UDC 681.322

# REVIEW OF METHODS FOR SMARTPHONE APPLICATION FOOT SIZE ESTIMATION FROM IMAGES

*I. ZAKHARAVA, R. BOHUSH*
Polotsk State University, Belarus

*The task of obtaining real-world coordinates of an object is quite challenging. In this paper proposed brief explanation of algorithms dedicated to overcome this challenge. They were analyzed including their advantages and disadvantages in application to possible implementation in smartphone app.*

Introduction. The Internet technologies grow rapidly. With that traditional approach from selling goods directly to the customer in offline switched to the online space. Nowadays more and more shops use Internet to perform commercial transactions. Online stores have a variety of products and it helps to the customers to purchase a product with the better price. Also there is no need to go out and all shopping can be performed right inside the house. The online retailers trying to make system even more comfortable and all payment procedure can take about 2 minutes.

But online shopping has some disadvantages too. For instance, a customer has to buy a product without seeing actually how it looks like. Customers may click and buy some product that is not really required by them. The electronic images of a product are sometimes misleading. The color, appearance in real may not match with the electronic images. If the size does not fit the customer want to return goods back. A big percentage of refund leads to business owner money loss.

The buyers also can choose wrong size that does not fit to their body parameters and get bad experience items purchasing in online stores. To avoid this experience some online shops propose free fitting when the currier deliver all purchased good and then taking back things that does not fit to the customer by the appearance or size. But it consumes a lot of time and money as well.

This problem became very crucial for the shoe shopping. Correct foot measurements requires a trained skills and special instruments like calipers for linear measurements that used for heights, lengths and widths calculation as well as measuring tape for girths. The accuracy of final results depends of skills level.

To make process of foot size estimation more comfortable for the customer a lot of companies propose image processing algorithms to acquire and measure surface of the foot. In this paper we will review approaches that can be used in application to this task.

Algorithms review. Nowadays, state-of the art approaches can be separated on the few categories. Some of approaches use additional sensor to achieve more information about object. For example, approach presented in [1] use RGB-D cameras. A RGB-D image is simply a combination of a RGB image and its corresponding depth image. A depth image is an image channel in which each pixel relates to a distance between the image plane and the corresponding object in the RGB image. The image of the foot pictured on the A4 sheet with the set of AR codes. The depth map obtained from an image converted to cloud points. After performed noise removal algorithm and performed a 2-Dprincipal component analysis (PCA) to find the long-axis and the short-axis of the 2-D foot print. Measurement error for proposed algorithm varies from the left and right foot. Also it depends on measurement type. In this paper were evaluated measurements of foot length, width and girth. For the left foot the minimal deviation in length measurements was 2.1 mm, in width it was 1.45 and for girth this value was 2.76. For the right foot the minimal deviation was 1,18 for foot length, 1.32 for width and 2,16 for girth. Proposed approach has a big inaccuracy that can be crucial for shoe size measurements. Also it uses hardware that cannot be used in average smartphone application.

The next algorithm presented in [2] also use RGB-D camera to measure an accurate distance to an object and from with that information obtain real world measurements. The RGB image from the camera goes to Convolutional Neural Network (CNN) from [3] for object detection. As output it produce bounding box that isolate object of interest from the background to perform more accurate measurements. Than was performed recalculation from image coordinates to real world measurements using e LIDAR/IR sensor. For hardware this algorithm use smartphone Lenovo TangoPhab 2 phone that collects point cloud of depth data. The main problem that phone can obtain RGB information at 30 fps(frames per second) and depth information at 5 fps. Also, there is a spatial shift between camera coordinates and depth sensor. Moreover, during experiments was found that IR sensor has missing spots in black or metallic surface. Proposed approach has deviation from 3 to 10 cm. These values are too big to take this approach into account for foot measurements.

In [4] presented another algorithm that uses depth information with image data to obtain real world measurements of on-tree mango fruits. For fruit detection was used cascade classifier with HOG features. Than was perdormid sematic segmentation of fruit area with Otsu[5] method in CIE Lab color space. After was performed filtration of mango peduncle with following dilatation. Than object geometrically filtered to check that the object has elliptical shape. The following operation is real world size calculation using the data from depth sensor. As mentioned earlier approach that use additional sensors to acquire depth information cannot be used by average smartphone owner.

Some of them use a template to compare image distance with real world measurements. In [6] authors presented algorithm that use A4 sheet of white paper as a pattern. The foot in a black sock is standing near or on the white pattern. The algorithm requires 3 photos with the different views of the foot. For each photo performed Canny edge detection to find edges of the pattern and foot. They are distinguished by the algorithm using geometrical information about form and size of template. The extracted template contour is refined with the Snakes algorithm from [7]. Than the foot size is recalculated using real world coordinates of the template. In paper mentioned that the accuracy is strongly associated with the shooting angle. If the angle is 10˚ than measurement error is equal to 1,5 mm. But if the angle if 15 degrees the error becomes 3,4 mm. Also, proposed approach do not use specific algorithm to extract points of a foot and there can be errors due to complicated background.

In [8] proposed an algorithm to estimate the size of rain drops. Authors take the videos of rain drops as input. All videosequence separates on frames that processed and each frame is threated as individual image. Each image goes to filtering and morphological operations to extract raindrop shape from background. For real world measurements performed static pixel to millimeter ratio. Proposed algorithm unsuitable in application to foot measuring via smartphone app due to static ration for coordinates recalculation and weak algorithms for object detection.

In [9] performed comparison of an app that use template and the booth with mirrors. An app uses 3 images of the foot on A4 white sheet. The paper used for camera calibration and real world size calculation. The 3 images should contain the foot pictured on the top and from both sides' also known as zenithal, lateral and median views. Than authors use 5 points landmark extraction to accurate measurements of foot length, toes girth, toes width, ball girth, ball width, instep girth and instep height. The booth for foot measurement contains mirrors that also present foot in 3 points of view. To distance between camera and booth bottom is a constant value and real world values are simply recalculated by it. The error for every measurement amounted from 0.5 mm to 2.2 mm for the app and from to 0,6 mm for 1.8 mm the booth.

Some researches focused their attention on stereo photogrammetry. In [10] is presented approach to obtain real world measurements of objects of interest for object tracking. The input image pictured by set of cameras. They are mounted on a stable bar with the stable distance of 413 mm. The object detected by fast and adaptive median background subtraction algorithm from [11]. Real world coordinates recalculated using camera rig distance. Authors performs static camera calibration using «chessboard» pattern. The error of measurements lays above the 10%. Proposed algorithm does not use any additional operations to reduce noise from an image which is leads to false positive result of the detector. Also, static calibration with the pattern requires skills and knowledge which makes it harder to potential implementation in smartphone app.

Another one algorithm presented in [12] uses the system of two cameras that mounted on the static distance. In paper proposed enhanced algorithm for tomatoes detection. That use active contour model from [7] with shape constant. As shape constant is used ellipse. On the detection algorithm output there are four proposals of tomato location and authors manually choose the best of them. For stereo image formation was used set of features. These features obtained by SIFT [13] descriptors. The best match between them is calculated using Euclidean distance measurement. The camera parameters are determined once, at the beginning of the season, by observing a calibration pattern at different positions and orientations in the scene. This method was first tested under ideal acquisition conditions and using manual segmentation. In this case, the percentage error between the actual radius and the estimated size was always less than 10 % with most (91 %) of the error less than 5 %, which demonstrate the robustness of radius estimation. The complete system was also applied to estimate the size of tomatoes cultivated in open fields for the agriculture season 2013. The percentage error was less than 10 % in most of the cases, despite the poor quality of images during this season (small size, pixelated images).Proposed approach also require static camera calibration and cannot be applied by average smartphone user.

In [14] reported that object of interest size also obtained using two cameras which are not movably fixed. The input images converted to gray colorspace. After, medial filtering for noise reduction was performed. Than performed simple background subtraction algorithm that assume that firs frame is background. That binarized image goes to morphological processing. This is followed by object detection using connected components algorithm from [15]. The real world measurements were calculated using the distance between two cameras. Proposed algorithms works very fast and achieved error deviation at ± 3 cm. This is very high value and proposed approach cannot be applied to our task. Also in paper presented images that does not contain complicated background and these algorithm unsuitable for real world scenarios.

Similar to previous work the authors [16] propose an algorithm of object detection and size estimation in underwater areas. Two camera located on the static distance between each other and both of the previously calibrated with pattern. Then performed object detection using background subtraction algorithm. For each object is calculated the centre of mass and these values used for the following calculation of real-world coordinates. The main disadvantages of proposed approaches are the same as previous: calibration is difficult to perform by the average app consumer and not accurate algorithm for object detection that almost unsuitable for real world scene.

Some approaches uses surprising algorithms and non-standard sensor types to measure the object. For example in [17] reported on their attempts to use a standard flatbed scanner - the type that would normally be found in the office environment for digitizing documents - to scan the foot sole and translate the output to a 3D form. The distance of the sole away from the scanner glass was estimated using the albedo of the sole surface and the pixel intensity of the resulting image, inspired by techniques used in the analysis of satellite images. The authors claim they are able to achieve an average error of <1 mm, in line with those achieved by more expensive scanning systems, however the system was tested using a foot model with a uniform color and it was noted that scanning a real foot, especially those with damage or injury could present problems for the reconstruction process. And proposed approach cannot use smartphone software as well.

Finally, at this moment a lot of attention focused on Neural Network (NN) approaches dedicated to this problem. There we will review not only algorithms dedicated to real-world object size estimation but and dense map calculation algorithms. As we mentioned earlier some approaches use RGB-D information to obtain object size. There we look at the state-of-the art algorithms that can help us get depth information without sensors.

In [18] presented algorithm that use stereo images. First of all images goes to stereo matching part. In this paper used algorithm mentioned in [19]. It uses SVM classification with Laplacian of Gaussian (LOG) transform and Euclidean distance correlation function. The Laplacian measures directed edge intensities over some area smoothed by the Gaussian. The depth map is obtained after this step. Then the depth map goes to three layer feed forward network that produce bounding boxes from it. This algorithm was tested on low resolution images. In contrast, smartphone cameras have much higher resolution that example of images that was proposed in paper. Consequently, there has to be much more inputs and hidden layers in NN architecture.

In [20] was used stereo matching algorithm too. There is stereo matching is performed using CNN. The architecture contains 8 layers. For the first 3 of them images processed independently. Than performed features concatenation. The final layer, projects the output to two real numbers that are fed through a softmax function, producing a distribution over the two classes (good match and bad match. The final stereo matching is performed by coss-based cost aggregation from [21]. Than is performed matching cost refinement by enforcing smoothness constraints on the disparity image. Proposed algorithm is more accurate than late approaches but it is too heavy for smartphone application.

In [23] was performed depth estimation using CNN architecture U-Net[24]. As input used original inage, lense parameters and binned depth map calculated like in [25]. As output they get simulation of sensor image and depth map. Proposed approach has the lowest values of distortion error. Unfortunately, it still heavy weighted approach for smartphone owners.

Conclussion. There is a lot approaches dedicated to digital foot measurements. Some of them has measurement accuracy even higher that mutual methods. At the same time automated approaches much faster and cheaper than hand crafted. Also the data obtain from this algorithm can be used for health insurance, shoes recommendations, deceases diagnostics and so on.

Proposed review showed that for digital measurements widely used different techniques that trying to maximize cost and accuracy , reduce processing time and learning period for average user. Bringing researchers in the field, scanning equipment manufacturers, orthotic, footwear companies, users and other stakeholders together to further explore these issues may result in cross disciplinary activity needed to resolve current needs and issues.

Currently almost every state-of-the art smartphone has more than one frontal camera that means that approaches that use stereo photogrammetry can be applied for our task. Also, algorithms that use foot landmarks to extract more contextual information from an image seem very promising too.

The CNN approaches mentioned there are heavy and cannot be implemented in smartphone app, but there is a chance to improve CNN architecture to make it more suitable for mobile devices.

REFERENCES

1. Yang-Sheng Ch. Estimation of 3-D Foot Parameters Using Hand-Held RGB-D Camera/ Ch.Yang-Sheng, Ch. Yu-Chun, K. Peng-Yuan, Sh. Sheng-Wen, H. Yi-Ping // In Proc of . ACCV 2014 Workshops, At Singapore, Volume: LLCN 9010. – 2015.-P. 407-418
2. Object Localization and Size Estimation from RGB-D Images. [Электронный ресурс].–Режим доступа: https://arxiv.org/pdf/1808.00641.pdf. – Дата доступа: 04.03.2020
3. Tensorflow Object Detection AP[Электронный ресурс].–Режим доступа:.https://github.com/tensorflow/models/tree/master/research/object_detection. – Дата доступа: 04.03.2020
4. Zhenglin W. On-Tree Mango Fruit Size Estimation Using RGB-D Images/ W. Zhenglin , W. Kerry // Sensors. – 2017 . – Vol.17. – P. 2738 - 2753
5. Otsu N. A threshold selection method from gray-level histograms/ N. Otsu // Automatica.- 1975.-Vol.11. – P. 23–27.
6. Беляев, С. Определение размеров стопы человека по фотографиям/ С. Ю. Беляев, В. Г. Шубников // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. – 2014. – №.2(193). – С. 23–30
7. Kass, M. Snakes: Active contour models/ M. Kass, P. Andrew,. W. Terzopoulos ,D. Terzopoulos // International Journal of Computer Vision. – 1988. – Vol. 1. – P. 321-331.
8. Thurai M. Measurements and Modeling of the Full Rain Drop Size Distribution/ M. Thurai, M. , V.N. Bringi, P. Gatlin, W. Petersen, M. Wingo // Atmosphere. – 2019 . – Vol.10. - №1. – P. 39-55
9. Ballester A. Low-cost data-driven 3D reconstruction and its applications/ A. Ballester, E. Parrilla, J. Vivas, A. Piérola, J Uriel, S.-A. Puigcerver-Palau, P. Piqueras, Paola,C. Solves, M. Rodríguez, J. González, S. Alemany // In Proc. of 6th International Conference and Exhibition on 3D Body Scanning technologies, At Lugano, Switzer-land, Volume: Technical Session 10: RGB-D Sensors & Low Cost Systems.- 2015
10. Kollmitzer Ch.Object Detection and Measurement Using Stereo Images / Ch. Kollmitzer// In Proc of Multi-media Communications, Services and Security: 5th International Conference, MCSS 2012. – 2012. – P. 287 – 297
11. Lo B. P. L.Automatic congestion detection system for underground platforms/ B. P. L. Lo , S. A. Velastin // In Proc. of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing. ISIMP 2001 (IEEE Cat. No.01EX489. – 2001. - P. 158-161.
12. Verma U.Segmentation and size estimation of tomatoes from sequences of paired images./ U.Verma, F. Rossant, I. Bloch. // EURASIP Journal on Image and Video Processing. – 2015.- Vol. 33
13. Lowe D. G. Object recognition from local scale-invariant features/ D. G. Lowe //In. Proc. of the International Conference on Computer Vision – 1999. - .Vol. 2. – P. 1150–1157.
14. Mustafah Y.M. Stereo vision images processing for real-time object distance and size measurements/ Y. M. Mustafah, R. Noor, H. Hasbi and A. W. Azma // 2012 International Conference on Computer and Communica-tion Engineering (ICCCE). - 2012 . - P. 659-663.
15. B. Horn Robot Vision, MIT Press, 1986, Chap. 4.
16. A real-time stereo vision system fordistance measurement andunderwater image restoration. [Электронный ресурс].–Режим доступа: http://www2.ene.unb.br/mylene/pubs/2016-08-JournaloftheBrazilianSocietyofMechanicalSciencesandEngineering.pdf. – Дата доступа: 04.03.2020
17. Martedi S. Shape measurement system of foot sole surface from flatbed scanner image/ S. Martedi, H. Saito, M. Servières// In Proc. of . MVA2009 IAPR Conference on Machine Vision Applications.- 2009. - P. 2-5.
18. Zhao L. Stereo- and Neural Network-Based Pedestrian Detection/ L. Zhao, Ch. Thorpe //. Intelligent Transpor-tation Systems, IEEE Transactions. – 2000.-Vol. 1. - № 3. -P. 148 - 154.
19. Small Vision Systems: Hardware and Implementation [Электронный ресурс].–Режим доступа:https://www.cs.cmu.edu/~motionplanning/papers/sbp_papers/integrated1/konolidge_stereo_vision.pdf. – Дата доступа: 04.03.2020
20. Computing the Stereo Matching Cost with a Convolutional Neural Network [Электронный ресурс].–Режим доступа: https://arxiv.org/pdf/1409.4326.pdf. – Дата доступа: 04.03.2020

21. Zhang, K. Cross-basedlocal stereo matching using orthogonal integral images/K. Zhang, J. Lu, G. Lafruit// .Circuits and Systems for Video Technology, IEEE Transactions.-2009.-Vol. 19. - №7. – P.1073–1079.
22. Hirschmuller H. Stereo processing bysemiglobal matching and mutual information / H. Hirschmuller //.Pattern Analysis and Machine Intelligence, IEEE Transactionson. – 2008. – Vol. 30.- №2 . – P. 328–341.
23. Chang J. Deep Optics for Monocular Depth Estimation and 3D Object Detection/ J. Chang,G. Wetzstein// . Conference: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). .-2019 . – P.10192-10201
24. Ronneberger O. U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, Ph. Fischer, Th. Brox//. In Proc of International Conference on Medical image com-puting and computer-assisted intervention. – 2015 . – P. 234–241
25. Hasinoff S,W. A layer-based restoration framework for variable-aperture photography/ S.W Hasinoff , K. N Kutulako// In Proc of Computer Vision, 2007. ICCV 2007. IEEE 11th In-ternational Conference. – 2017, - P. 1–8