Proceedings of the of the 17th Conference on Computer Science and Intelligence Systems pp. 531–534

DOI: 10.15439/2022F28
ISSN 2300-5963 ACSIS, Vol. 30

A Modified ICP Algorithm Based on FAST and Optical Flow for 3D Registration

Konrad Koniarski
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
Email: konrad.koniarski@gmail.com

Andrzej Myśliński
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
Email: andrzej.myslinski@ibspan.waw.pl

***Abstract*—This paper presents a modified Iterative Closest Point (ICP) algorithm based on a suitable selection of initial points and local optical flow to speed up registration of static scenes with high accuracy. The biggest disadvantages of using standard ICP algorithm are appropriate initialization and effective matching point step in each iteration. In the proposed modification we deal with these problems and optimize this method for Augmented Reality application. As this application uses RGB-D images sequence the changes between consecutive key-frames are small. Therefore only small subset of the source image key-points is selected using scale-space pyramid and FAST approaches. It leads to the significant reduction of the number of the processed image points. Since the point matching technique using local optical flow is applied, in each optimization step of ICP the costly point matching procedure can be abandoned. The proposed approach has been validated by the numerical examples.**

I. Introduction

THE reconstruction of the geometry of the environment from a movable camera is a well studied problem in the category of computer vision topics. In theory, the determination of the trajectory is already possible with the use of only a few reference points traced in real time [1]. However, the smaller number of tracked points makes the solution more sensitive to data noise. Recently, we have seen significant progress in the development of methods for dense 3D reconstruction from many images. Unfortunately, many of these proposed approaches are not able to work in real time [2]. In addition, they usually require a large number of calibrated images, making them unsuitable for live reconstruction from one movable camera. On the other hand, there are many approaches for the reconstruction of dense depth maps from pairs of images [3]. While these approaches have been shown to provide excellent results in dense depth estimation, they are usually computationally too expensive for real-time applications.

In this article we propose modification of the first block of Iterative Closest Point (ICP) method - looking for a match of the identities of the points in each iteration. It has been replaced with the previous step using local optical tracking. This approach, perhaps surprisingly, does not involve a significant loss of speed, but allows a significant reduction in the amount of data while increasing accuracy. The modified ICP method was designed for Augmented Reality (AR) applications. AR is the computer vision system integrating computer generated virtual information with the real world environment in the form of image, sound or video. The added information, usually virtual objects, has to be precisely aligned with the real world. Image registration is the key element of AR algorithm. The ICP method was first presented by Besl and Mackay in 1992 (see [4]). This method is more concept then solid algorithm. The first step, i.e., the initial selection of the points is the most important step to enforce ICP method to be convergent to global rather then to local minimum. There is always a trade off between using feature points or dense data. In literature ICP approach is based on feature extracting methods as in [5]. It leads to long computational time and low quality of point cloud to be augmented in final step of AR image generation. Therefore in this paper, we present different approach than proposed in literature for initial points selection and matching based on scale-space and Features from Accelerated Segment Test (FAST) [6], [7] approaches as well as local optical flow method. The selected set of points may contain outliers. However, in the proposed method there is no separate mechanism for removing such points. Outliers are removed in the optical tracking procedure if they do not meet the stability criteria of this procedure. As a result, we get a set of points that no longer contain outliers. Finally in the last step the error metric is minimized to find the six parameters of the transformation, i.e. the rotation matrix and the translation vector. The main focus is on the speed of convergence and the accuracy of the final transformation and the application it to construct AR image. The performance of the ICP method depends mainly on the data and proper initialization. If some a-priori assumption concerning the similarity of the frames can be made, the quality of matching is much better. Therefore in this paper, we assume that point clouds are constructed from two following frames of a video sequence. The pixel brightness is also assumed not to change significantly on those two consecutive frames.

II. Related Work

A. Geometric Registration

Most of the registration methods operate on candidate correspondences. Popular method use point to point matches based on local geometric descriptors [8], other defines correspondence on pairs or tuples of points [9]. When candidate correspondence are collected, alignment is estimated attractively from sparse subset to correspondence. This iterative process is typically based on variant of randomized algorithms like RANSAC [8], [9]. When the data is noisy and the surfaces only partially overlap, existing pipelines often require many iterations to sample a good correspondence set and find a good reasonable alignment. In many applications we have a priori knowledge that stream of consecutive data observation has relatively small transformations. This approach is know in literature as local refinement where rough initial alignment is known and the result is tight registration usually based on denser correspondence compared to global alignment [10]. ICP method and its modifications are popular for local refinement. The simplest algorithm of the ICP starts with initial alignment and alternates between establishing correspondence via find the closest point and recalculate the alignment based on the current correspondence set. ICP can give an accurate result when initiated near the optimal position, but it is unreliable without such initialization. In many papers [11], [12] are explored different approaches to increase ICP sensitivity to local optima. There are many modification of ICP method based on correspondence or transformation parameter estimation step modifications [5]. Park [11] proposed modification where he used geometry as well as intensity of RGB color value. This approach is valid when registration use data stream from the source when conditions are not changed like frame stream where camera intrinsic parameters are the same. The accumulated 3D model can be either in the form of a volumetric representation [12], a 3D point cloud [13] or a set of depth maps.

B. Optical flow of feature points

Optical flow is popular method of image processing when there is a need to know the movement (speed and direction) of the part of the image. According to Akpinar et al. [14], optical flow estimation algorithms can be grouped according to the theoretical approach while interpreting optical flow. These are differential techniques, region-based matching, energy-based methods and phase-based techniques. There are two groups of optical flow methods. The first is local optical flow introduced by Lucas-Kanade [15], and the second global optical flow introduced by Horn and Schunck [16]. In this paper we mostly focus on geometric aspect of registration then local optical flow is more suitable [7], [17]. Input data stream consists color RGB intensity frames and depth information.

III. Proposed Algorithm

A. Overview

Our goal is to calculate the rigid body motion transition T consisting of translation t and rotation R that minimizes the difference between the two sets of points O and M. In the next subsections, the process of locating, matching, and filtering the appropriate feature points is described and the following section presents the proposed pose estimation calculation.

B. Finding and Matching Visual Features

An RGB-D image consists of a color image I and a depth image D recorded in the same coordinate frame. We assume to have a pair of RGB-D (I_i, D_i) and (I_j, D_j) images and an initial T^0 transformation that roughly aligns (I_i, D_i) to (I_j, D_j). Also $p = (u, v)^T$ is the a pixel in (I_i, D_i) and $p' = (u', v')^T$ is the corresponding pixel in (I_j, D_j). The goal is to find the optimal transformation that densely aligns the two RGB-D images. Here we assume that the individual frames are not distant in time, so that the color intensity of the pixels does not change rapidly and the initial match T^0 can be equal to the identity matrix. The first step of registration is to select feature points using FAST method on RGB image. For image I FAST point detecting method gives set of feature points P. In order to avoid the problem with the scale and the high speed of moving individual points, a scale pyramid is built. The original RGB image I is used in first layer, then the size of image is divided by 2 and the resultant image, the same procedure is repeated until the image becomes too small. As a result the set $N = \{P_k\}$ of points is obtained where k is number of layers in pyramid. Then local optical flow method is used for tracking of selected points $N_{i->j}$ from image I_i on consecutive image I_j. Let $O_{i->j}$ be the set of successfully tracked points and $C_{i->j}$ theirs movement vectors. Then the projections π of the RGB-D image pixel $p = (u, v)^T, d = D(p)$ over 3D space is done using

$$\Pi(u, v, d) = [\frac{(u - c_x)d}{f_x}, \frac{(v - c_y)d}{f_y}, d, 1]^T, \tag{1}$$

where f_x and f_y are the camera focal lengths and (c_x, c_y) is the principal point. The inverse projection function π^{-1} is defined as follow

$$\pi^{-1}(x, y, z, 1) = (\frac{xf_x}{z} + c_x, \frac{yf_y}{z} + c_y, z)^T. \tag{2}$$

Using the projection π over points set $O_{i->j}$ the 3D point set is obtained.

$$W_{i->j} = \pi(O_{i->j}) \tag{3}$$

In practice, the depth component in an RGB-D image need not always be defined. Then such a pixel cannot be projected into 3D space. In that case such pixels are not used for 3D projection. Set $W_{i,j}$ is also called point cloud. The photometric objective is to find transformation T satisfying:

$$p = \pi^{-1}(T\pi(p', D(p'))). \tag{4}$$

Projection of pixels p and p' should be the same point in 3D space

$$\pi(p, D(p)) = T\pi(p', D(p')). \tag{5}$$

C. Pose estimation

The ICP consist of two steps correspondence estimation and transformation parameter estimation. In the first iteration we consider two consecutive image frames: current (I_i, D_i) and registered (I_j, D_j). The objective of correspondence estimation step of ICP is to build mapping function ϕ which define correspondence between I_i and I_j

$$p = \phi(p'). \tag{6}$$

In the proposed method function ϕ is defined by $C_{i->j}$ and it does not have to be recalculated at each loop step. The transformation parameters estimation is done by looking for transformation T that minimize objective function

$$\begin{aligned} T_{n+1} &= \operatorname{argmin}_T \sum_{i=1}^{O} \| \, \pi(p_i, D(p_i)) - \\ &\qquad T_n \pi(p'_i, D(p'_i)) \, \|^2 \\ &= \operatorname{argmin}_T \sum_{i=1}^{O} \| \, \pi(\phi(p'_i), D(\phi(p'_i))) - \\ &\qquad T_n \pi(p'_i, D(p'_i)) \, \|^2, p_i, p'_i \in O \end{aligned} \tag{7}$$

where n is iteration step counter. The computations may be completed when the predetermined number of iterations have been performed or when the estimate of T is sufficient. Optimization problem (7) is solved using Gauss-Newton method. In each iteration, we linearize T locally as a 6 elements vector $\xi = (\omega_1, \omega_2, \omega_3, t_1, t_2, t_3)$. ξ contains rotation component ω and a translation component t.

$$T \approx \begin{pmatrix} 1 & -\omega_3 & \omega_2 & t_1 \\ \omega_3 & 1 & -\omega_1 & t_2 \\ -\omega_2 & \omega_1 & 1 & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} T^n \tag{8}$$

Using the Gauss-Newton optimization scheme, we calculate ξ by solving the linear system

$$J_r^T J_r \xi = -J_r^T r \tag{9}$$

where r is the residual vector and J_r is Jacobian. In each optimization step n, r and J_r are calculated. Then T_n is obtained from ξ (see equation 8). Next T is updated by T_n using transformation to $SE(3)$ group.

IV. Numerical experiments

Publicly available sequenced RGB-D framesets were used for the qualitative evaluation of the method proposed in this paper. All datasets used for test were published by Sturm [10]. RGB-D frames were registered using Microsoft Kinect device. Along with the images, the proposed benchmark dataset also provides the real trajectory taken by the camera acquired by an external, high-precision motion interception system.

A. Performance comparison

The performance comparison is summarized in the Table I. Processing time and error were calculated for different RGB-D sequences. Processing time was calculated as average of the time for registration consecutive frames. Mean error was calculated as the difference between benchmark trajectory and method calculated trajectory. Proposed method was compared to standard ICP implementation.

Algorithm 1 RGB-D images alignment

```
Require: Pair of RGB-D images $(I_i, D_i), (I_j, D_j)$,
    initial transformation $T^0$
Ensure: $T$ registration transformation from frame $i$ to $j$
  Build scale pyramid for RGB images $I_i, I_j$
  Calculate feature points using FAST method $N = \{P_k\}$
  Calculate local optical flow for points $N$ and get $O_{i->j}$, $C_{i->j}$
  Project $N$ into 3D space using projection equation (1)
  while not converged do
    $r \leftarrow 0$, $J_r \leftarrow 0$
    Use $C_{i->j}$ as correspondence between points from frames $i$ and $j$
    Solve equation (9) to get $\xi$
    Update $T$ using equation (8) and map to $SE(3)$
  end while
```

TABLE I: Speed comparison of proposed method and ICP

Dataset	Proposed method		ICP method	
	Processing time [ms]	Mean error [m]	Processing time [ms]	Mean error [m]
fr1/xyz	0.123	0.251	0.182	0.511
fr1/rpy	0.128	0.262	0.195	0.25
fr2/xyz	0.131	0.17	0.152	0.19
fr2/rpy	0.190	0.291	0.211	0.31

B. Scene reconstruction

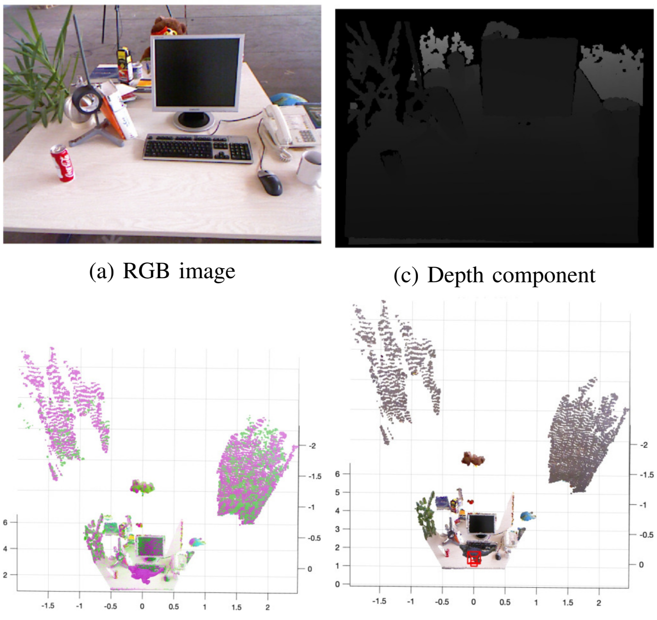
Single frame reconstruction is presented on Figure 1. Top row presents RGB image and depth component that is used for projection. Bottom row contains 3D projections of frame and camera location related to scene.

C. Augmented Reality application

Proposed method was used to build AR system as example of usage. Image 2 presents RGB frames and model position for different views as well as final AR image. In this case, the model image in the appropriate position is presented on the RGB image frame. The problem of obscuring an object added by scene elements is not considered.

V. Conclusion

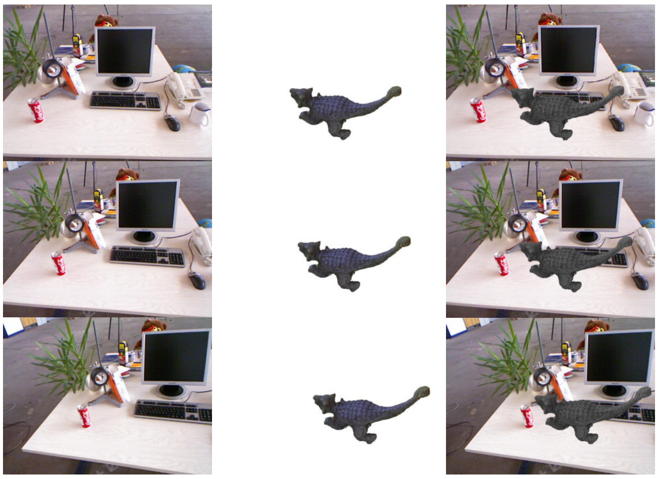
The article presents an approach to the problem of image registration in a multi-frame sequence. The proposed method uses information obtained by an RGB-D camera moving in a static environment and is compared to the ICP method, another popular approach often used in similar applications. Its performance was assessed in terms of accuracy and processing time on the benchmark data sets. The proposed method achieved an average accuracy of 12% better than ICP and the processing time on average 37% better than ICP. The proposed algorithm works for both consecutive images and multiple image frames. The novelty of the use of local optical flow allows for better results than in the case of the classic ICP algorithm. The application of the proposed method has been shown on the example of AR. Information from previous frames is accumulated in the form of a point cloud. This



(a) RGB image (c) Depth component

(b) 3D projection (d) Camera position

Fig. 1: Single frame projection into 3D scene.



(a) Frame (b) Model position (c) AR image

Fig. 2: AR system based on proposed registration method.

feature can be used in applications where 3D reconstruction plays an important role. In this article we not deal with the loop closure problem. However this method potentially could be used for simultaneous location and mapping algorithm (SLAM) application as well. This will be studied in future work.

References

[1] Q. Y. Zhou, J. Park, and V. Koltun, "Fast global registration," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9906 LNCS. Springer Verlag, 2016. doi: 10.1007/978-3-319-46475-6_47. ISBN 9783319464749. ISSN 16113349 pp. 766–782.

[2] A. W. Fitzgibbon, "Robust registration of 2D and 3D point sets," *Image and Vision Computing*, vol. 21, no. 13-14, pp. 1145–1153, 12 2003. doi: 10.1016/J.IMAVIS.2003.09.004

[3] C. Kerl, J. Sturm, and D. Cremers, "Dense visual SLAM for RGB-D cameras," in *IEEE International Conference on Intelligent Robots and Systems*, 2013. doi: 10.1109/IROS.2013.6696650. ISBN 9781467363587. ISSN 21530858 pp. 2100–2106.

[4] Y. He, B. Liang, J. Yang, S. Li, and J. He, "An Iterative Closest Points Algorithm for Registration of 3D Laser Scanner Point Clouds with Geometric Features," *Sensors 2017, Vol. 17, Page 1862*, vol. 17, no. 8, p. 1862, 8 2017. doi: 10.3390/S17081862. [Online]. Available: https://www.mdpi.com/1424-8220/17/8/1862/htmhttps://www.mdpi.com/1424-8220/17/8/1862

[5] E. Marchand, H. Uchiyama, and F. Spindler, "Pose Estimation for Augmented Reality: A Hands-On Survey," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 12, pp. 2633–2651, 12 2016. doi: 10.1109/TVCG.2015.2513408

[6] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," *Proceedings of the IEEE International Conference on Computer Vision*, vol. II, pp. 1508–1515, 2005. doi: 10.1109/ICCV.2005.104

[7] Koniarski, "Augmented reality using optical flow," *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015*, pp. 841–847, 10 2015. doi: 10.15439/2015F202. [Online]. Available: https://fedcsis.org/proceedings/2015/drp/202.html

[8] Mahesh and M. V. Subramanyam, "Automatic feature based image registration using SIFT algorithm," in *2012 3rd International Conference on Computing, Communication and Networking Technologies, ICCCNT 2012*, 2012. doi: 10.1109/ICCCNT.2012.6396024

[9] A. Fontes and J. E. B. Maia, "Visual Odometry for RGB-D Cameras," 3 2022. doi: 10.48550/arxiv.2203.15119. [Online]. Available: https://arxiv.org/abs/2203.15119v1

[10] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *IEEE International Conference on Intelligent Robots and Systems*, 2012. doi: 10.1109/IROS.2012.6385773. ISBN 9781467317375. ISSN 21530858 pp. 573–580.

[11] J. Park, Q. Y. Zhou, and V. Koltun, "Colored Point Cloud Registration Revisited," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017-Octob, pp. 143–152, 12 2017. doi: 10.1109/ICCV.2017.25

[12] R. A. Newcombe, R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," in *2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011*, 2011. doi: 10.1109/ISMAR.2011.6092378. ISBN 9781457721830 pp. 127–136. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.221.100

[13] K. Koniarski and A. Myśliński, "Feature Point Cloud Based Registration in Augmented Reality," *Lecture Notes in Networks and Systems*, vol. 364 LNNS, pp. 418–427, 12 2021. doi: 10.1007/978-3-030-92604-5_37. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-92604-5_37

[14] S. Akpinar and F. N. Alpaslan, "Optical flow-based representation for video action detection," *Emerging Trends in Image Processing, Computer Vision and Pattern Recognition*, pp. 331–351, 1 2015. doi: 10.1016/B978-0-12-802045-6.00021-1

[15] B. D. Lucas and T. Kanade, "Iterative Image Registration Technique With an Application to Stereo Vision." vol. 2, 1981, pp. 674–679. [Online]. Available: https://www.researchgate.net/publication/215458777_An_Iterative_Image_Registration_Technique_with_an_Application_to_Stereo_Vision_IJCAI

[16] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, 8 1981. doi: 10.1016/0004-3702(81)90024-2

[17] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods," *International Journal of Computer Vision*, vol. 61, no. 3, pp. 1–21, 2 2005. doi: 10.1023/B:VISI.0000045324.43199.43