

Multi-sensor Registration and Integration for Inspection and Navigation

Faysal Boughorbel, Andreas Koschan, Mongi Abidi

Department of Electrical and Computer Engineering, University of Tennessee, Knoxville, TN, 37996, fboughor@utk.edu

Abstract – *This paper describes a unified framework for addressing the problem of multi-sensor registration and integration in the context of inspection by autonomous platforms. The task is approached from an optimization angle by designing an effective general-purpose registration criterion that can be employed for both aligning 3D and 2D datasets. The resulting pipeline reconstructs full 3D representations of the scenes and objects of interest, including multi-spectral texture overlay. The obtained information rich representation is very useful in automating decision and reducing the cognitive load on human operators*

I. INTRODUCTION

Mobile and fixed robotic systems for inspecting hazardous environments that are equipped with multiple imaging devices are deployed in many facilities around the world. These systems rely mostly on human operators to interpret the sensory input and decide the appropriate action. To achieve further autonomy for these systems, in labor-intensive tasks such as nuclear waste sorting and critical components control, the robots need to acquire a more complete representation of the objects that are manipulated, or the environments in which they will operate. This can be achieved by integrating and fusing the information contained in the different onboard sensing devices.

Before starting the data fusion step, a registration stage that computes the spatial correspondence between the sensory inputs is needed. Our research addresses this problem of aligning different imaging sensors used for object and scene reconstruction, with the ultimate goal being the automation of the registration task in both the single modality and multi-modal cases. The sensors that are considered in this paper are range-mapping scanners, which acquire the 3D geometry of the scenes and objects of interest, high-resolution video cameras, and infrared thermal cameras.

The final output of our system is 3D models with registered texture overlays generated from the multi-spectral cameras. The actual objects and scenes of interest vary in scale from large rooms that need to be mapped to micro and nano-scale components that require quality control and inspection, passing by medium sized objects typical of radioactive waste-sorting stations. In all these cases the basic approach is the same, while the sensing technology employed usually changes. For instance, in the case of scene geometry acquisition, two approaches are commonly employed: passive stereo systems and active laser scanning devices. In the first case depth maps

are obtained from correspondence between pairs of images, while in the case of laser scanners time of flight, triangulation, and frequency modulation are used to sample the geometry of the scenes. Tradeoffs between accuracy and speed usually determine the choice the passive or active systems. In our system 3D geometry is the reference to which all other multi-spectral images will be registered giving a representation that can be manipulated using standard graphics tools and permitting the use of powerful scene analysis tools.

II. OVERVIEW OF APPROACH

To achieve our goal of automatic and accurate sensor registration we developed a unified framework based on a single matching criterion. The method was applied to the task of 3D-3D registration of multiple range maps, which is a necessary step for the reconstruction of scene geometry. It was also employed for aligning multi-spectral images with the reconstructed geometry, as well as for the registration of the multi-spectral 2D images. The first of these tasks is an important problem in many computer vision and graphics applications. The main issue is the recovery of the transformations that relate different 2½ D scans of a scene. Once these views registered, subsequent integration steps build the 3D model.

One of the most commonly used methods for this task is the Iterative Closest Point (ICP) algorithm [1][6], which is a locally convergent scheme that can perform well with a close enough initialization. Invariant feature extraction techniques [4] were also employed in two-step systems that commonly required the use of ICP as a final refinement stage. To enhance the region of convergence and overcome several additional drawbacks of current methods, we devised a registration criterion that can be used in a single stage system. Our approach operates at the point level and defines an energy function that

depends on the datasets to be registered and on the transformation parameters. The proposed criterion employs notions of feature Saliency to assign scalar weights to each 3D point. An attraction force between two points that depends on their weights is defined, and the matching criterion is obtained by integrating over all these forces. Thus defined, the registration criterion is explicitly expressed in terms of the sought transformations, and its derivatives easily obtained, allowing for the use of various powerful optimization schemes and extending the range of convergence.

The method is also applied to the task of registering 3D geometry with 2D (multi-spectral) intensity images. This is important for generating information-rich textured models, which are in turn employed for recognition and inspection. This problem is related to the pose estimation task and is also commonly encountered when dealing with model-based object recognition. In our framework we formulate the task as a 3D-2D edge-matching problem that will be solved using our criterion. Creases are extracted in the 3D models using robust voting techniques, and edge maps are obtained from the intensity images. The criterion is employed to detect the best overlap between the 3D and 2D edges and hence recover the geometry to image mapping.

The final task for which our registration method was employed is the alignment of visual and IR imagery. Here also the method is applied to extracted edge maps. Several motion models were tested to achieve accurate overlap. The fusion of these two modalities offers valuable information in inspection tasks about the material characteristics of the imaged objects.

III. THE REGISTRATION CRITERION

One important goal of this work is the design of a general point-sets registration criterion, which is differentiable and convex in a large neighborhood of the aligned position. The underlying motivation is overcoming the problems of standard point-set registration techniques, mainly the Iterative Closest Point method [1]. ICP was proven effective for registering datasets when starting from an initial position that is close to the registered one. One of the main reasons behind this limitation is the non-differentiable cost function used, which imposed local convergence. In most real applications the preliminary point-feature extraction and matching step is necessary before proceeding with the ICP refinement.

Recently, some effort went to designing approximations to non-differentiable matching and similarity measures including approximating Hausdorff distances. In our case we will use a straightforward criterion that is defined for general point-sets with associated attributes. A particular importance is given to extending as much as possible the region of convergence

of this function. In this context we have found that approximations of our criterion will lead to closed form recovery of rigid transformations that can be used for further iterative refinement. In addition, the criterion can be evaluated quickly using fast evaluation techniques.

Our basic idea is to use a Gaussian field to measure both the spatial proximity and visual similarity of two points belonging to the two datasets. Consider first the two pointsets

$$M = \{(P_i, S(P_i))\}_{i=1 \dots N_M} \text{ and } D = \{(Q_j, S(Q_j))\}_{j=1 \dots N_D},$$

with their associated attribute vectors. Those vectors can include curvature for smooth surfaces and curves, invariant descriptors as well as color attributes. The Gaussian measure of proximity and similarity between two points is given by:

$$F(P_i, Q_j) = \exp\left(-\frac{d^2(P_i, Q_j)}{\sigma^2} - (S(P_i) - S(Q_j))^T \Sigma^{-1} (S(P_i) - S(Q_j))\right) \quad (1)$$

with $d(P_i, Q_j)$ being the Euclidean distance between the points. The expression can be seen as a force field whose sources are located at one the point-sets and are decaying with distance in Euclidean and attribute space. We can now define an energy function that measures the registration of M and D as:

$$E(Tr) = \sum_{\substack{i=1 \dots N_M \\ j=1 \dots N_D}} \exp\left(-\frac{d^2(P_i, Tr(Q_j))}{\sigma^2} - (S(P_i) - S(Tr(Q_j)))^T \Sigma^{-1} (S(P_i) - S(Tr(Q_j)))\right) \quad (2)$$

Where Tr is the transformation relating the two point-sets. In our work we use rigid transformations to register 3D datasets, projective transformations to align geometry with images, and several non-rigid motion models for 2D-2D image registration [5][7]. In the case of 3D registration we use visual attributes that are invariant to rigid transformations. The parameter σ controls the decay with distance while Σ diagonal with small components punishes the difference in attributes. For both parameters very small E will count the number of points that overlap at a given pose. It can be shown that using this criterion we meet a rigorous definition of registration as maximization of both overlap and local similarity between the data.

The Gaussian energy function is convex around the registered position and is always differentiable, allowing for the design of efficient registration algorithms. Several powerful optimization techniques can be used for this task such as the Quasi-Newton and conjugate gradient algorithms. The parameter σ controls the size of the convex region of convergence. Therefore, we are interested in increasing σ as much as possible. This will be limited mostly by the decrease in the localization accuracy of our criterion, hence the tradeoff between large region of convergence and precision. Studying the cost function we found that the region of convergence can

be extended considerably in the case of sufficiently informative datasets and when using several local descriptors.

IV. LOCAL INVARIANTS FOR 3D REGISTRATION

If surfaces can be extracted from the datasets then a large number of invariant features can be used as attributes in our criterion [4]. Since we are focused mostly on the general case of point-sets registration, we will use 3D moment invariants. Additionally, we employ a local descriptor that we call saliency that was derived from the tensor-voting framework and can be seen as the dual of curvature in the point-sets case. The three moment invariants are commonly used for object recognition tasks and where also employed in registration algorithms such as in Sharp's extension of ICP [3]. These moments J_1 , J_2 , and J_3 are defined for a local neighborhood N around a point P by:

$$\begin{aligned} J_1 &= \mu_{200} + \mu_{020} + \mu_{002} \\ J_2 &= \mu_{200}\mu_{020} + \mu_{200}\mu_{002} + \mu_{020}\mu_{002} - \mu_{110}^2 - \mu_{101}^2 - \mu_{011}^2 \\ J_3 &= \mu_{200}\mu_{020}\mu_{002} + 2\mu_{110}\mu_{101}\mu_{011} - \mu_{002}\mu_{110}^2 - \mu_{020}\mu_{101}^2 - \mu_{200}\mu_{011}^2 \end{aligned} \quad (3)$$

with

$$\mu_{pqr} = \sum_{(X,Y,Z) \in N} (X - P^x)^p (Y - P^y)^q (Z - P^z)^r \quad (4)$$

In the case of noisy point-sets, where surfaces and curves are not well extracted, we devised a local measure of visual saliency based on the framework of Tensor Voting, introduced by Medioni et al [2]. We are considering non-oriented point-sets as our primary input. For our purpose we just use the first pass of the tensor-voting scheme. To evaluate the saliency at a site P_i we collect votes from neighboring sites P_j which cast the so-called stick tensor at P_i in the case of 2D voting, and the plate tensor for 3D. This tensor encodes the unit vector $t_{ij} = \frac{P_i - P_j}{\|P_i - P_j\|}$ lying on the line joining the two points using its covariance matrix for which the expression in 2D is given by

$$\begin{aligned} T_{ij} &= \begin{bmatrix} (t_x^i)^2 & t_x^i t_y^i \\ t_x^i t_y^i & (t_y^i)^2 \end{bmatrix} \\ &= \frac{1}{(P_i^x - P_j^x)^2 + (P_i^y - P_j^y)^2} \begin{bmatrix} (P_i^x - P_j^x)^2 & (P_i^x - P_j^x)(P_i^y - P_j^y) \\ (P_i^x - P_j^x)(P_i^y - P_j^y) & (P_i^y - P_j^y)^2 \end{bmatrix} \end{aligned} \quad (5)$$

These tensors are collected from sites in a small neighborhood around P_i using summation:

$$T_i = \sum_{j \neq i} T_{ij} \quad (6)$$

Our idea for obtaining a scalar measure of point saliency is to use the determinant of T_i which is the product of the principle axes of its eigen-system ellipsoid. This feature is similar to the other moments and behaves like Gaussian curvature in the case of smooth surfaces. It is invariant to rigid transformations, and accounts for degenerate tensors in flat surfaces where its value is zero.

The previous local descriptors will be summed in the expression of the energy function (2). Therefore, it is important to ensure proper scaling and de-correlation. For this purpose we use the same approach as in [3] where the covariance matrix of each moment is estimated in a nearly planar region of the datasets, then used to scale the vector of raw features into a set of uncorrelated invariant features with unit variance.

V. RESULTS

Our experimental work is mostly geared toward inspection and surveillance applications with multiple sensors mounted on robotic platforms. Inspecting vehicles in high security areas is an important practical problem that we are investigating. For this task we use a sensor package consisting of laser range scanners and high-resolution color cameras attached to a remotely controlled platform. Fig. 1 shows datasets acquired from a 14 passengers van. Range images (Fig. 1(b)) encode the depth as seen from the scanner's point of view. Using these depth maps, captured from several viewpoints partial 3D models of the van can be reconstructed (Fig. 1(c)). To assemble the different parts we used our registration criterion along with a fast optimization scheme. The final result forms a complete geometric representation of the vehicle (Fig. 1(d)). The following step was adding texture to the model. This requires registering the 3D object with its 2D perspective images. In this case, also, the Gaussian criterion is employed with projective transformations. The resulting textured van is rendered in Fig. 1(e).

In addition to this project, we are involved in inspecting very small-scale structures. The 3D reconstruction pipeline was applied to mechanical components of sub-millimeter dimensions such as the gear shown in Fig. 2. The range (Fig. 2(a)) and intensity (Fig. 2(b)) images are acquired using a Tunnel Microscope. The reconstructed 3D model of the gear is rendered in Fig. 2(c).

Finally, our registration and integration framework is being used to merge several multi-spectral datasets, which

are employed to robustly recognize objects and people in the context of securing sensitive facilities. Typical thermal and color images of a face are shown in Fig. 3 along with a composite image generated after the registration step.

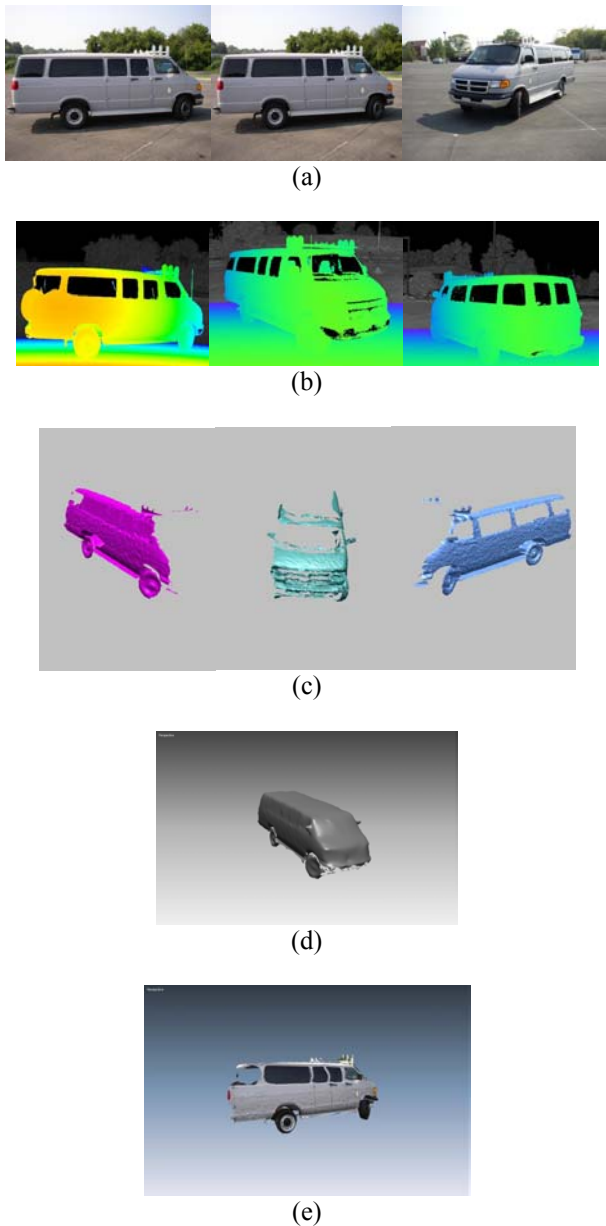


Fig. 1. Multi-modal datasets of a vehicle: (a) color images, (b) range maps, (c) reconstructed partial views, (d) complete model, and (e) model with texture.

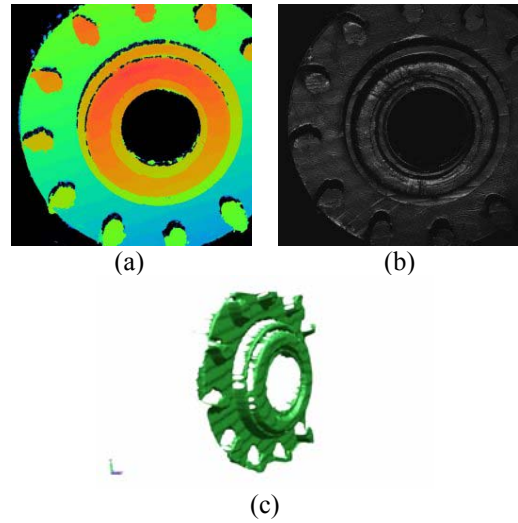


Fig. 2. Range (a) and intensity (b) images of a sub-millimeter gear obtained by Tunnel Microscopy. A 3D reconstruction of the gear is shown in (c).

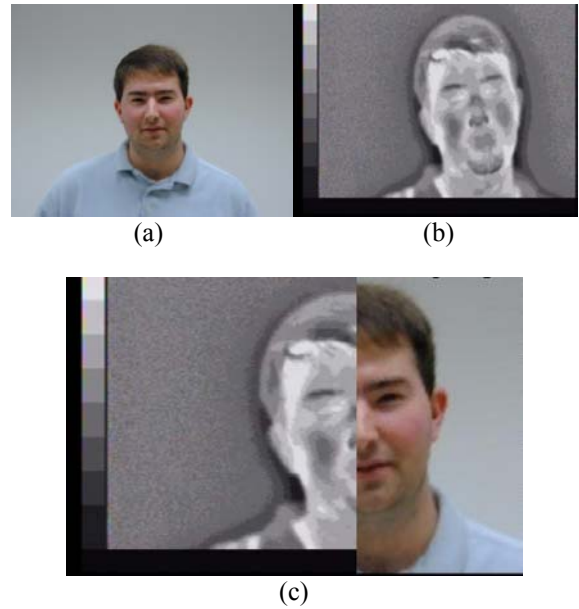


Fig. 3. Color and Thermal images of a face (a, b). The composite image (c) shows the result of registration.

VI. CONCLUSIONS

By using a unified method based on a single energy function, we addressed a wide variety of registration problems that are normally approached with largely different techniques. In our experiments the method showed robustness to partial overlap, noise, and cluttered

data. The use of single stage registration simplifies the scene modeling systems and reduces their computational cost; this is an important factor for implementing scene-understanding pipelines in the case of mobile robots. Furthermore, additional modules could be added to exploit the final scene representation, such as navigation and path planning software or robust multimodal object recognition components.

ACKNOWLEDGMENTS

This work was supported by the DOE University Research Program in Robotics under grant DOE-DE-FG02-86NE37968, by the DOD/TACOM/NAC/ARC Program, R01-1344-18, and by FAA/NSSA Program, R01-1344-48/49.

REFERENCES

1. P. J. BESL and N. D. MACKAY, "A Method for Registration of 3-D Shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239-256 (1992).
2. G. MEDIONI, M. S. LEE, and C. K. TANG, "A Computational Framework for Segmentation and Grouping", Elsevier, Amsterdam, 2000.
3. G. C. SHARP, S. W. LEE, and D. K. WEHE, "ICP Registration using Invariant Features", *IEEE Transactions on Pattern Analysis Machine Intelligence*, 24(1), 90-102 (2002).
4. R. CAMPBELL and P. FLYNN, "A Survey of Free-form Object Representation and Recognition Techniques", *Computer Vision and Image Understanding*, 81(2), 166-210 (2001).
5. F. L. BOOKSTEIN, "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6), 567-585 (1989).
6. S. RUSINKIEWICS, and M. LEVOY, "Efficient Variants of the ICP Algorithm", in Proc. of 3D Digital Imaging and Modeling, IEEE Computer Society Press, 145-152 (2001).
7. G. WOLBERG, "Digital Image Warping", IEEE Computer Society Press, 1990.