Lodz University of Technology

Faculty of Electrical, Electronic, Computer and Control Engineering

Karol Matusiak

Keypoint matching for object detection in 2.5D images

PHD THESIS

ABSTRACT

Supervisor:
prof. dr hab. inż. Paweł Strumiłło
Auxiliary supervisor:
dr inż. Piotr Skulimowski

Łódź 2018

Continuous development of computer image processing algorithms enables their broad applications in many fields, including industry, robotics, automatics and medicine. One of the crucial problems in computer image processing is creation of reliable methods for detection and recognition of objects in the registered scenes. These methods rely on low-level image processing techniques and employ statistical or geometrical image features like lines, corners or textures. The objective is to work out parameters characterising objects that would be invariant to image registering conditions (e.g. illumination contitions) and object orientation and position. One of the broadly used methods is detection of keypoints, which are defined as small image regions satisfying certain conditions, e.g. ubrupt brightness changes. Each keypoint is described by the defining a structure called descriptor, which enables keypoints distinguishing and matching. The set of detected keypoints and its descriptors enable building a simple representation of a scene object in the form of a template. Keypoints matching is a procedure of searching the nearest neighbor between the set of descriptors coming from the template and from the analyzed scene. As literature shows, there have been a numerous research efforts concentrationg on the problem of keypoints detection and matching in 2D images. Still, however, changing the registering conditions and object orientation in a scene constitute a huge challenge for the keypoint detection algorithms. In the conducted research an independent method evaluating repeatability of keypoints detection has been proposed [1]. Usage of depth sensors, which produce images whose points represent the scene depth (called depth map or 2.5D image), together with ordinary cameras registering 2D images proved to be especially promising [2]. To acquire depth images, stereovision cameras, as well as active cameras, which emit light into the scene and then measure reflected light parameters are used. As a result of the conducted research a new detection and keypoint description algorithm, which is based on keypoint location with respect to object edges defined in the depth map, has been developed. .
Following thesis were defined:
Thesis 1: A keypoint descriptor that consists of data about its localization versus object boundaries and data about depth improves the efficiency of keypoints matching.
Thesis 2: Removal of keypoints localized in the regions of object boundaries improves keypoints detection repeatability rate.
Presented algorithm is named as Depth-based feature transform (DBFT). The procedure allows to detect and localize the predefined object model in the analyzed scene. The proposed procedure is based on the SIFT [3] algorithm, but it can be applied to all keypoint detectors and descriptors that provide keypoint orientation information. Firstly, the SIFT keypoint detection and description algorithm is applied to the template 2D image. The next step is to perform the Canny edge detection on the template depth map [4]. A specific implementation of the Canny edge detector was employed that utilizes floating point values. We assume that the position of the keypoint in relation to object edges can be used to improve the performance of keypoint

detection and matching. To take this relation into account the keypoints that are located on the object boundaries are discarded from further calculations (the distance of the keypoint from a boundary must be larger than a given heuristic distance threshold equal to 3 pixels). The next step is to compute, for each keypoint P_1, the distances along the rays protruding from the keypoint location along the four following directions that are determined in relation to the orientation of the keypoint as specified in the SIFT algorithm:

1. ray along keypoint orientation,
2. ray rotated by 90 degrees vs. keypoint orientation,
3. ray rotated by 180 degrees vs. keypoint orientation,
4. ray rotated by 270 degrees vs. keypoint orientation.

The rotations are clockwise. The candidate length value d is incremented iteratively and the end of the ray in point P_2 is checked whether it is positioned on a depth edge:

$$x_{P_2} = x_{P_1} + d \cdot cos(\theta) \tag{1}$$

$$y_{P_2} = y_{P_1} + d \cdot sin(\theta) \tag{2}$$

where:
x_{P_1}, y_{P_1} - x and y position of analyzed keypoint P_1,
x_{P_2}, y_{P_2} - x and y position of analyzed point P_2, that is verified to contain depth edge,
θ - angle of keypoint orientation P_1 [rad],
d - distance of tested ray between P_1 and P_2.

The calculations are performed for each ray separately. All coordinate values are calculated with a sub-pixel precision. However, the binary map of depth edges is stored with a pixel precision. A pixel value is equal zero if there is no depth edge at a given point, and equal to one otherwise. Therefore, a sub-pixel bilinear interpolation is used, while loading coordinate of point P_2. For the case for which the computed ray reaches image boundary without intersecting with the depth edge, the given ray is tagged as undefined and will not be further used in descriptor matching. The final step of the algorithm for creating the object model is storing, for each of the remaining keypoints, the depth-based descriptor that is an 8-element matrix consisting of the two following values for each ray:

— depth value of keypoint position Z in meters,
— distance to the nearest depth edge d in pixels (explicitly -1 value represents not found edge intersection).

The next step is descriptor matching. At the beginning of this procedure, we apply the object model generation algorithm to the scene image. As a result, we obtain the set of keypoints with their SIFT descriptors along with additional depth information as proposed by our method.

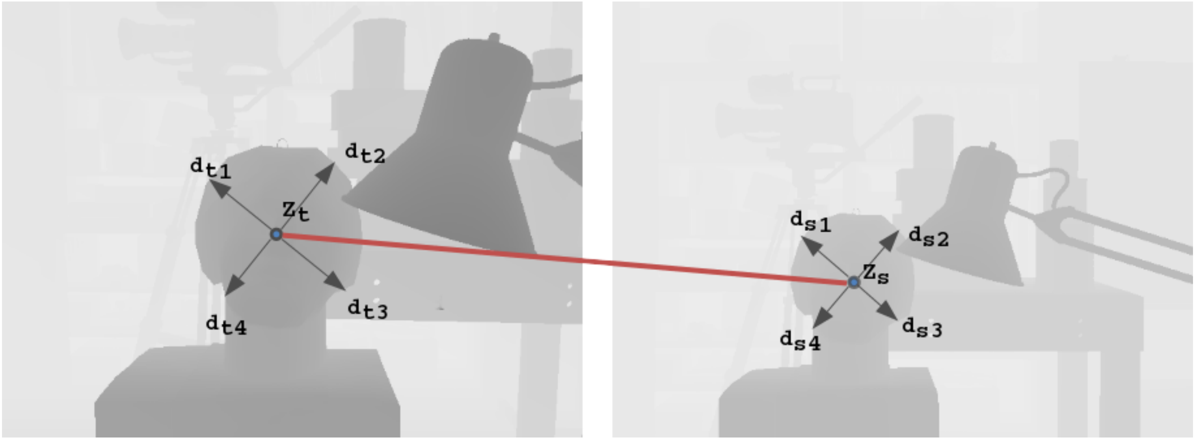

Figure 1: Comparison of distances d_t, d_s to the nearest edges for all rays determined for keypoints in the template image and scene images correspondingly. Disparity maps come from dataset [5]

Next, by a nearest neighbor search, pairs of the most similar SIFT keypoint descriptors between the template and the compared scene are identified. In the next step, the keypoints matches are filtered out with the use of depth and edge distance information. The following matching criterion is verified for all the keypoints and adjacent rays (relative to the main orientation of each descriptor):

$$|d_t - d_s \cdot Z_s/Z_t| < \epsilon \tag{3}$$

where:
d_t - distance from keypoint to the nearest depth edge in the template image [px],
d_s - distance from keypoint to the nearest depth edge in the scene image [px],
Z_t - depth of keypoint in the template image [m],
Z_s - depth of keypoint in the scene image [m].

If the corresponding difference in edge distances for a pair of rays is smaller than ϵ, rays similarity counter is increased. The counter is not incremented if the condition is not met or at least one ray from the pair is not defined. If rays similarity counter achieves a value of 50% of the total number of rays per keypoint, the keypoint's match is ranked as a good match and a bad match otherwise. Visualization of whole process is presented on the Fig. 1.

For evaluation purpose, three experiments were conducted:

— keypoints detection repeatability rate for SIFT and DBFT, with gathered data that contained 4 template object and 12 test scenes
— keypoints matching precision for SIFT and DBFT, with gathered data that contained 7 template objects and 100 test scenes
— keypoints matching precision and recall for PIFT, SIFT and DBFT with data presented in [6]

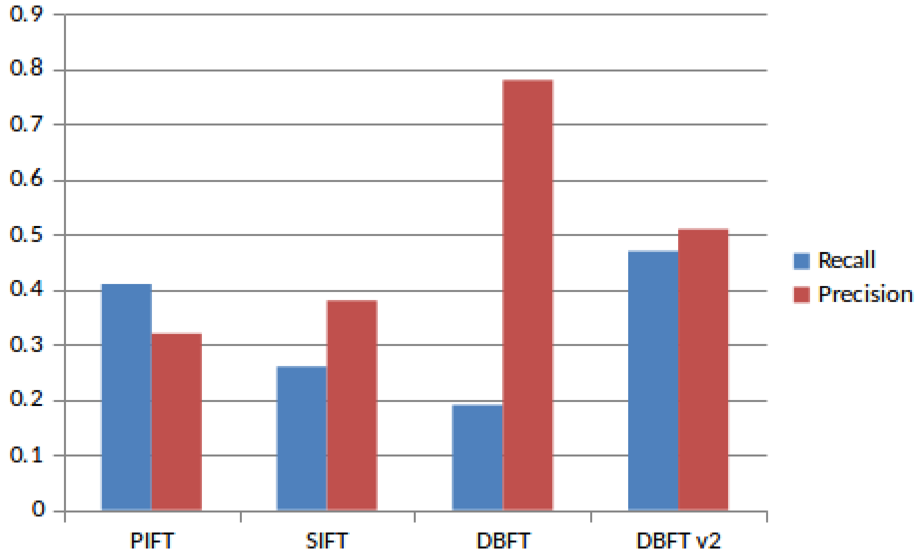

Figure 2: Recall and precision calculated for dataset [6]

Keypoints detection repeatability rate was calculated according the method presented in publication [1]. For the used data, the DBFT algorithm succeeded with average gain of 27% in comparison to the SIFT results. Precision is defined as a ratio of number of true positive matches to the total number of matches. Recall is defined as a ratio of number of true positive matches to all correspondences between images (possible matches). In the second experiment the precision of keypoints matching was improved by 12% in average by using DBFT instead of SIFT. The results of the final comparison are presented in Fig. 2. Usage of the SIFT algorithm with the procedure described in this work results in a significant difference of the recall value which is 63% of the PIFT outcome. On the other hand, the precision result of the SIFT increases in comparison to the PIFT from 0.32 to 0.38. Even higher gain of precision with the cost of decreasing the recall value can be observed for the standard DBFT configuration. The recall values is about 50% of PIFT results, whereas precision is almost 2.5 times larger. It is possible to change parameters of the proposed DBFT algorithm, so it is less restrictive in keypoints matching. For different settings of the DBFT algorithm parameters the result is marked in Fig. 2 as "DBFT v2". In this case the recall value increased by 5% in relation to the PIFT and precision was also at a higher level than the PIFT, in particular 60%.

To sum up, the main achievements of presented work are the following:

— Proposal of a novel approach termed unbiased keypoints detection evaluation procedure, which advances current state of the art in evaluating the performance of the keypoint detection algorithms,

— Design and verification of the keypoints selection algorithm based on the localization of keypoints in the depth map with respect to object edge; this algorithms has improved repeatability rate of the keypoint detection algorithm,

— Development of an algorithm for building and matching keypoints descriptors based on data including the position of the keypoint with respect to object boundaries and depth data. The proposed method has considerably improved the performance of keypoints matching for the tested data sets.

Bibliography

[1] Karol Matusiak, Piotr Skulimowski, and Pawel Strumillo. Unbiased evaluation of keypoint detectors with respect to rotation invariance. *IET Computer Vision*, 11:507–516(9), October 2017.

[2] K. Matusiak, P. Skulimowski, and P. Strumillo. Improving matching performance of the keypoints in images of 3d scenes by using depth information. In *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 1–5, April 2017.

[3] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[4] John Canny. A computational approach to edge detection. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 1986.

[5] Sara Martull, Martin Peris, and Kazuhiro Fukui. Realistic cg stereo image dataset with ground truth disparity maps. pages 40–42, January 2012.

[6] Qinghua Yu, Jie Liang, Junhao Xiao, Huimin Lu, and Zhiqiang Zheng. A novel perspective invariant feature transform for rgb-d images. *Computer Vision and Image Understanding*, 167(C):109–120, February 2018.