An Integrated Vision-Guided Robotic System for Rapid Vehicle Inspection

R. Fareh, P. Payeur, D. Nakhaeinia, R. Macknojia, A. Chávez-Aragón,
A.-M. Cretu, P. Laferrière, R. Laganière, R. Toledo
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, ON, Canada

***Abstract* — This paper presents the design and integration of a vision-guided robotic system for automated and rapid vehicle inspection. The main objective of this work is to achieve a seamless and efficient integration of several sensors and robotic components to rapidly acquire RGB-D data over the surface of a vehicle in order to efficiently navigate a robotic manipulator along the vehicle's surface and within regions of interest that are selectively identified. An efficient and accurate integration of information from multiple RGB-D sensors is proposed to achieve fully automated and rapid 3D profiling of automotive vehicles of various types and shapes. The proposed integrated system merges all components while taking into consideration strict requirements in the context of vehicle security screening. Experimental results at the different processing stages are presented and analyzed.**

Keywords* — *System integration, vision, RGB-D sensors, path planning, surface following, robot visual guidance.

I. Introduction

Detecting dangerous or prohibited substances in an automated and efficient manner became a critical asset to ensure the security of populations and properties worldwide. Public institutions such as government buildings, research centers, military bases, airports, and critical infrastructures require vehicle or individual screening systems at their periphery. While very efficient technologies exist to detect the presence of minute amounts of dangerous material particles, the process of collecting such particles safely and efficiently, without direct human intervention, remains a challenge. This research develops efficient and automated sampling procedures that build upon adaptive robotic technologies driven by multi-modal sensing systems to fully automate the particles collection while making it safer for the operators. The technology automatically adjusts the screening procedure to the diversity of shapes and sizes that characterize vehicles of various types, brands and sizes. It aims at providing authorities with pre-event screening mechanisms for prohibited substances that are versatile, safe for inspectors, easy to use, and permit automated scanning of large vehicles in time critical applications. This paper closely examines the integration considerations for the multi-modal sensory stages and the robotic equipment to perform efficiently and in a coordinated way while processing a full size automotive vehicle.

Vision-based robot systems have been the focus of significant research in both academia and industry. Manipulation for planetary exploration is presented in [1] using a well-calibrated system to achieve the required precision. The intrinsic and extrinsic parameters of a camera model are determined in an initial calibration step. Then, the second step involves kinematic calibration of vision with robotic devices. In this case, a stereo camera pair is used to determine the target range. Alternatively, several solutions for vision-guided robotic inspection make use of high-cost 3D profiling cameras, scanners, sonars or combinations of them, which often results in lengthy acquisition and slow processing of massive amounts of information. Alternatively, the ever growing popularity and adoption of the Kinect RGB-D sensor motivated its introduction in the development of vision-guided robotic systems. Numerous examples of application for the Kinect technology recently appeared in the literature [2-4]. For the development of the system presented here, the extreme acquisition speed of the Kinect technology and its low cost have been major selection criteria for this sensor to be used, given that rapidly acquiring color and 3D data over large volumes is instrumental. However, the technology still suffers from limited accuracy on depth measurements. Therefore, custom proximity and contact sensory stages [5] are also designed and integrated to further refine the robotic path planning and navigation and to compensate for errors in the rough 3D profile of a vehicle panels surface that can be generated from the low accuracy of the 3D data collected with fast Kinect sensors. Based on the surface shape acquired by a peripheral vision stage, various path planning strategies [6-8] are used to support efficient robot navigation. Using the extra proximity and contact sensory layer, adaptive path planning [9] is used to finely tune the navigation of a manipulator robot while performing close surface following.

The next sections detail the design of the overall systems, and overview the main components while paying attention to the interfacing between them and passing of information in compact, yet accurate representations. The final section reports on the overall performance of the integrated system, from automated vehicle category identification and surface shape acquisition, to adaptive control of a manipulator for close surface following during a particle samples collection procedure.

II. Application Description and Requirements

The aim of this work is to rapidly collect data over the surface of a vehicle in order to efficiently navigate a robotic manipulator toward regions of interest that are selectively identified as being critical for security screening. A set of

978-1-4799-2086-0/14/$31.00 ©2014 IEEE

Kinect sensors are placed conveniently to interact as a collaborative network of imagers to collect texture and 3D shape information over a vehicle within a short response time. The specifications initially set for the task were that the entire procedure of acquisition, modeling, and robotic interaction is completed within 2 to 3 minutes. Fig. 1 presents the layout of the vehicle scanning station. It is designed to allow a vehicle to be positioned in front of the vision system, without creating any occlusion. For that purpose, the robotic arm is placed at the extremity of its workspace outside the field of view of the sensors. Five Kinect sensors collect the color and depth information over one side of the vehicle while the vehicle is parked in the designated area. A duplication of the same setup can be deployed on the other side of the vehicle to acquire its complete surface. The color and depth information is then processed to construct a textured 3D model of the vehicle.

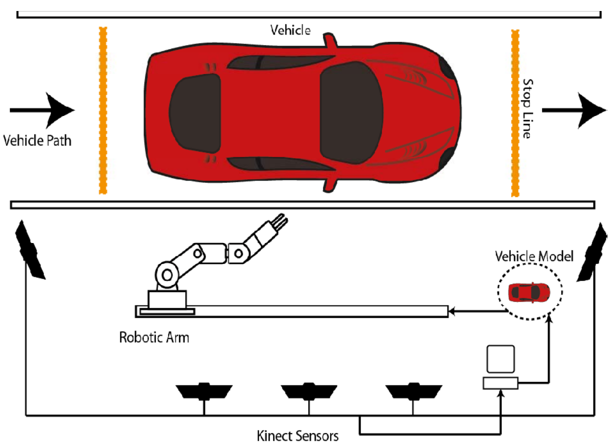


Figure 1: System layout for the proposed scanning system.

III. SYSTEM DESIGN

The complete assembly for the proposed vision-guided robotic platform for vehicle inspection is detailed in Fig. 2. The main components of the system are: a registration module using color and depth data from Kinect sensors to estimate the calibration parameters; a region of interest extractor which detects specific areas to inspect over the vehicle using an original visual detector of vehicle parts (VDVP) technique; an efficient surface mesh generator for 3D modelling; a global path planning algorithm that operates from a 3D surface mesh; and an adaptive path planning strategy that uses proximity and contact sensors embedded on the robot's end effector to achieve fine and precise robot control. These components are detailed in this section.

A. Vehicle category recognition

With the large number of vehicles circulating over our roads, the development of solutions for automated and efficient recognition and classification techniques for different vehicle categories became essential for a multitude of applications, such as optimizing available parking lots and spaces, balancing ferry loads, managing traffic and planning infrastructure or servicing vehicles. To meet these requirements, a vehicle classification system from images collected from 6 views has been proposed [10].

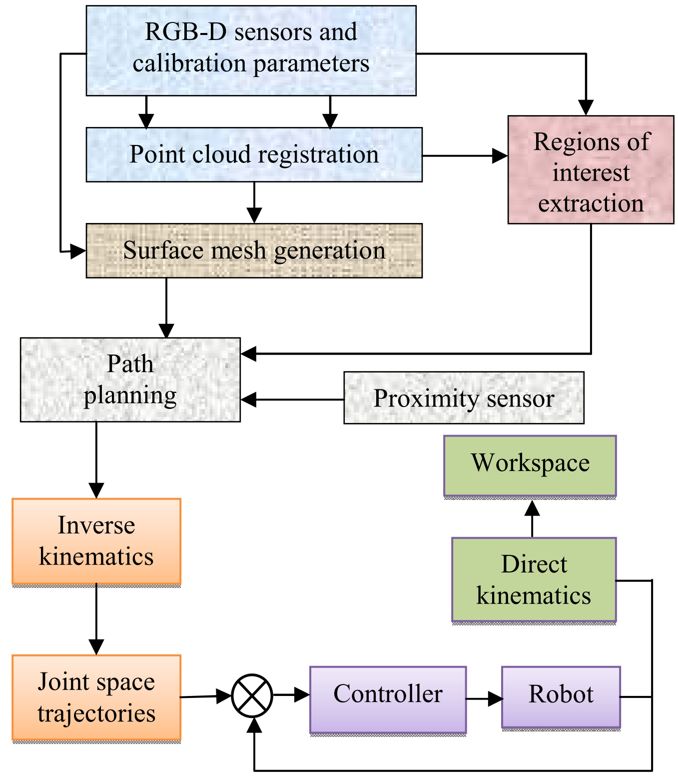


Figure 2. Block diagram of the comprehensive vision-guided robotic system for vehicle inspection.

Human visual attention mechanisms (VAM) and a series of binary support vector machines, complemented by a width and height-based (W-H) discriminating heuristic, are combined in order to identify a set of salient discriminative features and to achieve fully automated classification of vehicle categories. Our early work validated the method on vehicles of three different categories, Sedan, Sport car, and SUV. The overall recognition rates are reported in Table 1.

Table 1. Average classification rate

Salient feature	96%
Salient features + W-H information	99.13%

The approach demonstrated superior performance to alternative solutions from the literature, achieving over 99% correct classification when using the additional estimations on the width and height of the vehicles.

B. Cameras-manipulator calibration & data registration

In order to coordinate the movement of the manipulator with the color and depth data collected by every Kinect sensor, the corresponding textured 3D point clouds must be accurately registered with respect to the base of the robot. An extensive procedure to estimate the internal and external calibration parameters over a network of Kinect sensors, as well as their correspondence with the manipulator's reference frame was proposed in [11]. The method takes advantage of the depth and color information captured by RGB-D sensors. In order to meet the vehicle screening time requirements, the calibration procedure is performed offline and before the inspection station is put in operation. As such, calibration parameters initially estimated with the method detailed in [11] can be finely tuned with the use of an iterative closest point (ICP) algorithm as part the calibration process as it does not impact the execution time in production. The refined calibration parameters are later used to perform actual registration between the piecewise datasets that guide the manipulator from visual information, as shown in Fig. 3.

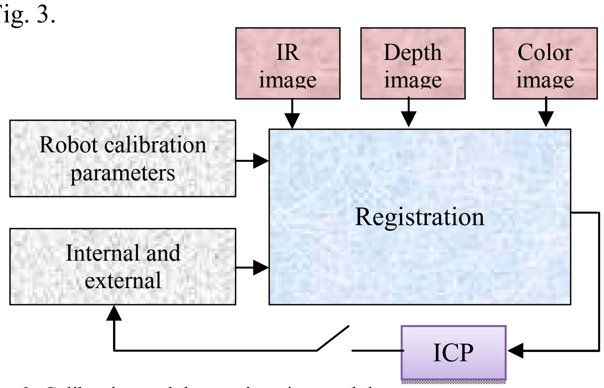


Figure 3: Calibration and data registration module.

C. Regions of interest extraction

Once the network of Kinect sensors is available and entirely calibrated, a visual detector of vehicle parts (VDVP) method is used to determine the location of the vehicle in the scene and subsequently the location of a set of significant vehicle components [12]. The specific areas of interest over a large object such as a vehicle are reorganised in order to speed up the modeling process and to facilitate the guidance of the robot arm. Moreover, acquiring knowledge about the location of dominant features over the vehicle according to its category, recognized as described in section III. A, significantly reduces the amount of time spent on scanning the vehicle surface profile as required to accurately drive the robot. This is achieved by focusing the operation only over selected areas of importance for the screening procedure. The VDVP method receives as an input a color image of a lateral view of the vehicle and automatically and efficiently determines the location of up to 14 vehicle characteristic parts. Different types of vehicles such as: 4-door sedan, 2-door sedan, 3-door hatchback, 5-door hatchback, SUV and pickup-trucks can be processed by the VDVP method. A sample of the result obtained after applying the VDVP method over a test image is shown in Fig. 4. Round areas indicate features detected by the classifiers, while square regions represent the locations of features that were inferred based on the distribution of other detected features that support more reliable differentiation and are therefore easier to detect.

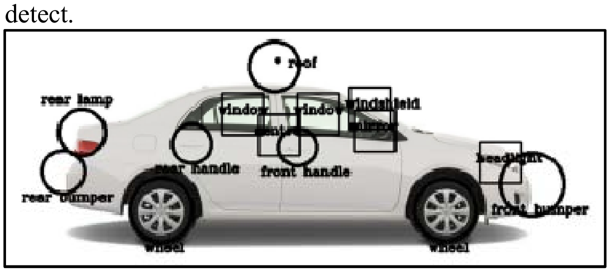


Figure 4. Detection of parts of interest over a 4-door sedan car.

D. Local and efficient surface mesh generation

A surface mesh is created from the registered point clouds using an accelerated meshing technique [12] that takes advantage of the structured organization of depth readings generated by Kinect sensors. Using the external calibration parameters previously calculated, the point clouds collected by the set of Kinect sensors are first aligned. Then, using the VDVP method, the points contained within specific discovered areas of interest are segmented from the whole point cloud to make the representation of the surface shape only over those regions of interest faster to compute, and easier to process for guiding the manipulator. For this purpose, the locations of the detected parts of interest in color images are mapped unto the 3D point cloud. Next, an individual triangle mesh is built over the corresponding groups of point, that is for each region of interest. A set of local 3D colored models that represent the shape and visual appearance of the surface within each region of interest is obtained. Rather than using the classical Delaunay triangulation technique [13], which tends to be computationally expensive, here the triangulation mesh is computed more efficiently by taking advantage of the structured information provided by Kinect sensors as detailed in [12]. The proposed meshing technique generates a mesh for the entire point cloud from one Kinect in about 0.1 second. A 3D mesh of a mock-up car door panel is shown in Fig. 5, after the Quadric Clustering decimation algorithm [14] is applied over the resulting triangle mesh to reduce the number of triangles for better visualization.

Figure 5. Triangular mesh of a door panel obtained with an accelerated meshing method.

E. Global path planning

The piecewise 3D models generated in the previous step are used as an input for a robotic path planner to move a manipulator robotic arm in close proximity of a vehicle to perform some inspection tasks. The global path planning strategy, introduced in [15], assumes that each zone of interest is bounded in a rectangular box as shown in Fig. 6. The planning strategy starts with the closest vertex to the upper-left corner of the rectangular bounding box. Then the robot moves horizontally towards the next point which is within a distance 2r, which corresponds to twice the circular area of radius r covered by the end effector of the manipulator, from the previous point along the Z-axis until it reaches to the surface edge on the right side of the rectangular bounding box (Zmax). In the present work r = 50 mm, which corresponds to the physical size of the tool used to validate the methodology. The surface edge is identified by checking the neighbour vertex at each step. The robot moves vertically down to the next position of the end-effector, until the neighbour vertices are no longer within the rectangular area. The next point is considered within a distance 2r from the previous location of the robot, but along the Y-axis. Then the robot again moves horizontally but in the opposite direction until it reaches the surface edge on the left side of the rectangular bounding area (Zmin). The process continues until the robot has scanned the entire area, creating a raster scanning pattern over the selected part of surface of the vehicle to inspect.

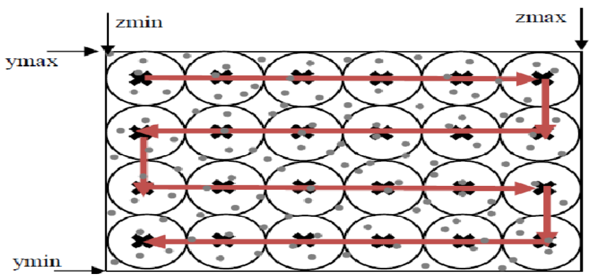
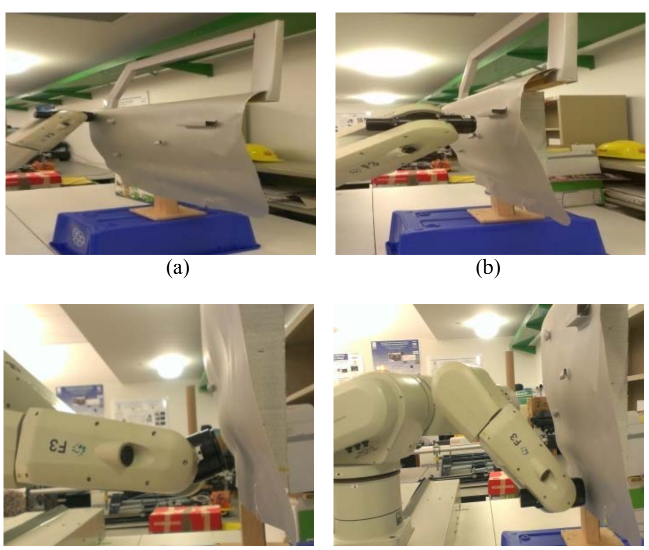


Figure 6. Global trajectory planning over a region of interest.

In order to ensure proper alignment of a close inspection camera or accurate contact over the vehicle bodywork, the orientation of the end effector must also match that of the local normal to the surface of the vehicle at every point along the global path. Therefore, the proper end-effector's orientation is computed and added to the definition of the trajectory, in correspondence with the local normal to the surface of the vehicle panel, which inherently exhibits smooth aesthetic curves. This is achieved by using a normal calculation technique that is further detailed in [15]. As a result, the region of interest over a vehicle body panel is accurately followed, within the precision made available by the RGB-D sensors used. Fig. 7 exemplifies the complete scan of a car door panel, including the adaptation to its curves that affect the orientation of the robot's end effector.



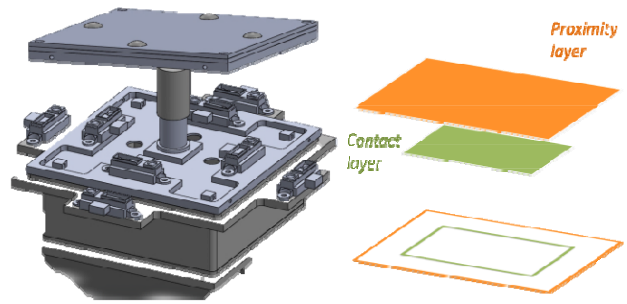
(a) (b)

(c) (d)

Figure 7. Robot performance at following the panel's curved surface and matching its position and orientation.

F. Proximity sensing for close interaction

Kinect sensors used to support for the global path planning provide only limited accuracy on the depth measurements and are not a priori accurate enough to drive fine and precise interaction of the robot with the vehicle. The robotic path planning is then further assisted by proximity and contact sensors mounted on the robot end effector. Any error due to the relatively low accuracy of the 3D model of the vehicle achieved with the Kinect sensors is compensated by these proximity and contact sensors in order to ensure more precise interaction between the robot and the object. In addition, the proximity sensors will intervene only over limited regions of the vehicle where screening is required. The custom device designed for this purpose is depicted in Fig. 8a. It is composed of two rigid plates separated by springs which provide enough freedom of movement for the upper plate to adapt to the orientation of the surface that the robotic wrist might touch. The resulting compliant instrumented wrist device embeds 8 infrared distance sensors organized in two layers. Each set of 4 sensors provides respectively the relative displacement between the bottom plate (the tool plate of the robot) and the compliant plate when contact occurs, or the distance to the vehicle body panel when the wrist is in proximity of the vehicle but does not touch it. (Fig. 8b). The information measured by each set of 4 sensors is converted into a transformation matrix that informs the robot on either how the reaction forces created by the vehicle in contact with the upper plate influence the orientation of the latter, or how close the end effector remains from the object while the robot approaches it. As a result, the two sensing layers embedded in the robotic wrist provide extra sensing capabilities to guide the robot while approaching the surface and eventually contacting with it. These go beyond the higher layer RGB-D visual stage described previously.



(a) (b)

Figure 8. CAD drawing and dual sensing layers of the compliant instrumented device.

G. Adaptive robot control

The robotic system used in this work is a 7-DOF CRS F3 manipulator. The 3D model of a vehicle's region of interest generated from the peripheral network of Kinect sensors is first used by the robot as a general guidance for navigation and interaction with the vehicle's surface to perform the inspection task. Because Kinect sensors only support a limited precision on their depth measurements, the proximity sensors presented in section III.F are used to provide more accurate local information during the approach phase. When the robot reaches close to the surface, the transformation matrix that represents the relative position and orientation of vehicle panel's surface with respect to the tool plate of the robot is computed. The orientation in 3D space is deduced from the normal vector to the surface which is estimated from the four normal vectors of the proximity infrared sensors embedded in the robotic wrist. The distance from the vehicle panel surface to the fixed bottom plate (end effector tool plate) is estimated by calculating the average of any pair of parallel distance measurements. The same principle also applies, but with the second set of wrist embedded sensors, when the robot is in contact with the surface.

The raster pattern motion trajectory previously defined by the global path planning (section III.E) becomes the inputs to the adaptive path planning. First, the robot moves to the starting point and follows the global path planning. Three alternative modes of navigation are used when the end effector is close to the vehicle surface. First the information provided by the peripheral network of RGB-D sensors about the surface covered by the end-effector is used. Once the local information provided by the proximity sensors differs from the one collected by the Kinect (beyond a predefined tolerance), the controller switches to a local mode and the trajectory is updated using the extra information from the proximity layer of sensors on the robotic wrist. Similarly, when the robot reaches contact with the inspected surface, the third control mode, which relies on the contact layer of sensors in the robotic wrist, replaces the proximity control mode. An adaptive robot controller, exploiting a mode switching strategy, is designed to allow each of the three sensing layers (peripheral vision, proximity, contact) to drive the navigation of the manipulator and closely monitor the interaction of the robot with the vehicle surface where appropriate, thereby overcoming the limited resolution of the RGB-D sensors adopted.

Finally, the adaptive controller also provides support for obstacles avoidance while the robot closely follows a surface. As illustrated in Fig. 9, by default the end effector continues to navigate in accordance with the global path while switching between peripheral vision (global), and proximity or contact (local) information. However, when an obstacle is detected by the sensors, a hit point, H_j, is defined. The end-effector temporarily follows the obstacle boundary while relying on the local sources of information, until the global path is met again at leaving point, L_j.

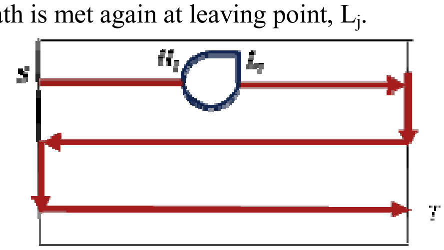


Figure 9. Adaptive path planning and collision avoidance.

IV. ARCHITECTURE DESIGN

The general overview of the system and its main components has been described in the previous section. This section details the interconnections between those components and the main computer as well as how they interact with each other. The complete system architecture is shown in Fig. 10.

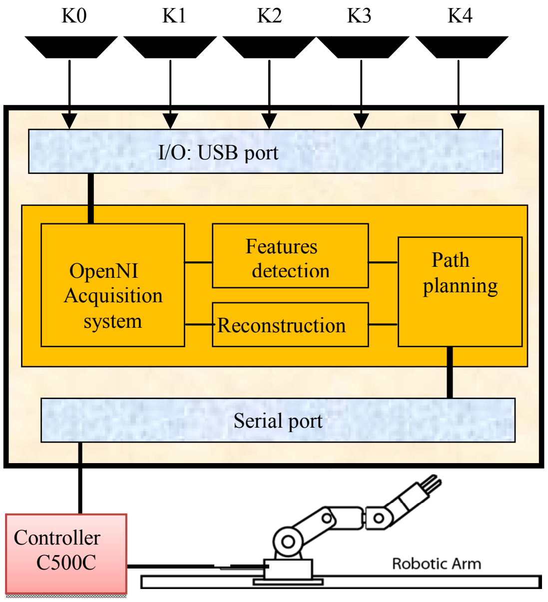


Figure 10. Robot-vision architecture design.

Five Kinect for Xbox 360 devices are connected to dedicated USB ports on the main computer. All Kinect sensors are operated using the OpenNI framework upon which custom calibration and acquisition software has been developed. The color and depth data provided by the Kinect sensors is collected. Then, the information is registered, based on the pre-estimated calibration parameters. The color data is used to detect regions of interest over the vehicle, that are mapped unto corresponding subsets of 3D point clouds. A local 3D model reconstruction is then computed within every region of interest and represented in the reference frame of the manipulator before being saved as a PLY format file. The path planning algorithm loads the PLY file for a given region of interest and uses the contained 3D model to navigate the robot over the patch of surface. The Cartesian space trajectory which defines the position and orientation for the end-effector is transformed to the joint space via the robot's inverse kinematics. The joint space trajectory is then sent to the controller to perform the surface following over the region of interest on the vehicle. The controller sends the appropriate torque commands to the robot in order to ensure a good tracking in the joint space. The robot controller is connected via a serial cable to the same computer where the RGB-D data is acquired and the model is generated.

V. EXPERIMENTS AND PERFORMANCE

The algorithms used for this work were developed in C++ and run on a computer with an Intel core i7 CPU and Windows 7. For the 3D surface shape reconstruction of a regular vehicle, the average time needed is 4.0 sec. per Kinect sensor view. When using a network of 5 Kinect sensors covering a 180^{o} lateral view of the vehicle, the network of sensors collects the information in two time slots to avoid interference; the initialization of each device takes between 1 and 2 seconds. As a result, the scanning and 3D textured modeling processes for the entire side of a vehicle are completed within 30 seconds. The calibration process is performed off-line and has no impact of the inspection rate. The automated selection of regions of interest allows for rapid extraction of subsets from the generated 3D model, which also expedites the triangulation process which is run only over bounded regions of interest.

The validation of the proposed vision-guided robotic system included lab tests and real-world application scenarios. The latter involved its installation in an underground parking garage where data was collected and 3D textured models were created over vehicles of different types and sizes (Fig. 11a). For the robotic part, the path planning algorithm was experimentally validated on a 7-DOF CRS F3 manipulator in the laboratory. The experimental results showed that the robot can accurately scan an entire region of interest while closely following the curvatures on the surface of an automotive vehicle body panel (Fig. 11b), within the set time requirements.

(a) (b)

Figure 11 : Experimental setup: a) capturing 3D data over a network of Kinect sensors, and b) robot performance evaluation at following a door panel's curved surface while matching its position and orientation.

VI. CONCLUSION

This work presented the design and implementation of a vision-guided robotic system for automated and rapid vehicle inspection. First, the proposed layout of the vehicle scanning station was defined. Then, each stage of the inspection system was presented with an emphasis on their integration into a coordinated and efficient system solution able to meet tight time requirements. An automated classification technique is used for recognizing different vehicle categories. A calibrated network of Kinect RGB-D sensors is operated to rapidly generate color and depth datasets on which an accelerated triangular mesh generation algorithm is applied that takes advantage of the intrinsic structure of range data provided by Kinect sensors to further speed up the 3D model generation. Subsets of the surface mesh are extracted that correspond to regions of interest automatically detected over a vehicle, which then support a robot path planning operation to ensure full coverage of the region of interest by a manipulator. A global trajectory planning method is used to automatically screen large objects such as automotive vehicles and to support close surface shape following with a robotic manipulator. Proximity and contact sensors are added to the end effector to provide locally accurate information and to update the path via a tri-modal adaptive control method. Experimental results demonstrated that the data acquired over the surface of a vehicle are rapidly collected and that the robot can successfully scan an entire region of interest while closely following the curved surface of an automotive vehicle.

REFERENCES

[1] S. W. Squyres, R. E. Arvidson, E. T. Baumgartner, J. F. Bell III, P. R. Christensen, and S. Gorevan, "Athena Mars Rover Science Investigation," *Journal of Geophysical Research,* vol. 108, p. 8062, 2003.

[2] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D Full Human Bodies using Kinects," *IEEE Transactions on Visualization and Computer Graphics,* vol. 18, pp. 643-650, 2012.

[3] A. Maimone and H. Fuchs, "Encumbrance-free Telepresence System with Real-Time 3D Capture and Display Using Commodity Depth Cameras," in Proc. of *10th IEEE International Symposium on Mixed and Augmented Reality,* pp. 137-146, 2011.

[4] P. Noonan, T. Cootes, W. Hallett, and R. Hinz, "The Design and Initial Calibration of an Optical Tracking System Using the Microsoft Kinect," in Proc. of *IEEE Nuclear Science Symposium and Medical Imaging Conference* pp. 3614-3617, 2011.

[5] M. Hashimoto and Y. Imamura, "Design and Characteristics of a Parallel Link Compliant Wrist," in *Proc. of IEEE Intl Conf. on Robotics and Automation,* vol. 3, pp. 2457-2462, 1994.

[6] E. U. Acar, H. Choset, Y. Zhang, and M. Schervish, "Path Planning for Robotic Demining: Robust Sensor-based Coverage of Unstructured Environments and Probabilistic Methods," *International Journal of Robotics Research,* vol. 22, pp. 441-466, 2003.

[7] X. Liao and L. Carin, "Application of the Theory of Optimal Experiments to Adaptive Electromagnetic-Induction Sensing of Buried Targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 26, pp. 961-972, 2004.

[8] J. R. Spletzer and C. J. Taylor, "Dynamic Sensor Planning and Control for Optimally Tracking Targets," *The International Journal of Robotics Research,* vol. 22, pp. 7-20, 2003.

[9] O. Meister, N. Frietsch, C. Ascher, and G. F. Trommer, "Adaptive Path Planning for a VTOL-UAV," in *IEEE/ION Position, Location and Navigation Symposium,* pp. 1252-1259, 2008.

[10] A.-M. Cretu, P. Payeur, and R. Laganière, "Salient Features Based on Visual Attention for Multi-view Vehicle Classification," in *Proc. of IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, Ottawa, ON, pp. 64-69, 2011.

[11] R. Macknojia, A. Chávez-Aragón, P. Payeur, and R. Laganière, "Calibration of a Network of Kinect Sensors for Robotic Inspection over a Large Workspace," in *Proc. of IEEE Workshop on Robot Vision,* Clearwater, FL, pp. 184-190, 2013.

[12] A. Chávez-Aragón, R. Macknojia, P. Payeur, and R. Laganière, "Rapid 3D Modeling and Parts Recognition on Automotive Vehicles Using a Network of RGB-D Sensors for Robot Guidance," *Journal of Sensors,* vol. 2013, 2013.

[13] M. De Berg, O. Cheong, M. van Kreveld, and M. Overmars, *Computational Geometry*: Springer, 2008.

[14] P. Lindstrom, "Out-of-core Simplification of Large Polygonal Models," in *Proc. of 27th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 259-262, 2000.

[15] D. Nakhaeinia, R. Fareh, P. Payeur, and R. Laganière, "Trajectory Planning for Surface Following with a Manipulator Under RGB-D Visual Guidance," in *Proc. of IEEE Intl Symposium on Safety, Security, and Rescue Robotics*, Linkoping, Sweden, pp. 1-6, 2013.