Summarizing Image/Surface Registration for 6DOF Robot/Camera Pose Estimation

Elisabet Batlle, Carles Matabosch, and Joaquim Salvi

Institut d'Informàtica i Aplicacions, University of Girona
Campus Montilivi, 17071 Girona, Spain

Abstract. In recent years, 6 Degrees Of Freedom (DOF) Pose Estimation and 3D Mapping is becoming more important not only in the robotics community for applications such as robot navigation but also in computer vision for the registration of large surfaces such as buildings and statues. In both situations, the robot/camera position and orientation must be estimated in order to be used for further alignment of the 3D map/surface. Although the techniques differ slightly depending on the application, both communities tend to solve similar problems by means of different approaches. This article is a guide for any scientist interested in the field since the surveyed techniques have been compared pointing out their pros and cons and their potential applications.

1 Introduction

Thus far, robot navigation has been focused on 2D mapping in flat terrains and usually restricted to indoor structured scenarios [34]. Recently, the need to explore complex and unstructured environments has increased [27]. The complexity of this sort of environments requires 6DOF movement due to the unevenness of natural terrains. Besides, the growing interest in 3D modeling of large objects such as buildings and statues has forced the scientific community to face new challenges with the aim of reducing the propagation error present in registration [33]. In both situations, the robot/camera pose is estimated in order to be used in a further alignment of the 3D map/surface. Although the techniques differ slightly depending on the application, both communities tend to solve similar problems by means of different approaches [11] [31].

In general, a good estimation of the initial position is always required independently of the approach or technique used. Hence, section 2 provides a classification of the most important methods used to obtain a coarse pose estimation, including inertial navigation, visual odometry and surface-to-surface matching, among others. Then, pair-wise registration approaches such as the Iterative Closest Point are used to refine the alignment between two clouds of points, see section 3. Finally, any error accumulated between correlated views is minimized by means of cycles and overlapping regions common among the acquired views. Hence, section 4 discusses a new classification of these techniques including analytic methods such as bundle adjustment and the well known ICP

J. Martí et al. (Eds.): IbPRIA 2007, Part II, LNCS 4478, pp. 105–112, 2007.

Table 1. Coarse one-to-one pose estimation techniques. R: Restricted (some DOF are constrained in a limited range); TOF: Time-of-flight; LT: Laser Triangulation; DLP: Digital Light Projector.

Technique				author	DOF	sensor	scene
Coarse one-to-one pose estimation	mechanical devices	sensors		Nüchter, 2004 [27]	6	TOF	outdoor
				Folkesson, 2003 [11]	6R	TOF	outdoor
		mechanisms		Pulli, 1999 [31]	6	LT	object
				Bernardini, 2002 [2]	6	LT	object
	Computer vision	Image to image	Feature to point	Huang, 1989 [18]	6	monocular	indoor
				Shang, 1998 [39]	6	binocular	indoor
				Davison, 2003 [9]	6	monocular	indoor
			Point to feature	Lowe, 1999 [23]	6	binocular	indoor
		Surface to surface	Point to feature	Chen, 1998 [6]	6	DLP	object
				Johnson, 1999 [20]	6	DLP	object
				Carmichael, 1999 [5]	6	DLP	object
				Chua, 1997 [8]	6	database	object
				Huber, 2003 [19]	6	LT	object
			Feature to point	Nister, 2004 [28]	6	monocular	outdoor
				Stamos, 2003 [33]	6	TOF	outdoor
				Wyngaerd, 2003 [38]	6	DLP	object
				Triebel, 2005 [36]	6R	TOF	outdoor

multi-view approach, and statistical methods such as Simultaneous Localization And Mapping (SLAM), among others. These techniques are compared and discussed analyzing their pros and cons and potential applications. The article ends with conclusions.

2 Coarse One-to-One Pose Estimation

The initial position is always required independently of the approach or technique used. The initial pose can be obtained using two well-known approaches: 1) Initial pose estimation by mechanical devices and 2) Initial Pose estimation by computer vision. The first technique is based on benefiting by using some sort of device: a) sensors, such as odometers, compasses or inertial systems [11]; or b) mechanisms, such as rotating tables, robot arms or conveyors [31] [2].When sensors or mechanical devices can not be used or when their measure is rough or inaccurate, an estimation of the initial position by means of computer vision may be a good choice. Therefore, the second technique is based on directly analyzing the visual images (given by cameras) or the surface views (given by scanners) looking for correspondences which are used to solve the alignment and consequently the pose. Although in this paper the final registration concerns 3D objects, the initial pose estimation can be achieved using both 2D or 3D views. Therefore, two main groups of pose estimation techniques using computer vision are proposed: a) Image-to-image correspondences and b) Surface-to-surface correspondences. Image-to-image techniques are based on 2D image-to-image matching using both discrete and differential epipolar constraint dealing with 2D images directly acquired by a stereo-head [18] or a moving camera [9]. Note that in the calibrated case the 3D is computed by triangulation. Besides, in uncalibrated systems the motion up to a scale factor is estimated by solving the well-known Kruppa equations computing a perspective reconstruction. The

Table 2. Fine one-to-one pose estimation techniques. R: Restricted (some DOF are constrained in a limited range); TOF: Time-of-flight; LT: Laser Triangulation; DLP: Digital Light Projector.

Technique		author	DOF	sensor	scene
Fine one-to-one pose estimation (Pair-wise)	Point to point	Besl, 1992 [3]	6	LT	outdoor
		Greenspan, 2001 [14]	6	DLP	object
		Jost, 2002 [21]	6	database	object
		Guidi, 2004 [15]	6	DLP	object
		Triebel, 2005 [36]	6R	TOF	outdoor
		Trucco, 1999 [37]	6	synthetic data	object
	Point to plane	Chen, 1991 [7]	6	DLP	object
		Gagnon, 1994 [13]	6	monocular	object
		Park, 2003 [29]	6	database	object

Euclidean reconstruction is obtained by taking any metric measure from the scene that allows the determination of the scale factor, usually a distance between two 3D features [9]. On the other hand surface-to-surface techniques deal with 3D features or clouds of points acquired by any 3D acquisition technique such as stereo [28], laser triangulation or time-of-flight lasers [33], among others. Here, the main difference is in the way of selecting the matching points.

All these methods process the 2D/3D points of the given images/surfaces to extract significant points which are used in the matching process. Hence, the techniques are classified according to: a) feature-to-point approach when the significant points are only those that satisfy a given feature [17] [33]; and b) point-to-feature approach when an arbitrary group of points are characterized obtaining a set of features that differ one to another depending on point neighborhood [23] [8] [5].

In summary, although coarse pose estimation methods based on mechanical devices provide good results in flat terrains, a combination of both mechanical and computer vision methods is usually required in the presence of rough and unstructured environments. Techniques based on the discrete epipolar geometry have been widely studied and nowadays robust solutions are available even in 6DOF. Besides, the differential movement estimators are quite sensitive to noise. Hence, these methods are, in general, adapted to the application constraining the number of DOF with the aim of reducing the error in the estimation. Therefore, surface-to-surface alignment is more adequate for complex 3D scenarios, but then we have to avoid symmetries in the views to obtain accurate registrations.

3 Fine One-to-One Pose Estimation

Once an initial 3D pose is estimated by any coarse registration technique, an iterative minimization should be applied to obtain a refined pose and hence a better alignment between both views. Herein, the methods are classified according to the minimization function, which is usually the distance between corresponding points (point-to-point) or the distance between points and their corresponding plane (point-to-plane). For instance, Point-to-point alignment, such as the Iterative Closest Point (ICP) [3], focus on finding the distance between point

correspondences. ICP is the most common point-to-point fine registration method and the results provided by authors are good [14] [36]. However, the method can not cope with non-overlapping regions because outliers are barely removed. In addition, this method usually presents problems of convergence, many iterations are required and, in some cases, the algorithm converges to local minima. The algorithm proposed by Chen [7] (Point-to-plane) is an alternative to ICP. Given a point in the first image, the intersection of the normal vector at this point with the second surface determines a second point in which the tangent plane is computed. The distance between this plane and the initial point is the function to minimize. Despite the difficulty of determining the cross point between a line and a plane in a cloud of points, some techniques such as the fast variant of ICP proposed by Park [29] and the method of Gagnon [13] are presented to speed this process up. Compared to ICP, this method is more robust to local minima and, in general, better results are obtained. Moreover, the method is less influenced by the presence of non-overlapping regions and usually requires less iterations compared to ICP.

4 Cycle Minimization

One-to-one alignment of views in a sequence causes a drift that is propagated throughout the sequence. Hence, some techniques have been proposed to reduce the propagating error benefiting from the existence of cycles and re-visited regions and considering the uncertainty in the alignment. This sort of techniques is classified into analytic and statistic, as shown in Table 3 and explained in the following paragraphs.

Analytic minimization: In order to minimize the propagating error, some authors have improved their algorithms by adding a final step that aligns all the acquired views at the same time. These approaches spread one-to-one pair-wise registration errors throughout the sequence of views. Early approaches proposed the aggregation of subsequent views in a single metaview, which is progressively enlarged each time another view is registered [7]. Here, the main constraint is the lack of flexibility to re-register views already merged in the metaview. Some modifications of metaview approach have been presented to improve the efficiency of the algorithm [31] [27]. A different multi-view approach proposes a multi-view registration technique based on the graph theory: views are associated to nodes and transformations to edges. Authors consider all views as a whole and align all them simultaneously [19] [32]. Analytic methods based on the metaview approaches present good results when initial guesses are accurate and the surface to be registered does not have a large scale. Otherwise, the method suffers a large propagation error producing drift and misalignments and its greedy approach usually falls in local minima. The use of methods based on graphs has the advantage of minimizing the error in all the views simultaneously but they usually require a previous pairwise registration step, which accuracy can be determinant in the global minimization process. Besides, closing the loop

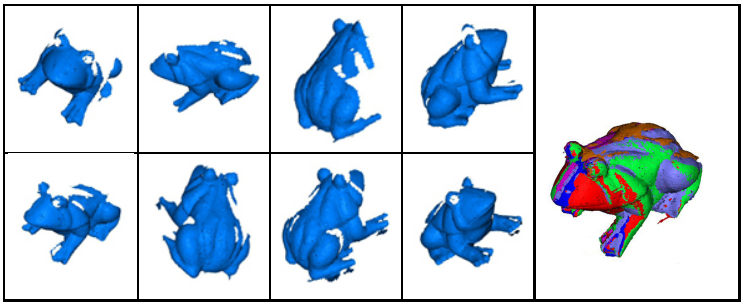

Fig. 1. Multi-view registration of multiple 3D views of a ceramic frog out in our lab

strategies provide trustworthy constraints for error minimization but require a huge amount of memory and usually involve a high computational cost.

Statistic minimization: The same problem of registering 3D views in a sequence has been also faced by means of a probabilistic approach (statistic techniques), especially in mobile robot navigation. The technique receives the name of Simultaneous Localization and Mapping (SLAM) since both the pose and the structure of the environment are estimated simultaneously. The main difference compared to analytic multi-view is that the uncertainty in the measure is not neglected. Hence, two main groups of techniques have been considered depending on the way of representing such uncertainty: a) Gaussian filters and b) non-parametric filters. Both Kalman Filter (KF) for linear systems and Extended Kalman Filter (EKF) for non-linear systems are undoubtedly the most well-known Gaussian filters. Both consist in two main steps: a) Prediction, which estimates the current state by using the temporal information of previous states; and b) Update, which uses the current information provided by robot on-board sensors to refine prediction. Whenever a landmark is observed by the on-board sensors of the robot, the system determines whether it has been already registered and updates the filter. Hence, when part of the scene is revisited, all the gathered information from past observations is used by the system to reduce the uncertainty in the whole mapping, strategy known as closing the loop. Besides, mobile robot localization and mapping has also been tackled by using non-parametric filters such as histogram filter or particle filter. The main advantage compared to Gaussian filters is the possibility of dealing with multimodal data distribution, so that multiple values (particles) are used to represent the belief [35] [9]. Nevertheless, note that Gaussian filters have a polynomic computational cost whereas the computational cost of a non-parametric filter may be exponential. In the presence of large environments in which tons of data are gathered, Gaussian filters state vectors increase considerably leading to inefficiency in terms of computational cost. Similar problems appear using non-parametric filters such as the particle filter. Hence, some authors have proposed different techniques to cope with computational cost and memory size [16] [22]. This drawback can be solved by using methods based on building submaps [4] which present more robustness against uncertainty compared to methods based

Table 3. Cycle minimization techniques. R: Restricted (some DOF are constrained in a limited range); TOF: Time-of-flight; LT: Laser Triangulation; DLP: Digital Light Projector.

Technique			author	DOF	sensor	scene
Cycle minimization	Analytic (Multiview)	Iterative lineal	Bergevin, 1996 [1]	6	monocular	object
			Huber, 2003 [19]	6	LT	object
			Pulli, 1999 [31]	6	LT	object
			Sharp, 2004 [32]	6	DLP	indoor
		robust	Nüchter, 2004 [27]	6	TOF	outdoor
			Masuda, 2001 [25]	6	LT	object
			Pollefeys, 2000 [30]	6	monocular	outdoor
	Statistic	Gaussian	Guivant, 2000 [16]	6	TOF	outdoor
			Martinelli,2005 [24]	6R	TOF	indoor
			Liu, 2003 [22]	6R	TOF	outdoor
			Bosse, 2003 [4]	6	TOF	outdoor
			Estrada, 2003 [10]	6R	TOF	outdoor
		Non Parametric	Davison, 2003 [9]	6	monocular	indoor
			Montemerlo, 2002 [26]	6R	TOF	outdoor

on a unique global map. Some methods impose global restrictions for global map joining, providing accurate solutions in the presence of short loops [12]. However, loop consistency constraints used in methods such as Hierarchical SLAM [10] can be essential in order handle larger loops and prevent inconsistency and misalignments in the final map.

In summary analytic methods are the most common in high-resolution object reconstruction by means of multi-view registration techniques. Although multi-view registration methods have demonstrated to provide accurate solutions, misalignments can appear in the presence of featureless environments, symmetries and smooth objects. Besides, statistical methods are the most used in 3D mapping in mobile robot navigation. The advantage of statistical methods is in their performance in the presence of less reliable sensors, complex environments and unstructured scenes with few features and landmarks. However, they are not recommended for handling tons of data since the manipulation of large state vectors derives to an inefficient computation.

5 Conclusion

This paper presents a state of the art of the most representative techniques for 6DOF pose estimation and 3D registration of large objects and maps. The most referenced articles over the last few decades have been discussed analyzing their pros and cons and potential applications.

The article is intended to be a guide for any researcher interested in the field. To the best of our knowledge, this article is the first that compares the techniques present in both robotics and computer vision communities, providing new classification criteria, discussing the existing techniques, and pointing out their pros and cons and potential applications.

Acknowledgments. This research has been supported by Spanish Project TIC2003-08106-C02-02.

References

1. Bergevin, R., Soucy, M., Gagnon, H., Laurendeau, D.: Towards a general multi-view registration technique. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(5), 540–547 (1996)
2. Bernardini, F., Martin, I., Mittleman, J., Rushmeier, H., Taubin, G.: Building a digital model of michelangelo's florentine pietà. IEEE Computer Graphics and Applications 22, 59–67 (2002)
3. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. Trans. on Pattern Analysis and Machine Intelligence 14, 239–256 (1992)
4. Bosse, M., Newman, P., Leonard, J., Soika, M., Feiten, W., Teller, S.: An atlas framework for scalable mapping. In: IEEE International Conference on Robotics and Automation, Amherst, MA, USA, vol. 2, pp. 1899–1906 (2003)
5. Carmichael, O., Huber, D., Hebert, M.: Large data sets and confusing scenes in 3-d surface matching and recognition. In: International Conference on 3-D Digital Imaging and Modeling, pp. 258–367, Ottawa, Ont. Canada (October 1999)
6. Chen, C.-S., Hung, Y.-P., Cheng, J.-B.: A fast automatic method for registration of partially overlapping range images. In: International Conference on Computer Vision, pp. 242–248, Bombay (January 1998)
7. Chen, G., ad Medioni, Y.: Object modeling by registration of multiple range images. Int. Conf. on Robotics and Automation 3, 2724–2729 (1991)
8. Chua, C.S., Jarvis, R.: Point signatures: A new representation for 3d object recognition. International Journal of Computer Vision 25(1), 63–85 (1997)
9. Davison, A.J., Mayol, W.W., Murray, D.W.: Real-time localization and mapping with wearable active vision. In: Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reallity, pp. 18–27 (2003)
10. Estrada, C., Neira, J., Tardos, J.D.: Hierarchical slam: real-time accurate mapping of large environments. IEEE Transactions on Robotics 21(4), 588–596 (2005)
11. Folkesson, J., Christensen, H.: Outdoor exploration and slam using a compressed filter. Int. Conf. on Robotics and Automation 1, 419–426 (2003)
12. Folkesson, J., Jensfelt, P., Christensen, H.I.: Graphical slam using vision and the measurement subspace. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3383–3390, Edmundton, Canada (August 2005)
13. Gagnon, H., Soucy, M., Bergevin, R., Laurendeau, D.: Registration of multiple range views for automatic 3-d model building. In: Computer Vision and Pattern Recognition, pp. 581–586 (June 1994)
14. Greenspan, M., Godin, G.: A nearest neighbor method for efficient icp. In: Third International Conference on 3-D Digital Imaging and Modeling, pp. 161–168, Quebec, Canada, May-June (2001)
15. Guidi, G., Beraldin, J.-A., Atzeni, C.: High-accuracy 3-d modeling of cultural heritage: The digitizing of donatello's "maddalena". IEEE Transactions on Image Processing 3, 370–380 (2004)
16. Guivant, J.E., Nebot, E.M.: Optimization of the simultaneous localization and map building algorithm for real time implementation. IEEE Transactions on Robotics 3(17), 242–257 (2000)
17. Harris, C.J., Stephens, M.: A combined corner and edge detector. In: Fourth Alvey Vision Conferences, pp. 147–151 (1988)
18. Huang, T.S., Faugeras, O.D.: Some properties of the e matrix in two-view motion estimation. Pattern Analysis and Machine Intelligence 11(12), 1310–1312 (1989)

19. Huber, D., Hebert, M.: Fully automatic registration of multiple 3d data sets. Image and Vision Computing 21(7), 637–650 (2003)
20. Johnson, A.E., Hebert, M.: Using spin images for efficient object recognition in cluttered 3d scenes. PAMI 21(5), 433–449 (1999)
21. Jost, T., Hugli, H.: A multi-resolution scheme icp algorithm for fast shape registration. In: First International Symposium on 3D Data Processing Visualization and Transmission, pp. 540–543 (2002)
22. Liu, Y., Thrun, S.: Results for outdoorslam using sparse extended information filters. ICRA, USA 1, 1227–1233 (2003)
23. Lowe, D.G.: Object recognition from local scale-invariant features. In: Int. Conf. on Computer Vision ICCV, pp. 1150–1157, Corfu, Greece (September 1999)
24. Martinelli, A., Tomatis, N., Siegwart, R.: Some results on slam and the closing the loop problem. In: IROS, pp. 2917–2922, Lausanne, Switzerland (August 2005)
25. Masuda, T.: Generation of geometric model by registration and integration of multiple range images. In: Third International Conference on 3-D Digital Imaging and Modeling, pp. 254–261 (May 2001)
26. Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B.: Fastslam: A factored solution to the simultaneous localization and mapping problem. In: National Conference on Artificial Intelligence, pp. 593–598, Vancouver, BC (July 2002)
27. Nüchter, A., Surmann, H., Lingemann, K., Hertzberg, J., Thrun, S.: 6d slam with an application in autonomous mine mapping. IEEE International Conference on Robotics and Automation 2, 1998–2003 (2004)
28. Nister, D., Naroditsky, O., Bergen, J.: Visual odometry. Computer Vision and Pattern Recognition 1, 652–659 (2004)
29. Park, S.-Y., Subbarao, M.: A fast point-to-tangent plane technique for multi-view registration. In: 3-D Digital Imaging and Modeling, pp. 276–283 (2003)
30. Pollefeys, M., Koch, M.R., Vergauwen, M., Van Gool, L.: Automated reconstruction of 3d scenes from sequences of images. Photogrammetry and Remote Sensing 55, 251–267 (2000)
31. Pulli, K.: Multiview registration for large data sets. In: International Conference on 3-D Digital Imaging and Modeling, pp. 160–168 (October 1999)
32. Sharp, G., Lee, S., Wehe, D.: Multiview registration of 3d scenes by minimizing error between coordinate frames. In: PAMI, pp. 1037–1050 (2004)
33. Stamos, I., Leordeanu, M.: Automated feature-based range registration of urban scenes of large scale. IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2, 555–561 (2003)
34. Tardos, D., Neira, J., Newman, P., Leonard, J.: Robust mapping and localization in indoor environments using sonar data. The International Journal of Robotics Research 21(4), 311–330 (2002)
35. Fox D Thrun, W., Burgard, S.: Probabilistic Robotics (2005)
36. Triebel, R., Burgard, W.: Improving simultaneous mapping and localization in 3d using global constraints. National Conference on Artificial Intelligence 3, 1330–1335 (2005)
37. Trucco, E., Fusiello, A., Roberto, V.: Robust motion and correspondences of noisy 3-d point sets with missing data. Pattern Recognition Letters 20(9), 889–898 (1999)
38. Wyngaerd, J.V.: Combining texture and shape for automatic crude patch registration. In: Int. Conf. on3-D Digital Imaging and Modeling, pp. 179–186 (2003)
39. Zhang, Z., Luong, Q.-T., Faugcras, O.: Motion of an uncalibrated stereo ring: Self-calibration and metric reconstruction. In: IEEE Transactions on Robotics and Automation, pp. 103–113 (1996)