Impact of Subpixel Paradigm on Determination of 3D Position from 2D Image Pair

L. Sroba, R. Ravas

The idea of subpixel feature detection is widely used in image processing area in these days and it has a high significance in many practical applications. This paper deals with impact of subpixel paradigm on accuracy of 3D coordinates determination using information related to corresponding 2D image pairs. In the other worlds, there is a study how the subpixel corner points detection could influence the accuracy of 3D reconstruction. For that reason our work contains this comparison for image pairs having various mutual position and resolution using data from given datasets. This contribution could answer the question if the paradigm of subpixel detection could be useful in VSLAM methods and other practical tasks where the precision is a key.

Introduction

The area of corner point's detection is well known and very often used in various areas of image processing and computer vision. The tasks such like motion tracking, object detection and recognition, robot navigation, stereo matching or 3D modelling are the illustrative examples.

There is no exact definition what the corner point is, but except many other statements it could be: point where at least two edges are intersected, point having the smallest radius of curvature or point around which is high change of brightness intensity in all directions.

As it is known, the smallest part of an image is a pixel. We usually cannot access information "between" pixels. But there is possibility to use some mathematical techniques to interpolate or approximate the brightness intensity and find the chosen features in subpixel accuracy.

One of the applications where the localization of corner points is crucial is 3D scene reconstruction. For that there is possible to use the theory of epipolar geometry and fundamental matrix [1].

The accuracy improving character of subpixel detection in case of homography determination is slightly described in [2].

In the next sections there will be the basic principles of pixel and subpixel corner detection described very briefly. Also the theory of fundamental matrix and its relation to obtaining the 3D coordinates will be slightly mentioned.

Pixel and subpixel corner detection

Many corner detectors were invented over the years and the Harris corner detector is one of the most famous. This detector was first time mentioned in [3].The main idea is to find the minimum of brightness intensity difference between chosen parts of an image (marked as W) and shifted parts of image in all directions. There is first-order Taylor series approximation for that purpose used. First step is determination of matrix M as it is shown in (1).

$$M(x,y) = \sum_{W} \left(\begin{bmatrix} I_x \\ I_y \end{bmatrix} \cdot \begin{bmatrix} I_x & I_y \end{bmatrix} \right) \tag{1}$$

The variables I_x and I_y represent the approximations of derivation (also known as differences) in horizontal and vertical direction. As soon as this matrix M is calculated for every pixel in image, the next step is the definition of values contained in matrix C. This matrix has the same size as tested image and is also known as cornerness map. There is a lot of ways how to calculate this matrix, for instance [4]. Last step is looking for the elements in matrix C having the highest values. These points are after global and local thresholding marked as found corner points.

As it is obvious, this algorithm can be used to find corner points in pixel accuracy. Here it will be shortly mentioned two ways how to obtain the subpixel coordinates of corner points. These two approaches were also implemented in our comparison. For some other algorithms of subpixel corner detection see [5].

Both methods using the previously found corner point in pixel accuracy as initial step. Once this point was detected, its position according of first approach [6] is refined to subpixel accuracy by fitting the 2D quadratic surface to the corner strength function in the local neighbourhood and its maximum is found. The equation of surface is following:

$$h(x,y) = ax^2 + bxy + cy^2 + dx + ey + f \tag{2}$$

When the coefficients are calculated, the assumption that the maximum of corner map corresponds to the first derivation of this function equals to zero could lead us to the final corner point subpixel coordinates very easily.

The second approach [7] is basically very similar to previous one. The only difference is that the subpixel shifts are determined for x and y direction separately using quadratic curve equation:

$$h(x,y) = ax^2 + by + c \tag{3}$$

The final subpixel corner point position is combination of both shifts and is calculated by using the same assumptions as before.

3D scene reconstruction

One of the ways how to obtain 3D world coordinates from 2D stereo images is to use the principles of epipolar geometry. The first step is to determine the fundamental matrix. This matrix has 3×3 size and describes a line (epipolar line) on which the corresponding point on the other image must lie. The 8-point algorithm [8] could be used for calculation. Based on this algorithm, there is necessary to find at least 8 pairs of corresponding points in both images. It there is more corresponding points detected and system is overdetermined, the solution can be obtained using the method of least squared for example.

Next step is the decomposition of fundamental matrix into rotation R matrix and translational t vector [9]. It gives us the information about relation between 3D world coordinates and 3D camera coordinates and it is also known as extrinsic camera parameters. If the 3D position could be localized, the intrinsic camera parameters also have to be stated. These parameters can be found by camera

$$\begin{bmatrix} u/W \\ v/W \\ W \end{bmatrix} = \begin{bmatrix} K \end{bmatrix} \cdot \begin{bmatrix} R|t \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4}$$

calibration [10]. The relation between image and 3D coordinates is shown in (4). The symbols u/W and v/W represents 2D image pixel coordinates, K is 3×3 intrinsic camera matrix, $R|t$ is 3×4 extrinsic camera matrix and X, Y, Z are 3D world coordinates of chosen point.

Because the 3D world coordinates are the ones we need to determine, for this purpose are triangulation methods used, the direct linear transformation (DLT) algorithm [11] for example.

Experimental tests

As it was already said, this paper deals with comparison of pixel and subpixel corner detection in case the determination of 3D position from image pair was taken into account. For that reason were the images from two datasets tested [12], [13]. The direct web links to these datasets and more information are stated in acknowledgement section.

There is a lot of ways how to compare and describe the accuracy of 3D reconstruction and camera calibration. We decided to use reprojection error criterion (dataset 1) and also direct comparison of found and ground truth translation vectors (dataset 2). Both comparisons are described in this section.

Let's start with first dataset and first comparison. One of the reasons why this dataset was chosen is because it contains chessboard segments where the X-corners are easily detected.

The whole process of testing was following: Firstly the pair of images containing the same scene was chosen from dataset. The example of tested image is shown in Figure 1. These two images have different angle of view and scene distance. Then the corner detection was applied and corresponding points in both images were found. We used not only traditional pixel but also both mentioned subpixel approaches. A next step was stating the fundamental matrix and extrinsic camera matrix. The intrinsic camera matrices were included to tested dataset. Last step was to determine the 3D world coordinates using triangulation as it was described in previous section.

As it was said, in this case, we decided to use reprojection error criterion. By definition the reprojection error is geometric error corresponding to the image distance between a projected point and a measured one. It is used to quantify how closely an estimate of 3D point recreates the point's true projection. In case of absenting ground truth information such as translation vector or rotation matrix, it is reasonable criterion for solution optimality evaluation.

In our case it means we have determined the 3D position of detected corner points (also extrinsic camera parameters were stated) and then we have reprojected these found 3D points back into 2D camera pixel coordinates. The example of reprojection the 3D points into 2D camera planes is shown in Figure 2. Let's say a

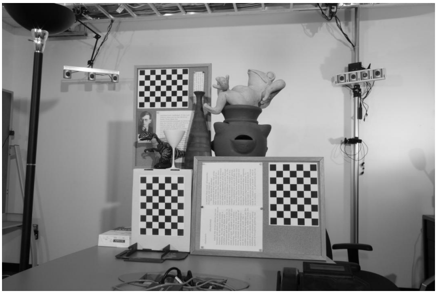

Figure 1. The example of tested image from dataset 1

little bit more about specifics of our experiment. We have chosen 24 tested images from dataset, where the both halves of this set showing the same scenes only the resolution is different. First half (12 images) had resolution 1068×712 pixels and other half (12 images) had resolution 4272×2848 pixels. These two subsets were divided into the 6 pairs, where two of them had relatively large mutual longitude angle (roughly $90°$), another two had relatively medium longitude angle (roughly $30°$) and last two pairs had relatively small longitude angle (roughly $10°$).For every from given pairs were 10 corresponding points detected and the reprojection error tests was performed as it was mentioned before. The results were statistically analysed and listed in the table.

In case of comparison based on dataset 2 the principle of the test is basically the same. There were again the image pairs for testing chosen, 10 corresponding points between images found and pixel and both subpixel algorithms tested. In the contrary of first comparison, this dataset contains ground truth information (translation vector and rotation matrix for every image) and the images are not consisting from chessboard segment. The examples of tested images from this dataset are shown in Figure 3.

Because of possibility to directly compare ground truth information from this dataset, we decided to evaluate our obtained and original ground truth translation vectors.

For this comparison we have chosen 9 tested image pairs (3 pairs for every from 3 different scene). These images have resolution 3072×2048 pixels (6 pairs), respectively 2048 × 1360 pixels (3 pairs) and different angle of view and scene distance. For every from these pairs were relative translation vectors computed

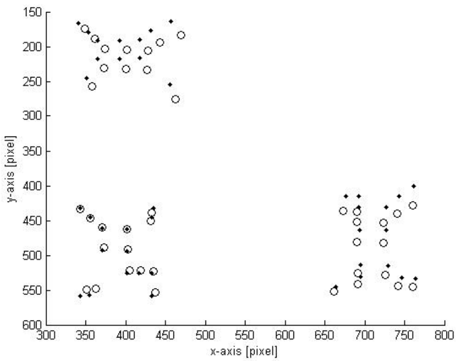

Figure 2. The example of reprojection the 3D points into 2D camera plan

Figure 3. The example of tested images from dataset 2

Experimental results

There are two tables presented in this section. In the first one is corresponding with first comparison and another one with second comparison.

Let's describe the values in first table a little bit more in detail. As you can see all results related to traditional pixel detection method have values equal to one. Subpixel method using quadratic curve is marked as method A and method using quadratic surface as method B respectively. For every single tested image the arithmetical mean (AM) and standard deviation (SD) of reprojection error (2D Euclidean distance between detected and reprojected points) were calculated. Then the ratio of subpixel and pixel results (arithmetic mean and standard deviation) for particular image was stated. Because there were multiple images in the same category (number of detected points, method, resolution, angle), these values were averaged. So there are averaged ratios of reprojection error arithmetic means and reprojection error standard deviations for particular parameters listed in Table 1.

As it is possible to see, subpixel corner detection approaches could not improve the results in every testing case we consider. But it is obvious, that the subpixel detection in case of small resolution and large or medium angle improved the reprojection error for every tested number of detected corners. On the other side, if we consider the small mutual longitude angle between the images, the results are worse if subpixel detection is involved. Using images with higher resolution, the results vary a little bit more so it is not possible to unambiguously see the influence of subpixel detection.

Table 1. Arithmetical mean and standard deviations ratios of reprojection errors

resulted ratios	method	small resolution			high resolution		
		large angle	medium angle	small angle	large angle	medium angle	small angle
AM ratio	pixel	1	1	1	1	1	1
	subpixel A	0.74	0.17	1.64	1.19	0.79	0.79
	subpixel B	0.74	0.51	1.50	1.29	1.77	0.57
SD ratio	pixel	1	1	1	1	1	1
	subpixel A	0.75	0.16	1.78	1.21	0.90	0.81
	subpixel B	0.74	0.65	1.27	1.31	3.32	0.60

The notation in case of second comparison and Table 2 is the same as before. As it was already mentioned, this test consisted of obtained and ground truth translation vectors comparison. Because the resulted translation vectors from our reconstruction are normalized (by the nature of used algorithm) and we usually do not know the exact units of ground truth datasets (plus the translation vectors are mostly calculated to be relative to chosen reference point in this kind of ground truth datasets), the relative translation vectors (between two corresponding images) coming from dataset were normalized (set the norm of vector to be equal to one) for proper comparison. It means that every vector elements (x, y and z) in every comparison case were divided by Euclidean norm of this same vector.

Table 2. Statistical data for the last years

tested pair	ED		
	pixel	subpixel A	subpixel B
1	0.1163	0.1266	0.1389
2	0.0792	0.0911	0.0853
3	0.1550	0.1339	0.1317
4	0.0582	0.0749	0.0741
5	0.0217	0.0315	0.0261
6	0.0786	0.0758	0.0736
7	0.0818	0.0769	0.0790
8	0.0647	0.0432	0.0453
9	0.0124	0.0146	0.0141
AM	0.0742	0.0743	0.0742
SD	0.0438	0.0402	0.0424

For every tested pair there was 3D Euclidean distance (ED) between found and original translation vector calculated and these values were listed in Table 2. The arithmetical means AM (what we can consider as systematic part of error) and standard deviations SD (random part of error) from these values were also stated.

As it is obvious from these results, the subpixel detection does not have any impact on accuracy of relative translation between two images determination. The values of both arithmetical means and standard deviations are almost identical and they are changing very slightly.

So the final conclusion based on these test is that subpixel corner detection does not have significant impact on accuracy of 3D position from image pair determination. The improvement of accuracy was more or less significant only in case of relatively small resolution images and the not relatively small displacement between them. This could be very often also the case of VSLAM methods. The reason why there were no improvements in comparison 2 might be the fact that in regular images (without chessboard segments) having highly different mutual position could be the right stating of corresponding points a little bit problematic or uncertain.

To finally prove or disprove the suitability of subpixel paradigm on 3D position determination or VSLAM methods is of course necessary to test this theory in more robust and thorough way.

Conclusion

This paper has dealt with impact of subpixel paradigm on determination of 3D position from 2D image pair. We have tried to answer the question, if the subpixel detection can increase the accuracy of 3D reconstruction. We have decided to use the Harris detector as traditional approach to detect the corner points and two subpixel corner detection methods, which were described in theoretical part of this paper.

We implemented the experiment, where the 3D positions of detected corner points using fundamental matrix and epipolar geometry theory were reconstructed. The details about this experiment are mentioned in experimental part of this contribution.

As the accuracy criterion were chosen the reprojection error and translation vectors comparison. The arithmetical mean and standard deviation of this reprojected error based on data from all involved corners were stated for every tested image. Because the tested images were divided into specific groups (resolution, angle, method), the obtained results were averaged and listed in the Table 1.

The Table 2 contains the 3D Euclidean distances between obtained translation vectors and ground truth translation vectors coming from dataset. The arithmetic mean and standard deviation from these values for particular method were also calculated and stated in the table. As far as it was found in this paper, the subpixel corner detection does not have significant impact on accuracy of 3D position determination under the circumstances we were discussed. The improvement of accuracy was more or less significant only in case of relatively small resolution images and the not relatively small displacement between them. This is very often exactly the case of VSLAM or visual servoing methods. The reason why there were no improvements in comparison 2 might be the problematic determination of corresponding corner points in case of chosen dataset images (no easy detected chessboard segments, highly different mutual position, different light conditions and so on).

The final statement subpixel detection effect on accuracy of 3D scene reconstruction requires the more robust and deep testing of course. It is one of our goals to implement the subpixel idea to one of VSLAM methods and evaluate the results in future.

Acknowledgement

The work presented in this paper was supported in part by Ministry of Education of the Slovak Republic under grant No. 1320 within the Framework of the program to Support Young Researchers: Teoreticka a prakticka studia vyuzitia SLAM metod.

This work was also supported by the Slovak Research and Development Agency under grant No. APVV-0469 – 12 and by the Slovak Grant Agency VEGA under grant No. 1/0963/12.

Special thanks to The Laboratory for Engineering Man/Machine Systems (LEMS) from Electrical Sciences faculty of the School of Engineering at Brown University, from where the tested dataset were obtained – *http : //vision.lems.brown.edu/* and also to Computer Vision Laboratory (CVLAB) from Ecole Polytechnique Federale de Lausanne (EPFL) for dataset *http : //cvlab.epfl.ch/data/strechamvs/* providing.

References

[1] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, pp. 239–259. Cambridge University Press, 2-nd ed., 2003.

[2] L. Sroba and R. Ravas, “Impact of subpixel paradigm on homography determination,” in *Proceedings of the ELITECH ’13, 15th Conference of Doctoral Students*, 2013.

[3] C. Harris and M. Stephens, “A combined corner and edge detectors,” in *Proceedings of the In Alvey Vision Conference*, pp. 147–152, 1988.

[4] J. Shi and C. Tomasi, “Good features to track,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593–600, 1994.

[5] L. Sroba and R. Ravas, “The subpixel x-corners detectors and their window shift robustness comparison,” in *Proceedings of the ELITECH ’12, 14th Conference of Doctoral Students*, 2012.

[6] M. Brown, R. Szeliski, and S. Winder, “Multi-image matching using multi-scale oriented patches,” in *Proceedings of the International conference on Computer Vision and Pattern Recognition CVPR2005*, pp. 510–517, 2005.

[7] M. Rea, D. M. Robbea, D. Elhawary, Z. Tse, M. Lamperth, and I. Young, “Sub-pixel localization of passive micro-coil fiducial markers in interventional mri,” in *Proceedings of the MAGMA*, 2009.

[8] F. Sur, N. Noury, M. Berger, and I. Grandest, “Grandest, computing the uncertainty of the 8 point algorithm for fundamental matrix estimation,” in *Proceedings of the British Machine Vision Conference*, 2008.

[9] W. Wang and H. Tsui, "A svd decomposition of essential matrix with eight solutions for the relative positions of two perspectives cameras," in *Proceedings of the 15th International Conference on Pattern Recognition*, pp. 362–365, 2000.

[10] Z. Zhang, *Emerging topics in computer vision*, ch. Camera calibration, pp. 4–43. Prentice Hall Professional Technical Reference, 2-nd ed., 2004.

[11] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, pp. 89–90. Cambridge University Press, 2-nd ed., 2003.

[12] C. Strecha, R. Fransens, and L. V. Gool, "Combined depth and outlier estimation in multi-view stereo," *Computer Vision and Pattern Recognition*, pp. 2394–2401, 2006.

[13] C. Strecha, R. Fransens, and L. V. Gool, "Wide-baseline stereo from multiple views: a probabilistic account," *Computer Vision and Pattern Recognition*, pp. 552–559, 2004.

Authors

Lukas Sroba — the 2nd year Ph.D. student, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia; E-mail: *lukas.sroba@stuba.sk*

Rudolf Ravas — the Associate Professor, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia; E-mail: *rudolf.ravas@stuba.sk*