Finding a Strong Key Point Correspondences in Large-Scale Images and Depth Maps

Andrey Dashkevich[1][0000−0002−9963−0998], Darya Vorontsova[1][0000−0001−7868−0067], and Sergii Rosokha[2][0000−0001−8239−6122]

[1] National Technical University "Kharkiv Polytechnic Institute", Kharkiv 61002, Ukraine
{dashkewich.a, dvorontso}@gmail.com
[2] NPP "SPECPOZHTEHNIKA" Ltd, Kharkiv 61017, Ukraine

Abstract. Depth map fusion is an essential task in multi-view stereo vision and structure from motion methods, but there are some issues of the process of depth maps matching of large-scale outdoor scenes, which are acquired by unmanned aerial vehicles. First, they can have low depth resolution due to long distance from a scene to a camera, in the second, as a consequence of above problem and some camera defects, they can have some noise. In our work we propose an approach of matching the depth maps based on a local shape and a color similarity correspondences. The proposed idea is to find similar regions by calculating local key points in small patches of two images by ORB feature detector. Then we find corresponding regions assuming spatial proximity of movement vectors of key points, this is done by the search of the main vector direction and length in feature space. We test our approach with the help of video sequences acquired through a consumer multicopter.

Keywords: Depth Map Fusion · Multi-view Stereo Vision · Structure From Motion · Unmanned Aerial Vehicles · Key Points · ORB Feature Detector · Spatial Similarity · Nearest Neighbors Search.

1 Introduction

At present there is an increase in quantity of stereo vision method applications in the production, robotics, in systems of virtual and augmented reality, geoinformation systems, computer games and simulating. This brings about the need for creation and development efficient methods and algorithms for solving the problems of stereo vision. The present paper represent a process of finding the effective ways to solve the problem of comparing large-scale depth maps obtained by unmanned aerial vehicles. At the same time, the obtained depth maps can have a low depth resolution due to the large distance from the camera to the scene and, as a result, such depth maps can have noise.

There are several approaches for solving the problem of finding strong correspondences of images and depth maps which aimed to increase accuracy of 3d model reconstruction. Among them are various variations of the iterative closest points (ICP) [1], robust point matching [2], methods based on probability

theory, correlation and Gaussian mixture models [3], coherent point drift [4], method of sorting the correspondence space [5]. In consequence of simpleness of realization and low computational complexity, various modifications of the ICP algorithm have found wide application [6]. A detailed comparative review and ways to increase the accuracy of algorithms for ICP matching can be found in [7]. In [8], the problem of sampling point's assortment for increase the accuracy of ICP algorithms was considered. As the ICP algorithms are the greedy optimization techniques, the initial matching has great influence on registration of the point sets. Therefore we need methods to find more precise initial approximation of surface matching to increase probability of convergence to the global minimum. The main contribution of presented work is the algorithm of finding of initial correspondences between two surfaces which leads to decreasing of time complexity of the ICP algorithms and increasing of ability to converge to global minimum.

2 Method Overview

Given a set of n consecutive frames in $F = f_1, f_2, ..., f_n$ and a set of $n-1$ depth maps $M = m_1, m_2, ..., m_{n-1}$, acquired from pairs of adjacent images in F. It is necessary to find such pairs of points $p, q \in R^3$ that for each point $p_j \in m_i$ and the corresponding point $q_k \in m_i + 1, i = 1..n - 1, j, k = 1..|m|, |\cdot|$ – number of pixels in the depth maps will be: $q_k = Ap_j + t$, where A is the 3×3 rotation matrix, t is the 3×1 transfer vector. Alongside strict correspondences between points of neighboring depth maps are unknown in advance. Thus, the task of finding the transformation A and t can be divided into two sub-tasks: finding the indices j, k for the corresponding pairs of points p, q on the depth maps and comparison of the obtained point sets.

In our work, we proceed from the assumption that adjacent frames have a small differences of the direction of movement of points, which consists of transference and rotation combination, while camera is significantly distant from the scene, we assume a rectilinear motion. Based on these assumptions, it can be expected that the movement of points in close sections of images will have almost similar values of the motion direction. At the same time, the initial point correspondences are found in the original RGB image space using the robust search of feature descriptions, such as scale-invariant feature transform (SIFT) [9] or oriented FAST and rotated BRIEF (ORB) [10].

We propose the following method for detecting strong correspondences between points on large-scale depth maps:

1. We take set of images F as input. Each image has size $W \times H$.
2. We set the parameters for dividing images into local areas, which include the width W and the height H of the image local part.
3. At each algorithm iteration, we take the images f_i and f_{i+1} from F and the corresponding to them depth maps m_i from M. Thus, we form the structure $data = f_i, f_{i+1}, m_i$.

4. We split f_i and f_{i+1} into t rectangular sections (patches) $p_j, j = 1..t$, $P_i = p_1, p_2, ..., p_t$, where $t = \lfloor W/w \rfloor \cdot \lfloor H/h \rfloor$, $\lfloor x \rfloor$ - integer part of x.
5. The next step is defining arrays of key points K_i and K_{i+1} for each segment p_{i+1} by the ORB algorithm, $K_i = k_j, j = 1..KN$ - number of key points.
6. Initial mapping of points from K_i and K_{i+1} is conducted. As a method of mapping can be used any method of search approximate nearest neighbors algorithm. We used the implementation from the FLANN library [11]. However, at this stage, there are many false-positive pairs, which have strongly influence on the further depth map matching process.
7. To remove incorrectly recognized correspondences of points, we propose a novel approach based on the dividing the parameter space of the key points into a regular grid and spatial indexing of the resulting grid for nearest points search.
8. The obtained exact matches are transferred to the depth maps and they are compared by the ICP method.

Let us demonstrate our approach to determine the exact matches of key points. For each pair of key points we set the vector $v_j = kp_{j+1} - kp_j$, where kp_j – coordinates of k_j in the patch p. Each vector v_j has an angle α and length l and indicates the direction of key point movement between frames.

Assumption 1. In small areas of the image, correctly mapped points will give close values of a and l, and false-positive pairs of points will give arbitrary corresponding values.

Based on this assumption, it is possible to construct a classifier that will find closely spaced points in the parameter space $\alpha - l$. To speed up the finding process of close elements in $\alpha - l$-space, we discretize it by splitting into a regular grid with the number of cells A along the $O\alpha$ axis and the number of cells L along the Ol axis, thus the size of the grid cell is:

$$\alpha_{size} = \frac{\max \alpha}{A} \tag{1}$$

$$l_{size} = \frac{\max l}{L} \tag{2}$$

subject to $\alpha, l \geq 0$

Then we build accumulator Acc array with size $(A \times L)$ and initialize each cell in it with zeros. For each v_j we discretize it α and l:

$$\alpha_{j_{discr}} = \lfloor \frac{\alpha_j}{A} \rfloor \tag{3}$$

$$l_{j_{discr}} = \lfloor \frac{l_j}{L} \rfloor \tag{4}$$

And then:

$$Acc[\alpha_{j_{discr}}][l_{j_{discr}}] = Acc[\alpha_{j_{discr}}][l_{j_{discr}}] + 1 \tag{5}$$

Summary of our approach is provided on the Fig. 1.

Input images and depth maps

Key points detection
(ORB or SIFT descriptors)

Initial matching
with FLANN lib

Movement vectors of
key points

Discretization step
(t=8)

Hash-table
building

cell with maximal
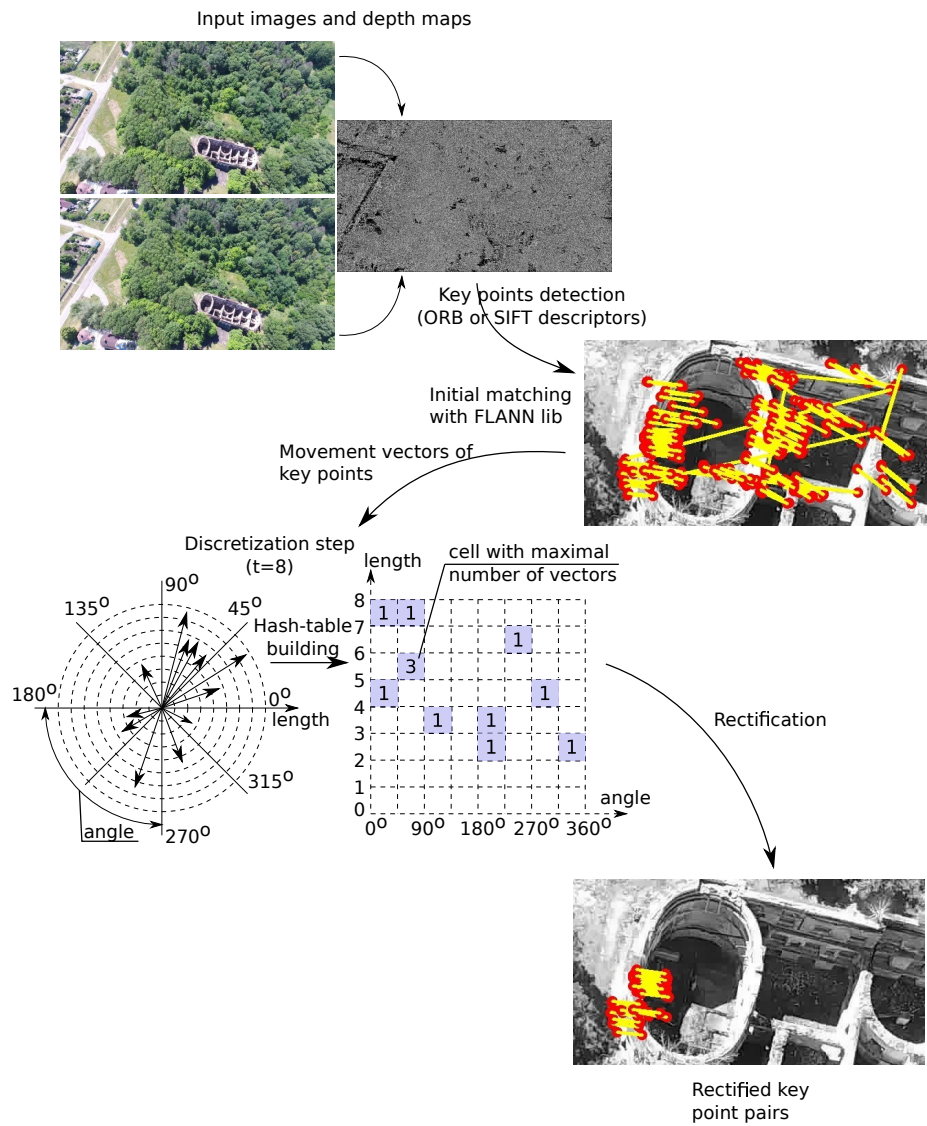number of vectors

Rectification

Rectified key
point pairs

Fig. 1. Illustration of proposed approach.

Due to Assumption 1 the cell with maximal value corresponds to the matched key point pairs and we mark such as a "good" matches. To accelerate computations we use a hash-table H to store indices of key points, the key value of the elements of H we calculate as follows:

$$hash = \alpha_{j_{discr}} \cdot L + l_{j_{discr}} \tag{6}$$

Therefore we can find strong correspondences of key points and their consecutive directions.

3 Experiments

We tested our algorithm on depth maps obtained from a video taken by a UAV with a large, relatively to the size of the seen, depth.

Image data and algorithm parameters:

- We have frames of video with 8 Mpx resolution, so size of our all images and depth maps is $(W \times H) = (4096 \times 2160)$ pixels each.
- Subdivision parameters: $w = W/8$, $h = H/8$.
- Number of key points for ORB detector $K_N = 100$.
- Number of discretized grid cells along axis: $A = L = 20$.

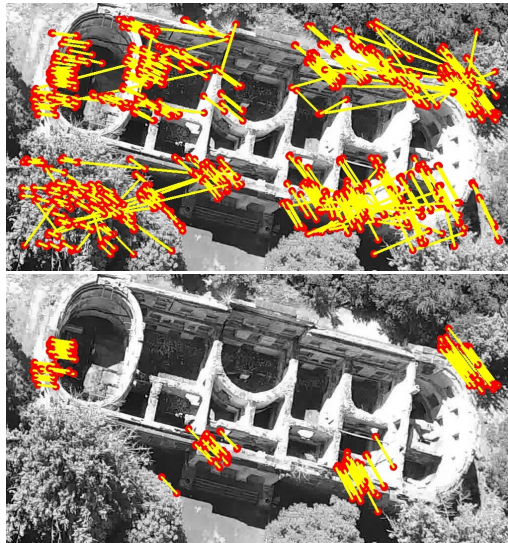Result of depth map rectification by our algorithm is demonstrated in Fig. 2.

Fig. 2. Key point pairs after FLANN matching (top) and after rectification (bottom).

4　Conclusions

In our paper we demonstrate an approach for estimation of strong correspondences between the key points in two sequential video frames. The proposed algorithm can eliminate false-positive occurrences from an array of key points. This is done by finding of major direction of key point movement in parameter space.

The limitation of our approach is that it can delete positive correspondences between the key points, thus, the future work is aimed to improving the quality of correct pairs matching. Another disadvantage is the ability of our algorithm to match key points only in the case of camera or scene movement.

References

1. Bouaziz, S., Tagliasacchi, A., Pauly, M.: Sparse Iterative Closest Point. Computer Graphics Forum. 32, 113–123 (2013). https://doi.org/10.1111/cgf.12178
2. Gold, S., Rangarajan, A., Lu, C.-P., Suguna, P., Mjolsness, E.: New algorithms for 2d and 3d point matching: pose estimation and correspondence. Pattern Recognition. 38 (8), pp. 1019–1031 (1998).
3. Evangelidis, G.D., Horaud, R.: Joint Alignment of Multiple Point Sets with Batch and Incremental Expectation-Maximization. IEEE Transactions on Pattern Analysis and Machine Intelligence. 40, 1397–1410 (2018). https://doi.org/10.1109/TPAMI.2017.2717829
4. Myronenko, A., Song, X.: Point-Set Registration: Coherent Point Drift. IEEE Transactions on Pattern Analysis and Machine Intelligence. 32, pp. 2262–2275 (2010). https://doi.org/10.1109/TPAMI.2010.46
5. Assalih, H.: 3D reconstruction and motion estimation using forward looking sonar. Doctoral dissertation, Heriot-Watt University. 208 p. (2013).
6. Park, S.-Y., Subbarao, M.: An accurate and fast point-to-plane registration technique. Pattern Recognition Letters. 24, 2967–2976 (2003). https://doi.org/10.1016/S0167-8655(03)00157-0
7. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: Proceedings Third International Conference on 3-D Digital Imaging and Modeling. pp. 145–152. IEEE Comput. Soc, Quebec City, Que., Canada (2001).
8. Gelfand, N., Ikemoto, L., Rusinkiewicz, S., Levoy, M.: Geometrically stable sampling for the ICP algorithm. In: Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. pp. 260–267. IEEE, Banff, Alberta, Canada (2003).
9. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. pp. 1150–1157 vol.2. IEEE, Kerkyra, Greece (1999).
10. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: 2011 International Conference on Computer Vision. pp. 2564–2571. IEEE, Barcelona, Spain (2011).
11. Muja, M., Lowe, D.G.: Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In: International Conference on Computer Vision Theory and Application VISSAPP'09). pp. 331–340. INSTICC Press (2009).