Reference Point Estimation Technique for Direct Validation of Subpixel Point Detection Algorithms for Internet of Things

Mariusz P. Wilk*, Brendan O'Flynn **
Tyndall National Institute
University College Cork
Cork, Ireland
mariusz.wilk@tyndall.ie* (corresponding author), brendan.oflynn@tyndall.ie**

***Abstract*—Subpixel point detection algorithms are important in many application spaces, especially those where limitations of the imaging device's resolution need to be overcome. Such algorithms help decrease the overall requirements of the given system. Many factors, such as power consumption and cost, are critical in the context of the Internet of Things. While these algorithms do offer an improvement in the precision of point detection, it is often difficult to directly determine their precision. The main reason for it is the lack of the point of reference that the outputs of subpixel point detection methods can be compared to. In this work, we present a novel method for finding the point of reference for validating the subpixel point detection algorithms directly. Its operation is demonstrated on an experimentally obtained sample dataset.**

Keywords—Subpixel, Point Detection, Precision Estimation, Algorithm, Technique, Internet of Things

I. INTRODUCTION

An important task in machine vision is the point detection. It is also referred to as interest point detection and is carried out as one of the first steps in many image processing algorithms. It is used extensively in tracking applications, where the coordinates of the given features, often individual points, need to be tracked over time. For example, the coordinates of the detected points in images become the input to the camera pose detection algorithms, such as those based on solving the Perspective-n-Point problem (PnP) [1]. Recently, point tracking has been particularly relevant for solving the PnP problem in the context of Virtual Reality (VR). It is used for pose tracking of the headset and hand controllers in 3D space, e.g. the Oculus Rift [2]. The Other application spaces include visual 3D ranging that relies [3-5]. It is also used in optical metrology systems that determine physical dimensions of objects using the locations of features detected in the images [6]. The Internet of Things (IoT) is one of the most promising areas for point detection using vision sensor technologies.

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund under Grant Number 13/RC/2077.

The relatively low cost and wide accessibility of imaging sensors often make them a preferred choice in many applications involving point detection. These advantages, however, come at a cost. The digital image sensors are fundamentally pixel arrays with mostly fixed pixel size and number. Regardless of the application, the precision of point detection depends mostly on these two parameters. The number of pixels on the pixel array is often considered the most significant limiting factor. The first possible solution to this problem that might come to mind would be to increase the resolution of the sensor. However, it is not always practical. High-resolution cameras are more expensive. Also, the computational requirements of the overall system increase exponentially with the increase in resolution, thus rendering it impractical for some applications, especially those in the context of IoT. If the system is not resource-constrained, even cameras with the highest resolution available on the market may not be precise enough at pixel level. Therefore, system designers often decide to overcome this limitation with subpixel point detection techniques. It is possible to estimate the location of a point in an image at a precision that is greater than the resolution of the image sensor. Typical optical systems rarely capture rays of light in a single pixel. Even the thinnest ray of light, that hits the image sensor, tends to be spread over several pixels. For instance, the pixel intensity profile of a single source of light, e.g. Light Emitting Diode (LED), along any of the two axes of the sensor forms a distribution of pixel intensities with a peak on the pixel that is at the smallest distance from the true peak intensity. Each pixel takes a discrete measurement of this distribution. The subpixel point detection algorithms estimate the location of the true peak using several intensity measurements from the pixels adjacent to the peak at pixel level. We made these assumptions in our recent work in which we proposed a computationally efficient subpixel point detection technique that used three pixel intensity measurements at the pixel-level peak [7]. We assumed that the pixel intensity profile of the point had the properties of a Gaussian distribution with a certain range of standard deviation, as shown in Fig. 1. It was a fair assumption, because in most

cases a point of light, e.g. an LED, does form such a distribution on the pixel array. Furthermore, many point tracking applications use such point sources of light as the tracking objects.

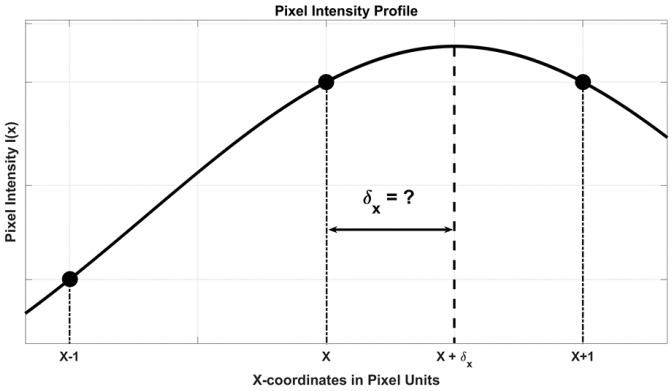


Fig. 1: Pixel Intensity Profile along x-axis

Subpixel point detection techniques are certainly effective at overcoming the fundamental precision limitations of the image sensors' resolution. Thus, the intensity peak at subpixel level can be found, e.g. the subpixel offset along the x-axis δ_x shown in Fig. 1. However, it is not always an easy task to quantify the actual precision of such algorithms. One of the common approaches in the literature involves making an indirect measurement. The precision of the algorithm can be determined by looking at the improvement it makes in each algorithm when compared against a benchmark tool. For example, in the case of a 3D ranging application the distance measurement in the real world can be used as the benchmark, as described in our earlier work [5]. This approach is not suitable for all situations. Sometimes it is beneficial to make a direct precision estimation. While it seems to be an obvious option, it is not a straight forward process. There is one fundamental challenge with it. To directly determine the precision of a given algorithm, a reference point is required, i.e. the location of the true peak. This reference point must be much closer to the location of the true peak than the output of any of the subpixel point detection algorithms under the evaluation. Otherwise, the result may be invalid.

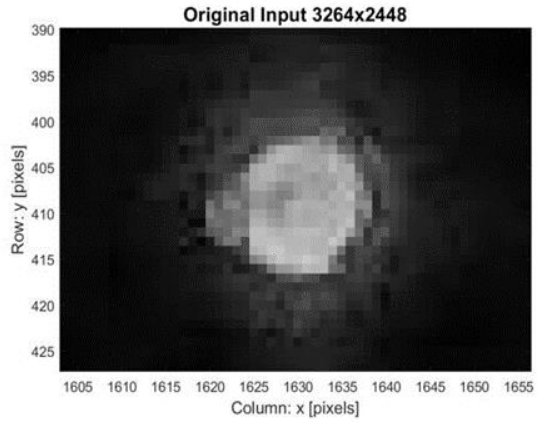
The direct validation of a subpixel point detection algorithm can be achieved using a single imaging device. A high-resolution camera can be used in this task. For example, a high-resolution camera can be used to acquire *N* input frames, which then can be down-sampled by a certain factor, of say 10, and used as inputs to the subpixel point detection algorithms. The coordinates of the peaks in the original high-resolution image can be used as the reference points in the validation process; after appropriate scaling. While it is an effective approach, its reliability or repeatability cannot be trusted. Some empirical testing can show that the pixel intensities at the peak's location can vary greatly from frame to frame. Thus, the location of the peak may shift from frame to frame. It is unacceptable, because the location of the reference point should not vary during the experiment if both camera and the point source of light are stationary. This problem is particularly apparent in low-light conditions with Infrared (IR) LED and cameras fitted a matching IR filter. For instance, our experimental setup consisted of an IR LED and an 8 Megapixel camera with a matching IR filter, both of which were stationary in controlled laboratory conditions. The fluctuations in pixel intensity were observable with naked eye when zooming in onto the point. This problem leads to a need for a more reliable way of determining the point of reference.

The problem of fluctuations in pixel intensities in the successive frames could be tackled with statistical methods, e.g. by computing mean peak location over all *N* input frames. One way of finding the peak's location can involve using the Circular Hough Transform (CHT), which can find the circle centres in images [8]. The circle centres should be coincident with the location of the points' centres; assuming a symmetric Gaussian distribution of the pixel intensities around the peak. The mean of the circle centres from the entire set of *N* input images can be computed and used as the reference point. However, it may prove to be insufficiently accurate. Although the intensity peak does generally have a form of a circle, it is not always an ideal circle. Also, the pixel intensity profile in low-light conditions does not have pure Gaussian properties, as shown in Fig. 2 (a). That is, the circle centre does not always coincide with the intensity peak. An alternative approach can involve calculating the mean of peak intensity locations over all *N* input frames. Again, it may not be a reliable measure as the standard deviation of this metric would be high, given the fact that the high-intensity pixels can be spread over a relatively large area, as it can be seen in Fig. 2 (a).

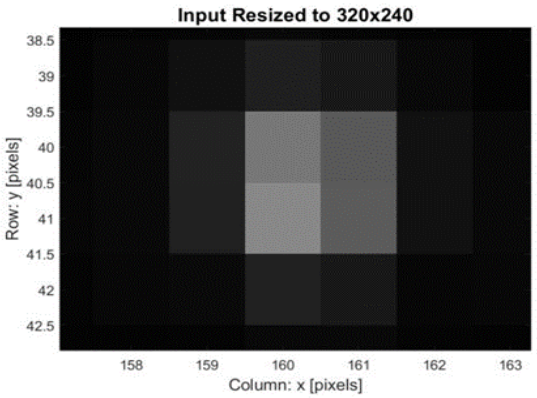
In this work, we propose a novel algorithm for finding the reference point that can be used for a direct validation of subpixel point detection algorithms; and solves the above problems. Our method is an iterative process, that uses a set of high-resolution images of a point, such as that shown in Fig. 2 (a) and processes it in a loop until it converges to the best achievable result for a given set of input frames. In part, it incorporates the mean CHT and the peak intensity methods mentioned above. The paper is organised as follows. The design of the algorithm and the experimental setup for its demonstration are described in Section II. The results and conclusions are shown in sections III and IV, respectively.

Main contributions of this work include:

- Novel technique for finding the point of reference for the direct validation of the subpixel point detection algorithms, described in Section II, A.
- Comparison between the methods based on mean circle centre and mean maximum pixel intensity and our proposed algorithm, in Section III.



(a)



(b)

Fig. 2: Original High-resolution Input Frame (IR LED and camera with matching IR Pass Filter): (a) and Down-sampled Input frame: (b)

II. METHODOLOGY

The proposed algorithm and the experimental test procedure are described in this section in detail. Firstly, each step of the algorithm is explained. It is followed by the description of experimental setup that was used to test the algorithm.

A. Algorithm

The proposed algorithm was designed as a multi-step iterative process. The location of the peak was estimated based on a combination of mean CHT and mean peak pixel intensity over all N input frames. The estimate of the true peak's location $\tilde{P}$ was updated in each iteration of the loop. The algorithm was run over n iterations until the $\tilde{P}$ no longer changed, i.e. the best achievable solution for this method was determined. The algorithm is shown in Listing 1. All variables used in this algorithm are two-element row vectors with the elements corresponding to the x- and y-axis, as shown below in (1):

$$\tilde{P} = [x\ y] \tag{1}$$

The first step of the algorithm involved the acquisition of a relatively large set of N input frames from the high-resolution camera in an experimental setup. It is assumed that the experimental setup is placed in a controlled laboratory environment. It is critical to ensure that there are no external light intensity fluctuations originating from uncontrolled ambient light. Secondly, there must not be any mechanical vibration present in the environment during the data acquisition process. Any mechanical distortion that could cause motion of either the camera or the source of light should be avoided, e.g. a slamming door, air drafts caused by motion or air conditioning systems, or even loud talking near the camera.

The second step involved finding an initial estimate of the peak's location $\tilde{P}$. It served as the initial input to the algorithm's loop to be refined over n iterations. It was determined by finding the location of the peak intensity. It could also be hand-picked by manual image inspection.

In the third step, the Region of Interest (ROI) was set. The centre of the ROI was set to the current value of $\tilde{P}_n$. The size of the ROI could be set manually. It depended on the resolution of the camera and the size of the point on the pixel array. The size of the ROI had one main requirement. Its area had to be greater than the area of the peak on the pixel array. The specific settings depended on camera's resolution and the size of the point on the pixel array. In this work, the size of the ROI was set to 50 pixels along x-axis and 35 pixels along the y-axis, as shown in Fig. 2 (a).

The next three steps were aimed at finding the best location estimate for the given iteration of the loop, n. Steps four and five were focused on finding the mean peak locations using two different methods. Firstly, the mean circle centre $\bar{C}$ was computed from all N input frames in the ROI. The result should be coincident with the peak's location since its 2D pixel intensity distribution forms an approximate circle around it. The CHT algorithm proposed by Atherton et al. was used in this step [8]. Subsequently, the mean pixel intensity $\bar{P}_{intens}$ was computed within the ROI. Finally, the peak estimate, for the given loop iteration n, $\bar{P}$ is the mean of $\bar{C}$ and $\bar{P}_{intens}$. It is also the mid-point between these two values. Optionally, a weighted mean could be considered, if either of the two parameters was considered more important, or accurate, in the calculations.

The final stage of the algorithm was aimed at determining whether the mean estimate $\bar{P}$ was less than half a pixel away from the current estimate $\tilde{P}_n$. If so, the algorithm's work was complete. It could proceed to the next steps, i.e.: setting the peak's final position estimate $\tilde{P}$, appropriately downscaling it, and down-sampling all N frames in the input set. Otherwise, the process must: go back to Step 3, adjust the ROI_{centre} with new, more accurate position estimate $\tilde{P}_n$, and execute the next loop, $n + 1$. The execution continued until the condition in Step 7 was false.

Listing 1: Point of Reference Estimator

1. Acquire N high-resolution input frames
2. Get initial peak estimate $\tilde{P}_n; n = 0$
3. Set $ROI_{centre} = \tilde{P}_n$
4. For all N frames, find $\bar{C}$
5. For all N frames, find $\bar{P}_{intens}$
6. Find mean estimate $\bar{P} = Mean(\bar{C}, \bar{P}_{intens})$
7. $If\ Abs(\tilde{P}_n - \bar{P}) > 0.5\ pixel$
   a. $n = n + 1$
   b. $\tilde{P}_n = \bar{P}$
   c. $GOTO$ Step 3
8. $\tilde{P}_n = \bar{P}$
9. Down-scale $\tilde{P}$
10. Down-sample all N input frames

B. *Experimental Setup*

The algorithm was tested in an experimental setup. It consisted of a high-resolution camera, an IR LED and a matching IR Longpass Filter that was attached to the camera. The OV8865 made by OmniVision, Inc [9] was selected as the camera. It was set to its greatest resolution, i.e. 3264x2448. It was used with a matching IR Longpass filter made by Edmund Optics, Inc [10] and the IR LED made by Vishay Intertechnology as the point-source of light [11]. The distance between the camera and the IR LED was set to 1.5 metre. The intensity of the IR LED was set to such a level that the maximum pixel intensity did not exceed 50 % of the sensor's range. The experimental setup was designed in such a way that the image frames were suitable for the subpixel point detection algorithms mentioned in the Introduction section. The algorithm execution and result analysis tasks were carried out offline in MATLAB software package. The images were acquired using Microsoft Surface Pro 4 computer (which incorporates the OV8865 camera sensor). The complete experimental setup is shown below in Fig. 3.

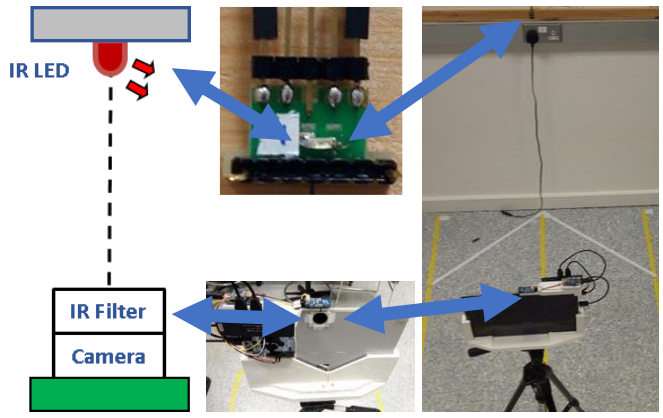


Fig. 3: Experimental Setup – Camera-IR Filter – IR LED

III. Results and Discussion

A series of experimental test runs was carried out to evaluate the performance of the proposed algorithm. It was done in two stages: data acquisition and offline processing.

Once the experimental setup had been carefully prepared for the data acquisition, the proposed algorithm was applied to it, as described in Listing 1. Its Step 1 was executed first. A total of $N \in (1000,1300)$ frames were acquired. The choice of this relatively low value was the result of a trade-off between a number that can yield a statistically significant result, and a number that our PC could process. For example, $N = 1000$ implies $\approx 2.5\ GB$ in memory usage. The computer that was used in this experiment tended to crash during the data acquisition process, when N was significantly larger than 1000. The acquired dataset was then used in the remaining steps of the algorithm. These remaining steps were executed offline. The algorithm was tested on several separate sets of input frames. In all cases, it converged on the solution, i.e. the condition in Step 7 was false, within $n \leq 5$ iterations.

The results obtained from one of the input datasets are superimposed on the last input frame from the dataset, $N - 1$. Whereas Fig. 4 shows the result on the original, high-resolution, input frame, Fig. 5 shows the same result on the resized image. The results were appropriately scaled down by the same factor as that of the image resizing (the result of the downscaling of a different input frame is also shown in Fig. 2 (b), as compared to the original high-resolution input, Fig. 2 (a)). These results show significant discrepancies between the different stages of the proposed algorithm. Also, the locations of the peak pixel intensity $P_{intens_{N-1}}$ along with the circle centre C_{N-1} calculated in this last frame, $N - 1$, are provided for comparison. This visualisation clearly shows the impact of the distortions present in the experimental setup. It shows that it differs substantially from the theoretical modelling of a symmetric 2D Gaussian distribution of the pixel intensity profile.

Fig. 5 focuses on the resized image and emphasizes the level of precision that each stage achieved. This figure exposes the location of the true peak, and how the different stages estimated it. Interestingly, the CHT method yielded a result that was very close to the true peak estimate $\tilde{P}$, also referred to as the mid-point, computed by our proposed method. However, it was only a coincidence, that applied to this particular input frame. It can be seen than the mean circle centre $\bar{C}$ is located farther away.

The quantitative results for the selected input dataset are shown in Table 1. They are the same as those visualised in Fig. 5. This table lists the results of the different stages of the proposed algorithm. It shows how the output, $\tilde{P}$, differs from the outputs of the intermediate stages. These results can be also represented in real-world units, e.g. metres, as the dimensions of the pixels in the camera, as well as the scaling factor in the resized images, are known.

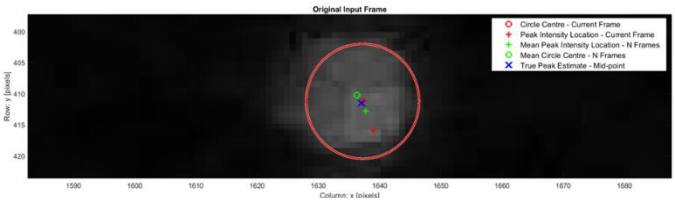


Fig. 4: High-Resolution Input Frame with the Main Output along with the Outputs of the Individual Substages of the Algorithm (Last Frame, *N-1*)

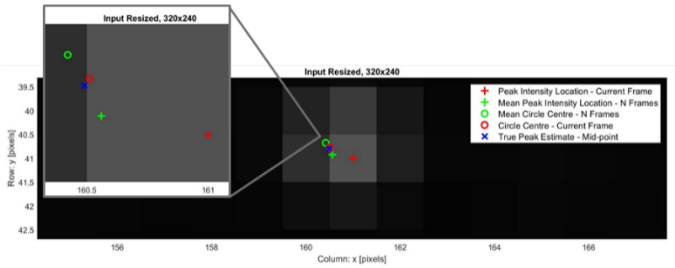


Fig. 5: Resized Input Frame with the Main Output along with the Outputs of the Individual Substages of the Algorithm (Last Frame, *N-1*)

Table 1: Results in Pixel Units - Visualised Fig. 5

Output	**Column – x**	**Row - y**
Peak Intensity – Current Frame	161	41
Mean Peak Intensity – N Frames	160.559	40.924
Mean Circle Centre – N Frames	160.421	40.673
Circle Centre – Current Frame	160.512	40.771
True Peak Estimate $\tilde{P}$	160.490	40.799

IV. Conclusions and Future Work

This work describes the proposed method for finding the point of reference, that is necessary to determine the precision of subpixel point detection algorithms in camera sensors directly. It shows how a high-resolution camera can be utilised in this process. Our proposed algorithm leverages the high information content of the high-resolution input frames and the two separate peak detection techniques, i.e. the peak pixel intensity and circle centre with the CHT. It finds the best achievable solution with an iterative statistical procedure applied to the entire input dataset. Its precision depends on the size of the input dataset, i.e. the N. In this work, $N \in (1000,1300)$ input frames were evaluated. The resultant estimate of the peak's location can be then down-scaled by the same factor as the input images that the subpixel point detection algorithms under evaluation are applied to, as shown in Fig. 4 and Fig. 5, respectively. Future work will involve validating it using larger datasets. It will also involve the validation of several existing methods present in the literature, including the Simplified Linear Interpolation method described in our previous work [7].

V. Limitations of This Study

The performance of the proposed method is associated with the size of the input dataset, i.e. the value of N. This is the main limitation of this study. A larger N would increase the statistical significance of the determined reference point's location.

VI. zReferences

[1] G. Xiao-Shan, H. Xiao-Rong, T. Jianliang, and C. Hang-Fei, "Complete solution classification for the perspective-three-point problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 930-943, 2003.

[2] S. M. LaValle, "Virtual Reality," in *Virtual Reality*: Cambridge University Press, 2016, pp. 254-265.

[3] L. Abraham, A. Urru, M. P. Wilk, S. Tedesco, and B. O'Flynn, "3D ranging and tracking using lensless smart sensors," in *SSI 2017: International Conference and Exhibition on Integration Issues of Miniaturized Systems*, 2017: Verlag Wissenschaftliche Scripten.

[4] L. Abraham, A. Urru, N. Normani, M. Wilk, M. Walsh, and B. O'Flynn, "Hand tracking and gesture recognition using lensless smart sensors," *Sensors*, vol. 18, no. 9, p. 2834, 2018.

[5] L. Abraham, A. Urru, M. P. Wilk, S. Tedesco, M. Walsh, and B. O'Flynn, "Point tracking with lensless smart sensors," in *SENSORS, 2017 IEEE*, 2017, pp. 1-3: IEEE.

[6] A. Criminisi, I. Reid, and A. Zisserman, "Single View Metrology," *International Journal of Computer Vision*, journal article vol. 40, no. 2, pp. 123-148, November 01 2000.

[7] M. P. Wilk, A. Urru, S. Tedesco, and B. O. Flynn, "Sub-pixel point detection algorithm for point tracking with low-power wearable camera systems: A simplified linear interpolation," in *2017 28th Irish Signals and Systems Conference (ISSC)*, 2017, pp. 1-6.

[8] T. J. Atherton and D. J. Kerbyson, "Size invariant circle detection," *Image and Vision computing*, vol. 17, no. 11, pp. 795-803, 1999.

[9] I. OmniVision Technologies. (2018, 23.02.2018). *OV8865*. Available: http://www.ovt.com/sensors/OV8865

[10] E. O. Inc. (2018, 19.02.2018). *1" Diameter, Optical Cast Plastic IR Longpass Filter*. Available: https://www.edmundoptics.com/optics/optical-filters/longpass-edge-filters/1quot-diameter-optical-cast-plastic-ir-longpass-filter/

[11] I. Vishay Intertechnology. (2018, 19.02.2018). *High Speed Infrared Emitting Diode, 940 nm, VSMB11940X01*. Available: http://www.farnell.com/datasheets/2245170.pdf?_ga=2.184115750.1623767235.1519051613-1677736993.1519051613&_gac=1.15708484.1519051613.Cj0KCQiAiKrUBRD6ARIsADS2OLlYcKGvXIt139L6uYIypwMFTqdQy6r-30rp9fSntK9OXyQTsLnW8b8aAm_VEALw_wcB