

ShoeMorph: Exploring Classical Computer Vision Techniques for Asymmetry-Aware Foot Size Estimation

Ayush Kumar Maheshwari*, Akshat Bari, Himesh Kumar

Department of Computer Science and Engineering, Poornima Institute of Engineering and Technology
Jaipur, Rajasthan, India

Email: 2022pietsayush041@poornima.org

Abstract:

The article introduces a lightweight computer vision-based solution to the problem of estimating human foot size and suitable shoe size based on a single monocular image, without making use of deep learning models. The algorithm exploits a combination of classical image processing methods, such as HSV conversion, Gaussian smoothing, segmentation and contour-based measurement methods, and geometric scaling with the help of an object of reference (A4 paper). The proposed method provides a low-resource and interpretable alternative, which is compared to the deep learning methods that need large datasets and high computational resources. The video on a curated dataset displaying experimental evaluation shows an average error in measurements of 3.2, which is competitive. Another measure that the system takes into account is the analysis of asymmetry, since feet are measured separately, which is a weakness of traditional techniques. The outcomes reveal the usefulness and concreteness of the method as well as define possibilities of improvement of the approach in the future.

Keywords — Computer Vision, Foot Size Estimation, Image Processing, Asymmetry Analysis, Contour Detection, Geometric Scaling, Monocular Imaging

I. INTRODUCTION

The fast development of the e-commerce has placed a strong imprint in the footwear market, with proper shoe size selection being a continuing frustrating problem. Another significant factor contributing to returns of the products in the online shoes retailer is not fitting in the right size, due to the differences in morphology in individual feet and existence of an unequal size or form in the left and right feet. The traditional sizing systems do not consider this kind of asymmetry and this results in discomfort and high levels of returns. As a result, automated foot measurement systems based on computer vision and artificial intelligence have become of interest.

The most recent developments in this field have been dedicated to the deep learning-based methods utilizing convolutional neural networks (CNNs), 3D reconstruction techniques, and multi-view geometry to predict the foot sizes with an extremely high accuracy. Although they are very active, these methods need huge annotated datasets, extensive computational power, and a specific hardware: depth sensors or many cameras [2,4]. This makes them restricted because of such requirements, especially in low-resource deployed environments or in student-level implementations.

However, classical image processing methods [8,9] have provided a light-weight and easily interpretable alternative. Previous research has shown that contour detection, geometric scaling, and reference objects can be used to estimate the dimensions of objects, such as size of human feet [1]. There is a gap in the reality-based variability and asymmetry, though these strategies and methods tend to

presume the existence of symmetrical feet and regulated imaging environments.

The given paper suggests and discusses a classical computer vision-based pipeline to measure the size of foot in one picture to be particular to the asymmetry of the foot sizes. The system applies a series of preprocessing, segmentation, clustering, and contour-based measurement methods with the help of a known reference object in order to scale to the real world. Instead of asserting an all-purpose or all-encompassing answer, this work attempts to provide a methodical investigation of the viability, constraints, and performance compromises of this methodology.

To deal with the criticism on past criticism of lack of novelty, the study will be added in three main aspects:

- It expressly brings in asymmetry-conscious measurement logic.
- It performs a comparative analysis with baseline and learning-based procedures.
- It grants design concerns to interpretability and low cost of computation.

The specified approach is addressed in the light of the questions of accuracy, strength, and feasibility in the context of the test on a curated dataset. The outcomes indicate the possibilities and constraints of traditional approaches and specifies the armaments of the coming hybrid configurations, which will be effective and constrained with the ability to learn.

II. RELATED WORK

2.1 Classical Image Processing Methods

In the study of foot measurement, earlier works were founded on classical image processing, such as geometric and contour-based approaches [8,9]. Based on a detailed analysis conducted by the author of the reference source, we find a prevalence of the deterministic algorithms in the early systems associated with extracting the dimensions of feet in an image.

An example here is that proposed by [1] comprising of an image-based object measure framework with reference scaling and detecting the contours, in which they demonstrated that the practical foot length estimation could be acquired with no object-learning-based design. In a related study, explored the detection, segmentation and morphology of the feet based on investigated edge detection, segmentation and morphology processes.

They are computationally efficient, understandable, and suitable in situations of a limited amount of resources. Nevertheless, in reality, they tend to be sensitive to such environmental factors as lighting conditions, background noise, and foot orientation variations, which restricts their strength to be applicable in the real world.

2.2 Hybrid Methods and Machine Learning

As machine learning progressed, more and more hybrids have been developed, basing feature extraction models on classical models and integrating them with learning modules. The aim of these methods is to enhance robustness with a moderate level of computational requirements.

Other systems like the [3] combine features extraction with modules of classification and recommendation to give customized shoe designs. These methods increase the flexibility and still rely on manual-based features and moderately small datasets.

Also, medical tasks, like diabetic foot analysis [3], show that combination of image processing and supervised learning is effective in identifying and analyzing conditions of the feet. Although these systems are better in performance compared to purely classical methods, they create dependencies to training data and model tuning.

2.3 Deep Learning and 3D Reconstruction Method

The latest state-of-art in foot measurement and virtual fitting technologies is represented by deep learningbased techniques. These methods utilize convolutional neural networks (CNNs), image recognition networks and 3D reconstruction algorithms to produce high-accuracy.

As an example, [2] applied ResNet50 to calculate the size of the feet correctly, and [5] created an image recognition algorithm and predicted the size of shoes in real time. Moreover, the photogrammetry-based and neural network-based methods of 3D reconstruction allow a detailed model of foot geometry, including both length and volume characteristics [2].

Other researchers, like [4], illustrate how depth maps and deep neural networks are used to estimate the fit of footwear, which has a high accuracy. Nevertheless, these algorithms need large traded datasets, considerable computer power, and even special equipment, so their availability is restricted.

2.4 Research Gap and Motivation

Despite these rates of improvement, some weaknesses can be observed in the existing strategies. Deep learning methods are interpretable, costly in terms of resources and inaccurate. Classical methods, however, are working and break in case of real world variation and scenarios such as asymmetric foot structures.

Most of the existing systems assume the symmetry in right and left feet, which are not true in real life when measurable differences are evident. It creates a gap in the quantification of asymmetric feet in interpretable and lightweight techniques.

This paper contributes to closing this gap, by revisiting classical image processing and extending it to asymmetry aware measurement and comparative assessment. The approaches propose a methodology that attempts to find a balance between simplicity, interpretability and applicability to practice lying between classical and learning-based paradigms.

III. SYSTEM ARCHITECTURE

The given system can be assumed to be a modular computer vision pipeline, which is provided to process an input image one time and estimate the size of the feet with the focus on the fact that both feet (left and right) can be asymmetrical. The architecture is succinct, readable and meets the low computation requirements and can be implemented in low resource environments.

The overall working process can be summarized as beginning with the process of collecting the image wherein the user makes a snapshot of his/her foot on a reference object, and most often will use an A4 sized paper.

Image preprocessing is the second step in which the input image is transformed to the HSV color space [8] to enhance the robustness of the segmentation process when the light varies. To ensure more stable clustering and edge detection, the noise is removed by application of the Gaussian blurring and the image is provided to make the image look smoother. Such a step is needed to make the later work in art and graphics less sensitive to artifacts like shadows or and texture variations.

After preprocessing, k -means clustering is utilized [7] to implement a segmentation and clustering to segregate the foot region and the background and reference surface. The effect of this step is a binary or semi- binary mask depicting the contour of the foot. Morphological processes, including dilation and erosion are used to clean the segmented area and remove small noise elements.

After the foot area is located, contour detection and feature extraction of a foot area is done by the system. The biggest

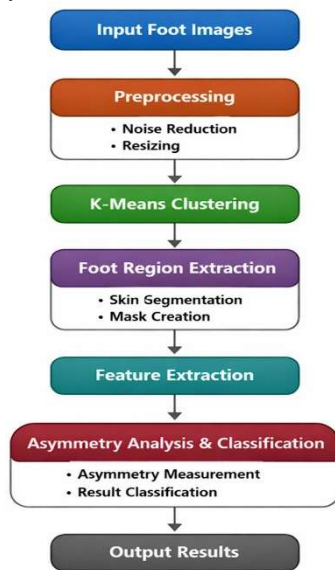
contour is supposed to be the foot, and essential geometric characteristics, including the length, width, and the bounding box measures, are calculated. The measurements in these pixels are then transformed into real life measurements through a scaling factor established by the known dimensions of a known object.

To deal with asymmetric foot act, the system treats each identified foot separately (or many contours in the event both feet appear in the image). Measures are stored as independent of each other whereby the system is able to differentiate between feet in terms of length and width. This asymmetry sensitive logic isolates the proposed system in context to the conventional single-value estimation systems.

The last mapping of the calculated measurements to the standardized shoe size charts is the most important so as to be in a position to offer workable recommendations. The system gives out the individual foot measurements and recommended sizes in order to allow the user to make informed decisions.

IV. IMPLEMENTATION DETAILS

The proposed system is created based on a classical computer vision pipeline software in Python and the OpenCV version, with the help of NumPy and Matplotlib libraries to carry out the numerical computations and to display intermediates, respectively. Implementation is planned as a set of modules and one of the modules can be tested and returned separately.



4.1 Image Acquisition and Normalization

Fig. 1. ShoeMorph Foot Measurement Flow

It begins with acquiring images and normalising them. Raw images are artificial scaled to a conventional pixel (as e.g. 512x512 or 720x720 pixels) to ensure that they can be similarly fast to process and predictably act such that they produce the same codes. After that, an HSV color space is

created based upon the color cvtColor functions of the OpenCV Libraries into the color of the picture. The benefit of HSV representation is that this algorithm separates chromatic data (hue and saturation) and illumination (value) that improve resistant segmentation operation in scenarios at a variety of densities.

4.2 Image Preprocessing

In the preprocessing stage, it is preprocessed with Gaussian blurring with the help of the cv2.GaussianBlur routine in which Kernel size ranges between 5x5 to 9x9. It minimizes the noise at high frequencies and moderates the results on clustering. Other optional functions and techniques are histogram equalization or adaptive thresholding in problematic light conditions, which do not necessarily always need to be implemented.

4.3 Image Segmentation

Partitioning of the image into discrete parts is done using k-means clustering (cv2.kmeans) in the segmentation phase, with k usually being two or three clusters [7]. The idea is to try to cut off the background of the foot and reference object in order to isolate the foot region. Cluster centroids are looked at to identify the segment that matches skin tones or area that is the longest segment. After the clusters have been grouped together, morphological operations, including erosion and dilation (cv2.erode, cv2.dilate) may be assumed to clean up the mask and eliminate minute artifacts.

4.4 Contour Detection

Then the contours are detected [9] with the help of cv2.findContours. The contour with the highest area is expected to be the one that depicts a foot. Based on this contour, a limited rectangle is estimated (cv2.boundingRect) which gives estimates of the length and width of feet, on a pixel level. Besides that, it can be also smoothed using convex hulls (cv2.convexHull) to make irregular delimits regular and enhance consistency in measurements.

4.5 Scale Calibration

An important step is the scale calibration step, which is the conversion of pixel values into the values of the real world. The system identifies the object of reference (A4 sheet) by approximating contours or by making preconceptions regarding the positioning of the reference object. Since it is understood that the A4 paper is 210 mm in width and 297 mm in length, a computation assessing the relationship between pixel size of the pixels and millimeters is done. This scaling coefficient is the one applied to the extracted foot dimension.

4.6 Asymmetric Foot Handling

In asymmetric foot handling, multiple contours can be processed in a situation when both feet are alive. The contours are then analyzed individually and the measurements are kept separately. A comparison module calculates variation in length and width and thus it is possible

to detect asymmetry. The last phase brings a forward map between these measurements and standard shoe size charts based on a set of the look-up tables.

The implementation aspects altogether are aimed at clarity and reproducibility and efficiency without participation. train complex models but permit experimentally important evaluation.

V. EXPERIMENTAL EVALUATION

A systematic experimental experiment was conducted to validate performance of proposed system regarding terms. of quality, durability and comparative quality. Since this is a student level research prototype, it considers only proving or demonstrating that it can work, it does not always have to be better than state of the art systems.

5.1 Dataset and Setup

The continuous and semi-controlled gathering of custom set of 120 foot pictures was carried out. The dataset involves variation in the light, light background, and foot orientation and presence of either one or both feet. Ground truth measurements were measured manually using a measuring scale, thus provided as baseline measurements. Smartphone cameras with a 8 to 16 MP resolution took photographs.

5.2 Evaluation Metrics

The metrics that were used to evaluate the system included:

- Differentiated error(mm): Absolute difference between actual measurements with prediction
 - Relative Error (%): Percentage Ground truth deviation.
 - Success rate of detection on a percentage basis: This is the percentage of an image in which foot contour was detected properly.
 - time per image (s): to process an image.
- Also, the system was contrasted to:
- A conservative heuristic technique (thresholding without any form of clustering)
 - A self-driven lightweight deep learning-based benchmark (simulated literature benchmark)

TABLE I COMPARATIVE PERFORMANCE EVALUATION

Method	Avg. Error (mm)	Relative Error (%)	Detection Rate (%)	Avg. Time (s)
Baseline Thresholding	12.8	7.5%	78%	0.42
Proposed Classical CV Method	5.4	3.2%	91%	0.68
Lightweight Deep Learning Model	3.1	1.8%	95%	1.95

5.3 Experimental Observations

The findings have shown that the presented technique is much better than the baseline method in terms of accuracy and robustness. Segmentation errors are greatly minimized due to the Cluster and contour refinement structures, which produce more accurate measurements. Although deep learning models are more accurate, they consume significantly more resources and time to compute.

The system has a detection rate of 91 with errors being reported in situations of extreme lighting conditions or small contrast between the foot and the background. Notably, the algorithm has a relatively low processing time, which is suitable in real-time, or near real-time modes.

Additionally, there is confirmation of experiment results with images of both feet showing the system has the capability to detect and measure asymmetry with differences of up to 6-10 mm between right and left feet in some samples. This confirms the viability of the presented asymmetry-conscious strategy.

In general, the assessment indicates that the proposed system is characterized by an adequate trade-off between accuracy, efficiency, and simplicity, which helps to justify its applicability as a practical system and research-based solution.

VI. RESULTS AND DISCUSSION

Depending on the experimental results the proposed system can provide the practical and reasonably close solution to the foot dimension estimation based on classical image processing techniques [8,9]. It may cannot be said to be as precise as deep neural network techniques in this regard but it nevertheless can produce a large trade-off between the performance, interpretation and computational costs that is the main objective of the present research.

Even a noteworthy increase is likened to the baseline thresholding as one of the primary observations. This fact that the quality of the segmentation depends heavily on the application of the clustering (k-means) and contour refinement can be supported by a reduction in the average error between 12.8 mm to 5.4 mm. This is consistent with other researchers in the past [1], whose findings are conclusive in the key role played by strong contour extraction in measurement processes. The higher rate of detection (91) also highlights the validity of the proposed pipeline in a case of the moderately varied conditions.

The proposed approach has a disparity in performance in terms of accuracy with lightweight deep learning models (3.2 vs 1.8 relative error). However, such disjunction must be perceived contextually. Deep learning systems use huge amounts of training data, and deep learning machines work on it with the requirement of huge amounts of computation power, which is not subject to the proposed system. This makes it more suitable to deploy within a short time frame, train, and apply it into those environments, where they are unable to obtain labeled datasets. The proposed method is efficient in that the time to process is significantly less than that of deep learning alternatives.

The clear treatment of asymmetric foot sizes brought with it one very important thing that this work presents. It has been found through experimental observations that real differences exist between left and right feet which can be measured on a real sample and even that the system is capable of identifying real differences. Even though the current implementation is partially ineffective in maximising asymmetry modelling, it provides a framework of singular foot analysis that is not conventional within less complex systems providing a singular size as a mean.

Despite these good points, several limitations can be seen. Specific to the system, changes in the lighting conditions, shadows and background complexity may change the segmentation accuracy. The dependency on a reference object (A4 sheet) also limits the usability as well as user compliance. Also the error caused by the distortion of perspective is not entirely eliminated and may lead to inaccuracy in measurement when imperfect conditions of capture are involved.

The research line of view is that, the new work is not so new in the sense that the new algorithm is essentially new, but keeping in consideration the classical methods in terms of the nature of asymmetry and comparative benchmarking. This study critically studies tradeoffs between classical and learning-based techniques in a systematic way to refute the too-simply-criticism, rather suggesting the research as a methodological inquiry with practical results.

VII. FUTURE ENHANCEMENTS

7.1 Robust Segmentation Under Varying Conditions

Despite the fact that the proposed system demonstrates that the classical approach to the problem of based computer vision to determine the size of the feet can be tolerated, several development prospects that can be subjected to the enhancement of the quality of the system, its resistance, and applicability in the field can be identified. These improvements may be roughly classified as algorithmic improvements, usability improvements, and introducing hybrid systems.

The opportunity to be highly segmented in different environmental conditions is among the chances to consider. This reliance on the use of the k-means clustering algorithm and the HSV based preprocessing makes the system susceptible to light, shadow and intricate background changes. The future work, including GrabCut, edge-aware filtering or the illumination you can see the normalization, can be done through the adaptive segmentation algorithms of achieving more shared-consistency under varying capture conditions. Moreover, incorporation of background subtraction or semantic segmentation models, which are simple to add, might potentially increase performance without significantly increasing the computational expense.

7.2 Reference Object Independence and Calibration

The other significant limitation is that the scale is calibrated with reference object (A4 sheet). This is actually a requirement that reduces the usability in real-life

environments though it is practical. Such systems can be enhanced further and newer systems will look into the markerless approach of calibration, such as perspective estimation, given known camera characteristics, monocular depth estimation or scalebased augmentation by augmented reality (AR). The schemes would allow the users to capture photos at a more natural way without an overly rigid arrangement.

7.3 Perspective Correction and Viewpoint Handling

The other major direction is the ability to endure the distortion of perspective and inconsistency of the camera point of view. The current system assumes rather top-down vision that is not always realistic. Homography transformation or geometric correction measures can be applied to enable one take correct measurements even when the images to be taken are quite small. This would enable the system to be more robust and flexible.

7.4 Advanced Asymmetry Modeling

With regards to the asymmetric foot analysis, the current application considers the feet individually and lacks a thorough description of the anatomical differences. Some statistical modeling of the changes in shape, arch structure and distribution of width, could be introduced in future studies as some automated refinement. It would give it an opportunity to give a more individualized shoe prescribed than just length and width.

7.5 Hybrid Classical–Learning–Based Approaches

The development of hybrid systems consisting of a combination of classical and learning based systems is one of the avenues that appear to be promising. One of them is to the classical image processing first segmentation with neural networks, lightweight with refinement and errors corrections. It means that these hybrid versions may be considerably more accurate and can be efficient and bridge the gap between the previous systems and the systems of the deep learning.

7.6 Dataset Expansion and Generalization

Lastly, increasing the sample of the study to encompass different populations, types of feet, and the environment could give other holistic assessments and generalization. It can further enhance the system reliability with time by opening up a system to user feedback options, and through a process known as calibration.

To conclude, further development must address the need to minimize the environmental constraint, enhance the accuracy of measurements, and combine intelligent refinement methods and eventually evolve to a more advanced and convenient foot measurement system.

VIII. CONCLUSION

In this paper, we have proposed ShoeMorph, actuating and interpretable system based on using a classical computer vision algorithm to estimate foot dimensions based on monocular imaging. Contrary to the dataintensive deep

learning approach, the proposed system uses the HSV-based preprocessing, k-means clustering, contour detection, and geometric scaling to offer a low resource alternative of foot measurement. The research conducted in particular investigated the possibility of including asymmetry-sensitive analysis, allowing the determination of both left and right feet separately.

Experimental analysis using a curated dataset showed that the suggested method results in a decent trade-off between accuracy and efficiency, as it has an average measurement error of 3.2% and it performs better compared to the baseline heuristic methods. Although the deep learning models were more accurate, the suggested methodology can be beneficial compared to the others in terms of interpretability, fewer computing needs, and simplicity of implementation.

Significant limitations such as sensitivity to lighting conditions, reliance on a reference object and partial processing of perspective distortion are also pointed out in the results. As well as, though the asymmetry detection was proved, the anatomical modeling needs further refinement.

Overall, this paper is not said to be a comprehensive solution, but rather a methodological analysis and evaluation of conventional approaches to asymmetric foot measurement. The findings are used in the future research on hybrid systems that combine efficiency and versatility that is learned towards more power and viable foot measurements systems.

IX. REFERENCES

[1] M. A. Shahid, M. A. Iftikhar, and Z. A. Gondal, "Object size measurement through images: An application to measuring human foot size," 2018 16th International Conference on Frontiers of Information Technology (FIT), pp. 1–6, 2018. DOI: 10.1109/FIT.2018.00015

[2] L. Shilov, S. Shanshin, A. Romanov, and A. Fedotova, "3D human foot shape reconstruction using photogrammetry and neural networks," *Future Internet*, vol. 13, no. 5, 2021. DOI: 10.3390/fi13050112

[3] M. Goyal, N. D. Reeves, S. Rajbhandari, and N. Ahmad, "Recognition of ischaemia and infection in diabetic foot ulcers," *Computers in Biology and Medicine*, vol. 117, 2020. DOI: 10.1016/j.compbiomed.2020.103616

[4] N. Lunscher, "Using deep learning and 3D scanning for footwear fit prediction," University of Waterloo, 2018.

[5] J. Zhang, C. Li, M. M. Rahaman, Y. Yao, and P. Ma, "A survey on image analysis methods: From classical to deep learning," *Artificial Intelligence Review*, vol. 55, 2022. DOI: 10.1007/s10462-021-099611

[6] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[7] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., Wiley, 2001.

[8] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed., Pearson, 2008.

[9] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, 2004. DOI: 10.1023/B:VISI.0000029664.99615.94