho \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \\ t_{41} & t_{42} & 1 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \quad (1)$$

where the 3-D world depth points are represented as $[x_i, y_i, z_i]$ in homogeneous coordinates, $[u_i, v_i]$ are the perceived structured-light pattern points from the image plane, t_{MN} are the projection-matrix coefficients, and ρ is a variable associated with the scaling factor. Equation (1) can be simplified as

$$\rho \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} \cdot U \quad (2)$$

where $U = [u_i, v_i, 1]^T$ and T_j , $1 \leq j \leq 4$ are the line vectors of the projection matrix.

When ρ is normalized, it results in

$$x_i = \frac{T_1 \cdot U}{T_4 \cdot U} \quad y_i = \frac{T_2 \cdot U}{T_4 \cdot U} \quad z_i = \frac{T_3 \cdot U}{T_4 \cdot U} \quad (3)$$

that can be rewritten as

$$\begin{aligned} T_1 \cdot U - x_i T_4 \cdot U &= 0 \\ T_2 \cdot U - y_i T_4 \cdot U &= 0 \\ T_3 \cdot U - z_i T_4 \cdot U &= 0. \end{aligned} \quad (4)$$

To determine the projection matrix (T_1 , T_2 , T_3 , and T_4), Chen and Kak proposed an acquisition process that requires the projected structured-light pattern to fall on known points. This procedure, which can be automated, also requires that a simple object, the calibration target, is placed in the path of the structured light, such that its edges would form discontinuities in the structured light and would be detectable by the camera system. The main limitation of the Chen and Kak's method is the requirement that the 3-D edge points of the calibration target need to be fully known and characterized by the intersection of planes defined in the world coordinate frame, before being substituted in (4). The structured-lighting system is then positioned at different locations by the movement of the robotic arm away from the calibration target, and edges are extracted from the images such that U in (4) is sampled. A minimal set of four unique coplanar but noncollinear calibration points is required to produce a set of 12 linear equations to solve for the 11 coefficients that define the calibration matrix.

By expanding the normalized version of (1), when $\rho = 1$, a set of linear equations based upon the number of sampled points used for the calibration is generated. The expansion of the first row for N sample points gives

$$\begin{aligned} t_{11} \cdot u_1 + t_{12} \cdot v_1 + t_{13} &= x_1 \\ t_{11} \cdot u_2 + t_{12} \cdot v_2 + t_{13} &= x_2 \\ &\dots\dots \\ t_{11} \cdot u_N + t_{12} \cdot v_N + t_{13} &= x_N. \end{aligned} \quad (5)$$

By substituting (5) and other expanded rows into (4), the end matrix is formed such that

$$Aq = b \quad (6)$$

where the expressions for A , q , and b in (6), which are expanded at the bottom of the page, can be used to solve for q . Theoretically, the minimal number of coplanar but noncollinear calibration points required is four. But in practice the actual number of calibration points should be larger, due to the sampling errors or the difficulties in extracting the segment discontinuities.

In order to eliminate the requirement on the known edge points on the calibration target, an adaptation of the Chen and Kak's procedure is proposed, which takes advantage of the depth data that is perceived by the laser range finder. The calibration target is a triangular-shaped piece of cardboard, with its face placed directly in front of the path of the striped structured light, as shown in Fig. 2. Instead of determining the geometric intersecting planes within the world coordinate frame, these inaccurate manual measurements are replaced with the high-resolution samplings from the laser range finder $[X_i, Y_i, Z_i]$. By extracting the edge points that the structured-light plane forms with the calibration target $[u_i, v_i]$, both the laser range finder and the camera systems can easily detect the matching discontinuities in the line scan. As a result, this adaptive method of calibration requires no *a priori* knowledge of the calibration target and of its position in space.

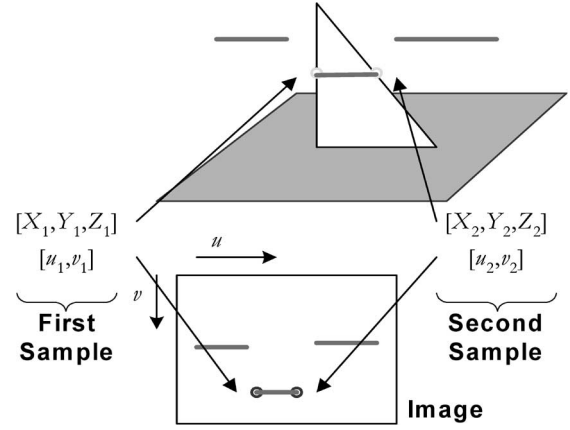


Fig. 2. Structured light and laser-range-finder sampling.

Using a simple triangular-shaped calibration target placed in front of a planar background, the structured-light pattern is extracted by determining the structured-light segments emitted on the target and on the background. The start and the end of the structured-light pattern are valid only when the following pattern is detected: background light segment—foreground light segment—background light segment, as shown in Fig. 2. This guarantees that the laser range finder can distinguish the same pattern, and the discontinuity of the structured light can be determined. Both the left and right structured-light systems (based on both stereovision cameras) simultaneously sample the same feature points produced by the calibration target.

The specific shape of the calibration target is not required to be triangular. Any object that produces two discontinuities of the striped structured light can be used as a calibration target. However, the edges of the object must not be collinear, or the object must be scanned in such a way that the set of sample calibration points is not collinear.

These modifications to the Chen and Kak's approach not only reduce error from human intervention but also simplify the

$$A = \begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -u_1 \cdot x_1 & -v_1 \cdot x_1 \\ u_2 & v_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -u_2 \cdot x_2 & -v_2 \cdot x_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_N & v_N & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -u_N \cdot x_N & -v_N \cdot x_N \\ 0 & 0 & 0 & u_1 & v_1 & 1 & 0 & 0 & 0 & -u_1 \cdot y_1 & -v_1 \cdot y_1 \\ 0 & 0 & 0 & u_2 & v_2 & 1 & 0 & 0 & 0 & -u_2 \cdot y_2 & -v_2 \cdot y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & u_N & v_N & 1 & 0 & 0 & 0 & -u_N \cdot y_N & -v_N \cdot y_N \\ 0 & 0 & 0 & 0 & 0 & 0 & u_1 & v_1 & 1 & -u_1 \cdot z_1 & -v_1 \cdot z_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & u_2 & v_2 & 1 & -u_2 \cdot z_2 & -v_2 \cdot z_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & u_N & v_N & 1 & -u_N \cdot z_N & -v_N \cdot z_N \end{bmatrix}$$

$$q = [t_{11} \quad t_{12} \quad t_{13} \quad t_{21} \quad t_{22} \quad t_{23} \quad t_{31} \quad t_{32} \quad t_{33} \quad t_{41} \quad t_{42}]^T$$

$$b = [x_1 \quad x_2 \quad \dots \quad x_N \quad y_1 \quad y_2 \quad \dots \quad y_N \quad z_1 \quad z_2 \quad \dots \quad z_N]^T$$

procedure by eliminating the need to determine the complicated intersecting planes defining the edges of the calibration target. Full advantage is taken from the multimodality of the sensing system, even during the calibration phase. In addition, this adaptation holds the ability to directly estimate the 3-D coordinates with respect to the laser-range-finder reference frame.

B. Stereovision and Laser-Range-Finder Intercalibration

Only one subsystem-to-subsystem calibration is required to complete the design of the proposed multimodal sensor, which is a calibration between the stereovision to the laser-range-finder reference frames. A calibration mechanism is proposed, which is similar to the one used for the structured-lighting-system calibration and which utilizes the same calibration target.

Unlike the Pless and Zhang's approach for the camera to laser-range-finder calibration, which uses a linear closed-form solution and a regressive nonlinear optimization approach to correlate between the camera frame of reference and the laser range finder [16], the proposed strategy requires only the linear closed-form solution, as advantage can be taken from the structured light. The complexity of the stereovision to laser-range-finder-subsystem calibration is reduced when the structured light is projected on the scene.

To correlate points between the laser-range-finder and stereovision subsystems, the approach discussed in Section IV-A is used, where a triangular-shaped calibration target is placed in the direct path of the laser range finder, structured light, and the stereovision systems. The laser range finder and the stereovision systems detect the discontinuities in the structured-light pattern. The points located at the discontinuities in the image planes facilitate the matching required to generate a 3-D point in the stereovision subsystem's reference frame. The discontinuities from the stereovision reference frame are then correlated with the discontinuities from the laser-range-finder reference frame to produce an estimate of the calibration between the two subsystems.

The following standard transformation equation allows relating the stereovision reference frame to the laser-range-finder reference frame as long as their extrinsic properties remain unchanged, which is ensured by the fixed assembly of the sensor's components on the robot's end effector. A 3-D coordinate observed by the laser range finder and the stereovision subsystem is denoted, respectively, by P_{LRF} and P_{SV} . The stereovision system is related to the laser range finder by a rotational matrix and a translation vector denoted by R_{S2L} and T , respectively, such that

$$P_{\text{LRF}} = R_{\text{S2L}}P_{\text{SV}} - T. \quad (7)$$

Since the laser-range-finder depth perception points are within the field of view of the stereovision system, the equation that relates the laser range finder to the stereovision system can be rewritten as

$$P_{\text{SV}} = R_{\text{L2S}}(P_{\text{LRF}} + T) \quad (8)$$

where $R_{\text{L2S}} = R_{\text{S2L}}^{-1}$.

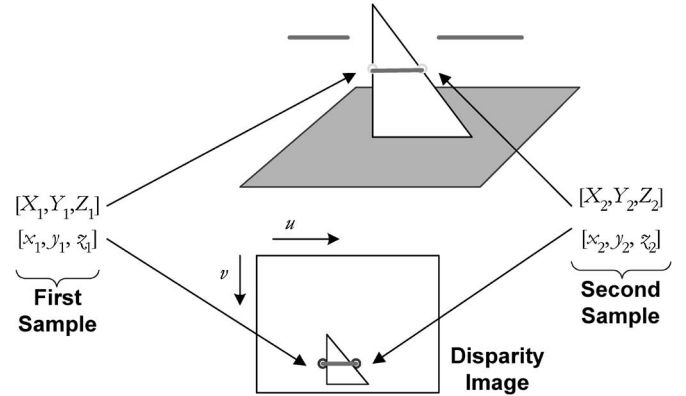


Fig. 3. Intersubsystem calibration—stereovision and laser range finder.

Given the fact that the laser range finder scans a single line, the laser-range-finder depth values are 2-D, such that all the points are found along a common plane $Y = 0$. Therefore, using a compact notation, P_{LRF} can be redefined to encode all the sampled depth values, as $P_{\text{LRF}} = [X, Z, 1]^T$ in the homogeneous coordinates. Equation (8) becomes

$$P_{\text{SV}} = R_{\text{L2S}} \begin{pmatrix} 1 & 0 & \\ 0 & 0 & T \\ 0 & 1 & \end{pmatrix} P_{\text{LRF}} \quad (9)$$

where $T = [t_X, t_Y, t_Z]^T$ and R_{L2S} is a (3×3) rotational matrix. Solving (9) is then simplified to the identification of the calibration matrix M between the stereovision pair and the laser range finder

$$P_{\text{SV}} = M \cdot P_{\text{LRF}}. \quad (10)$$

The same procedure for extracting the correlated points in structured lighting is used to calibrate the stereovision and laser-range-finder subsystems. The structured-light discontinuities within a single foreground (made by the calibration target) and a background scene, as shown in Fig. 3, are used to create a disparity image for the stereovision system. Using a minimal set of three correlated sample points from the disparity image, the relationship defined in (10) can be solved. The calibration matrix M is then inverted and applied to transform the stereovision 3-D points into the laser-range-finder reference frame, to achieve a unified dataset.

C. Calibration and System Overview

The entire calibration process of each multimodal system can be completed within a single instance, given that the structured lighting and the stereovision to laser-range-finder calibrations both use the same mechanism and reuse the sample points for either system. Fig. 4 illustrates the collection of the calibration samples, with the laser stripe intersecting the triangular calibration target in front of a planar background.

Stereovision calibration can also be integrated within the same calibration space by adding a calibration pattern on the triangular target used by the structured-lighting system. The only requirement for the calibration procedure to succeed is to have the multimodal system capture the sampling points after

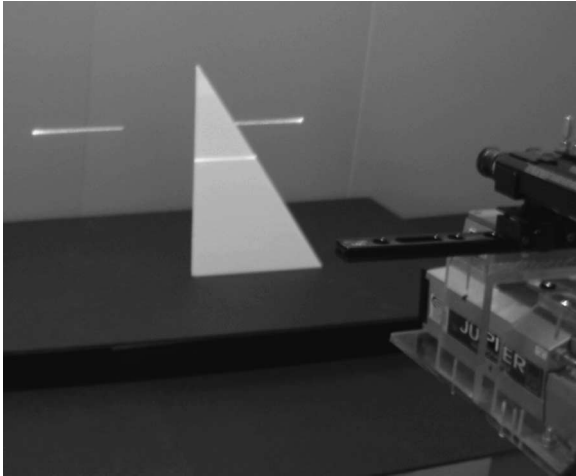


Fig. 4. Sample-points collection for the multimodal-system calibration.

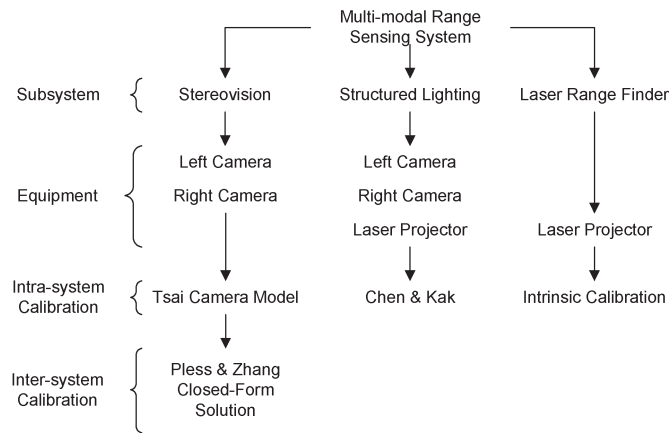


Fig. 5. Multimodal system's structural overview.

the different world locations. Thus, a single calibration target can completely calibrate the entire multimodal system. The chart in Fig. 5 characterizes the multimodal system currently implemented, the subsystems used, the equipment, and the intrasystem and intersystem calibrations required.

V. INTEGRATED ROBOTIC SENSOR

The multimodal sensing system is comprised of a robotic arm, a laser range finder, and a stereovision-camera pair integrated to a single computer via their respective controllers, as shown in Fig. 6. This setup allows for the flexibility of data acquisition as a result of the maneuverability of the robotic arm. The robot used in this design is a CRS F3T serial manipulator with six revolute joints mounted upon a 2-m track, providing a seventh degree of freedom. A CRS C500C digital programmable controller is connected to the seven degree-of-freedom robot and communicates via an in-house driver through an RS-232 asynchronous link. The driver interface, built from the RAPL-3 language, provides an interface to a higher application software by translating the requests for translations and rotations and has the ability to simultaneously control all the robotic joints and end effectors.

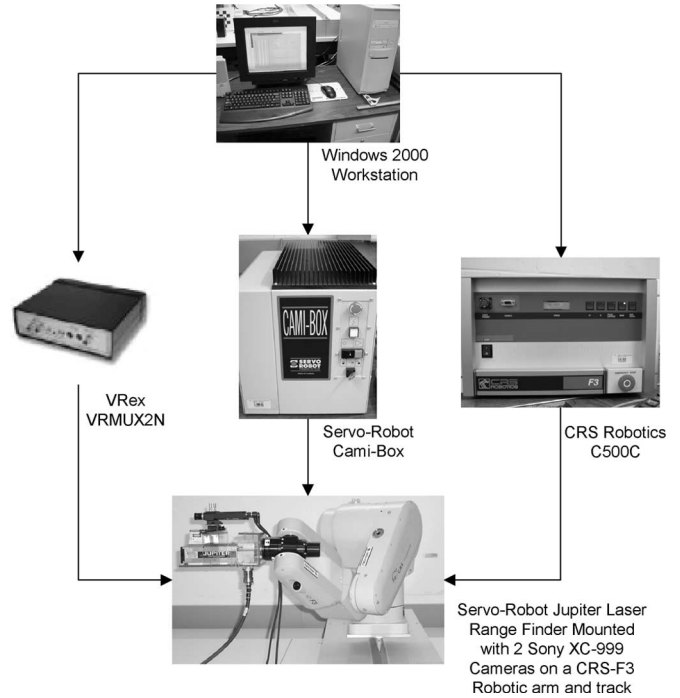


Fig. 6. Diagram of the integrated sensing system.

Attached to the end effector of the F3T manipulator is an in-house constructed Plexiglas chassis housing a Jupiter laser range finder, which is manufactured by Servo-Robot Inc. The Jupiter laser range finder is capable of acquiring a maximum of 512 sampling points per scan through the emission of a single visible red-line-striped structured-light pattern onto the scene using a 150-mW laser emitter. The Servo-Robot Cami-Box is used to control the Jupiter laser line scanner and offers an RS-232 asynchronous link to an external interface.

The Jupiter laser range sensor was selected for this design due to its relatively large field of view and its ability to provide high resolution and precise scans of the objects between 30 cm to 1 m. In addition to its functional capabilities, the Jupiter laser range finder has a compact construction and a lightweight design, whose payload is favorable to the manipulator.

Mounted on top of the Plexiglas chassis is a metal bracket holding the stereo cameras and allowing the baseline to be manually adjusted. A pair of Sony XC-999 CCD cameras, which is the key component of the VRex CAM-3000C product, is used as the stereovision system. To synchronize the stereovision-image acquisition, the use of the VRex VRMUX2N is coupled to a Matrox Orion video card permitting real-time image sensing. Given that the robotic arm is restricted by the payload of its end effector, the Sony XC-999 CCD cameras, which are lightweight and require little power to operate, can be successfully integrated into the range-sensing system.

The camera control, which is accomplished by the VRex VRMUX2N unit, ensures simultaneous stereo-image acquisition within a period of an internal or provided external-clock source. This eliminates the need of constructing software to perform the synchronization between individual cameras and eases the processing required for automation purposes. To reduce the equipment and overhead costs, the VRex VRMUX2N

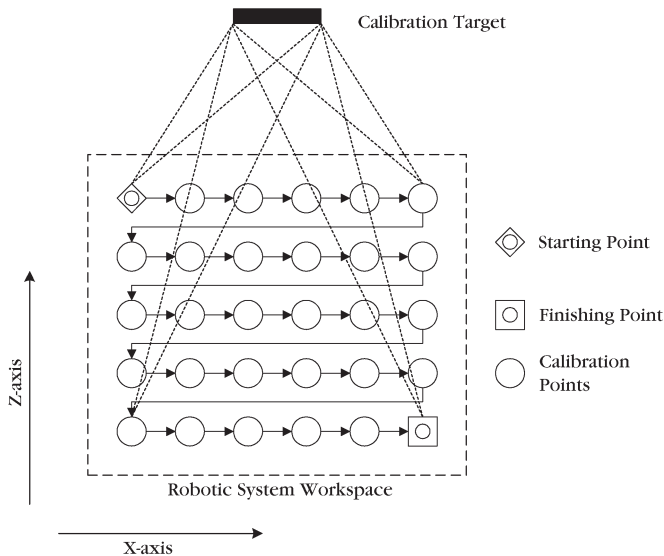


Fig. 7. Automated calibration path and sampling positions.

provides a multiplexed single-interlaced video feed of both stereovision-camera signals. For video and image processing, the Matrox Orion video frame grabber deinterlaces the multiplexed feed using its on-board processor, thus, alleviating the resource consumption for the other real-time processes. The deinterlaced processes produce two raw 320 pixel \times 240 pixel-dimensional images.

To operate all the three systems (Jupiter laser range finder, VRex Stereovision System, and CRS Robot), a custom interface was designed to integrate their functionalities. This interface binds into an RAPL-3 application that accepts the destination-joint positions as an input to control the robot arm. A modified inverse-kinematics solution, inspired from [19], for this specific 7-DOFs manipulator was also developed for the robot-controlling module of the application. The module to control the laser range finder relies on the low-level internal commands provided by the manufacturer to interact with the Cami-box controller. The particular interest for this application was the implementation of the commands to enable and disable the laser, to activate the acquisition of the data using the scan command, and to download the information stored in the sensor's controller memory. The matrox-imaging library (MIL) [20] is used to acquire the segregated left and right images from the stereovision system, and the open computer-vision library (OpenCV) [21] provides the necessary intrinsic camera calibration and stereovision-system disparity algorithms.

Using the high-precision CRS-F3T robotic system, the multimodal end effector is placed in various positions within the workspace or a designated calibration area. As a result, the automated calibration process of the multimodal system is then properly controlled. A series of positions are programmed to the CRS-F3T robotic system, which in turn define the path positions where the multimodal system samples the calibration points. An example of such a sampling path consists of positioning the laser range finder at a constant altitude and moving to various latitude and longitude positions, as shown in Fig. 7, to produce a gridlike calibration path.

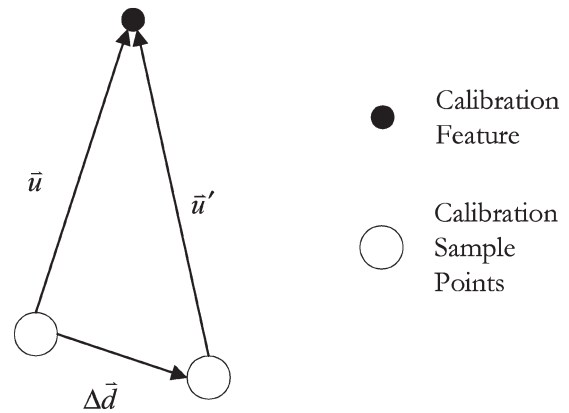


Fig. 8. Relative change from one calibration-sample position to another.

Although the positioning of the robotic end effector is not crucial during the multimodal-system calibration, its precision positioning allows itself to validate the calibration results. As the system samples the key features from the calibration target, a relative validation of each calibration-sample point is performed using the knowledge of the initial-calibration position of the multimodal system to the target and its displacement from the previous sampling grid position, as shown in Fig. 8. At each new calibration position, the distance to a same feature point is denoted as $\vec{u}' = \vec{u} - \Delta\vec{d}$, where \vec{u} is the determined distance of the feature point from the last sampled position, and $\Delta\vec{d}$ is the distance between the last calibration position and the new calibration position.

VI. EXPERIMENTAL RESULTS

To validate the operation of the multimodal range sensor named Multi-Modal System, Stereovision, Structured lighting, and Laser range finder (M²S-SSL), sequences of images were taken by the sensor on various objects, following the intracalibration and intercalibration of the system using the proposed approach. A series of horizontal scan lines was taken in a nontextured environment, after M²S-SSL was subjected to 20 calibrations positions.

To determine the sufficient number of the sampling points required to ensure the convergence of the intracalibration and intercalibration matrices, a series of tests were performed to determine the performance of the subsystems' reconstruction as a function of the number of the calibration samples used. Performance is evaluated as the maximum percentage error on the relative size of the reconstructed objects, which is defined as follows:

Maximum Percentage Error

$$= \max \left[\frac{|L_{\text{actual}} - L_{\text{experimental}}|}{L_{\text{actual}}} \cdot 100\% \right] \quad (11)$$

where $L_{\text{experimental}}$ is the length of the object from a subsystem 3-D reconstruction, and L_{actual} is the length across the *a priori* known object.

Fig. 9 presents the maximum percentage error achieved, respectively, on the reconstruction by a laser active triangulation and stereovision of a fixed *a priori* known object for the

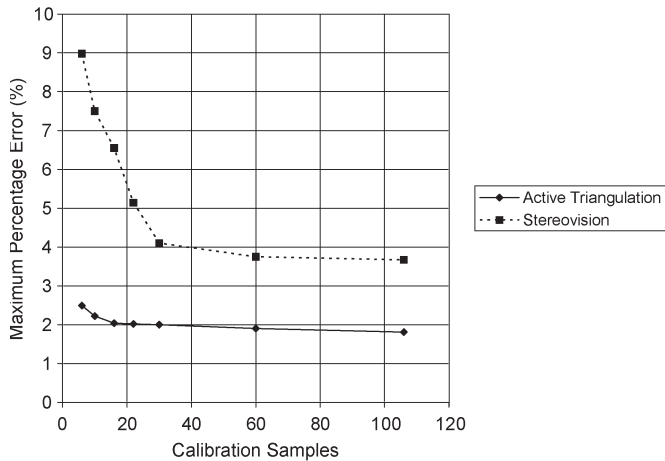


Fig. 9. Reconstruction performance over a number of the calibration samples.

different numbers of calibration samples. From these results, it is easily deduced that 20 calibration samples are sufficient for the calibration matrices to converge.

Figs. 10–13 present the experimental results for a rectilinear vase whose rim is slightly tilted, a chair made of sponge, a tractor toy, and a wooden-house frame, respectively. Each dataset from the structured-lighting systems (left and right), the laser range finder, and the stereovision subsystem have been transformed, such that the views are seen from the same perspective. Scans were performed at a granularity of 5 mm between the successive scan lines of the laser range finder.

From the visual inspection of the datasets, the difference in the precision and accuracy provided by each system is noticeable. For example, the scans resulting from the stereovision system identify the vase but do not clearly distinguish that the vase is positioned such that its corner edge is the closest to the scanner. In the example of the wooden house, only the left side of the house is detected by the disparity algorithm. Yet, overall, the four subsystems can easily detect that an object exists within their field of view, as shown in Fig. 11.

The disparity algorithm used is based upon the Birchfield and Tomasi approach [22], which is included as the only disparity algorithm in the current version of the OpenCV library. Birchfield’s disparity algorithm is unique to the traditional stereovision-disparity techniques in the sense that it uses a rough disparity map to produce crisp discontinuities instead of using the discontinuity knowledge to compute the disparity. A pixel dissimilarity algorithm performed on a per scan line basis is used to determine these discontinuities. Yet, if a complex surface object with a little or no pixel dissimilarity is scanned, the algorithm fails to distinguish the contour features of the object. An example of this limitation can be seen in Fig. 10(b), where the edge contour of the vase is detected as a flat surface. In some scanning scenarios, objects that have little pixel dissimilarity to the background cannot be detected and are treated as a background object, as perceived in Fig. 13(b).

On the other hand, the results obtained after calibration, from both structured-lighting subsystems, consistently provide datasets similar to each other and visually comparable to the

high-accuracy measurements of the laser range finder. These results demonstrate the validity and the accuracy that can be achieved with the proposed calibration scheme. Such an example can be seen from the results presented in Fig. 12(c) on the tractor toy where the left structured-lighting system found more features in the rear wheel than the laser range finder.

Experimentation conducted during the development of the multimodal sensor provided an opportunity to observe the sensitivity of the various modes of acquisition. For example, stereovision was revealed to be highly dependent upon an illuminated environment, which ensures that the scan target is detected and distinguished from the other foreground and background objects. The structured-lighting system appeared to be dependent upon the ability to extract the structured-light pattern emitted onto the scene. A dimmed-light setting was found to be appropriate for this range sensor to perform. These two contrasting environmental conditions provide a supplementary challenge if simultaneous calibration is desired.

For the purpose of development, the proposed multimodal-range-sensor system was placed in an accommodating environment where each subsystem can fully operate and correlate features between each other. Lighting is adjusted such that the stereovision can easily differentiate the textured objects and ensures that the structured-lighting systems accurately extract the laser-light patterns without an over-saturated lighting. To establish what environment lighting is sufficient, a method has been developed that consists of placing the triangular calibration target directly in front of the multimodal system. A simple test procedure is used to determine whether the laser range finder can detect a single foreground object, while the striped structured-light segments can be extracted by the structured-light subsystems, and the stereovision can also compute the disparity from the calibration target.

Both the structured-lighting and stereovision subsystems are dependent upon the precision of the cameras used. If the granularity of the image dimensions is coarse, the accuracy of the extracted features from the image is reduced. As objects are placed further away, there is no distinguishable difference in the image if the object is slightly moved close or further away from the camera. The structured-light pattern that is projected onto the scene would be observed at the same pixel location in the image reference plane. To determine how drastic these errors are, the precision of the calibration is validated by placing an *a priori* object at incremental distances away from the multimodal system. Fig. 14 depicts the surmounting errors from the structured-lighting and stereovision subsystems as *a priori* volume becomes distant from the multimodal system. Note that although the stereovision performs less accurately than the other subsystems, its use provides correlated and complementary data.

The proposed M²S-SSL automated calibration procedure takes approximately 10 min to perform using 20 different calibration-path points distributed on a sampling grid similar to that of Fig. 7, with five rows and four equidistant samples per row. This provides a total of 40 structured-light sampling points (left and right sides of the target) for each subsystem and 40 sampling points for the stereovision and laser-range-finder-intersubsystems calibration.

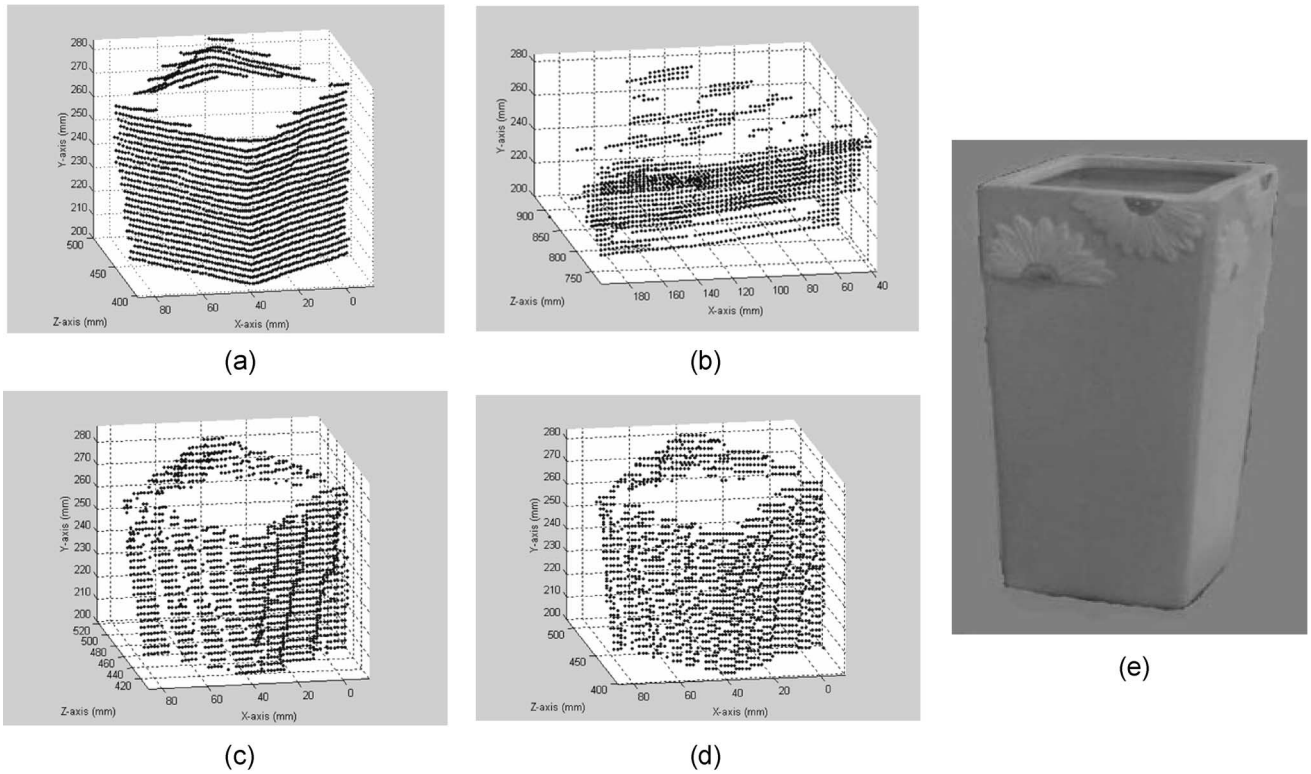


Fig. 10. Point cloud rendering of a rectilinear vase scanned with (a) laser range finder; (b) stereovision; (c) structured lighting using the left camera; (d) structured lighting using the right camera; and (e) picture of the vase.

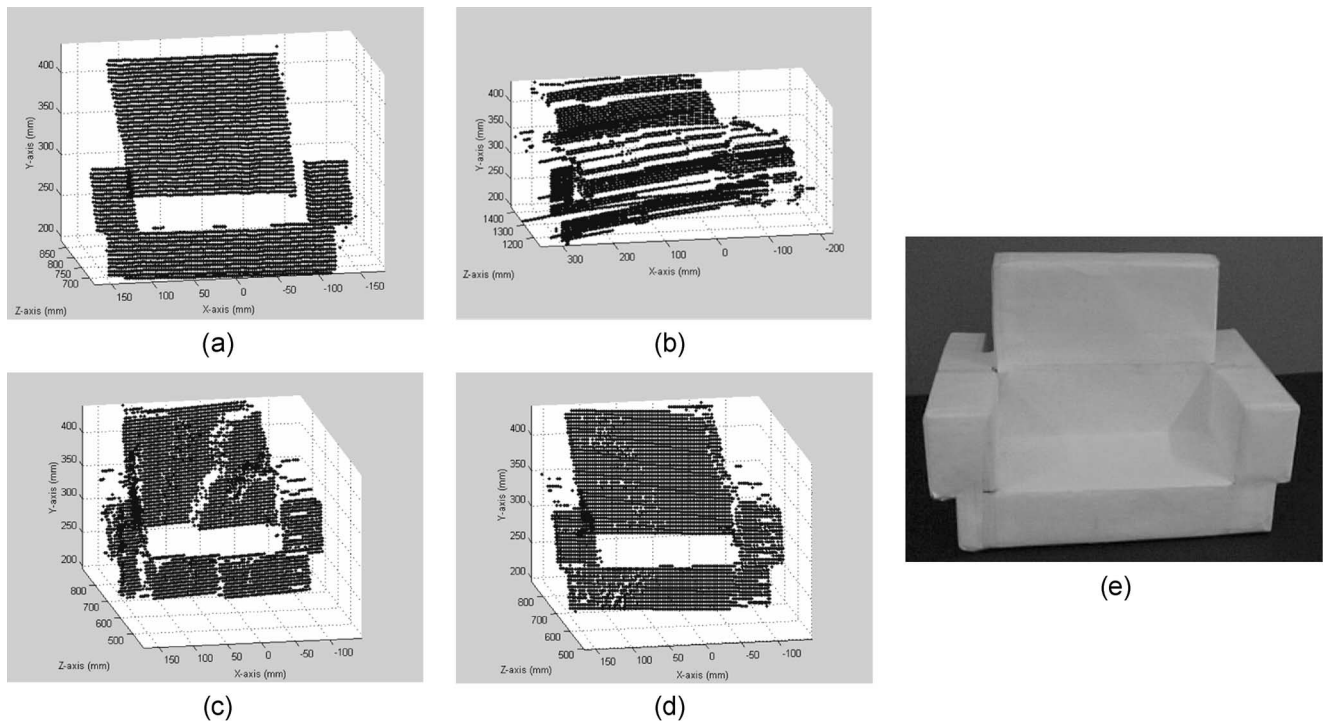


Fig. 11. Point cloud rendering of a sponge chair scanned with (a) laser range finder; (b) stereovision; (c) structured lighting using the left camera; (d) structured lighting using the right camera; and (e) picture of the sponge chair.

Once the calibration phase is completed, the scanning procedure can be performed from any viewpoint, given that the sensor assembly preserves the registration between the sensing

modalities. When using a legacy robotic sensing system that only relied on the laser range finder, the acquisition time of 64 scan lines from three different viewing areas took

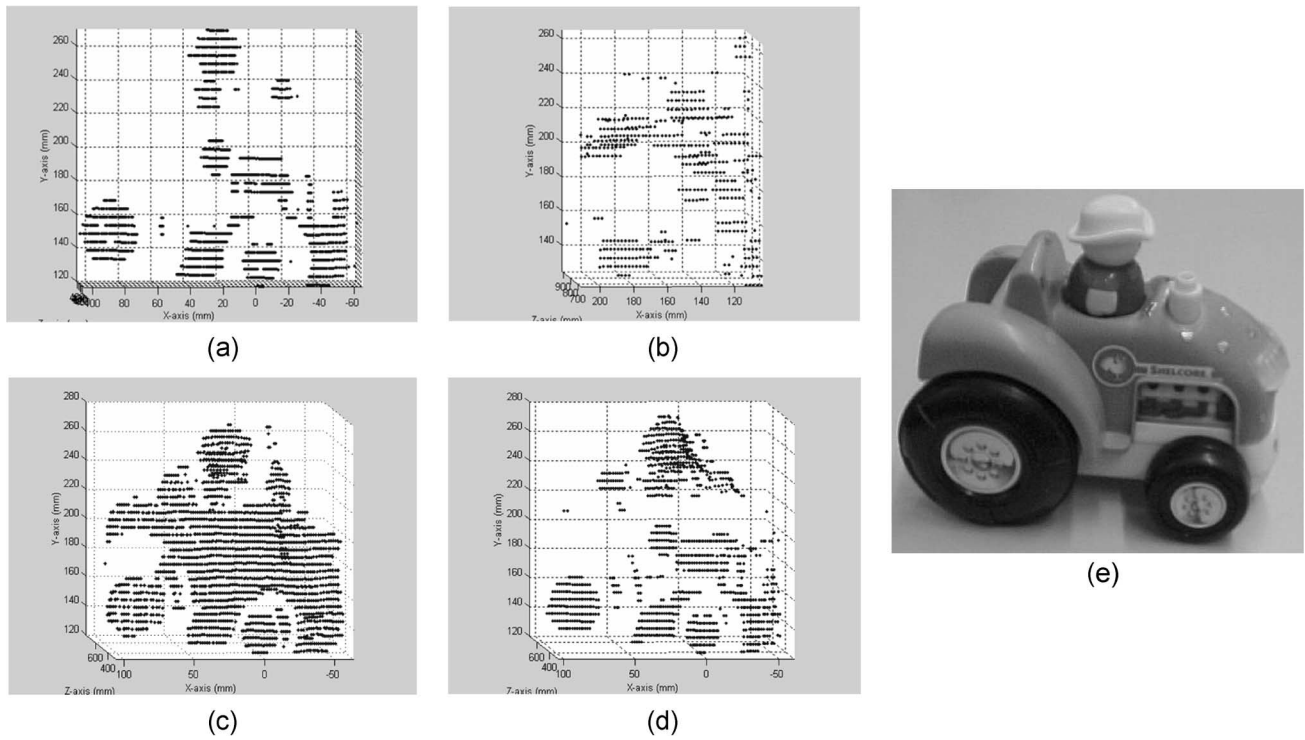


Fig. 12. Point cloud rendering of a tractor toy scanned with (a) laser range finder; (b) stereovision; (c) structured lighting using the left camera; (d) structured lighting using the right camera; and (e) picture of the tractor toy.

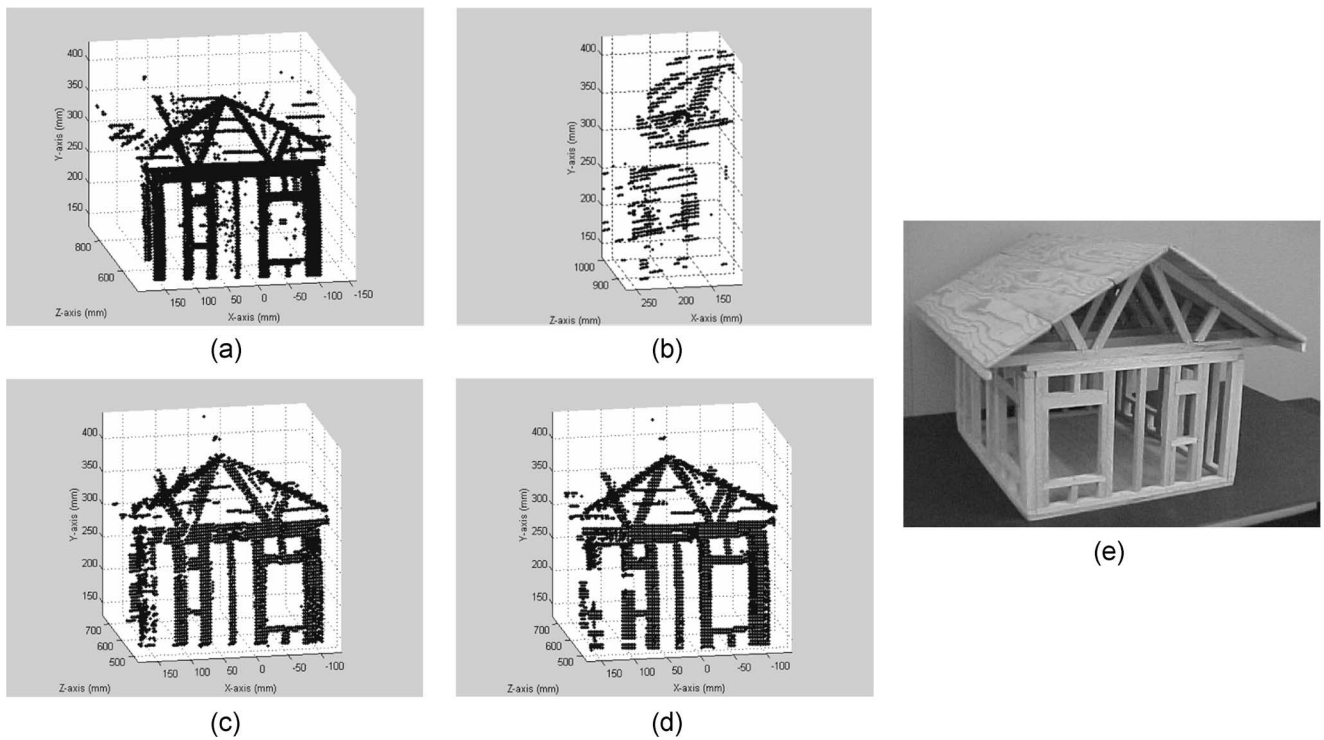


Fig. 13. Point cloud rendering of a wooden-house frame scanned with (a) laser range finder; (b) stereovision; (c) structured lighting using the left camera; (d) structured lighting using the right camera; and (e) picture of the house frame.

approximately 3 h. With the integrated robotic system, including the multimodal sensor, scanning the same 64 scan lines from three different viewing areas, the acquisition time

is drastically reduced to 30 min, which corresponds to about 9 s per scan line of 512 range points for the laser range finder. This reduction is due to the advantage of fully programmed

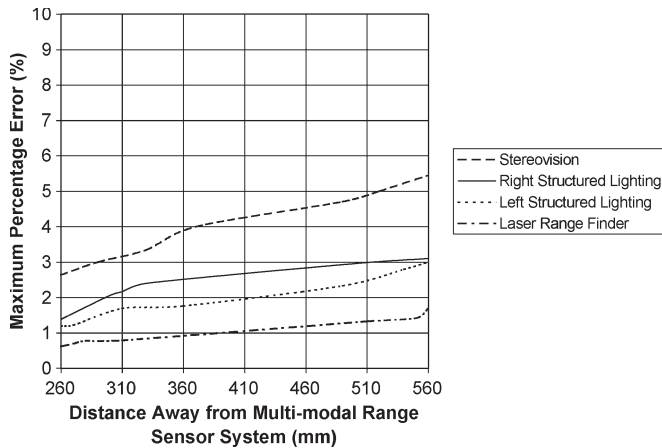


Fig. 14. Multimodal-subsystem performance as a function of the distance to the sensor.

sensor displacements and the precision of the integrated computer-controlled solution. Moreover, four complementary streams of range data are now made available simultaneously from a single scanning procedure performed within the same amount of time. The experimentation demonstrated that the main limitation on the scanning speed comes from the control loop of the robot manipulator currently used to move the multimodal sensor.

VII. CONCLUSION

An original calibration scheme has been introduced to determine the intracalibration and intercalibration parameters for a multimodal range-sensing device that advantageously combines the respective strengths of an active laser range finder, a dual structured-lighting system, and a dense stereovision approach. Experimental results demonstrated the validity of the approach and the accuracy that can be achieved. This paper also led to the development of generic guidelines for the design of multimodal sensing systems.

The experimentation put in evidence the strict requirements that must be applied to ensure that the calibration is performed in suitable conditions. Using the defined multimodal-calibration process with the proposed multimodal range sensor, applications that were traditionally dependent upon single-mode range sensors can be easily replaced by multimodal range sensors that provide supportive results and improve the accuracy of the entire system.

The integrated robotic solution has made the acquisition process flexible and less arduous. It provides a significant reduction of the acquisition time over the operation of traditional integrated systems, while human intervention is no longer required and repeatability is inherently improved. The proposed fully automated calibration process of a multimodal sensing system increases the accuracy of the fusion of each scan within a consistent 3-D dataset.

Further investigation will study the amalgamation of the various datasets produced by the multimodal system into a single dataset, which advertently provides the statistically optimal representation of the scanned scene. Other improvements

considered will be to provide additional information on coloring, texturing, and other essential attributes otherwise not detected by single-mode range sensors. Such an enhancement to the multimodal range-sensing system will lead to a more sophisticated model of the captured environment. Finally, the data are to be used in creating occupancy models and in serving higher level robotic applications such as safe path planning and manipulation.

REFERENCES

- [1] National Aeronautical and Space Administration—Jet Propulsion Laboratory, *NASA Facts—Mars Pathfinder*, Sep. 10, 2004, Pasadena, CA: California Institute of Technology. [Online]. Available: http://www.jpl.nasa.gov/news/fact_sheets/mpf.pdf
- [2] R. O. Duda, D. Nitzan, and P. Barrett, "Use of range and reflectance data to find planar surface regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 3, pp. 259–271, Jul. 1979.
- [3] R. A. Jarvis, "A laser time-of-flight range scanner for robotic vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 5, pp. 505–512, Sep. 1983.
- [4] K. S. Fu, R. C. Gonzalez, and C. S. G. Lee, *Robotics: Control, Sensing, Vision, and Intelligence*. New York: McGraw-Hill, 1987.
- [5] U. R. Dhond and J. K. Aggarwal, "Structure from stereo—A review," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1489–1510, Nov./Dec 1989.
- [6] S. D. Cochran and G. Medioni, "3-D surface description from binocular stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 10, pp. 981–994, Oct. 1992.
- [7] P. Curtis, C. S. Yang, and P. Payeur, "An integrated robotic multi-modal range sensing system," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, Ottawa, ON, Canada, May 2005, pp. 1991–1996.
- [8] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- [9] M. Hebert, "Active and passive range sensing for robotics," in *Proc. IEEE Int. Conf. Robot. Autom.*, San Francisco, CA, 2000, vol. 1, pp. 102–110.
- [10] G. Dudek, P. Freedman, and I. M. Rekleitis, "Just-in-time sensing: Efficiently combining sonar and laser range data for exploring unknown worlds," in *Proc. IEEE Int. Conf. Robot. Autom.*, Minneapolis, MN, 1996, vol. 1, pp. 667–672.
- [11] D. Laurent, M. El Mustapha, P. Claude, and V. Pascal, "A mobile robot localization based on a multisensor cooperation approach," in *Proc. IEEE Int. Conf. Ind. Electron., Control and Instrum.*, Taipei, Taiwan, R.O.C., 1996, vol. 1, pp. 155–160.
- [12] J. Miura, Y. Negishi, and Y. Shirai, "Mobile robot map generation by integrating omnidirectional stereo and laser range finder," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, vol. 1. Lausanne, Switzerland, 2002, pp. 250–255.
- [13] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Autom.*, vol. RA-3, no. 4, pp. 323–344, Aug. 1987.
- [14] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, MA: MIT Press, 1993.
- [15] C. Chen and A. Kak, "Modeling and calibration of a structured light scanner for 3-D robot vision," in *Proc. IEEE Int. Conf. Robot. Autom.*, Raleigh, NC, 1987, vol. 4, pp. 804–815.
- [16] R. Pless and Q. Zhang, "Extrinsic calibration of a camera and laser range finder," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2004, vol. 3, pp. 2301–2306.
- [17] E. Trucco and B. Fisher, "Acquisition of consistent range data using local calibration," in *Proc. IEEE Int. Conf. Robot. Autom.*, San Diego, CA, 1994, vol. 4, pp. 3410–3415.
- [18] Servo-Robot Inc., *Jupiter 3-D Laser Vision Camera Installation and Operation Manual*, 1996, St-Bruno, QC, Canada.
- [19] W. M. Keck, "Virtual Factory Lab, Virtual Robot System version 2," School Ind. Systems Eng., Georgia Inst. Technol., Atlanta, GA, 2002. Technical Report.
- [20] Matrox Electronic Systems Ltd., *Matrox Imaging Library Version 6.1 User Guide*, Mar. 2000, Dorval, QC, Canada.
- [21] Intel Corporation, *Open Source Computer Vision Library—Reference Manual*, 2001, Santa Clara, CA.
- [22] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," in *Proc. IEEE 6th Int. Conf. Comput. Vis.*, Bombay, India, 1998, pp. 1073–1080.



Christopher S. Yang (M'04) received the M.A.Sc. degree in electrical engineering from the School of Information Technology and Engineering (SITE), University of Ottawa, Ottawa, ON, Canada, in 2006.

His primary research interests are in range-sensing technologies and their multimodal capabilities via calibration.



Phillip Curtis (S'03) received the M.A.Sc. degree in electrical engineering from University of Ottawa, Ottawa, ON, Canada, in 2005. He is currently working towards the Ph.D. degree at the same university.

His research interests include registration algorithms, modeling, and machine vision.



Pierre Payeur (S'90–M'98) received the Ph.D. degree in electrical engineering from Laval University, Quebec, Canada, in 1999.

In 1998, he joined the University of Ottawa, Ottawa, ON, Canada, as an Assistant Professor in the School of Information Technology and Engineering (SITE) and cofounded the Vision, Imaging, Video, and Autonomous Systems Research Laboratory (VIVA). His current research interests are volumetric 3-D modeling, range-data processing, robot guidance, teleoperation, and integration of computer

vision in autonomous systems control.

Dr. Payeur is a member of the IEEE Robotics and Automation Society, the IEEE Instrumentation and Measurement Society, and the Ordre des Ingénieurs du Québec.