Identifying Competition Characteristics of Athletes Through Video Analysis

Yuzhong Liu[1], Tianfan Zhang[2], Zhe Li[3], Mengshuang Ma[4]
School of Physical Education, Hubei Engineering University, Xiaogan, China[1]
School of Mathematics and Statistics, Hubei Engineering University, Xiaogan, China[2, 4]
School of Computer and Information Science, Hubei Engineering University, Xiaogan, China[3]

Abstract*—The vast repositories of training and competition video data serve as indispensable resources for athlete training and competitor analysis, providing a solid foundation for strategic competition analysis and tactics formulation. However, the effectiveness of these analyses hinges on the abundance and precision of data, often requiring costly professional systems for existing video analysis techniques. Meanwhile, readily accessible non-professional data frequently lacks standardization, compelling manual analysis and experiential judgments, thus limiting the widespread adoption of video analysis technologies. To address these challenges, we have devised an intelligent video analysis technology and a methodology for identifying athletes' competition characteristics. Initially, we employed target detection models, such as You Only Look Once (YOLO), renowned for their ease of deployment and low environmental dependency, to perform fundamental detection tasks. This was further complemented by the intelligent selection of standardized scenes through customizable scene rules, leading to the formation of a standardized scene dataset. On this robust foundation, we achieved classification and identification of competition participants as well as sideline recognition, ultimately compiling a comprehensive competitive dataset. Subsequently, we constructed an athlete posture estimation method utilizing OpenPose, aimed at minimizing interference caused by obstructions and enhancing the accuracy of feature extraction. In experimental validation, we gathered a diverse collection of table tennis competition video data from the internet, serving as a validation dataset. The results were impressive, with a detection success rate for standardized scenes exceeding 94% and an identification success rate for competitors surpassing 98%. The accuracy of posture reconstruction for obstructed individuals exceeded 60%, and the effectiveness of identifying athletes' main features exceeded 90%, convincingly demonstrating the effectiveness of the proposed video analysis method.

***Keywords*—Video analysis technology; scene recognition method; athlete identification; posture reconstruction; table tennis competition; feature extraction**

I. Introduction

Table tennis, characterized by its rapid pace, high technical demands, and the necessity for swift reactions, increasingly requires sophisticated methods for tactical evaluation. Traditional approaches to technical and tactical analysis have predominantly relied on literature reviews, quadruple indicators, and video observations [1], often dependent on manual observation and subjective judgment. In the systematic study process, Wu, H. discussed the statistical methods for analyzing technical and tactical applications in table tennis competitions in Statistical Methods of Table Tennis Records, utilizing basic indicators such as serve and attack, receiving, counterattacking, and looping. This marked one of the first international, systematic discussions on this topic, introducing a three-segment indicator statistical method [2]. Lames proposed describing a table tennis match using a transition probability matrix for a given match state, employing Markov chains to calculate the winning probabilities for both sides [3]. Wenninger and Lames [4] aimed to ascertain the impact of different tactical behaviors on the probability of winning in table tennis by capturing the temporality of matches through high-dimensional numerical derivation, thereby determining the correlation of tactical behaviors. Utilizing a logistic regression model, Wu et al. [5] enhanced the scientific rigor and effectiveness of table tennis technical and tactical analysis. With additional data support, the model could be further refined. Zhao and Tang [6] applied the Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) to analyze the quality of table tennis matches, avoiding the influence of opponents' strengths and tactical performances, thus enabling accurate evaluations of match quality. Song et al. [7] diagnosed table tennis matches using a hybrid algorithm based on Long Short-Term Memory-Backpropagation Neural Networks (LSTM-BPNN).

To minimize reliance on manual, subjective evaluations, computer-assisted analysis techniques have incrementally gained ground in the technical and tactical analysis of ball games. In 1999, the Swiss-developed Dartfish tactical analysis system revolutionized the field, enabling analysts to capture, scrutinize, and disseminate video footage from training sessions and competitions. This system boasts a comprehensive toolkit for video data analysis and processing, standing as the most sophisticated and widely implemented competition analysis system to this day [8]. Rahmad et al. [9] delved into the utilization of video-based intelligent systems for recognizing sports actions, introducing a video-centric action recognition framework and discussing the merits of deep learning in sports action recognition. Their research advocated a versatile method for classifying actions across diverse sports, considering varying backgrounds and characteristics, pointing the way for future studies. Manafifard et al. [10] surveyed the current state of player tracking in football videos, contrasting the strengths and weaknesses of various techniques and putting forth evaluation metrics to steer future research. Thomas et al. [11] dissected computer vision applications and related research avenues in sports, outlining commercial systems such as camera and player tracking solutions, and introducing consolidated sports datasets. Harnessing video analysis technology, critical data such as player positions can be distilled from match videos through image processing, facilitating quantitative analysis and evaluation. This approach not only elevates the precision and speed of evaluations but also aids coaches and athletes in gaining deeper insights into match scenarios and opponent traits, laying a scientific foundation for future training and competitive strategies. Wu et al. [12] developed iTTVis, a representative work in table tennis match data visualization analysis, offering a comprehensive and intuitive visualization system for technical and tactical analysts.

The dependency of match data visualization analysis on its data sources is paramount, as the level of detail and comprehensiveness of the data significantly affects the analytical outcomes. For example, the NBA's Sport VU system employs at least six portable high-definition cameras in each stadium, capturing intricate player data like spatial coordinates, timestamps, and player IDs at various frame rates. This sophisticated backend processing system tracks each player's passes, shots, and every on-court action, providing robust data for NBA match visualization analysis [13]. However, such sophisticated data levels are still a rarity in most other ball sports, especially at regional and university training and competition venues. This limits the generalized application and widespread adoption of related analytical technologies [14]. While Ma Jianhong et al. (2020) introduced a big data platform utilizing wireless sensor networks to establish a table tennis match database for real-time updates and historical data retrieval, and some scholars assessed offensive striking quality through the analysis of three-dimensional ball trajectories using computer vision, these advancements still involve significant upfront costs, which hinder the widespread use and application of video analysis and intelligent analytical technologies.

This study discusses the utilization of video analysis technology to evaluate key technical and tactical factors in table tennis matches, with a focus on the application of object detection and tracking technologies in the assessment of table tennis techniques and tactics. By processing and analyzing match video data, the positions of key players can be extracted and combined with match rules and tactical requirements for in-depth quantitative analysis and evaluation. This is anticipated to guide coaches and athletes in training and competition, enhancing their technical and tactical levels as well as athletic performance. The main tasks include:

1) The execution of most basic detection tasks can be accomplished using common target detection models, characterized by ease of deployment, low environmental dependency, and low cost.

2) Intelligent sample preprocessing technology allows for the automatic selection of standard scenes, enabling users to define custom selection rules; automatic identification of athletes and classification of matches is facilitated.

3) An athlete posture estimation method based on OpenPose [15] accurately reconstructs and estimates athlete postures, reducing interference from factors such as personnel and venue equipment obstructions, thereby enhancing the accuracy of feature extraction.

II. Table Tennis Technique Analysis and Target Characteristics Through Video Analysis

Based on the "three-segment indicator statistical method" highlighted in existing research as pointing to key statistical elements, the main data indicators to be focused on in this work have been determined. The key lies in "extracting important indicators from easily obtainable, less detailed, and accurate data". Match live broadcasts, online replays, and general camera systems can conveniently provide competition image data; however, such data hardly support the application of refined analysis like the minute movements of the table tennis player’s hand during serve and receive, or the trajectory of the table tennis ball. Nevertheless, the more significant motion characteristics of athletes, such as relative positions, posture changes, and the extent of these changes, are observable.

A. Static Basic Data

Combining general video recognition applications, a basic structured expression of the table tennis competition scene can be established, and the task targets for subsequent feature extraction can be determined, as shown in Fig. 1.

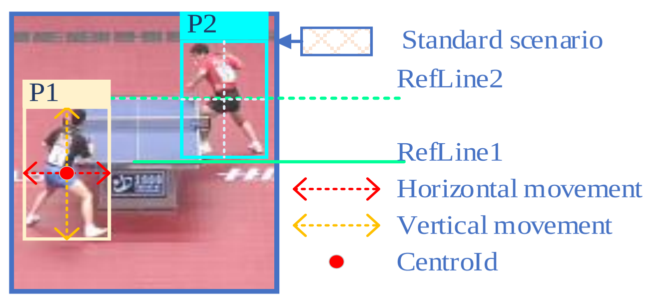


Fig. 1. Schematic diagram of structured data for athlete movement characteristics.

The scene includes three key objects, namely, the table tennis table, and the players P_1 and P_2 at both ends of the table. The position of the table tennis table remains unchanged. To facilitate subsequent quantitative analysis, two reference lines, RefLine1 and RefLine2, are constructed based on the upper and lower edges of the table tennis table. Each athlete is represented by a Bounding Box (BBox), each BBox is defined by a CentroId (C(x,y)), width, and height, i.e., BBox (C, width, height) = BBox (x, y, width, height). In addition, each image frame is accompanied by a relative timestamp t.

Furthermore, thanks to 2D posture estimation and recognition methods, more complex and detailed athlete posture features can also be obtained [16], as illustrated in Fig. 2.

B. Temporal Features

From these easily obtained "coarse" features, a series of video-based analyses can be conducted. For example, as depicted in Fig. 3, the changes in the positions of athletes from time t_1 to t_2 are showcased, which are invaluable for analyzing match scenarios and formulating tactics. Temporal information drives the change of all data related to BBox, which can be defined as the characteristic changes of athletes during the competition, including changes in movement direction, speed, and magnitude.

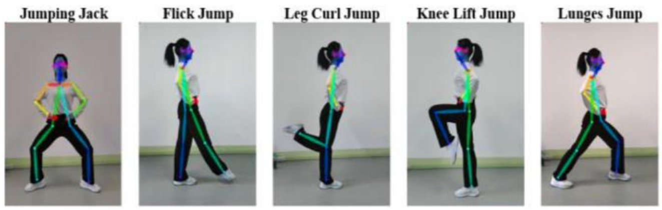


Fig. 2. Athlete posture features identified using 2D estimation methods [16].

Fig. 3. Changes in the positions of athletes from time t_1 to t_2.

C. Data Acquisition and Preprocessing

Owing to technical and tactical training data often being core assets of teams and players, obtaining publicly standardized sample data poses a challenge. Consequently, this work has opted for videos of public competitions as the data source. Although television broadcasts and online samples are readily accessible, they present a series of issues when compared to internal private data, as illustrated in Fig. 4.

- Videos comprise multiple perspectives, placing the video samples under non-uniform conditions, such as close-ups and replays.
- Changes in the original positions of the motion due to changing courts.
- Data proportion changes caused by image distortions, among others.

To address these issues, a series of preprocessing steps is necessary to achieve data standardization, as shown in Fig. 5. Initially, appropriate sample frames are selected based on the angle of view. Generally, a 45° overhead view is preferable, filtering out all transition videos. Subsequently, athletes on either side of the match (far and near ends) are divided according to the competition setup. Lastly, due to the perspective transformation caused by the filming angle, athletes' proportions and positions undergo changes. Thus, image correction is performed using the table tennis table as a known reference.

Considering the complexity of the samples and the low efficiency of manual filtering, an auxiliary selection application was developed based on YOLO: a) Using the table tennis table as a standard reference object to build a training set, train and obtain the basic features of the "standard viewpoint"; b) Define standard viewpoint rules based on basic features, i.e., "a table tennis table matching the features is detected in the center of the court, with one player at each end of the table tennis table", thereby achieving standard viewpoint filtering; c) On this basis, construct an unsupervised two-center classifier for players, clustering player samples with the same features together to achieve player identification; d) Divide the court side according to the players' positions relative to the table tennis table (upper and lower relative positions).

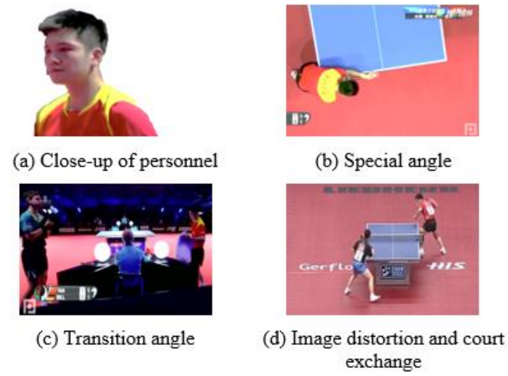


Fig. 4. Several issues in the original samples.

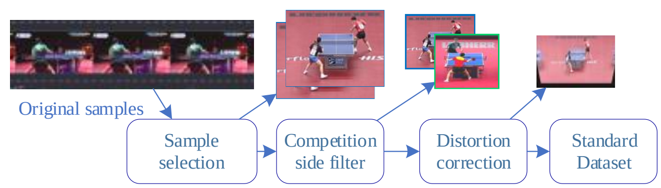


Fig. 5. Schematic diagram of the sample set preprocessing process.

III. Method for Extracting Athlete Motion Features Based on Video Analysis Technology

To effectively support the preprocessing of original samples to form a unified sample set and to efficiently extract athlete features on this basis, a video analysis and feature extraction framework based on YoLo was constructed, as shown in Fig. 6.

The entire process is divided into two major parts: 1) data standardization; and 2) feature extraction. Among them, the first part consists of three sub-parts: 1.1) Table Tennis Table Identification; 1.2) Player Detection; and 1.3) Scene Standardization.

A. Basic Target Detection Method

Although slightly less performant than YoLo v4 and other SOTA models, YoLo v5 excels in flexibility and processing speed. It is easy to deploy and compatible with a variety of platforms, including smart devices [17], which led to the selection of YoLo v5 as the basic detection model. YoLo primarily serves two functions:

- Based on the pre-trained base model, athlete detection is achieved with an accuracy exceeding 98% on the sample set described in Section II(C), fulfilling the needs for analysis.
- Another role is to assist in identifying the table tennis table and aid in filtering the original videos to obtain a standardized dataset.

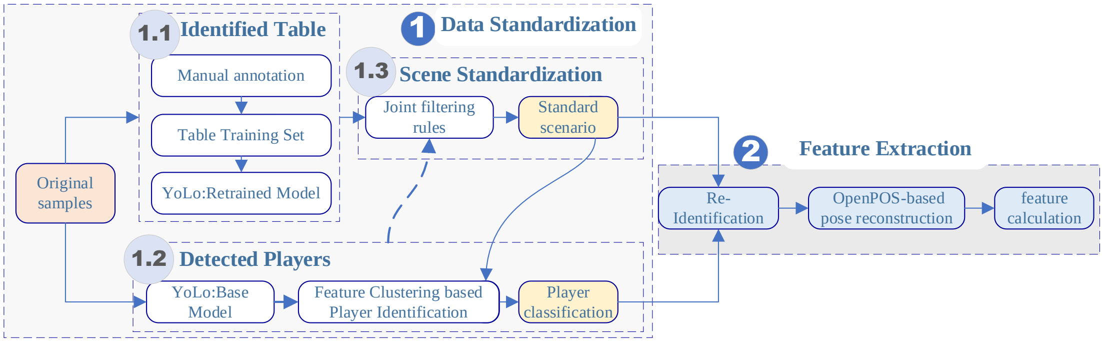
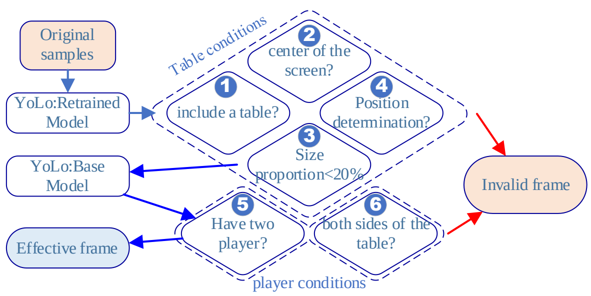


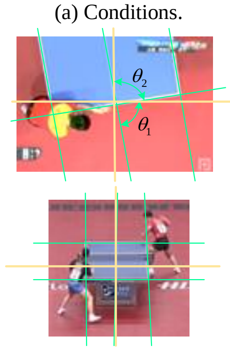
Fig. 6. Video analysis and feature extraction framework based on YoLo.

B. Data Standardization

1) Standard frame filtering: Generally, an overhead view of approximately 45° across the entire court ensures a basic viewing effect. Thus, this angle has been designated as the "standard viewpoint", characterized by several typical features, namely, an overhead view at approximately 45°, the table tennis table positioned in the middle of the field of view, with the table's outline resembling a rectangle or trapezoid, and players located at the (upper and lower) ends around the table tennis table.



(a) Conditions.



(b) Position determination.

Fig. 7. Standard frame filtering flowchart.

As shown in Fig. 7(a), assuming the frame to be determined is marked as $f(C(x,y),w,h)$, it must undergo filtering through two sets of conditions, i.e., table and player conditions. a) If a table is detected within f, its region of interest (ROI) is marked as $B(C(x,y),w,h)$; b) Whether it is located in the center of the image, i.e., $\Delta d=(f_x-B_x,f_y-B_y)\leq\sigma$ (σ is the eccentricity threshold, generally <5%); c) The size of the table's ROI relative to the entire image frame $((B_w*B_h)/(f_w*f_h))\leq\delta$ (δ is the proportion threshold, generally <20%); d) Although conditions a-c can determine a table of appropriate size, they cannot distinguish whether the table is located in the correct direction. To solve this issue, a method used in autonomous driving for lane detection has been adopted [18], as illustrated in Fig. 7(b). That is, a correctly positioned table has its top and bottom lines essentially horizontal, while the left and right lines are nearly vertically parallel (parallelism is assumed based solely on the detection angle $\leq\theta\approx5°$); e) Detection of whether two athletes P_1 and P_2 are included in the scene; f) P_1 and P_2 are located on either side of the table, either $C_x^{P_1} < B_x \,\&\&\, C_x^{P_2} > B_x$ or $C_x^{P_1} > B_x \,\&\&\, C_x^{P_2} < B_x$, respectively.

2) Athlete identification Based on feature clustering: While standard scene filtering is performed, personnel in invalid frames are also filtered out, leaving behind a dataset of personnel features (ROI), as illustrated in Fig. 8.

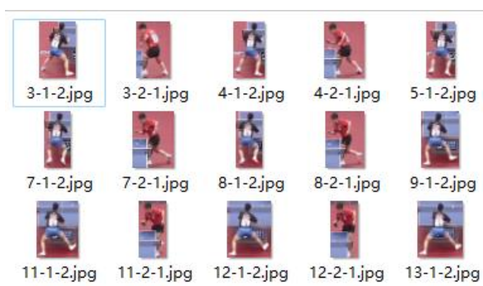


Fig. 8. Collection of athlete ROIs extracted based on the fifth rule.

Although athletes can relatively easily be divided into the correct court sides according to the sixth rule, it remains necessary to correctly differentiate the athletes, as they will switch sides to "battle from the opposite side". Considering that the preprocessing has significantly reduced the candidate information, simple feature information can be utilized to achieve feature clustering, thereby distinguishing athletes.

The method proposed by Zhang et al. [19] is employed, using sparse clustering to transform the ROIs of candidate samples into image features, which are then differentiated through clustering, as shown in Fig. 9. Subsequently, the positions of athletes P_1 and P_2 relative to the table are differentiated according to condition f), thereby indirectly achieving differentiation of match scenarios.

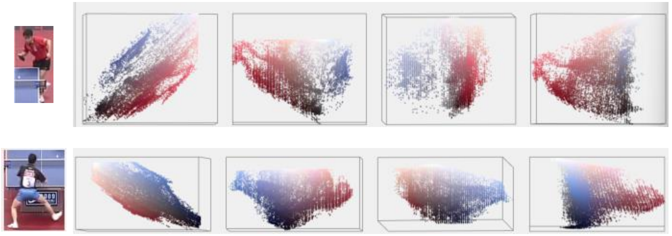


Fig. 9. Differentiating athletes using a sparse feature clustering method.

C. Scene Perspective Transformation and Personnel Re-Identification

To eliminate potential perspective distortion in the original samples, an inverse perspective transformation process is required, which ensures that lines within the image remain straight after projection. In match videos, the table tennis table should appear as a standard rectangle. Due to the camera's shooting angle and positioning, a rectangular table may appear trapezoidal. Hence, it is opted to project this trapezoidal area into a rectangle, setting four endpoints of the rectangle at appropriate positions on the horizontal axis of the right image. The generic formula for perspective transformation is:

$$[x',y',w'] = [u,v,w]\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{1}$$

where, (u,v) is obtained from the left original picture after transformation through the transformation matrix $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. The transformation matrix consists of four parts, with $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ representing linear transformations including scaling, shearing, and rotation. $\begin{bmatrix} a_{31} & a_{32} \end{bmatrix}$ is used for translation, and $\begin{bmatrix} a_{13} & a_{23} \end{bmatrix}^T$ produces a perspective transformation.

Through the inverse perspective transformation (see Fig. 10), the motion amplitude and trajectory on the horizontal position, including the distance, i.e., the athlete's trajectory on the x, y coordinate plane, can be calculated. After the perspective transformation, the candidate feature areas obtained during the preliminary preprocessing also undergo certain deformations. Therefore, a re-execution of personnel detection is necessary to obtain updated ROI samples.

D. Athlete Feature Reconstruction Method Based on OpenPose

Due to the inability of traditional detection methods to detect limbs obstructed by the table tennis table, which affects the generation of *BBOX*, directly affecting the subsequent evaluation and modeling of athlete features. As shown in Fig. 11, the same athlete of P_2 and P'_2 is affected by the table's obstruction, causing significant changes in the ROI. Therefore, before proceeding with modeling analysis, it is necessary to reconstruct the obscured limb parts of the athlete.

As illustrated in Fig. 12, a posture-guided feature decoupling Transformer network based on OpenPose [15] was designed in this study. This network utilizes known posture information to decouple human and joint components and uses posture information to guide the separation of non-obstructed and obstructed features, thus reconstructing the obscured athlete image (features).

The encoder was constructed in this study based on the Transformer classification model [20]. For a given athlete image $x \in \mathrm{R}^{H\times W times C}$, where H, W, C represent the image's height, width, and channel dimensions, respectively. Since the Transformer encoder requires only sequences as input, the input image x is first divided into N blocks of equal size using a sliding window, with each image block sized P, and the sliding window stride set to S. Then N can be expressed as:

$$N = \left\lfloor \frac{H}{S-P} \right\rfloor \times \left\lfloor \frac{W}{S-P} \right\rfloor \tag{2}$$

where, $\lfloor \bullet \rfloor$ denotes the floor function. When $S<P$, the generated image blocks overlap, but this can mitigate the loss of image spatial neighborhood information. Through a linear projection function $f(\bullet)$, the flattened blocks are mapped to D dimensions, where $f(\bullet)$ is trainable. Thus, the embeddings for blocks, termed as $E \in \mathbb{R}^{N\times D}$, are obtained, i.e., $E_i = f(x_i)$, with i=1,2,…,N. Then, a scientific series classification tag x_{class} is added to E_i, which is used as the encoder's global feature representation f_{global}. The final input sequence E_{input} can be expressed as:

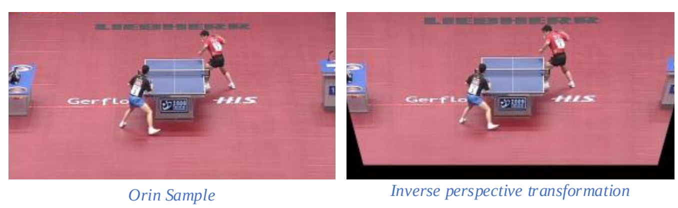


Fig. 10. Inverse perspective transformation of the original sample.

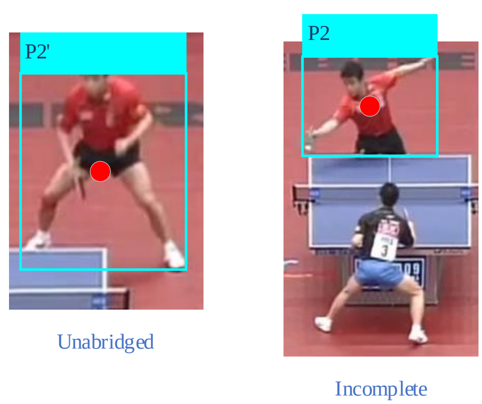


Fig. 11. The athlete obstructed by the table: affecting feature extraction.

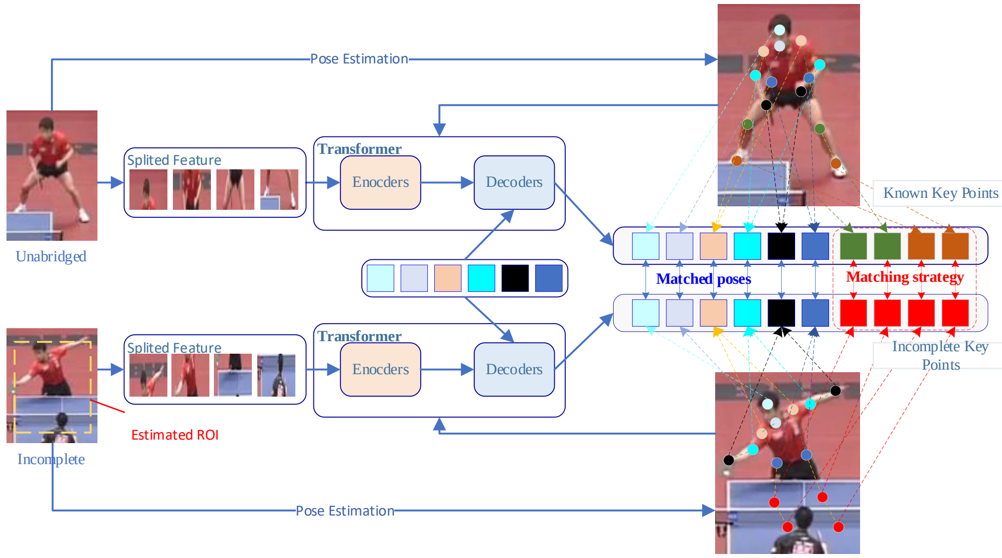


Fig. 12. Framework for the posture reconstruction method.

$$E_{input} = \{x_{class}; E_i\} + PE + cm \cdot C_{id} \tag{3}$$

where, PE is the positional embeddings, $C_{id} \in \mathbb{R}^{(N+1)\times D}$ is the camera embeddings, which remains the same for the same image. cm is a hyperparameter that balances the weight of the camera embeddings. Then the input embeddings E_{input} are processed by m Transformer layers. The final output of the encoder, $f_{en} \in \mathbb{R}^{(N+1)timesD}$, can be divided into two parts of global and part features, denoted as $f_{gb} \in \mathbb{R}^{1\times D}$ and $f_{part} in \mathbb{R}^{NtimesD}$. To learn more about the distinctive features of human body parts, part features f_{part} are sequentially divided into K groups, with each group sized $(N//K)\times D$.

$$\begin{aligned} L_{en} = L_{id}(P(f_{gb})) + \frac{1}{K}\sum_{i=1}^{K} L_{id}(P(f_{i_{gp}})) + L_{tri}(f_{gb}) \\ + \frac{1}{K}\sum_{i=1}^{K} L_{tri}(f_{i_{gp}}) \end{aligned} \tag{4}$$

where, $P(\cdot)$ represents the probability prediction function.

Images of athletes obstructed by objects suffer from performance degradation due to the reduced availability of body information and potential ambiguities in the non-body parts. Therefore, a posture estimator was constructed in this study to extract key point information from images.

Initially, the estimator extracts M landmarks from the input image x. These landmarks are then used to generate a heatmap $H = [h_1, h_2, ..., h_M]$. For larger-sized x, each heatmap is downscaled to a size of $(H/4)\times(W/4)$, effectively enhancing computational speed. Through filtering of landmarks using threshold τ, the landmarks with the highest and lowest confidence are obtained: $landmark_{\max}$ and $landmark_{\min}$, where $landmark_{\max}$ corresponds to a joint on the person.

When the number K of groups for key body parts equals the number M of landmarks and heatmaps, i.e., $K = M$, the posture feature information of various parts can be integrated. Thus, for grouped local features $f_{group-part}$, a fully connected layer can be introduced on the basis of H to obtain the same H' as $f_{group-part}$. Then by element-wise multiplication of $f_{group-part}$ with H', posture-guided features $P = [P_1, P_2, ..., P_M]$ are obtained, representing the feature information of each joint in the human posture. This allows for posture reconstruction using known complete P features and currently incomplete features P' due to obstruction.

To enhance the accuracy of reconstructing missing features, it is necessary to utilize history P to construct a set $\{f^1_{group-part}, f^2_{group-part}, \cdots, f^k_{group-part}\}$ for training. During reconstruction, the set is sorted by similarity. P_i, which has the highest match with the current known parts of P' to be constructed, can be used for reconstructing the obstructed parts. As shown on the right side of Fig. 12, "matched poses" with the highest similarity can determine the candidate reconstruction parts, referred to as the "matching strategy".

Of course, if precise posture estimation is not required and the estimation is utilized to obtain lower-accuracy ROI information, a simplified method is to use P_i to substitute P', given that the ROI information between the two is nearly identical. This significantly enhances the accuracy of the obtained ROI.

IV. Experiment and Result Analysis

A. Data Preparation

Data sets were constructed from three major table tennis events, namely, the World Table Tennis Championships (WTTC), the Table Tennis World Cup (TTWC), and the Table Tennis Match in the Olympic Games (TTMOG). These events are considered the most prestigious in the world of table tennis, attracting top athletes from various countries.

Athletes Ma Long (with a score of 4810, ranked third as of February 2024) and Fan Zhendong (with a score of 7455, ranked first as of February 2024) [21] were selected. Ma Long is the first male player to achieve a Super Grand Slam, having won singles championships at the Olympics, World Championships, World Cup, Asian Games, Asian Championships, Asian Cup, Tour Finals, and National Games. Fan Zhendong is a Chinese male table tennis player who has won multiple championships at the World Junior Championships in singles, mixed doubles, and team events, as well as runner-up in singles and doubles, along with several ITTF and World Cup titles. Ma Long and Fan Zhendong were chosen due to their high prestige and achievements within the Chinese table tennis community. Ma Long is the first male athlete to complete a Grand Slam, consistently performing at the top level. Fan Zhendong represents the excellence of the younger generation, having won numerous international competitions.

Video samples were sourced from publicly available match videos on the internet, which are accessible to all researchers. The selection included 21 videos of Ma Long participating in WTTC, TTWC, and TTMOG matches from 2007 to 2022, and 13 videos of Fan Zhendong from 2013 to 2023 for analysis.

By choosing match videos of top athletes as the data set, an in-depth analysis of their performance and technical characteristics in these competitions was facilitated. These videos offer a wealth of material for studying specific athletes' match strategies, stroke techniques, and tactical thinking. Through the editing, annotation, and analysis of these videos, followed by model training, relevant sports data could be extracted.

The compiled samples are shown in Table I and Table II, totaling 102,211 seconds (approximately 1700 minutes), and 2,638,635 frames. This demonstrates that just 34 matches of two athletes in major events accumulate such a vast amount of data, which would be challenging to process manually.

B. *Result Analysis*

The data, following preprocessing and initial analysis, are presented in Table III and Table IV. From an overall distribution perspective, the analysis method proposed in this paper effectively processes the original samples, accurately filters to obtain standard scenes, identifies athletes, achieves the reconstruction of obstructed personnel postures to a certain extent, and correctly identifies all matches. It is also observed that when the original sample resolution is low, the corresponding recognition indicators fall below the average. This suggests that the detection accuracy decreases when the sample size of the subject, such as the athlete, is too small. The impact is most significant on posture reconstruction, which requires more clear known postures to establish prior information.

TABLE I. SAMPLE SET: MA LONG

No.	Competition	Top rank	Competitors	Competitor top rank	Duration (s)	Resolution	FPs	Matches	Total frames
1	2007 WTTC	4	Joo Se-Hyuk	11	1316	960*540	29.9	6	39374
2	2008 TTWC	3	Glinka	14	648	384*288	15	5	9720
3	2009 TTWC	2	Samsonov	6	696	1440*1080	30	7	20880
4	2009 WTTC	2	Wang Hao	1	2781	1280*720	25	5	69525
5	2010 WTTC	1	Jun Mizutani	8	1646	960*540	25	3	41150
6	2011 TTWC	1	Zhang Jike	2	4364	960*540	25	7	109100
7	2011 WTTC	1	Wang Hao	1	4043	480*360	25	6	101075
8	2012 TTWC	1	Boll	5	1966	864*486	25	4	49150
9	2012 TTWC	1	Gao Ning	14	2737	1280*760	25	3	68425
10	2013 WTTC	1	Wang Hao	3	2986	480*360	25	6	74650
11	2014 TTWC	1	Zhang Jike	4	3689	480*270	25	7	92225
12	2015 TTWC	1	Fan Zhendong	2	2581	480*272	25	4	64525
13	2015 WTTC	1	Fang Bo	8	3905	480*270	25	6	97625
14	2016 TTMOG	1	Zhang Jike	4	2351	1280*720	25	4	58775
15	2017 TTWC	1	Boll	5	5360	1916*1080	25	7	134000
16	2017 WTTC	1	Fan Zhendong	2	4445	1280*716	25	7	111125
17	2019 TTWC	3	Lin Yun-Ju	7	929	1920*1080	30	7	27870
18	2019 WTTC	3	Falck	7	3627	1280*720	25	5	90675
19	2020 TTWC	3	Tomokazu Harimoto	5	4053	1280*720	24	7	97272
20	2021 TTMOG	2	Ovtcharov	7	4928	864*486	25	7	123200
21	2022 WTTC	2	Fan Zhendong	1	4175	1280*720	25	7	104375

TABLE II. SAMPLE SET: FAN ZHENGDONG

No.	Competition	Top rank	Competitors	Competitor top rank	Duration (s)	Resolution	FPs	Matches	Total frames
1	2013 WTTC	5	Zhang Jike	2	2216	480*360	25	4	55400
2	2015 TTWC	2	Ma Long	1	2581	480*272	25	4	64525
3	2015 WTTC	2	Koki Niwa	11	2386	1920*1080	29.9	5	71508
4	2016 TTWC	2	Xu Xin	3	2950	480*272	25	5	73750
5	2017 WTTC	2	Ma Long	1	4445	1280*716	25	7	111125
6	2018 TTWC	1	Boll	1	2692	864*486	25	5	67300
7	2019 TTWC	1	Tomokazu Harimoto	3	2342	1280*718	25	5	58550
8	2019 WTTC	1	Liang Jingkun	7	3515	1280*720	25	6	87875
9	2020 TTWC	1	Ma Long	3	4797	1920*1080	29.9	7	143766
10	2021TTWC	1	Tomokazu Harimoto	4	2342	1280*718	25	5	58550
11	2021 WTTC	1	Masataka	31	1712	1440*1080	30	4	51360
12	2022 TTWC	1	Ovtcharov	6	2488	864*480	30	5	74640
13	2023 WTTC	1	Wang Chuqin	1	4519	1280*720	30	6	135570

TABLE III. SAMPLE SET: MA LONG

No.	Original			After calculation			Recognition accuracy (%)					
	Duration (s)	FPs	Total frames	Duration (s)	FPs	Total frames	Scenario	Players	Incomplete player	Incomplete ROI	Frame filtering	Matches
1	1316	29.9	39374	515	25	12875	98.4	96.7	62.7	87.9	67.3	√
2	648	15	9720	291	15	4365	89.3	96.9	56.0	82.4	55.1	√
3	696	30	20880	359	25	8975	95.7	96.9	61.0	86.0	57.0	√
4	2781	25	69525	365	25	9125	96.9	99.0	60.3	88.2	86.9	√
5	1646	25	41150	312	25	7800	90.6	98.2	62.1	84.4	81.0	√
6	4364	25	109100	629	25	15725	89.9	97.4	59.9	89.7	85.6	√
7	4043	25	101075	483	25	12075	88.8	97.3	57.2	82.7	88.1	√
8	1966	25	49150	204	25	5100	90.4	98.3	65.9	85.2	89.6	√
9	2737	25	68425	311	25	7775	96.6	99.9	57.3	85.2	88.6	√
10	2986	25	74650	479	25	11975	88.7	97.9	57.4	86.4	84.0	√
11	3689	25	92225	568	25	14200	91.3	98.2	58.7	86.2	84.6	√
12	2581	25	64525	348	25	8700	98.4	97.6	55.8	88.7	86.5	√
13	3905	25	97625	446	25	11150	92.3	96.8	57.4	86.9	88.6	√
14	2351	25	58775	328	25	8200	97.8	99.7	68.0	88.9	86.0	√
15	5360	25	134000	601	25	15025	95.0	98.2	60.1	90.1	88.8	√
16	4445	25	111125	381	25	9525	98.7	99.8	56.5	83.0	91.4	√
17	929	30	27870	365	30	10950	96.6	99.5	62.0	93.1	60.7	√
18	3627	25	90675	261	25	6525	97.2	98.1	70.7	95.3	92.8	√
19	4053	24	97272	388	24	9312	99.2	98.6	65.6	89.5	90.4	√
20	4928	25	123200	429	25	10725	92.4	98.1	57.7	95.0	91.3	√
21	4175	25	104375	413	25	10325	95.0	100.0	66.8	92.5	90.1	√

TABLE IV. SAMPLE SET: FAN ZHENGDONG

No.	Original			After calculation			Recognition accuracy (%)					
	Duration (s)	FPs	Total frames	Duration (s)	FPs	Total frames	Scenario	Players	Incomplete player	Incomplete ROI	Frame filtering	Matches
1	2216	25	55400	300	60	18000	91.6	96.5	56.6	85.0	67.5	√
2	2581	25	64525	265	60	15900	89.0	95.4	57.8	88.5	75.4	√
3	2386	29.9	71508	275	60	16500	99.1	98.8	70.0	93.6	76.9	√

4	2950	25	73750	335	60	20100	91.7	98.8	60.5	94.8	72.7	√
5	4445	25	111125	505	60	30300	96.9	99.3	55.8	85.5	72.7	√
6	2692	25	67300	337	60	20220	91.0	95.2	56.2	83.2	70.0	√
7	2342	25	58550	288	25	7200	97.7	99.2	63.7	91.2	87.7	√
8	3515	25	87875	493	25	12325	94.2	99.6	66.7	92.5	86.0	√
9	4797	29.9	143766	553	25	13825	96.4	97.9	61.0	89.3	90.4	√
10	2342	25	58550	307	25	7675	95.0	99.7	65.9	95.1	86.9	√
11	1712	30	51360	255	25	6375	95.4	99.2	59.2	90.7	87.6	√
12	2488	30	74640	373	25	9325	91.5	96.8	56.9	88.8	87.5	√
13	4519	30	135570	414	25	10350	93.7	99.3	67.5	95.8	92.4	√

Table V and Table VI display the statistical results of processing two sets of sample collections. A total of 102,211 frames were processed, with a filtering ratio reaching 81.82%, indicating that approximately 80% of the samples sourced from the internet were either invalid or could only provide limited information. The recognition rate for standard scenes was 94.17%, suggesting that besides a few scenes that were not correctly identified, the majority of effective frames were accurately filtered.

The accuracy for athlete identification was higher than 98%. A small number of non-recognitions or false detections were attributed to some samples having low resolution and the presence of numerous distracting objects in the scene. Although the accuracy for reconstructing the posture of obstructed players was only around 60%, the average accuracy for coarse-grained ROI features obtained based on this posture estimation reached 89%, enhancing the accuracy of subsequent analyses.

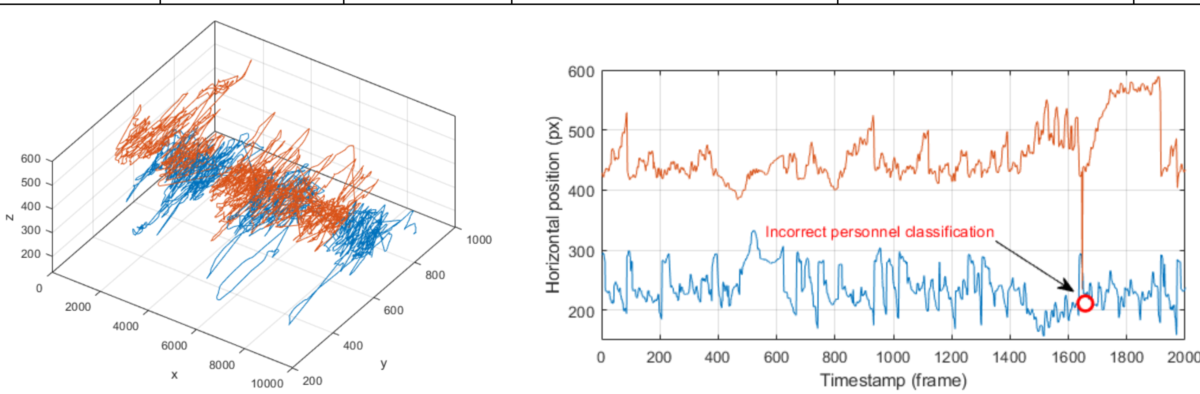
Fig. 13 showcases the match data between Ma Long and Joo Se-Hyuk at the 2007 WTTC. It is observable that the confrontation between the two athletes was fiercely competitive. As seen in Fig. 13(a), the blue player (Joo Se-Hyuk) exhibited a higher frequency and amplitude of movement than the red player (Ma Long), which could have been a contributing factor to Ma Long's defeat in this match. Additionally, Fig. 13(b) also reveals a significant misidentification point, where the classifier erroneously assigned Ma Long to the blue side (opposite side of the table tennis table). Through this data visualization, the error could be quickly identified and corrected.

TABLE V. STATISTICAL INDICATORS

Data Set	Duration (s)	Total frames	Duration (s)	Total frames	Frame filtering
Ma Long	63226	1584716	8476	210427	82.60%
Fan Zhendong	38985	1053919	4700	188095	81.05%
Total	102211	2638635	13176	398522	81.82%

TABLE VI. STATISTICAL INDICATORS: RECOGNITION ACCURACY

Data Set	Scenario	Players	Incomplete player	Incomplete ROI	Matches
Ma Long	94.25%	98.24%	60.92%	87.97%	1
Fan Zhendong	94.10%	98.13%	61.36%	90.30%	1
Average	94.17%	98.18%	61.14%	89.14%	1



(a) Three-dimensional temporal graph. (b) Horizontal direction motion temporal graph

Fig. 13. Visualization of Ma Long's match data in the 2007 WTTC (Sample 1).

V. Conclusion

This study investigates and designs an intelligent method for analyzing sports videos, capable of standardizing low-resolution, non-professional video samples such as television broadcasts and online videos at a lower cost. It accurately identifies standard scenes under appropriate conditions, competitive athletes, and key match features. Through the use of collected table tennis match videos as test subjects, the effectiveness of the proposed method for standardizing and preprocessing competition videos, as well as extracting features, was verified. The analysis method presented in this paper can be applied to video analysis of similar competitive sports, demonstrating significant potential for broader application.

As technical and tactical analysis is a highly challenging task, the work presented in this article focuses on collecting extensive data from non-standardized environments and attempting to standardize and structure them, which lays the foundation for subsequent complex technical and tactical analysis. The technical and tactical aspects of table tennis are exceedingly complex, encompassing various aspects such as serving, receiving, attacking, defending, and stalemate, each encompassing multiple techniques and tactics. Accurately identifying and distinguishing these techniques and tactics during data analysis poses a significant challenge. In table tennis competitions, athletes' technical and tactical choices are often influenced by multiple factors, such as their opponents' technical characteristics, the progress of the match, and their mental state. These factors are difficult to capture through a single data point and require comprehensive consideration of contextual information in the analysis. Identifying effective technical and tactical patterns from vast amounts of data and making predictions for future matches are crucial objectives of technical and tactical analysis. However, due to the complexity and uncertainty of table tennis competitions, this goal is often difficult to achieve. Furthermore, table tennis matches are conducted in real-time, and athletes' technical and tactical choices are dynamic. The challenge lies in acquiring and analyzing data during the match in real-time to provide coaches and athletes with immediate feedback and suggestions. These issues are all issues that need to be gradually addressed in future research.

References

[1] B. Dou, H. Li, and X. Wang, "An analysis of the techniques and tactics of japanese table tennis player hina HAYATA taking the Hangzhou Asian games as an example," Bulletin of Sport Science and Technology, vol. 31, no. 12, pp. 46-50, 2023.

[2] Y. Zou, "The refinement and improvement study of the three-stage-analysis method in table tennis," Ph.D. dissertation, Chengdu Sport University, Chengdu, China, 2015.

[3] H. Zhang, "Research on the mathematical simulation diagnosis system (SIMSS) for ball games: A case study of table tennis," in 4th National Conference on Youth Sports Science, Shanghai, China, 2005.

[4] S. Wenninger and M. Lames, "Performance analysis in table tennis - stochastic simulation by numerical derivation," International Journal of Computer Science in Sport, vol. 15, pp. 22-36, 2016.

[5] F. Wu, T. L. Liu, and X. Y. Zhang, "The application of logistic regression model in table tennis technique analysis: based on the matches between Wang Hao and Zhang Jike," J. Beijing Sport Univ., vol. 39, no. 5, pp. 96-102, 2016.

[6] X. Zhao and J. Tang, "Quality Assessment of Table Tennis Matches Based on TOPSIS: Illustrated by the Case of MA Long and FAN Zhendong," J. Capital Univ. Phys. Educ. Sports, vol. 29, no. 3, p. 5, 2017.

[7] H. Song, Y. Li, X. Zou, P. Hu, and T. Liu, "Elite male table tennis matches diagnosis using SHAP and a hybrid LSTM–BPNN algorithm," Sci. Rep., vol. 13, p. 11533, 2023.

[8] I. Daniel-Andrei, M. Mircea-Dan, M. Claudiu, S. Zenovia, M. George-Dănuț, and O. Ilie, "Quantifying the functional diagnosis in the rehabilitation of postural problems of biomechanical junior female players in table tennis," Balneo Res. J., vol. 12, no. 1, pp. 53-60, 2021.

[9] N. A. Rahmad, M. A. As'ari, N. F. Ghazali, N. Shahar, and N. A. J. Sufri, "A survey of video based action recognition in sports," Indones. J. Electr. Eng. Comput. Sci., vol. 11, pp. 987–993, 2018.

[10] M. Manafifard, H. Ebadi, and H. A. Moghaddam, "A survey on player tracking in soccer videos," Comput. Vis. Image Underst., vol. 159, pp. 19–46, 2017.

[11] G. Thomas, R. Gade, T. B. Moeslund, P. Carr, and A. Hilton, "Computer vision for sports: Current applications and research topics," Comput. Vis. Image Underst., vol. 159, pp. 3–18, 2017.

[12] Y. Wu, J. Lan, X. Shu, C. Ji, K. Zhao, J. Wang, and H. Zhang, "iTTVis: Interactive visualization of table tennis data," IEEE Trans. Vis. Comput. Graph., vol. 24, no. 1, pp. 709-718, 2018.

[13] Z. X. Yang, J. Yang, J. Bai, and L. X. Liu, "Research on data analysis of NBA based on big data technologies," China Sport Sci. Technol., vol. 52, no. 1, pp. 96-104, 2016.

[14] A. Cao, H. Zhang, and Y. Wu, "A survey on visual analysis of ball games," Sport Sci. Res., vol. 42, no. 3, pp. 26-36, 2021.

[15] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using part affinity fields," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, pp. 172-186, 2018.

[16] Y. Liu, T. Zhang, Z. Li, and L. Deng, "Deep learning-based standardized evaluation and human pose estimation: A novel approach to motion perception," Traitement Signal, vol. 40, no. 5, pp. 2313-2320, 2023.

[17] C. Dewi, and H. J. Christanto, "Automatic medical face mask recognition for COVID-19 mitigation: Utilizing YOLO V5 object detection," Rev. Intell. Artif., vol. 37, no. 3, pp. 627-638.

[18] N. J. Zakaria, M. I. Shapiai, R. B. Ghani, M. N. Yassin, M. Z. Ibrahim, and N. Wahid, "Lane detection in autonomous vehicles: A systematic review," IEEE Access, vol. 11, pp. 3729-3765, 2023.

[19] T. Zhang, Z. Li, Q. Yuan, and Y. Wang, "A spatial distance-based spatial clustering algorithm for sparse image data," Alex. Eng. J., vol. 61, no. 12, pp. 12609-12622, 2022.

[20] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al. "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.

[21] International Table Tennis Federation, "ITTF TABLE TENNIS WORLD RANKING (2024 Week #8)," 2024. Available: https://www.ittf.com/wp-content/uploads/2024/02/2024_8_SEN_MS.html