

DEVELOPMENT OF A REAL-TIME AUTOMATED WRESTLING MATCH ANALYSIS SYSTEM BASED ON YOLO26X-POSE AND ATHLETE RE-IDENTIFICATION

¹Tassybayev D.^a, ²Dadabayev O., ³Savitskii V.

¹ Astana IT University / SIC «Industry 4.0», Astana, Kazakhstan

² Research Institute of Physical Culture and Sport under the Sports Academy of the Republic of Uzbekistan, Tashkent, Uzbekistan

³ Directorate for Sports Development of the Ministry of Tourism and Sports of the Republic of Kazakhstan, Astana, Kazakhstan

Corresponding author: Tassybayev D.B. tassybayev.kostanay@gmail.com

Abstract. In this work a real-time wrestling match analysis system built on the YOLO26x-Pose detector is described. The system runs four modules in sequence: two-dimensional pose estimation for both wrestlers; appearance-based re-identification (ReID) using MobileNetV3-Small embeddings with Hungarian assignment and three-speed exponential moving average (EMA) updates; a heuristic finite-state machine (FSM) that assigns one of five match states: SEPARATION, CLINCH, TAKEDOWN_ATTEMPT, PAR_TERRE, DANGER from skeletal geometry; and a homography-based top-down minimap. The ReID module includes an entanglement-aware EMA freeze, anchor-drift protection, per-zone passivity timing, and per-frame CSV logging. On a standard GPU, without any wearable sensors, the system achieves a mean Identity Preservation Rate (IPR) of 90.1%, FSM accuracy of 85.0%, and 27.7 FPS throughput.

Keywords: intelligent system, wrestling, computer vision, pose estimation, re-identification, finite-state machine, real-time analytics.

Introduction. Tracking two wrestlers through grips, throws, and par terre holds is harder than standard multi-person tracking. Their bounding boxes overlap for extended periods, their appearances become indistinguishable mid-clinch, and the brief separations between contact phases give trackers little time to recover lost identity. Referees face the same perceptual load, which is why automated analysis tools could add practical value in training and officiating.

Computer vision research in sports has concentrated heavily on football, basketball, and hockey [1, 15]; combat sports have attracted comparatively little attention. Wearable sensor systems [2] offer one path to athlete monitoring but degrade under high-contact conditions where sensors shift or readings saturate. A smart mat approach [3] addressed exercise recognition in isolation but not the problem of tracking two bodies in contact across a shared surface.

BlazePose [4] and YOLO-Pose [5] can localize 17 body keypoints at real-time rates. ReID pipelines designed for pedestrian surveillance [6] transfer poorly to wrestling: when two wrestlers clinch, their bounding boxes merge and their crops look nearly identical. Lightweight CNN descriptors can still maintain accuracy under these conditions at acceptable compute cost [7].

Technical action recognition in combat sports has seen limited systematic work. Multi-view video synchronized with IMU data [8] can resolve ambiguities that single-camera systems miss. Temporal action spotting methods from team sports [9] show that learned detectors can localize brief events, though the open-field setting differs substantially from two bodies in constant contact. Top-down mat visualizations exist for team sports [10] but not, to our knowledge, for wrestling with full perspective correction. Passivity detection that is flagging when a wrestler avoids engagement in the outer zone has not been addressed computationally.

This study builds an integrated system combining YOLO26x-Pose [11], MobileNetV3-Small-based ReID [13], a directed FSM, and a projective minimap.

Materials and Methods. The system integrates several interconnected modules into a unified video processing pipeline as shown in Figure 1. On each frame, all people on the mat are detected with simultaneous body keypoint extraction. The resulting detections are matched against three pre-registered reference embeddings (wrestler 0, wrestler 1, referee) via ReID. Updated wrestler tracks are passed to the FSM module, and their positions are projected onto the minimap via homography. All data are recorded to a CSV file and displayed in the graphical interface.

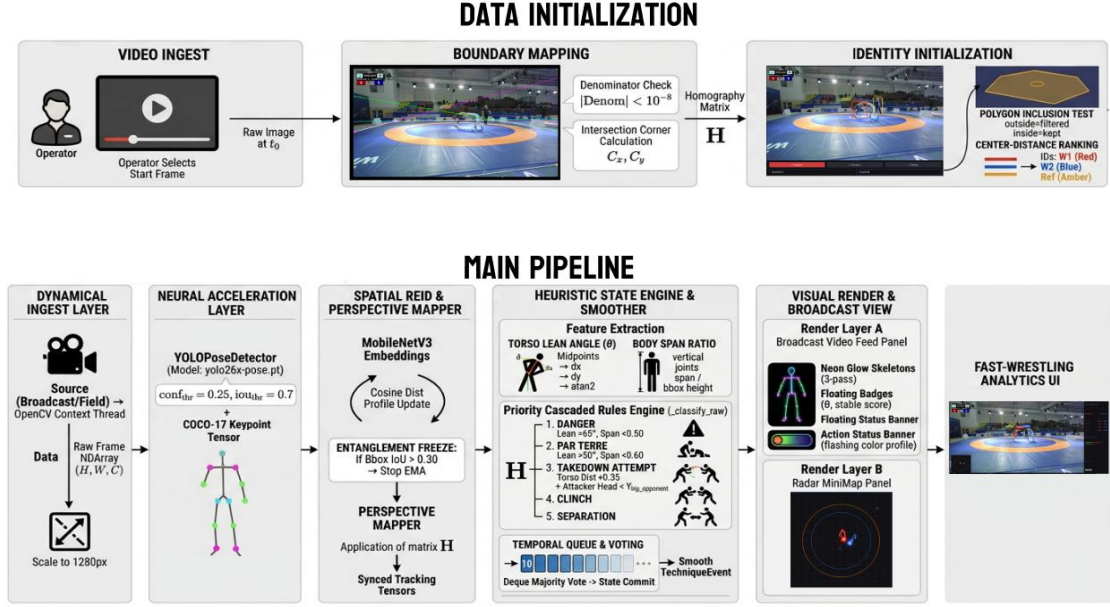


Figure 1. Architecture diagram of the real-time wrestling analysis system.

For person detection and pose estimation, the YOLO26x-Pose model (Ultralytics) [11] is used in a single forward pass. The video stream is processed at a resolution of 1280 pixels. For each detection, a bounding box and an array of 17 COCO body keypoints are formed:

$$\mathcal{K} = \{k_i = (x_i, y_i, c_i)\}_{i=0}^{16}, c_i \in [0, 1], \quad (1)$$

where c_i is the keypoint confidence score. Keypoints with $c_i < 0.3$ are considered unreliable and excluded from subsequent computations.

Before analysis begins, the operator performs a one-time mat calibration in the *MatCalibrationDialog* window by clicking twice on each of the four mat sides (top, right, bottom, left), defining 8 reference points in total. For each pair of points on one side, a line equation is determined, and the four mat corner points are computed as pairwise intersections of adjacent lines:

$$\begin{bmatrix} P_{tl} \\ P_{tr} \\ P_{br} \\ P_{bl} \end{bmatrix} = \begin{bmatrix} \text{intersect}(L_{top}, L_{left}) \\ \text{intersect}(L_{top}, L_{right}) \\ \text{intersect}(L_{bottom}, L_{right}) \\ \text{intersect}(L_{bottom}, L_{left}) \end{bmatrix}, \quad (2)$$

From the four corner points src and a target rectangle dst of size $s \times s$ pixels, the 3×3 perspective transformation matrix is computed.

Matrix H is applied frame-by-frame to transform wrestler pixel coordinates into minimap coordinates, providing a top-down view regardless of camera position. The real mat size follows the UWW standard of $12 \times 12\text{m}$.

After mat calibration, the user selects the match starting frame via the *PlaybackDialog*. The system scans ± 30 frames around the selected moment and chooses the frame with the maximum number of detections whose ankle keypoints lie inside the mat contour.

In the *WrestlerConfirmDialog*, the operator assigns three slots: wrestler 0 (W0), wrestler 1 (W1), and referee (REF). The system automatically proposes the three detections closest to the mat center; the operator may correct the assignment. After confirmation, a reference embedding for each slot is computed via a single batch forward pass through MobileNetV3-Small [13]:

$$f_{\text{slot}} = \text{embed}(I, b_{\text{slot}}), \text{slot} \in \{\text{W0}, \text{W1}, \text{REF}\}, \quad (3)$$

References are stored as L2-normalized vectors of dimension 576, and their copies are fixed as immutable anchors to protect against drift.

On each frame, all N detections are embedded in a single batch forward pass through MobileNetV3-Small - a feature extractor operating in fp16 on CUDA yielding a set of candidate descriptors $\mathcal{C} = \{c_j\}_{j=0}^{N-1}$.

The cosine distance matrix is constructed for all slot–candidate pairs:

$$D(r, j) = 1 - f_r \cdot c_j, \quad (4)$$

The optimal assignment is found via the Hungarian algorithm [14]:

$$(r^*, j^*) = \arg \min_{\sigma} \sum_r D(r, \sigma(r)), D(r, \sigma(r)) < \theta_{\text{cost}}, \quad (5)$$

where $\theta_{\text{cost}} = 0.30$ is the acceptance threshold.

After assignment, the reference descriptor is updated via a three-speed exponential moving average (EMA):

$$f_r \leftarrow (1 - \alpha) f_r + \alpha c_{j^*}, \quad (6)$$

where α is selected by the rule:

$$\alpha = \begin{cases} \alpha_{\text{full}} = 0.04, & \text{if } D(r, j) < 0.12 \\ \alpha_{\text{slow}} = 0.01, & \text{if } 0.12 \leq D(r, j) < 0.30, \\ \alpha_{\text{min}} = 0.002, & \text{if } D(r, j) \geq 0.30 \end{cases} \quad (7)$$

If the IoU of the two wrestlers' bounding boxes exceeds 0.30, their images are mutually contaminated, making reference updates unreliable. In this case, α is forced to zero for all wrestler slots:

$$\begin{aligned} \text{entangled} = \text{IoU}(\text{bbox}_{W0}, \text{bbox}_{W1}) > 0.30 &\Rightarrow \alpha_{W0} = \\ \alpha_{W1} &= 0, \end{aligned} \quad (8)$$

After each update, the cosine distance between the updated reference f_r and the fixed anchor a_r is computed:

$$\text{drift}_r = 1 - f_r \cdot a_r, \quad (9)$$

If $\text{drift}_r > 0.20$, the reference is reset to the anchor: $f_r \leftarrow a_r$. This prevents irreversible reference degradation during prolonged entanglement.

The *TechniqueRecognizer* finite-state machine assigns one of five match states per frame based on geometric features of the two wrestlers, evaluated in descending priority:

DANGER > PAR TERRE > TAKEDOWN ATTEMPT > CLINCH > SEPARATION

For each wrestler i , the following features are computed: trunk lean angle lean_i - the angle from vertical of the mid-shoulders-to-mid-hips vector in degrees ($0^\circ = \text{upright}$, $90^\circ = \text{horizontal}$); vertical span coefficient $\text{span}_i = \frac{\text{max}_y - \text{min}_y}{h_{\text{bbox}}}$ (10), where h_{bbox} is the bounding box height; normalized torso-center distance $\text{torso}_{\text{dist}} = \frac{\|t_{W0} - t_{W1}\|}{w_{\text{frame}}}$ (11).

State transition rules, in descending priority:

$$\text{State} = \begin{cases} \text{DANGER} & \text{if } \exists i: \text{lean}_i > 65^\circ \wedge \text{span}_i < 0.50 \\ \text{PAR_TERRE} & \text{if } \exists i: \text{lean}_i > 50^\circ \wedge \text{span}_i < 0.60 \\ \text{TAKEDOWN_ATT.} & \text{if } \text{torso}_{\text{dist}} < 0.35 \wedge \text{lean}_{\text{att}} > 40^\circ \wedge \text{head}_{\text{att}_y} > \text{hip}_{\text{def}_y} \\ \text{CLINCH} & \text{if } \text{torso}_{\text{dist}} < 0.35 \wedge \text{lean}_{W0} < 35^\circ \wedge \text{lean}_{W1} < 35^\circ \\ \text{SEPARATION} & \text{otherwise} \end{cases} \quad (12)$$

Majority voting in a sliding window of 10 frames suppresses isolated detector errors: the FSM state is updated only when the leading class in the window changes. A state transition generates a *TechniqueEvent* recording the subject (attacking wrestler index, or *None* for bilateral events) and the frame number.

The pixel position of each wrestler is computed as the mean of ankle coordinates (or hip joints if ankles are absent):

$$\text{pos}_i = \frac{\text{ankle}_L + \text{ankle}_R}{2}, \quad (13)$$

The projected minimap position $\text{mm}_i = H \cdot \text{pos}_i$ (14) is smoothed over 5 frames to reduce keypoint jitter. The minimap marks three zones per UWW regulations: central circle ($r = 0.5$ m), competition zone ($r = 3.5$ m), and passive zone ($r = 4.5$ m).

Per-frame statistics logged for each wrestler: cumulative travel in meters, with a 0.10–1.50 m/frame dead zone to filter keypoint noise and identity-switch artifacts; time in the passive zone (frames where $3.5 < r \leq 4.5$ m, converted to seconds); current zone label - Center ($r \leq 0.5$ m), Competition ($0.5 < r \leq 3.5$ m), or Passivity ($r > 3.5$ m); and distance from mat center. All fields are written to CSV each frame.

Results. The implementation uses Python 3.10 with Ultralytics ≥ 8.4 , PyTorch ≥ 2.0 / torchvision ≥ 0.15 , PyQt6 ≥ 6.6 , OpenCV ≥ 4.9 , Supervision ≥ 0.26 , and SciPy ≥ 1.10 , on an NVIDIA RTX-class GPU. ReID runs in fp16; YOLO uses GPU device 0.

The dataset comprised 10 wrestler pairs (20 athletes: 12 men, 8 women) at three locations: a sports hall with overhead camera and uniform artificial light (A); a multi-purpose hall with mixed lighting and a side camera (B); a university training hall with variable lighting and a slight camera tilt (C). Each pair completed 3 bouts of 3–6 minutes - 30 recordings, approximately 243,000 annotated frames at 30 fps. Two certified referees labeled FSM states and wrestler identities independently; inter-annotator agreement was $\kappa = 0.88$.

Location A gave the best results (IPR = 93.7%, FSM accuracy = 89.4%), as shown in Table 1. At B and C, changing ambient light and occasional bystanders in frame pulled both metrics down. Averaged across all three sites: IPR = 89.1%, FSM accuracy = 85.0%.

Table 1 - Overall system performance by location

Location	Shooting conditions	ReID IPR, %	FSM acc., %	FPS	Latency, ms
A	Uniform artificial light	93.7	89.4	28.4	35.2
B	Mixed lighting	88.2	84.1	27.8	36.0
C	Variable light, side camera	85.4	81.6	26.9	37.2
Mean	-	89.1	85.0	27.7	36.1

Per-class FSM results are shown in Table 2. SEPARATION achieves the highest F1 (95.7%) and DANGER the second highest (92.5%), both having clear geometric signatures. TAKEDOWN_ATT. yields the lowest F1 (78.9%) due to its brief duration (0.3–0.8 s) relative to the 10-frame voting window. Mean macro F1 is 87.1%.

Table 2 - FSM technical action classification results

FSM State	Precision, %	Recall, %	F1, %	Typical misclassifications
SEPARATION	95.1	96.4	95.7	Early CLINCH on approach
CLINCH	82.3	79.8	81.0	Confusion with TAKEDOWN_ATT.
TAKEDOWN_ATT.	80.6	77.2	78.9	Short phase duration
PAR_TERRE	88.4	86.7	87.5	Confusion with DANGER on lean
DANGER	91.2	93.8	92.5	False positives on fall
Macro avg	87.5	86.8	87.1	-

The contribution of the entanglement-aware EMA freeze is shown in Table 3. At full entanglement ($\text{IoU} \geq 0.30$, present in 13.5% of frames), the freeze raises IPR from 71.3% to 84.1% (+12.8 p.p.). Frame-weighted mean IPR increases from 85.8% to 90.1% (+4.3 p.p.).

Table 3 - ReID accuracy under varying overlap (IoU) conditions

IoU overlap condition	Frames, %	Without EMA freeze, %	With EMA freeze, %	Gain, p.p.
No overlap (IoU < 0.10)	58.4	97.8	97.8	-
Partial ($0.10 \leq \text{IoU} < 0.30$)	28.1	88.4	88.4	-
Entanglement (IoU ≥ 0.30)	13.5	71.3	84.1	+12.8
Mean (frame-weighted)	100	85.8	90.1	+4.3

Detailed pipeline latency is given in Table 4. The dominant component is YOLO26x-Pose (22.4 ms, 62% of cycle). Total latency is 36.1 ± 2.3 ms, yielding 27.7 FPS.

Table 4 - Pipeline component latency

Pipeline component	Mean, ms	Std, ms	Min–Max, ms	Cycle share, %
YOLO26x-Pose (inference, GPU)	22.4	1.8	18.1–28.6	62.0
ReID batch, MobileNetV3-Small (fp16)	4.8	0.4	3.9–6.2	13.3
Hungarian algorithm (3 slots)	< 1	-	-	2.8
FSM + homography + statistics	8.9	0.6	7.1–11.4	24.7
Total (1 frame)	36.1	2.3	29.1–47.2	27.7 FPS

Comparison with two baseline methods on the same 30-bout test set is presented in Table 5. Baseline 1 (BlazePose [4] + IoU tracker, no ReID) achieves IPR = 67.3% and FSM accuracy 70.8%: IoU tracking loses identity at every overlap event. Baseline 2 (YOLO-Pose [5] + cosine ReID without EMA or Hungarian algorithm) improves IPR to 79.8% and FSM accuracy to 79.3%, but reference descriptors drift during prolonged clinches. The full proposed system achieves IPR = 90.1%, FSM accuracy 85.0%, and 36.1 ms latency.

Table 5 - Comparison with baseline methods (30 bouts, mean over locations A–C)

Method	Pose detector	ReID module	IPR, %	FSM acc., %	FPS	Latency, ms
Baseline 1: BlazePose [4] and IoU tracker	BlazePose [4]	None (IoU tracker)	67.3	70.8	24.1	41.5
Baseline 2: YOLO-Pose [5] and cosine ReID	YOLO-Pose [5]	Cosine (no EMA, no Hungarian)	79.8	79.3	26.4	37.9
Proposed w/o EMA freeze (ablation)	YOLO26x-Pose	Full and no freeze	85.8	83.2	27.4	36.5
Proposed system (full)	YOLO26x-Pose	Full and EMA freeze	90.1	85.0	27.7	36.1

Mean wrestler positioning error on the minimap was 0.31 ± 0.14 m. Mean absolute error (MAE) of passivity time versus referee annotation was 1.8 ± 0.7 s per bout, mainly due to zone-transition latency near the passive-zone boundary.

Discussion. The system reaches 90.1% IPR and 87.1% macro F1. The ablation numbers are instructive: swapping YOLO26x-Pose for BlazePose costs 22.8 IPR points; removing the EMA freeze costs another 4.3. The two components address different failure modes and their gains do not overlap - which also explains why the full pipeline at 36.1 ms is worth the extra 5.4 ms over Baseline 1.

The FSM's advantage over a learned classifier is legibility: every state transition traces back to a specific threshold on lean angle or torso distance. That matters in officiating contexts, where a result needs to be explainable, not just accurate. The weak point is TAKEDOWN_ATT. (F1 = 78.9%) - the action typically lasts 0.3–0.8 s, which gives the 10-frame voting window little to work with. A 5-frame window would likely recover some recall, at the cost of more false positives elsewhere; that trade-off has not been tested and should be.

The anchor rollback fires when cosine drift exceeds 0.20, snapping the reference back to its initial state. The snap is abrupt: after clinches longer than roughly 8 s, identity occasionally flips for a frame or two before stabilizing. Soft re-initialization - blending current and anchor rather than replacing outright - is the natural fix. On passivity timing, 1.8 s MAE per bout is acceptable for coach feedback but not for a system that would trigger an official warning.

The main constraint on all these results is data. Thirty bouts from 20 athletes at three locations do not cover the lighting conditions, camera angles, and body types that vary across weight classes and competition levels.

Conclusion. The system tracks two wrestlers in real time, classifies their match state, and maps their positions to a calibrated top-down minimap - from a single fixed camera, no wearables, at 27.7 FPS. The entanglement-aware EMA freeze is the core technical contribution: it stops identity corruption during the clinch phases that break standard trackers. The geometric FSM is a deliberate trade-off - lower ceiling than a learned model, but every decision is auditable. Three things would meaningfully extend this work: a larger dataset spanning more competition contexts, multi-camera fusion to handle par terre occlusions, and a passivity detector tight enough for live officiating use.

References

- 1 Naik B.T., Hashmi M.F., Bokde N.D. A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. Applied Sciences. 2022. Vol. 12. Article 4429. <https://doi.org/10.3390/app12094429>
- 2 De Fazio R., Mastronardi V.M., De Vittorio M., Visconti P. Wearable Sensors and Smart Devices to Monitor Rehabilitation Parameters and Sports Performance: An Overview. Sensors. 2023. Vol. 23. Article 1856. <https://doi.org/10.3390/s23041856>
- 3 Fu B., Jarms L., Kirchbuchner F., Kuijper A. ExerTrack - Towards Smart Surfaces to Track Exercises. Technologies. 2020. Vol. 8. Article 17. <https://doi.org/10.3390/technologies8010017>
- 4 Bazarevsky V. et al. BlazePose: On-device Real-time Body Pose Tracking. arXiv:2006.10204. 2020. <https://doi.org/10.48550/arXiv.2006.10204>
- 5 Maji J., Nagori S., Mathew M., Poddar D. YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss. Proc. IEEE/CVF CVPR Workshops. 2022. <https://doi.org/10.1109/CVPRW56347.2022.00297>
- 6 Ye M., Shen J., Lin G., Xiang T., Shao L., Hoi S.C.H. Deep Learning for Person Re-Identification: A Survey and Outlook. IEEE TPAMI. 2022. Vol. 44. No. 6. P. 2872–2893. <https://doi.org/10.1109/TPAMI.2021.3054775>
- 7 Luo H. et al. A Strong Baseline and Batch Normalization Neck for Deep Person Re-Identification. IEEE TCSVT. 2022. Vol. 32. No. 10. P. 6896–6908. <https://doi.org/10.1109/TCSVT.2022.3146833>

8 Kwolek B., Michalczuk A., Krzeszowski T., Switonski A., Josinski H., Wojciechowski K. Calibrated and Synchronized Multi-View Video and IMU Dataset for Evaluation of Gait Recognition Algorithms. *Multimedia Tools and Applications*. 2019. Vol. 78. P. 32437–32465. <https://doi.org/10.1007/s11042-019-07945-y>

9 Cioppa A. et al. Context-Aware Loss Function for Action Spotting in Soccer Videos. *Proc. CVPR*. 2020. P. 7451–7460. <https://doi.org/10.1109/CVPR42600.2020.00748>

10 Zheng F., Al-Hamid D.Z., Chong P.H.J., Yang C., Li X.J. A Review of Computer Vision Technology for Football Videos. *Information*. 2025. Vol. 16. Article 355. <https://doi.org/10.3390/info16050355>

11 Ultralytics. Pose Estimation. Ultralytics YOLO Docs. <https://docs.ultralytics.com/tasks/pose/> Accessed: 12.05.2026.

12 Lin T.-Y. et al. Microsoft COCO: Common Objects in Context. *arXiv:1405.0312*. 2014. <https://doi.org/10.48550/arXiv.1405.0312>

13 Howard A. et al. Searching for MobileNetV3. *Proc. ICCV*. 2019. P. 1314–1324. <https://doi.org/10.1109/ICCV.2019.00140>

14 Kuhn H.W. The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*. 1955. Vol. 2. No. 1–2. P. 83–97. <https://doi.org/10.1002/nav.3800020109>

15 Bonidia R.P. et al. Computational intelligence in sports: A systematic literature review. *Advances in Human-Computer Interaction*. 2018. Vol. 2018. Article 3426178. <https://doi.org/10.1155/2018/3426178>

YOLO26X-POSE ЖӘНЕ БАЛУАНДАРДЫ ҚАЙТА ТАҢУ НЕГІЗІНДЕ КҮРЕС БЕЛДЕСУЛЕРІН НАҚТЫ УАҚЫТ РЕЖИМІНДЕ АВТОМАТТЫ ТАЛДАУ ЖҮЙЕСІН ӘЗІРЛЕУ

¹Тассыбаев Д. Б.^а, ²Дадабаев О., ³Савицкий В.

¹*Astana IT University / «Industry 4.0» FZO, Астана қ., Қазақстан*

²*Өзбекстан Республикасы Спорт академиясы жанындағы Дене шынықтыру және спорт ғылыми-зерттеу институты, Ташкент қ., Өзбекстан*

³*Қазақстан Республикасы Туризм және спорт министрлігінің Спортты дамыту дирекциясы, Астана қ., Қазақстан*

Хат алмасу үшін автор: Тассыбаев Д. Б. tassybayev.kostanay@gmail.com

Андатпа. Бұл жұмыста YOLO26x-Pose детекторы негізінде әзірленген, күрес белдесулерін нақты уақыт режимінде талдау жүйесі ұсынылған. Жүйе ретімен орындалатын төрт модульден тұрады: қос балуанның дене қалпын екі өлшемді анықтау; сырт келбет сипаттарына негізделген қайта тану (ReID) - бұл үшін MobileNetV3-Small эмбеддингтері, Мажар алгоритмі және деректерді жаңартуға арналған үш жылдамдықты экспоненциалды жылжымалы орташа мән (ЕМА) қолданылады; қаңқа геометриясына сүйене отырып, белдесу барысын бес күйдің біріне (АРАҚАШЫҚТЫҚ, КЛИНЧ, ПАРТЕРГЕ ТҮСІРУ ӘРЕКЕТІ, ПАРТЕР, ҚАУІПТІ ЖАҒДАЙ) жатқызатын эвристикалық шекті автомат (FSM); сондай-ақ гомография әдісімен жасалған кілемнің жоғарыдан төмен қарай көрінетін шағын картасы. Қайта тану (ReID) модуліне балуандардың өзара айқасып қалған сәттерін ескере отырып ЕМА мәндерін тоқтату, тірек нүктелерінің ауытқуынан қорғау, боз кілемдегі аймақтар бойынша пассивтілік уақытын есептеу және әр кадр сайын CSV форматында деректерді тіркеу функциялары енгізілген. Графикалық процессорда (GPU), тағылатын сенсорларсыз-ақ бұл жүйе иденттілікті сақтаудың орташа дәлдігін (IPR) 90,1%, белдесу күйін анықтау тиімділігін (FSM) 85,0% және секундына 27,7 кадр (FPS) өнімділігін көрсетті.

Түйін сөздер: интеллектуалды жүйе, күрес, компьютерлік көру, дене қалпын анықтау, қайта тану, шекті автомат, нақты уақыттағы аналитика.

РАЗРАБОТКА СИСТЕМЫ АВТОМАТИЧЕСКОГО АНАЛИЗА БОРЦОВСКИХ ПОЕДИНКОВ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ НА ОСНОВЕ YOLO26X-POSE И ПОВТОРНОЙ ИДЕНТИФИКАЦИИ СПОРТСМЕНОВ

¹Тассыбаев Д. Б.^а, ²Дадабаев О., ³Савицкий В.

¹*Astana IT University / НИЦ «Industry 4.0», г. Астана, Казахстан*

²*Научно-исследовательский институт физической культуры и спорта при Спортивной академии Республики Узбекистан, г. Ташкент, Узбекистан*

³*Дирекция по развитию спорта Министерства туризма и спорта Республики Казахстан, г. Астана, Казахстан*

Автор для корреспонденции: Тассыбаев Д. Б. tassybayev.kostanay@gmail.com

Аннотация. Предложена система видеоанализа борцовских поединков в реальном времени, построенная на детекторе YOLO26x-Pose. Конвейер состоит из четырёх компонентов. Первый оценивает позы обоих борцов в каждом кадре. Второй - модуль повторной идентификации (ReID) на эмбедингах MobileNetV3-Small - назначает треки через венгерский алгоритм и обновляет опорные дескрипторы трёхскоростным EMA; при сплетении спортсменов обновление замораживается, при дрейфе дескриптора - автоматически откатывается к якорю. Третий компонент, конечный автомат (FSM), распознаёт пять состояний поединка по геометрии скелета: разрыв дистанции, клинч, попытка перевода, партер, опасное положение. Четвёртый строит мини-карту с видом сверху через гомографию и фиксирует время пассивности по зонам ковра. Все данные записываются в CSV покадрово. На стандартном GPU, без носимых датчиков, система достигает IPR 90,1%, точности FSM 85,0% и 27,7 FPS.

Ключевые слова: интеллектуальная система, борьба, компьютерное зрение, оценка позы, повторная идентификация, конечный автомат, аналитика в реальном времени.

Авторлар туралы ақпарат // Информация об авторах // Information about the Authors

Tassybayev Damir – master’s student, Machine Learning Developer, Astana IT University / SIC «Industry 4.0», Astana, Republic of Kazakhstan.

Тассыбаев Дамир Булатович – магистрант, машиналық оқыту әзірлеушісі, Astana IT University / "Industry 4.0" ҒЗО, Астана қ., Қазақстан Республикасы.

Тассыбаев Дамир Булатович – магистрант, разработчик машинного обучения, Astana IT University / НИЦ "Industry 4.0", г. Астана, Республика Казахстан.

e-mail: tassybayev.kostanay@gmail.com

ORCID iD: [0009-0004-3180-6207](https://orcid.org/0009-0004-3180-6207)

Dadabayev Odiljon – researcher, Research Institute of Physical Culture and Sport under the Sports Academy of the Republic of Uzbekistan, Tashkent, Republic of Uzbekistan.

Дадабаев Одилжон Жалолидинович – зерттеуші, Өзбекстан Республикасы Спорт академиясының жанындағы дене шынығу мәдениеті ғылыми зерттеу институты, Ташкент қ., Өзбекстан Республикасы.

Дадабаев Одилжон Жалолидинович – научный сотрудник, Научно-исследовательский институт физической культуры и спорта при Академии спорта Республики Узбекистан, г. Ташкент, Республика Узбекистан.

e-mail: o.dadabaev70@mail.ru
ORCID iD: [0009-0008-0968-9066](https://orcid.org/0009-0008-0968-9066)

Savitskii Valentin – master’s degree, Performance Hub, Directorate for Sports Development of the Ministry of Tourism and Sports of the Republic of Kazakhstan, Astana, Republic of Kazakhstan

Савицкий Валентин Сергеевич – магистр, Performance Hub, Қазақстан Республикасы Туризм және спорт министрлігінің спортты дамыту дирекциясы, Астана қ., Қазақстан Республикасы

Савицкий Валентин Сергеевич – магистр, Performance Hub, Дирекция по развитию спорта Министерства туризма и спорта Республики Казахстан, г. Астана, Республика Казахстан

e-mail: sergeevich84@mail.ru
ORCID iD: [0000-0001-8490-9155](https://orcid.org/0000-0001-8490-9155)