A Hierarchical Approach to Landmarks Detection in Taekwondo Poomsae Videos

Munther Abualkibash, James Gedney, Yongbom Kim, Jeongkyu Lee
1Department of Computer Science and Engineering, 2Department of Martial Arts
University of Bridgeport, CT USA
{mabualki, jgedney, ybkim, jelee}@bridgeport.edu

Abstract—Taekwondo Poomsae performance is a series of basic movements for offense and defense techniques. Despite the high popularity and long history of Taekwondo, there has been less effort to systemize Taekwondo Poomsae competition. In this paper, we propose a hierarchical approach to landmarks detection in Taekwondo Poomsae videos, which is a significant change in a series of movements. First, we propose a kinematic model for basic Poomsae movements based on the anatomic analysis of player’s body parts. Second, we measure a player’s movement from a Poomsae video using changed pixels. Third, we segment a Poomsae video into a number of movements, each of which contains the same semantic, i.e., basic Poomsae movement. Since the initial segments are usually over-segmented, we classify the segmented movements into higher level that represent significant movements of Poomsae performance. Finally, we identify landmarks from the created movement hierarchy. The experimental results show that the 70% of landmarks are detected correctly.

Keywords—Landmark detection;Taekwondo video;Poomsae

I. INTRODUCTION

Over 177 countries, more than five million people worldwide practice Taekwondo as their martial art style. Specifically, the Poomsae (called as form, kata, or hyung) is a series of basic movements in Taekwondo for offense and defense techniques. Despite the high popularity and long history of Taekwondo, there has been less effort to systemize Taekwondo Poomsae competition, which may cause several issues, such as fair judging and accurate scoring, which are caused by the subjective decision in Poomsae competition.

In order to address the judging issues, various techniques are adapted into many sports games. The most popular technique is instant replay [1,2] using video technology to correct referee’s mistake in NFL (National Football League), MLB (Major League Baseball), and tennis games. Another approach is to utilize electric devices. World Taekwondo Federation has been developing a new electric guard that can automatically detect punch and kick in two players’ sparring match [3].

Taekwondo Poomsae cannot be free from the issue of referee’s judgments. Therefore, it is highly desired to model and systemize a Poomsae performance using a captured video. In this paper, we propose a hierarchical approach to detection of landmarks in Taekwondo Poomsae videos. A landmark in Taekwondo Poomsae videos can be defined as follows:

Definition 1. A landmark of Taekwondo Poomsae video is continuous frames that indicate a boundary of two different movements with significant change.

The examples of landmarks in Taekwondo Poomsae videos are “ready stance”, “turning (changing direction)”, and “combination of several movements (kick and punch)”. Since a landmark is a significant change along a series of movements in a Poomsae performance, it is highly desired to detect and identify such landmarks from a video, which can be used for further analysis of Poomsae performance.

In order to detect landmarks from Taekwondo Poomsae videos, we employ a hierarchical method, i.e., bottom-up approach. The proposed approach consists of two main components, i.e., characterization and analysis, and four sub-components as follows: (1) Movement modeling: Based on the anatomic analysis of player’s body parts, a kinematic model is proposed for basic movements of Poomsae performance; (2) Movement measurement: Using a frame-based pixel difference technique, we measure a player’s movement from a Poomsae video that is captured by a camera; (3) Movement Hierarchy: Using the measurement, we segment a Poomsae video into a number of movements, each of which contains the same semantic. Since the initial segments are usually over-segmented, we classify more significant movements into higher level; and (4) Landmarks Detection: We identify landmarks from the upper level of created movement hierarchy.

Our contributions in this paper are as follows: (1) We propose a new kinematic model of basic movements in Taekwondo Poomsae based on the anatomic analysis of player’s body; (2) We characterize a player’s movement of Taekwondo Poomsae video based on the kinematic model; and, (3) We propose a hierarchical approach to detect landmarks from a Poomsae video, which can be used for basic unit of Poomsae video analysis.
II. BACKGROUND AND RELATED WORK

In this section, we present background of Taekwondo Poomsae, and review related work with motion analysis of sports videos, which inspires our proposed approach.

A. Background of Taekwondo Poomsae

In Taekwondo, there are two types of major competitions, i.e., Kyorugi (sparring with two players), and Poomsae (performance of forms by single, pairs or groups). Specifically, the Poomsae is a series of basic movements in Taekwondo for offensive and defensive techniques, which can be applied to Kyorugi.

Figure 1 shows the Poomsae line and movements of Taeguek I Jang that is the most basic Poomsae in Taekwondo consisting of walking and basic movements such as Makki (block), Jureugi (punching), and Chagi (kick). In Figure 1, Taeguek I Jang has 16 movements which are numbered from 1 to 16, while 0 is Joon-bi, i.e., ready poom. It has three combined movements, i.e., 5 (5-1), 12 (12-1), and 14 (14-1). From the observation of various Poomsae performances, Taekwondo Poomsae has some important characteristics as follows: (1) Pause between movements; (2) Symmetric pattern of Poomsae line (blue bar); (3) Combination of basic movements, i.e., block, punch, and kick; and (4) Balance of each movement. Therefore, the observed characteristics will be considered when we model and measure Poomsae movement in this paper.

B. Motion Analysis in Sports Videos

There has been a lot of research on motion analysis in sports videos [6-12], which focus on identifying and classifying either human or object motion. These researches can be classified into two areas motion segmentation, and motion classification. Shot boundary detection (SBD) is a basic in any video processing. Motion segmentation is a fundamental task of motion analysis. Therefore, many approaches and their results have been presented in the literatures [6-9]. The segmented motions can be classified further for indexing motions [10], and searching [9].

None of aforementioned approaches deals with a Taekwondo Poomsae video where a single player performs a series of designed movements for accuracy, balance and power.

III. CHARACTERIZATION OF POOMSAE MOVEMENT

In this section, we model and characterize Poomsae videos to segment them based on the modeled Taekwondo movements.

A. Modeling Taekwondo Movement

Due to the huge number of different movements in Taekwondo, it is almost impossible to model each and every movement separately. Instead, we employ anatomic analysis of Taekwondo player’s body to model the movements. Figure 2 illustrates our proposed kinematic model of basic Taekwondo movements.

A human body can be divided into upper and lower body since Taekwondo movements are either hand or foot techniques. The upper body is decomposed into head, arm, and torso while lower body is into pelvis, leg, and foot; however, only arms, legs, and feet can be considered as specialized parts of the kinematic model in Taekwondo movements. Then, the movement methods and applied body parts are linked.

The kinematic model should be enhanced and adjusted to video processing for Poomsae videos in this paper. In particular, we will consider the following challenges for the rest of paper: (1) the body parts that can be characterized by video descriptors need to be highlighted; (2) based on the highlighted body parts, the basic movements will be segmented; and (3) each landmark in a Poomsae video should be explained by building a movement hierarchy.

B. Measurement of Poomsae Movements

Taekwondo Poomsae is a combination of 20 ~ 40 basic movements. Each has its own characteristic techniques. In order to capture such characteristics, we measure a player’s movement, and build a movement graph, and then we use it for movement segmentation and landmarks detection.

Unlike existing methods in kinematic analysis in [4], our proposed measurements utilize a variety of image and video processing techniques without the help of any devices.

1) Pre-processing of measurement: Before we measure a player’s movement, we need to process each frame for more accurate measurement, i.e., (i) intensity equalization, (ii) detecting top-edge of a mat, (iii) finding a player’s feet, and (iv) depth of a player.

Intensity equalization: First, we compute intensity values, i.e., [z], of all pixels in RGB domain in a Poomsae video, since a dominant color of Poomsae video is black and white, i.e., player’s belt and uniform color, respectively. In addition, a histogram equalization technique is applied to each value to increase a global contrast.
Top-edge of a mat: A top-edge of a playing mat is the basis of position data. We regard a dominant color that is an average of the bottom most 20 rows as a mat color, and compare it with each row. Then, the top-edge is the first row having a sharp color change.

Player’s feet: To detect player’s feet, we compute the average values of all pixels underneath top-edge of a mat, and compare it with each row.

Depth of a player: Due to the limitation of optical lens, a lot of information is distorted from a real. In addition, the captured data cannot represent a depth of an object. Figure 3 illustrates a concept for a depth of a player with different positions. In Figure 3 (b) \( h \) is a height of a camera, while \( d_n \) is a distance between a camera and the edge of a mat. Therefore, a vertical angle between the edge of a mat and camera position is computed as \( \theta_n = \arctan(d_n/h) \), where \( h \) and \( d_n \) can be measured before processing. In addition, an angle between ‘a top-edge of a mat’ and ‘bottom of feet’ \( \theta_f \) can be computed using the ratio of the total number of rows to the camera’s vertical angle of view. Finally, a depth of a player’s position, i.e., a distance \( d_y = h_x \times \tan(\theta_n-\theta_f) \).

Movement measuremen: For each frame, i.e., two dimensional image \( I \in \mathbb{R}^{N \times L} \), where \( N \) and \( L \) are the height and width of the image, we consider the image intensity value equalized. To measure a player’s movement, we exploit frame-by-frame difference scheme, which is frequently used for general video processing algorithms [5]. A movement in the \( n \)-th video frame, \( M_n \), can be measured by:

\[
M_n = \sum_{x,y} c(x,y)
\]

if \( I_n(x,y) = I_{n+1}(x,y) \) then \( c(x,y) = 0 \), otherwise \( c(x,y) = 1 \)

where \( I_n(x,y) \) is the equalized image intensity value of the \( n \)-th image at \( x \) and \( y \) position, and \( n > 0 \). Therefore, \( M_n \) indicates the total number of changed pixels between two consecutive frames, \( n \) and \( n+1 \). The higher value of \( M_n \) the more movements in the frame \( n \).

Figure 4 (a) shows an example of a movement graph of Taegeuk 1 Jang Poomsae video in Figure 1 computed by \( M \) in Equation (1). The \( x \)-axis indicates frame number while the \( y \)-axis shows the measurement of Poomsae movement \( M_n \). Each red bracket matches for one movement in Taegeuk 1 Jang Poomsae, while the height indicates the size of form.

![Figure 4. Examples of movement graph for Taegeuk 1 Jang Poomsae.](image-url)
3) Corrected Movement measurement: Since the movement \( M \) is measured by frame-by-frame comparison, it may include lots of noise data. For example, other moving objects in a video will be considered as a movement, which makes it hard to detect each Poomsae movement. In addition, due to the limitation of capturing camera, the same movement may be measured differently depending on the distance between a player and a capturing camera. To address the errors, we conduct a correction step for the movement measurement. The correction step applies three techniques; (1) Minimum boundary box; (2) Belt elimination; and (3) Movement normalization.

**Minimum boundary box:** In order to reduce the noise of \( M \), we restrict a comparison area by defining a minimum boundary box of a player that includes a whole body. The minimum boundary box \( B \) can be determined by following the steps:

- **bottom:** is the last row containing the bottom of the player’s feet detected in Section III.B.1.
- **left/right:** Assuming 6 feet width of a boundary box, we identify left and right sides of the box by computing pixel positions using a player’s depth.
- **top:** In Figure 3, a vertical angle of side of a box and a camera \( \theta = \arctan(d_x/(h_y-h_s)) \), where a virtual height of a box \( h_s = 7 \) feet. Using a computed \( \theta \), we can determine a top of a box in each frame by using a ratio of the number of rows and \( d_s \).

In Equation (1), \((x, y) \in \mathbb{R}^{N_{xt}}\) is replaced by \((x, y) \in \mathbb{R}^{N_{yt}}\), where \( S \) and \( T \) are the height and width of the minimum boundary box that includes a player.

**Belt elimination:** Although there is a little bit pause between two movements, \( M \) values during the pause are not zero. Such errors are caused by unexpected belt movements. To correct that, we eliminate any movements of belt by using a color. Since we consider a black belt, it is straightforward to remove a black belt movements from the minimum boundary box \( B' \).

**Movement normalization:** Since a movement measurement \( M \) is based on pixels in an image captured by a camera lens, there is a distortion depending on a distance between a player and a camera. Figure 3(c) shows two image \( f_1 \) and \( f_2 \) captured from different distances \( d_1 \) and \( d_2 \), respectively. As shown in the figure, two images for the same objects have different size depending on a distance. Therefore, we need to normalize \( M \) to an edge of a mat to avoid the distortion. In Section III.B.1, we computed a distance between a player and a camera for each frame \( d_t \). The normalization factor for the \( n \)-th frame \( v_n \) can be computed as \( v_n = d_{n}^* / d_{n}^* \), where \( d_{n}^* \) is a distance between a player and a camera in the \( n \)-th frame, and \( d_{n}^* \) is a distance between top-edge of a mat and a camera.

Considering all three correction steps, a corrected Poomsae movement \( M' \) is defined as follows:

\[
M'_n = v_n \sum_{x,y} c(x,y)
\]

if \( B'_n(x,y) = B'_{n+1}(x,y) \) then \( c(x,y) = 0 \), otherwise \( c(x,y) = 1 \)

where \( B'_n(x,y) \) is the image intensity value of the minimum boundary box in the \( n \)-th frame at \( x \) and \( y \) position, and \( v_n \) is a normalization factor of the \( n \)-th frame. Figure 4 (b) is a corrected movement graph by applying \( M' \) in Equation (2). Peaks of a movement graph in Figure 4 (a) are increasing or decreasing depending on a player’s position. The corrected graph now has stabilized peaks and less noise data in valleys.

**Upper vs. Lower movement graphs:** We measure the upper \((M^{Up})\) and lower \((M^{Low})\) body movements separately. Since the upper and lower body of the anatomic model match with hand and foot techniques in Taekwondo movements, \( M^{Up} \) and \( M^{Low} \) will be useful to differentiate kick-based and hand-based movements. To measure \( M^{Up} \) and \( M^{Low} \), a detected black belt is used to divide body into two parts.

C. Player’s Position

Since Poomsae movement are either hand/foot based techniques or their combination, it is important to determine where a player performs a Poomsae. A player’s position in the \( n \)-th frame is identified by intersecting the ‘bottom of the feet’ and ‘middle point between left and right boundaries’. The identified position can be used to determine the movement’s direction, turning movements, and landmarks detection.

Figure 5 illustrates the minimum boundary box (blue rectangle box covering a player), top-edge of a mat (yellow horizontal line), a belt (dotted line and small rectangle), and a player’s position (white circle dot).

![Figure 5. Example of minimum boundary box, top-edge of a mat, belt, and a player’s position.](image)

IV. HIERARCHICAL LANDMARKS DETECTION

In this section, using the movement graph we segment a video into a number of movements, and construct a movement hierarchy that illustrates the hierarchical content of a Poomsae video, which can be used for landmarks detection.

A. Movement Segmentation

The pause between two consecutive movements may not be obvious for movement detection. To address this issue we employ threshold technique for pause based detection.

First, if \( M' \) is less than a certain threshold value \((T_h)\), the corresponding frame is considered as a pause. Otherwise, the frame belongs to a movement. Then, a movement can be segmented whenever the first pause frame is detected. In this...
paper, we use 10% of the maximum values of $M'$ based on our empirical result.

### B. Landmarks Detection using Movement Hierarchy

The movement segmentation approach mentioned above often over-segmented a Poomsae movement. In order to detect landmarks, we exploit a hierarchical approach to build movement hierarchy. Each node in the hierarchy is a movement node, labeled as $MN^i$, where the subscript denotes a movement, and the superscript indicates a level of the node in the hierarchy. A movement hierarchy, called scene tree [13, 14], is created as follows:

1. Create a movement node $MN^i$ for each Movement, (Section 4.1).
2. Set $i ← 1$.
3. Check a player’s position (Section 3.3) if Movement, has the same direction to Movement$_1$,..., Movement$_i$ in descending order. The comparison stops when a related movement, say Movement$_i$, having the same direction, is identified. If no related movement is found, we create a new empty node, connect it as a parent node to $MN^i$, and proceed to step 5.
4. We connect all movement nodes, $MN^i$ through $MN^j$ to a parent node of for $MN^0$.
5. If more movements, set $i ← i + 1$, and go to Step 3.
6. For each movement node at upper level, we select the boundaries as landmarks of a Poomsae video.

Figure 6 shows a part of the movement hierarchy of Taeguek 1 Jang Poomsae video. In this figure, there are 7 movements detected from $M'$. Let Movement$_1$ and Movement$_2$ have the same moving direction, i.e., to right, based on a player's position, while Movement$_3$ and Movement$_4$ move to left. Movement$_5$ and Movement$_6$ are moving forward to a camera, and Movement$_7$ is static. We have total four movement nodes in the upper level, i.e., $MN^0$, $MN^1$, $MN^2$and $MN^3$. The dotted lines in the figure may have a significant change, such as a player’s turning around, i.e., three landmarks in the example.

![Figure 6. Example of movement hierarchy and detected landmarks of Taeguek 1 Jang.](image)

For the specialized movement hierarchy, we can replace a movement graph $M'$ by specialized movement graph, such as $M'^{Up}$ or $M'^{Low}$.

### V. Experimental Results

To assess our proposed approach, we conducted the experiments using eight Taekwondo Poomsae videos. For the data set, first we captured a video for each Poomsae. Then, the captured videos are digitized with 320×240 pixels and 15 fps resolution. Second, the specialist examined the eight videos, and labeled the following information: basic movements, combined movements, and landmarks (i.e., ready, ending, and turning points). Table 1 describes a data set used in this experiment. The third column shows the number of movements counting combined movements separately, while the fourth column is counting the combination of movements as one. The last column is the total number of landmarks in each video. For the evaluation metrics, we employ ‘recall’ and ‘precision’ in Information Retrieval.

#### Table 1. Description of a Data Set

<table>
<thead>
<tr>
<th>Poomsae</th>
<th>Duration (seconds)</th>
<th># of movements</th>
<th># of landmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Non combination</td>
<td>Combination</td>
</tr>
<tr>
<td>Jang 1</td>
<td>430</td>
<td>256</td>
<td>180</td>
</tr>
</tbody>
</table>

| Jang 2   | 427                | 256           | 180           | 80           |
| Jang 3   | 425                | 256           | 180           | 80           |
| Jang 4   | 423                | 256           | 180           | 80           |
| Jang 5   | 421                | 256           | 180           | 80           |
| Jang 6   | 420                | 256           | 180           | 80           |
| Jang 7   | 419                | 256           | 180           | 80           |
| Jang 8   | 418                | 256           | 180           | 80           |

| Total    | 838                | 256           | 180           | 80           |

#### A. Performance of Movement Graph

The performance of movement graph is important for landmarks detection, since the proposed hierarchy utilizes the measured movements for its lower level. In order to evaluate the performance of movement graph, we perform movements segmentation using three measurements, i.e., $M'$, $M'^{Up}$ and $M'^{Low}$. For each measurement, we evaluate it by two ways: (1) with considering the combination of movements, i.e., counting as one movement, and (2) without considering the combination of movements, i.e., counting each movement separately. All experiments of movements segmentation are performed with $Th = 0.1 \times \max\{M\}$.

The results of movement segmentation are given in Table 2. We use ‘Recall’ ($Re$) and ‘Precision’ ($Pr$) mentioned above to verify the performance of those techniques. The higher recall indicates a higher capacity of detecting correct movements, which is more important than higher precision to build a hierarchy. In addition, the values in (●) are without considering the combination. The results given in the table show that the averages of ($Re$ and $Pr$) with considering the combination using $M'$, $M'^{Up}$ and $M'^{Low}$ are (94%, 35%), (95%,
31%) and (91%, 20%), respectively. Among three movement measurement techniques, $M'$ and $M''$ are better than the other in terms of recall, so we use either of them for the lower level construction of movement hierarchy. However, $M'''$ is the best when we count combined movements separately, since the combined movements may include kick technique that can be captured by $M'''$ easily. In terms of precision, since we use very generous threshold ($Th$), it is inevitable to have such a low precision values. Such a low precision can be addressed by merging when a movement hierarchy is constructed.

**TABLE 2. RESULTS OF MOVEMENT SEGMENTATION.**

<table>
<thead>
<tr>
<th>Poomsae (Taeguek)</th>
<th>$M'$</th>
<th>$M''$</th>
<th>$M'''$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Re</td>
<td>Pr</td>
<td>Re</td>
</tr>
<tr>
<td>1 Jang</td>
<td>1.00</td>
<td>0.91</td>
<td>0.41</td>
</tr>
<tr>
<td>2 Jang</td>
<td>1.00</td>
<td>0.92</td>
<td>0.36</td>
</tr>
<tr>
<td>3 Jang</td>
<td>0.90</td>
<td>0.69</td>
<td>0.28</td>
</tr>
<tr>
<td>4 Jang</td>
<td>0.95</td>
<td>0.75</td>
<td>0.40</td>
</tr>
<tr>
<td>5 Jang</td>
<td>0.90</td>
<td>0.82</td>
<td>0.37</td>
</tr>
<tr>
<td>6 Jang</td>
<td>0.88</td>
<td>0.87</td>
<td>0.34</td>
</tr>
<tr>
<td>7 Jang</td>
<td>0.89</td>
<td>0.74</td>
<td>0.44</td>
</tr>
<tr>
<td>8 Jang</td>
<td>0.92</td>
<td>0.74</td>
<td>0.30</td>
</tr>
</tbody>
</table>

* (*) is recall or precision by counting combined movements separately.

B. Landmark Detection

To assess the proposed movement hierarchy for Taekwondo Poomsae videos, we compare the results of the upper level in the movement hierarchy with the annotated information provided by a specialist. Since the upper level of movement hierarchy is built by merging the movements having the same direction, the boundaries, i.e., landmarks, represent sudden changes in a Poomsae video.

Table 3 shows the results of landmarks detection using movement hierarchy. As seen in table, the proposed scheme can detect 70% of landmarks on average with 45% of precision. Due to a limitation of a capturing camera, several true landmarks are missed by the low accuracy of forward/backward.

**TABLE 3. RESULTS OF LANDMARKS DETECTION USING MOVEMENT HIERARCHY.**

<table>
<thead>
<tr>
<th>Poomsae (Taeguek)</th>
<th>Total (#) Landmark</th>
<th>Detected (#) Landmark</th>
<th>Correct (#) Landmark</th>
<th>Recall (C/T)</th>
<th>Precision (CD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Jang</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>0.80</td>
<td>0.53</td>
</tr>
<tr>
<td>2 Jang</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>0.80</td>
<td>0.53</td>
</tr>
<tr>
<td>3 Jang</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>0.80</td>
<td>0.57</td>
</tr>
<tr>
<td>4 Jang</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>0.60</td>
<td>0.42</td>
</tr>
<tr>
<td>5 Jang</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>0.70</td>
<td>0.50</td>
</tr>
<tr>
<td>6 Jang</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>0.70</td>
<td>0.46</td>
</tr>
<tr>
<td>7 Jang</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>0.50</td>
<td>0.33</td>
</tr>
<tr>
<td>8 Jang</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>0.60</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>121</td>
<td>55</td>
<td>0.69</td>
<td>0.45</td>
</tr>
</tbody>
</table>

VI. CONCLUDING REMARKS

Although Taekwondo is the most popular Martial Arts worldwide, it is struggling from fair judging because of its lack of systemized analysis to avoid the subjective decision of referees. In order to address the problem, we propose a hierarchical approach to landmarks detection in Taekwondo Poomsae videos, which are significant changes in a series of movements. We first characterize Poomsae movements by measuring a player’s forms from a captured video. The Poomsae measurement is used for constructing movement hierarchy, which provides different views of a segmented video. Finally, we classify the segmented movements into higher level in the hierarchy, i.e., landmarks. Our experimental results show that the 70% of landmarks are detected correctly with 45% of precision. The detected landmarks will be used for automatic judging system for the future study.

VII. REFERENCES


