UDC 004.932

# 3D HUMAN POSE RECOGNITION SYSTEM: METHODS, TECHNOLOGIES, AND REAL-TIME APPLICATIONS

***Suleimenova Alua***

Master's student, Applied data Analytics,
Department of Computational and Data Science
Astana IT University, Astana, Kazakhstan

*This article discusses the development of a system for automatic analysis and classification of three-dimensional human poses based on a real-time video stream. The proposed approach includes determining the main body positions using the coordinates of key points and dynamically tracking changes in posture in order to identify potentially dangerous or uncomfortable conditions. As part of the work, a time control mechanism is implemented: the system records the duration of being in a dangerous pose and generates a warning when a set threshold is exceeded. The developed algorithm demonstrates high sensitivity to critical changes in body position and can be integrated into intelligent behavior monitoring systems in various application areas, including industrial safety, medical surveillance and real-time personnel management. Special attention is paid to the issue of minimizing computational costs while maintaining high recognition accuracy, which makes the proposed solution suitable for use in conditions of limited resources. The relevance of the study is determined by the increasing requirements for automatic human condition monitoring systems in a dynamically changing environment.*

**Keywords:** 3D human pose, pose recognition, motion analysis, real-time monitoring, automatic detection, security control, intelligent systems.

## *Introduction*

3D human pose estimation is a key technology with a wide range of applications, including augmented and virtual reality, healthcare, sports analysis, and human-computer interaction. Accurate and real-time reconstruction of human poses from videos or images remains a challenging task, requiring solutions that combine high accuracy, computational efficiency, and adaptability to different environments.

A number of methods have been developed in recent years that have significantly advanced the field of 3D pose estimation. However, despite impressive progress, existing models often face limitations related to robustness in real-world conditions, latency during inference, and the need for further optimization for lightweight or mobile platforms.

This work focuses on two key objectives. The first is to review state-of-the-art approaches to 3D human pose estimation. The second is to develop and implement improvements to the existing solution aimed at improving the accuracy and robustness of pose reconstruction while maintaining the speed required for real-time applications. By analyzing current methods and proposing targeted improvements, this study contributes to the continuous development of efficient and practical systems for 3D human pose estimation that are suitable for use in real-world settings.

***Literature review***

In recent years, the task of 3D human pose estimation has attracted significant attention from researchers. This interest is driven by the development of technologies in augmented reality, healthcare, sports, and other fields where accurate and fast pose recovery in real-time is crucial. Against this backdrop, numerous solutions have emerged, each addressing specific technical challenges. One of the early notable works in this area was presented by Pavlakos et al. (2017), who proposed a method of step-by-step refinement of 3D pose predictions using volumetric representations. [1] Their Coarse-to-Fine Volumetric Prediction approach built a spatial model by gradually refining joint positions. This strategy helped address the ambiguity problem inherent in reconstructing 3D coordinates from 2D images. However, despite its accuracy, the model demanded considerable computational resources, limiting its applicability for real-time systems. The search for greater efficiency led researchers to reconsider rigid localization methods based on heatmaps. In this context, an important milestone was the work of Sun et al. (2018), who introduced Integral Human Pose Regression. Instead of simply locating peaks in heatmaps [2], this method used weighted averaging, significantly improving localization accuracy without increasing computational cost. Such an approach became especially relevant for systems intended to run on resource-constrained platforms like mobile devices. Nonetheless, many direct coordinate regression methods still struggled with anatomically implausible predictions. To overcome this challenge, Kolotouros et al. (2019) introduced an innovative method called SPIN, which combined direct prediction of human body model parameters with optimization based on observed data.[3] This hybrid approach greatly enhanced the physical plausibility of 3D poses and improved robustness to noise and recognition errors. At the same time, it became clear that analyzing single images limited the potential of pose estimation systems. A natural next step was leveraging motion information from frame sequences. A prominent example of this direction was the work of Kocabas et al. (2020), who introduced the VIBE method.[4] By utilizing recurrent networks for video analysis, VIBE achieved not only higher accuracy but also much smoother pose predictions, which is critical for applications in animation and virtual reality. As demands for accuracy and robustness continued to grow, more powerful architectures were required to model spatial-temporal dependencies. In this regard, the work of Dong et al. (2021), who introduced MotionBERT, marked a major advancement.[5] By

bringing transformer principles into 3D pose estimation tasks, they significantly improved prediction quality through modeling complex relationships between joints over time. This allowed for accurate recovery of movements even in dynamic and challenging scenes. While academic research advanced, developing practical, real-time solutions remained an essential direction. A noteworthy contribution here was MediaPipe Pose by Google Research (Lugaresi et al., 2019).[6] Unlike heavy research prototypes, MediaPipe was designed from the outset for lightweight integration into mobile devices and browsers. It delivered high-speed performance with reasonable accuracy, making it one of the most popular tools in the industry. Finally, the role of datasets in training and evaluating these systems cannot be overlooked. While Human3.6M (Ionescu et al., 2014) with its laboratory conditions had long been the standard, recent years saw the rise of 3DPW (von Marcard et al., 2018), a dataset collected in real-world environments.[7] The diverse scenarios in 3DPW enabled models to better generalize to real-life situations, where lighting, motion, and interactions with the surroundings are significantly more complex. Thus, the evolution of 3D human pose estimation methods has been shaped by the step-by-step resolution of key challenges — from improving joint localization accuracy to ensuring realistic and temporally stable reconstructions. Today, achievements in this field already make it possible to talk about practical real-time applications of 3D pose estimation technologies across various industries.[8]

***Analysis of existing models***

One of the most popular approaches to pose analysis are models based on convolutional neural networks, such as OpenPose, which can extract coordinates of key body points in real time. The PoseNet model is also widely used, providing high speed on mobile devices with sufficient accuracy. Modern methods often combine pose information with facial expression or action data, using multimodal networks (e.g., HRNet, BlazePose). At the same time, models are distinguished by a balance between accuracy and performance, which allows you to choose solutions depending on specific tasks - from video surveillance to sports analytics and medicine (Table 1).

Table 1 - Comparison of existing models of human pose recognition

| **Model** | **Features** | **Accuracy** | **Speed** | **Application Area** |
|---|---|---|---|---|
| OpenPose | Multi-point body, hand, and face tracking | High | Medium | Sports, healthcare, video surveillance |
| PoseNet | Lightweight, works on mobile devices | Medium | High | AR applications, mobile systems |
| HRNet | Maintains high accuracy at different resolutions | Very High | Medium | Medical research, motion control |
| BlazePose | Optimized for smartphones, tracks 33 key points | Medium | Very High | Fitness apps, real-time applications |

***Implementation of human pose analysis methods***

To improve the efficiency of standard systems for determining 3D human poses, a module was developed that extends the functionality of the MediaPipe Pose library by enabling real-time action recognition. The system starts with processing the video stream and extracting key body points using MediaPipe. After capturing a frame, the image is converted to RGB format and the pose model is processed, which allows obtaining the coordinates of the human joints.[9]

The analysis of movements is based on tracking the vertical coordinates of the hips and knees in successive frames (Figure 1). The average position of the left and right hips is calculated for each time mark, and then the rate of change in position is determined:

```
left_hip = landmarks[mp_pose.PoseLandmark.LEFT_HIP]
right_hip = landmarks[mp_pose.PoseLandmark.RIGHT_HIP]
left_knee = landmarks[mp_pose.PoseLandmark.LEFT_KNEE]
right_knee = landmarks[mp_pose.PoseLandmark.RIGHT_KNEE]

hip_y = (left_hip.y + right_hip.y) / 2
knee_y = (left_knee.y + right_knee.y) / 2

if frame_number in previous_positions:
    prev_hip_y = previous_positions[frame_number]
    speed = abs(hip_y - prev_hip_y)
```

Figure 1

Based on the calculated movement speed, basic activities are classified: walking, running, sitting, or active movements such as dancing. Speed thresholds allow differentiating activity types with minimal computational effort (Figure 2).

```
    if speed > 0.05:
        predicted_label = "Running"
    elif speed > 0.02:
        predicted_label = "Unknown action"
    elif abs(hip_y - knee_y) < 0.1:
        predicted_label = "Sitting"
    else:
        predicted_label = "Dancing"
else:
    predicted_label = "Walking"
```

Figure 2

In the absence of significant changes in hip position, the distance between the hips and knees is additionally analyzed, which allows for increased accuracy in detecting static poses, such as sitting.

Recognized actions are visualized in real time on the frame, facilitating understanding of human behavior without the need for complex post-processing (Figure 3).

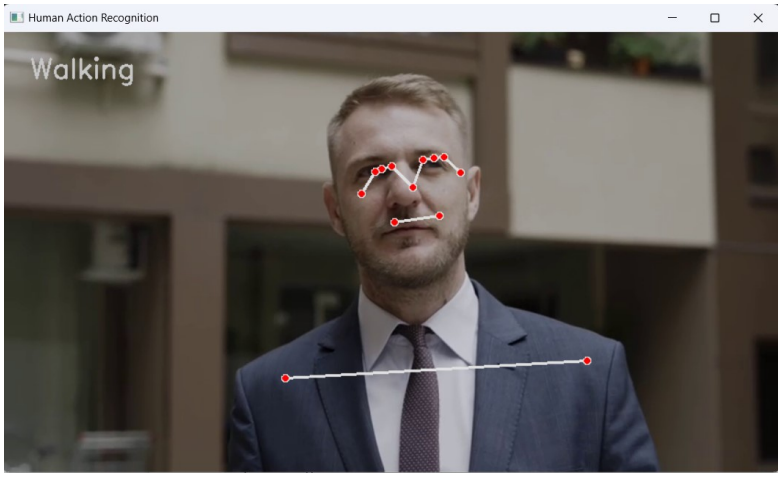


Figure 3

This solution differs from the basic MediaPipe in that it adds an initial level of semantic interpretation of movements, which makes the system not only capable of determining pose, but also recognizing behavioral patterns. Thanks to this, the module can be integrated into the tasks of monitoring activity, analyzing behavior, and creating interactive applications.

For further development of the system, it was important not only to recognize static poses, but also to take into account the factor of time spent in a potentially dangerous position. At this stage, additional logic for tracking the duration of dangerous poses was integrated into the development.

The main task of the next block of code is to classify basic human states using the MediaPipe library. Recognition of three types of poses was chosen:

1. Dangerous pose (for example, raised arms - a potential sign of a threat),
2. Discomfort pose (the person is bent, which may indicate a fall or fatigue),
3. Neutral pose (no deviations).

The key classification function is as follows (Figure 4):

```
if wrist_y < shoulder_y:
    label = "Dangerous pose (raised hands)"
    color = (0, 0, 255)
elif hip_y - shoulder_y > 0.3:
    label = "Discomfort (the person is bent over)"
    color = (0, 165, 255)
else:
    label = "Neutral posture"
    color = (0, 255, 0)
```

Figure 4

Particular attention was paid not only to the fact of the appearance of a dangerous pose, but also to its duration. For this, the start_time variable was used, which records the moment of the start of detection (Figure 5):

```
if pose_label == "Dangerous pose (raised hands)":
    if start_time is None:
        start_time = time.time()
    elif time.time() - start_time > dangerous_pose_duration:
        cv2.putText(frame, "Warning! Too long in dangerous pose", (30, 150), cv2.FONT_HERSHEY_SIMPLEX, 1, (0, 0, 255), 2)
else:
    start_time = None
```

Figure 5

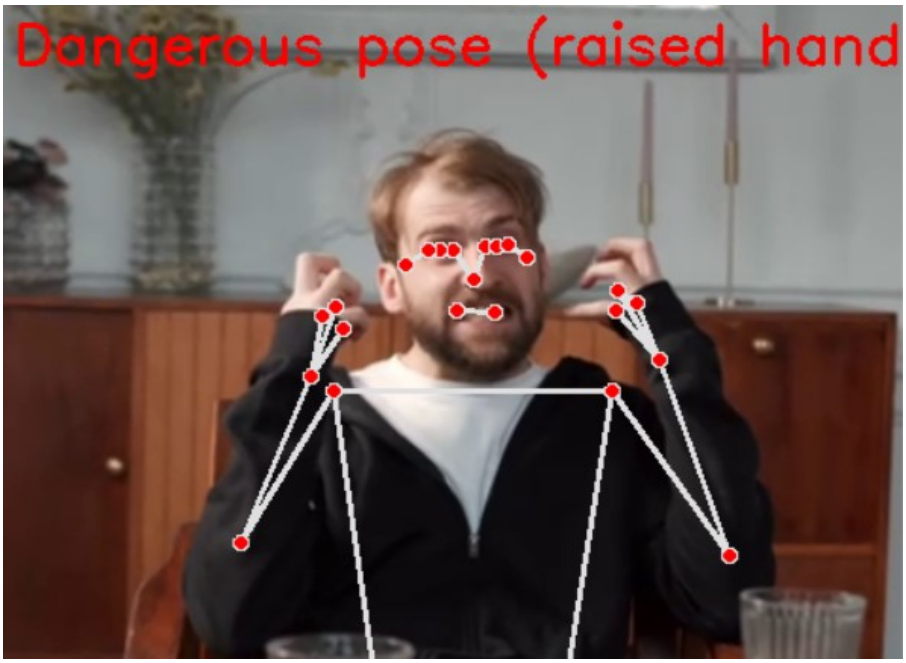


Figure 6

The image below demonstrates the operation of the developed system for automatic analysis of human poses (Figure 6). Based on key points of the skeletal model extracted from the video stream, a pose was identified in which the wrists are located above shoulder level. According to the specified heuristic rules, such a position of the hands is classified as a dangerous pose ("Dangerous pose (raised hands)"), which may indicate an unusual or potentially risky situation. The algorithm automatically places markings on the frame in real time, which allows for a quick response to such changes in the behavior of the observed object. The implementation of the algorithm demonstrated the stability of work on video at first, an adequate response to the duration of dangerous situations (with the output of warnings when a specified time threshold is exceeded), as well as the ability to scale for the tasks of detecting various human states.

The use of neural network approaches to posture analysis is combined with the ease of achieving real equipment and the prospects for expanding the functionality, for example, for reading gestures, analyzing posture or assessing stress levels. The development is especially relevant in the twentieth development of "smart" spheres, medical assistants, as well as in global healthcare, education and security. Thus, the presented method of observing and classifying human body states based on video analytics shows efficiency, adaptability to various scenarios and potential for further improvement.

**Conclusion**

In this paper, a system for analyzing human poses on video is developed that is capable of classifying neutral, uncomfortable, and dangerous body positions in real time. The implementation has shown high efficiency with low computational costs, which makes it promising for practical application. Development prospects include expanding the number of recognizable states, implementing facial expression analysis, using three-dimensional pose models, and optimizing for mobile and embedded platforms. The system has potential for use in security monitoring and adaptive human-machine interaction interfaces.

## REFERENCES

[1] Pavlakos, G., Zhou, X., Derpanis, K.G., Daniilidis, K. (2017). *Coarse-to-Fine Volumetric Prediction for Single-Image 3D Human Pose*. CVPR 2017.

[2] Sun, X., Xiao, B., Wei, F., Liang, S., Wei, Y. (2018). *Integral Human Pose Regression*. ECCV 2018.

[3] Kolotouros, N., Pavlakos, G., Black, M.J., Daniilidis, K. (2019). *Learning to Reconstruct 3D Human Pose and Shape via Model Fitting in the Loop*. ICCV 2019.

[4] Kocabas, M., Athanasiou, N., Black, M.J. (2020). *VIBE: Video Inference for Human Body Pose and Shape Estimation*. CVPR 2020.

[5] Dong, J., Jiang, Y., Huang, W., van Gool, L. (2021). *MotionBERT: Jointly Learning Motion Prediction and 3D Pose Estimation*. NeurIPS 2021.

[6] Lugaresi, C., et al. (2019). *MediaPipe: A Framework for Building Perception Pipelines*. arXiv preprint arXiv:1906.08172.

[7] von Marcard, T., Henschel, R., Black, M.J., Rosenhahn, B., Pons-Moll, G. (2018). *Recovering Accurate 3D Human Pose in The Wild Using IMUs and a Moving Camera*. ECCV 2018.

[8] Kanazawa, A., Black, M.J., Jacobs, D.W., Malik, J. (2018). "End-to-end Recovery of Human Shape and Pose." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[9] Zanfir, M., Marinoiu, E., Sminchisescu, C. (2018). "Monocular 3D Pose and Shape Estimation of Multiple People in Natural Scenes —

The Importance of Multiple Hypotheses." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

## СИСТЕМА 3D РАСПОЗНАВАНИЯ ПОЗЫ ЧЕЛОВЕКА: МЕТОДЫ, ТЕХНОЛОГИИ И ПРИЛОЖЕНИЯ В РЕАЛЬНОМ ВРЕМЕНИ

***Сулейменова Алуа Нурлановна***

*В данной статье рассматривается разработка системы автоматического анализа и классификации трехмерных поз человека на основе видеопотока в реальном времени. Предлагаемый подход включает определение основных положений тела с использованием координат ключевых точек и динамическое отслеживание изменений позы с целью выявления потенциально опасных или дискомфортных состояний. В рамках работы реализован механизм контроля времени: система фиксирует длительность нахождения в опасной позе и выдает предупреждение при превышении установленного порогового значения. Разработанный алгоритм демонстрирует высокую чувствительность к критическим изменениям положения тела и может быть интегрирован в интеллектуальные системы мониторинга поведения в различных прикладных областях, включая промышленную безопасность, медицинский надзор и управление персоналом в реальном времени. Особое внимание уделено вопросу минимизации вычислительных затрат при сохранении высокой точности распознавания, что делает предлагаемое решение пригодным для использования в условиях ограниченных ресурсов. Актуальность исследования определяется возрастающими требованиями к системам автоматического мониторинга состояния человека в динамически изменяющейся среде.*

**Ключевые слова:** 3D поза человека, распознавание поз, анализ движения, мониторинг в реальном времени, автоматическое обнаружение, контроль безопасности, интеллектуальные системы

## 3D АДАМ ПОЗАСЫН ТАНУ ЖҮЙЕСІ: ӘДІСТЕР, ТЕХНОЛОГИЯЛАР ЖӘНЕ НАҚТЫ УАҚЫТТЫ ҚОЛДАНБАЛАР

***Сулейменова Алуа***

*Бұл мақалада нақты уақыттағы бейне ағыны негізінде адамның үш өлшемді позаларын автоматты талдау және жіктеу жүйесін әзірлеу қарастырылады. Ұсынылған тәсіл негізгі нүктелердің координаталарын*

*пайдалана отырып, негізгі дене позицияларын анықтауды және ықтимал қауіпті немесе ыңғайсыз жағдайларды анықтау үшін позадағы өзгерістерді динамикалық бақылауды қамтиды. Жұмыс аясында уақытты бақылау механизмі енгізілді: жүйе қауіпті жағдайда болу ұзақтығын тіркейді және белгіленген шекті мәннен асып кеткен кезде ескерту береді. Әзірленген алгоритм дене қалпындағы сыни өзгерістерге жоғары сезімталдықты көрсетеді және әртүрлі қолданбалы салаларда, соның ішінде өнеркәсіптік қауіпсіздік, медициналық бақылау және нақты уақыттағы персоналды басқару салаларында мінез-құлықты бақылаудың интеллектуалды жүйелеріне біріктірілуі мүмкін. Жоғары тану дәлдігін сақтай отырып, есептеу шығындарын азайту мәселесіне ерекше назар аударылады, бұл ұсынылған шешімді шектеулі ресурстар жағдайында қолдануға жарамды етеді. Зерттеудің өзектілігі динамикалық өзгеретін ортада адам жағдайын автоматты бақылау жүйелеріне қойылатын талаптардың артуымен анықталады.*

**Кілт сөздері:** 3D адам позасы, позаны тану, қозғалысты талдау, нақты уақыттағы бақылау, автоматты анықтау, қауіпсіздік мониторингі, интеллектуалды жүйелер.