Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 26 (2023), No. 3, pp. 685–693
DOI : 10.47974/JDMSC-1741

Computer vision, machine learning based monocular biomechanical and security analysis

Arun Kumar [#]
Department of Electronics and Communication Engineering
New Horizon College of Engineering
Bengaluru
Karnataka
India

Himanshu Sharma [†]
Shruti Mathur [§]
Dimpal Sharma [‡]
Girraj Khandelwal [@]
Gajanand Sharma *
Department of Computer Science & Engineering
JECRC University
Jaipur
Rajasthan
India

Abstract

Modern computer vision technologies have served to bridge the gap between contemporary scientific analysis and machine learning assisted digital processing. Within the field of biomechanics, applied strategies incorporating both conventional and machine assisted means have shown great success in augmenting the observations of electromyogram graphical sensors; albeit within the constraints of specialized, multiple-source arrays. The ongoing study represents an endeavor to utilize several distinct technologies to achieve

[#] *E-mail:* dr.arunk.nhce@newhorizonindia.edu
[†] *E-mail:* Himanshu.sharma@jecrcu.edu.in
[§] *E-mail:* smsavi@gmail.com
[‡] *E-mail:* dimpal286@gmail.com
[@] *E-mail:* girraj13@gmail.com
* *E-mail:* Gajanan.sharma@gmail.com (Corresponding Author)

similar results with the application of a single optical sensor. This paper documents the research, development, and implementation of a monocular feature extraction pipeline, designed for the intended use of supplementing modern athletic biomechanical analysis and the security of the data. We will examine the techniques presented by related works, and how we have implemented these strategies into our framework.

Subject Classification: *Primary 68T45, Secondary 68M25.*

Keywords: *Computer vision, Machine learning, Biomechanics, Motion tracking.*

1. Introduction

Over the past decade, advances within the study machine learning have paved the way for groundbreaking applications and novel pattern recognition and analysis strategies. The aim of this study is to discuss our efforts in developing a modular framework of computer vision technologies to generate kinematic data efficiently and accurately with the application of a singular sensor. This framework is devised with the intent for use in the field, capable of functioning with real-world, sports and/or health science footage. The benefits of such a tool have been discussed at length within the sports and health sciences community [1]. Currently, to obtain sufficiently accurate optical tracking data of physical activity requires the use of an array of sources, consistently calibrated to provide spatially and temporally consistent results. In addition, most capable systems necessitate the application of markers to the observed subjects' joints, to correct for optical distortions and/or outlier frames. Although remarkably consistent, the classical approach is therefore confined to observation within an unnatural setting [2]. By providing a means to accomplish comparable observation in the field, the proposed pipeline would serve to augment research capabilities by lessening the impact of such factors. It is hypothesized by the members of this study, that such a goal is immediately attainable with several recent developments within applied computer vision. To accomplish this task, development will incorporate several existing technologies, forming a modular and efficient processing pipeline structure [3]. The subset priority of each integrated module will be independently discussed in section 3a.-c. With the improvements presented in future iterations, development aims to generate optimized and lightweight prototypes to improve measurement capabilities in the field of biomechanics.

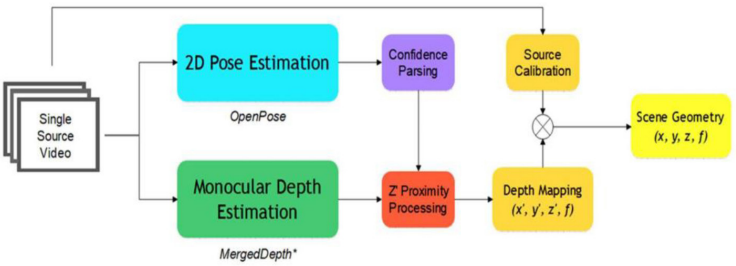

Figure 1

Monocular Biomechanics Pipeline

2. Related Work

Figure 1. Monocular Biomechanics Pipeline provide estimations on the occluded instances incurred with a single source projection. Within the study of human biomechanics, unencumbered athletic and rehabilitative motion tracking techniques have become vital to candid observation [4]. Conventional hardware requirements limit the reach of capable motion tracking arrays to single sites, often impeding the anticipation of unknown variables in real- world situations. As the role of computer vision has developed, studies have achieved great success in this regard, supplementing smaller stereo sensors with image processing and machine learning frameworks to obtain viable kinematic data [5]. Such technologies have already begun to demonstrate their unparalleled utility in the field, particularly in the way of sports and exercise science [6], presenting new avenues to observe and measure athletic prowess. Intel True View has demonstrated effective information securing, with these technologies setting milestones within these fields, it may be possible to reach a bit further [7]. Our method, as shown in Figure 1, proposes a pipeline of existing technologies through which biomechanical data can be obtained with the use of a single optical source. By coordinating the efforts of two distinct neural architectures acting upon the same source, pseudo-3-dimensional joint data can be estimated from a single projection. The resulting points, existing within a projection respective of the source, can be filtered for spatial accuracy [8].

2.1 *Contribution*

As a baseline argument for the validity of our hypothesis, we will examine the strategies, successes, and shortcomings of recent and

similar endeavors. Regarding the proposed single source human pose estimation, several modern studies have demonstrated their success with the implementation of affinity field supplemented Convolutional Neural Networks (CNN). The methods used to temper confidence map output layers with part affinity predictions. Furthermore, the authors show how such methods can offer consistent and stable multi-person 2D pose segmentation. Their model (OpenPose) has been utilized here as the foundational pose estimation branch within our pipeline. Regarding depth estimation, traditional methods for obtaining reliable depth maps incur the use of multiple synchronized sources for feature point and extrinsic transformation correlation [9]. With only a single source however, the problem becomes much more complex. In the absence of imaging parallax, attempts at refining the conventional strategy propose the use of encoder-decoder networks, subpixel geometry, and intrinsic source parameterization in segmentation [10]. In the cited works, these authors have shown that most developmental difficulty arises from the effective training of segmenting models, as correcting for spatial occlusion and temporal inconsistency requires a specialized supervision implementation. In the following sections describing the depth estimation branch of our pipeline, we will discuss such algorithms and our findings in the scope of this application.

3. Processing Pipeline

3.1 *Secure Process of Source Image Sequence*

To conceptualize the suggested workflow, we will discuss the relevant mathematics of processing, classifying, and interpreting data sets as applied within this study. Characteristically, raw data enters the system as image sequences, indexed temporally. Consequentially, this set (S_n) comprises all algorithmic inputs of four-dimensional arrays, whose elements denote spatial (x, y), color (c), and frame (f) values for any source (s_n), as well as the extrinsic (M(f)) and intrinsic (K(f)) parameterization of the source (S_n). For our purposes, source parameterization will not change between framesets. Pipeline input data can be described:

$$S_n \ni [s_n(x, y, c), f), s_n\ M_n\ K_n \quad (1)$$

where:

$$M_n = (R\,t\ ), \forall f \quad (2)$$

$$K_n = [\alpha_x\ \gamma\ x_0\ 0\ \alpha_y\ y_0\ 0\ 0\ 1\ ], \forall f \quad (3)$$

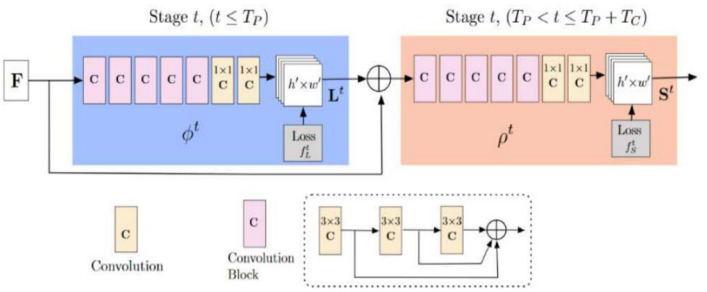

Figure 2

***Open-Pose* primary model**

3.2 *Pose Estimation Branch*

Once prepared, the frames (*sn*) are passed through the first 10 layers of an initial feature extraction CNN (VGG- 19, *PoseNet*[]), from which the returned feature maps (*F(sn)*) are separated by source and frame indices:

$$F(s_n) = PoseNet(s_n)\Big|_{c_1}^{c_{10}}$$
$$\{VGG(s_n[:,:,:,1]), VGG(s_n[:,:,:2]), \ldots VGG(s_n[:,:,:,f])\} \quad (4)$$

Each frame of the feature map array is input to the OpenPose [4] network, delivering the map to the first-stage input and second stage combined input as shown in Figure 2.

OpenPose employs the use of adaptive affinity field driven supervision, modifying the utilized loss function to evaluate pixel regions rather than conventional pixel-wise cross entropy. In our pipeline, joint positions ($\boldsymbol{J}(s_n)$) are received via *OpenPose*'s output JSON structure functionality, containing spatial coordinates and confidence metrics for each member of the "BODY_25" joint model. Parsed from the network output, the data can be described as:

$$J(s_n) = OpenPose(F(s_n)) = S'(L'(F(s_n)) + F(s_n)) \quad (5)$$
$$J\hat{}(s_n) = [\text{"}j\{body\_25\}\text{"}, (x, y,\ confidence(0:1))] \quad (6)$$

4. Security Depth Estimation Branch

4.1 *MonoDepth2*

To estimate scene depth, each sequence (***SP***) was passed through the *PyTorch* implementation of the Monodepth2 network. This work aims to provide a practical route to monocular depth estimation, which in the case

for the success of this project could neither be ignored nor superseded. In their paper, Godard et al. describe the U-NET architecture used to gather feature maps, and the separate pose correspondence encoder network used to cross check geometric anomalies and verify the efficacy of the unobscured features [11].

4.2 *Post-Processing*

Once the key-point positions ($J(s_n)$) have been extracted and the depth maps ($M(s_n)$) have been processed, each estimated joint position can be given a 3- dimensional coordinate for every unobstructed frame. The accuracy and stability of the resulting model at this stage is highly dependent upon the applied resolution of the OpenPose network; occluded joints or temporal noise maintained for any significant number of frames cause irrevocable information loss. To account for such error, the x-y joint data is parsed to flag and hold any joint below the confidence threshold, denoting some type of distortion or exclusion.

$$DC_{Pos}(s_n) \ni M([J\hat{}(s_n) : conf > 0.5]), \forall\ "j\{body\_25\}" \tag{7}$$

$$DC_{Null}(s_n) \ni M(J\hat{}(s_n) : conf < 0.5]), \forall "j\{body\_25\}" \tag{8}$$

With the fitted and flagged x-y joint data obtained, placement in the relative z-dimension can commence. To accommodate for any temporal flickering and spatial disparity, the coordinates of each non flagged joint-frame are used to gather a neighborhood (p_n*p_n) region of the depth map at the desired point. Sharp edges within these neighborhoods are detected and avoided via Canny filtering, and the values of the surrounding region are averaged with distance-to-center respective weight. The resulting value is then assigned as the effective z- dimension for the joint. With all high confidence 3- dimensional data mapped, the positions of each flagged joint-frame are be interpolated by fitting the existing information. To further improve the effectiveness of this interpolation, it can be demanded that the lengths of each joint-link pair are maintained with a margin proportional to the expected "unobstructed" error, as these metrics ideally will not change. Secondly, the data from each joint is then modeled outside of the discrete frame interval with the MATLAB Curve Fitting toolbox to further reduce spatial noise. Correlation success rates vary with implementation and preprocessing techniques applied to the input tensors, however the resulting models proved exceptionally helpful in isolating sources of macroscopic error. Confidence values are scaled as

point weights for this interpolation, and the regression method is tuned to deviate only from lapses in confidence between frames.

5. Results

The monocular computer vision pipeline was applied to two different human activities. The back squat exercise and manual wheelchair propulsion were video recorded using four GoPro cameras and five USB3 machine cameras while an eight camera Vicon motion analysis system simultaneously tracked the reflective markers shown in Figure 3. The Vicon system was used as a validation source since it has been shown to accurately analyze motion with a high degree of consistency.

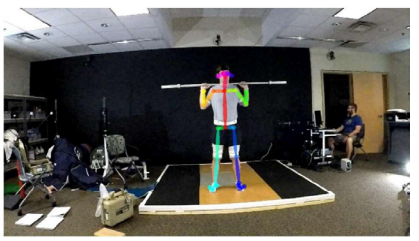

Figure 3

***Open-Pose* output visualization**

Figure 4

***Merged-Depth* output visualization**

Each single source video was analyzed individually by the computer vision pipeline to validate the calculated three-dimensional joint coordinates with the joint coordinates generated from the Vicon system indicated in Figure 4. A Pearson-correlation calculation was used to report the strength of the relationship between the single source and

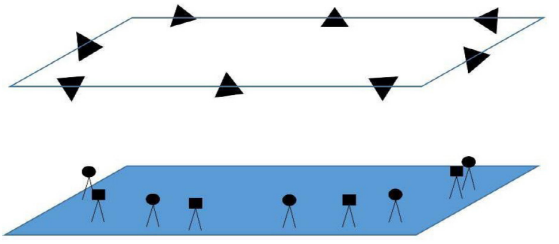

Figure 5

Camera setup for data collection

◀ = Vicon camera; ● = USB3 camera; and ■ = GoPro 4 camera

multi-source generated data set. To test the robustness of the computer learning pipeline the cameras were positioned at different optical angles to determine the effect of camera location on the accuracy of determining joint kinematics shown in Figure 5.

6. Conclusion

Although currently the spatial pseudo-data implies there remains some active sources of noise propagating through the framework, initial implementations have laid the groundwork for the progress of the monocular biomechanics pipeline. Plans for future work include further optimization of module interaction, and the construction of a specialized network for depth estimation, to be trained with outputs from the current pipeline. Furthermore, observations made during development demonstrate an opportunity for further isolation of the biomechanical signals themselves, via disturbance modeled interpolative signal processing, and Kalman estimation. Development in this regard will focus on minute corrective elements; a sort of "aim small, miss small" approach.

Conflict of Interest:

Funding: There is no funding support in this work.

Compliance with ethical standards

Conflict of interest: All authors do not have any conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors

References

[1] Xiangyun Zha. Intelligent correction algorithm for chromatic aberration of led spectral image under multiple-beam interference. *Journal of Discrete Mathematical Sciences and Cryptography* 21:2, pages 327-332 (2018).

[2] Arun Kumar, Pragati Tripathi and Alaknanda Ashok "Novel scheme of k-SVM analysis using PCA and NN for detection of MRI brain images", *Journal of Interdisciplinary Mathematics*, Vol. 23, No. 5, pp. 967-976 (2020). DOI : 10.1080/09720502.2020.1723923.

[3] Lee, W., Chen, H., Chen, M., Shen, I., & Chen, B. (2017, October 13). High-resolution 360 Video Foveated Stitching for Real-time VR. Retrieved September 25, 2020.

[4] Z. Cao, G. Hidalgo, T. Simon, S. Wei, Y. Sheikh, OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, Cornell University (2019).

[5] He, Kaiming et al. "Deep Residual Learning for Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 770-778.

[6] G. Refai-Ahmed, Hoa Do, A. Raghupathy, R. Kadam and J. Gillis, "Thermal design of monocular vision system used in automotive application," 2017 18th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), pp. 1-7 (2017), doi: 10.1109/EuroSimE.2017.7926243.

[7] Anil Kumar & Sandeep Kumar Sharma. Information cryptography using cellular automata and digital image processing, *Journal of Discrete Mathematical Sciences and Cryptography*, 25:4, 1105-1111 (2022), DOI: 10.1080/09720529.2022.2072437.

[8] J. Dong, X. Cheng and T. Ikenaga, "Multi-physical and Temporal Feature Based Self-correcting Approximation Model for Monocular 3D Volleyball Trajectory Analysis," 2021 17th International Conference on Machine Vision and Applications (MVA), pp. 1-4 (2021), doi: 10.23919/MVA51890.2021.9511408.

[9] Teoh, S.S., Bräunl, T. Symmetry-based monocular vehicle detection system. Machine Vision and Applications 23, 831-842 (2012). https://doi.org/10.1007/s00138-011-0355-7.

[10] Boeing, A., Braunl, T.: ImprovCV: Open component based automotive vision. In: Intelligent Vehicles Symposium, 2008 IEEE, pp. 297-302 (2008).

[11] Azhand, A., Rabe, S., Müller, S. et al. Algorithm based on one monocular video delivers highly valid and reliable gait parameters. Sci Rep 11, 14065 (2021).