

SPORTS VIDEO ANALYTICS USING DEEP LEARNING

TANU ANEJA



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
JULY 2025

© **Indian Institute of Technology Delhi (IITD), New Delhi, 2025**

SPORTS VIDEO ANALYTICS USING DEEP LEARNING

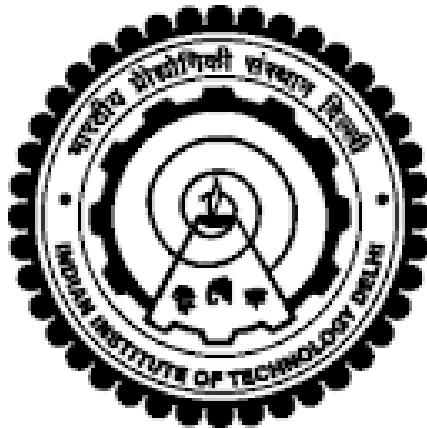
by

TANU ANEJA

Department Of Electrical Engineering

Submitted

*in fulfilment of the requirements of the degree of Doctor of Philosophy
to the*



INDIAN INSTITUTE OF TECHNOLOGY DELHI
JULY 2025

To my parents and family

Certificate

This is to certify that the thesis entitled "**Sports Video Analytics using Deep Learning**" being submitted by **Ms.Tanu Aneja** to the **Department of Electrical Engineering, Indian Institute of Technology Delhi**, for the award of the degree of **Doctor of Philosophy** is the record of the bona-fide research work carried out by her under our supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted either in part or in full to any other university or institute for the award of any degree or diploma.

(Prof. Brejesh Lall)

Department of Electrical Engineering

Indian Institute of Technology Delhi

Delhi

Hauz Khas, New Delhi-110016, India

India

(Prof. Santanu Chaudhry)

Department of Electrical Engineering

Indian Institute of Technology

Hauz Khas, New Delhi-110016,

Acknowledgments

First and foremost, I would like to express my profound gratitude to my supervisors Prof. Brejesh Lall and Prof. Santanu Chaudhary for their constant support, encouragement, and immense knowledge during the tenure of my Ph.D program. I am really blessed to have such supporting and experienced supervisors.

I would like to thank my SRC committee members: Prof. S. D. Joshi, Prof. Prem Kalra, and Prof. Sumantra Datta Roy for their insight and valuable feedback during my research.

I am indebted to Prof. Brejesh Lall, who has provided me constant motivation to excel in the field of research. The genesis of this thesis would not have been possible without his guidance, critical review and encouragement. His intelligent ideas and comprehensive understanding helped me to establish the overall direction of the research, got me past many significant challenges and always motivated me to strive for excellence.

I sincerely admire the contribution of all my peers, lab incharge and most importantly my friends for extending their unstinting support and cooperating attitude during the course of this study.

I would like to thank my parents for their love, prayer, support and encouragement at each and every step of my life. They are the one reason I wanted to do well in the field of study and research, and I thank them from core of my heart for their belief in me. The one name

that I want to mention the most is that of my husband Shubhrangam Malaviya, without his support it would not have been possible to complete this journey. In the last, I thank my kids for their sacrifices and unwavering support.

Above all, all the praises and thanks to the Almighty, for blessing me this opportunity and for his continuous grace and mercy.

Tanu Aneja

Date:

Place:

Abstract

Sports video analytics is a challenging but extremely benefiting task, as game action analysis can aid players in immensely improving their responses. Game action analytics represented in measurable manner can even help coaches to train players. They can learn from past actions of best of player responses and help players achieve their highest potential. In Cricket, shot analytics is complex as shot response depends on many factors such as batsman response, ball speed, direction and trajectory etc.

One of the areas for evaluating batsman Performance is batsman pose analysis during shots. Through pose analysis, Sports analysts can help players in analysing their responses and if needed help in recommending corrective actions. Also pose analysis can help in understanding the weaknesses of opponents and adjust their game plans on the basis of those findings. Scope of work is to analyse batsman pose and propose novel solutions for batsman action analysis using deep learning architectures.

A first attempt as a part of this effort was Optimal representation of batsman pose. Batsman pose as a precursor of shot action is an important and crucial aspect. This helps in determining shot action and enriched batsman pose-based features immensely help in pose analysis. Thereby optimal and enriched batsman pose representation is important for improving shot recognition accuracy and overall shot analytics.

For this, MSCOCO format [1], most popular standard for 2D human pose representation format was modified for cricket video analytics. Due to non-availability of benchmark cricket dataset for shot recognition, the focus was on creating feature rich annotated and labelled temporal pose keypoints dataset specifically for cricket for accurate pose estimation of

batsman. Shot sequences consist of 2D batsman pose annotated and labelled 15 frames per shot. The data was collected from internet highlights videos of cricket matches. During data selection, it was ensured that selected shot should represent mostly expected behaviour of batsman for a shot. The shots where unexpected outcomes were produced due to reasons such as unusual bat tilt, uncommon body pose etc, those shots were rejected. To apply uniformity to data collection, shots segments are taken from video angles where camera covers batsman from frontal view.

To represent batsman 2D pose, original MSCOCO [1] format (2D human pose representation format) is enriched by adding additional keypoints for bat. As movement of bat is part of comprehensive representation of batsman action, it aids in better batsman pose representation. Through exhaustive experimentation, we demonstrate that accuracy of proposed frameworks increases significantly with added keypoints.

For accurate batsman pose detection, pretrained OpenPose [27] pose detector was fine-tuned for our dataset. State of the art 2D pose detectors are widely available and predict human pose keypoints with high accuracy. We leverage one the popular OpenPose [27] pre-trained pose detector and fine-tune it for cricket dataset for accurate batsman pose estimation. We achieved average accuracy of 92.2% for batsman keypoints.

Our first work utilizes 2D batsman pose keypoints estimation along with additional features for the temporal shot sequences. LSTMs help in capturing temporal dependency between the frames and along with final layer consisting of fully connected layer followed by softmax layer, are trained for cricket shot recognition task. Best accuracy achieved by proposed architecture is 96.2% whereas for baseline architecture, best case accuracy is only 86%.

Major contribution of the thesis is through the work where we explore the possibility to represent impact of batsman joints for a cricket shot in measurable manner. It is evident from

visual analysis of cricket shots, that some joint movement may have more impact for certain shots. Hence to explain batsman action in terms of joint impact variability for varied shots, we propose an architecture that uses this directed graph representation for batsman pose along with conditional spatio-temporal graph convolution networks to capture this information in measurable manner.

The proposed architecture based on Conditional spatio-temporal graph convolution networks [78] help in training the conditional connections between joints, such that different poses can adopt appropriate connections to exploit varied non-local dependence. The visualization of final layer provides insight into conditional dependency of joints for particular shot. The empirical evaluation of final layer matrix helps in understanding underlying impact of joints on shot action and provide distinctive markers for joint impact assessment.

In summary, we propose mechanism and framework for optimal batsman pose representation. Through series of work, we propose architectures for shot recognition and studying batsman pose to understand underlying impact of joints for cricket shots. This solution can easily be adapted for variety of applications such as pose correction, shot comparison etc. It can also be part of complex solution for overall analytics of shot where other factors such as ball trajectory, speed etc. are also considered.

सारांश

स्पोर्ट्स वीडियो एनालिटिक्स एक चुनौतीपूर्ण लेकिन अत्यंत लाभकारी कार्य है, क्योंकि गेम एक्शन विश्लेषण खिलाड़ियों को उनकी प्रतिक्रियाओं में अत्यधिक सुधार करने में सहायता कर सकता है। मापने योग्य तरीके से प्रस्तुत गेम एक्शन एनालिटिक्स भी कोचों को खिलाड़ियों को प्रशिक्षित करने में मदद कर सकता है। वे खिलाड़ियों की सर्वश्रेष्ठ प्रतिक्रियाओं के पिछले कार्यों से सीख सकते हैं और खिलाड़ियों को उनकी उच्चतम क्षमता हासिल करने में मदद कर सकते हैं। क्रिकेट में, शॉट एनालिटिक्स जटिल है क्योंकि शॉट की प्रतिक्रिया कई कारकों पर निर्भर करती है जैसे कि बल्लेबाज की प्रतिक्रिया, गेंद की गति, दिशा और प्रक्षेपवक्र आदि।

बल्लेबाजों के प्रदर्शन के मूल्यांकन का एक क्षेत्र शॉट के दौरान बल्लेबाजों की मुद्रा का विश्लेषण है। पोज़ विश्लेषण के माध्यम से, खेल विश्लेषक खिलाड़ियों को उनकी प्रतिक्रियाओं का विश्लेषण करने में मदद कर सकते हैं और यदि आवश्यक हो तो सुधारात्मक कार्रवाई की सिफारिश करने में मदद कर सकते हैं। काम का दायरा बल्लेबाज मुद्रा का विश्लेषण करना और गहन शिक्षण आर्किटेक्चर का उपयोग करके बल्लेबाज कार्रवाई विश्लेषण के लिए उपन्यास समाधान प्रस्तावित करना है।

इस प्रयास के एक भाग के रूप में पहला प्रयास बल्लेबाज की मुद्रा का इष्टतम प्रतिनिधित्व था। शॉट एक्शन के अग्रदूत के रूप में बल्लेबाज की मुद्रा एक महत्वपूर्ण और निर्णायक पहलू है। यह शॉट एक्शन को निर्धारित करने में मदद करता है और समृद्ध बल्लेबाज मुद्रा-आधारित सुविधाओं से मुद्रा विश्लेषण में काफी मदद मिलती है। इस प्रकार शॉट पहचान सटीकता और समग्र शॉट विश्लेषण में सुधार के लिए इष्टतम और समृद्ध बल्लेबाज मुद्रा प्रतिनिधित्व महत्वपूर्ण है।

इसके लिए, एमएससीओसीओ प्रारूप [1], 2डी मानव मुद्रा प्रतिनिधित्व प्रारूप के लिए सबसे लोकप्रिय मानक को क्रिकेट वीडियो एनालिटिक्स के लिए संशोधित किया गया था। शॉट पहचान के लिए बेंचमार्क क्रिकेट डेटासेट की अनुपलब्धता के कारण, बल्लेबाज के सटीक पोज़ अनुमान के लिए विशेष रूप से क्रिकेट के लिए सुविधा संपन्न एनोटेटेड और लेबल किए गए टेम्पोरल पोज़ कीपॉइंट डेटासेट बनाने पर ध्यान केंद्रित किया गया था। शॉट अनुक्रमों में 2डी बल्लेबाज़ की मुद्रा अंकित होती है और प्रति शॉट 15 फ्रेम लेबल किए जाते हैं। डेटा क्रिकेट मैचों के इंटरनेट हाइलाइट वीडियो से एकत्र किया गया था। डेटा चयन के दौरान, यह सुनिश्चित किया गया कि चयनित शॉट को शॉट के लिए बल्लेबाज के अधिकतर अपेक्षित व्यवहार का प्रतिनिधित्व करना चाहिए। वे शॉट जहां असामान्य बल्ले के झुकाव, असामान्य शारीरिक मुद्रा आदि जैसे कारणों से अप्रत्याशित परिणाम उत्पन्न हुए थे, उन शॉट्स को अस्वीकार कर दिया गया था। डेटा संग्रह में एकरूपता लागू करने के लिए, शॉट्स सेगमेंट को वीडियो कोण से लिया जाता है जहां कैमरा बल्लेबाज को सामने से कवर करता है।

सटीक बल्लेबाज पोज़ डिटेक्शन के लिए, हमारे डेटासेट के लिए पूर्व-प्रशिक्षित ओपनपोज़ पोज़ डिटेक्टर को ठीक किया गया था। अत्याधुनिक 2डी पोज़ डिटेक्टर व्यापक रूप से उपलब्ध हैं और उच्च सटीकता के साथ मानव पोज़ के मुख्य बिंदुओं की भविष्यवाणी करते हैं। हम एक लोकप्रिय OpenPose पूर्व-प्रशिक्षित पोज़ डिटेक्टर का लाभ उठाते हैं और सटीक बल्लेबाज पोज़ अनुमान के लिए इसे क्रिकेट डेटासेट के लिए फ़ाइन-ट्यून करते हैं। हमने बल्लेबाजों के प्रमुख बिंदुओं के लिए 92.2% की औसत सटीकता हासिल की।

हमारा पहला काम टेम्पोरल शॉट अनुक्रमों के लिए अतिरिक्त सुविधाओं के साथ-साथ 2डी बैट्समैन पोज़ कीपॉइंट अनुमान का उपयोग करता है। LSTM फ्रेम के बीच अस्थायी निर्भरता को पकड़ने में मदद करते हैं और अंतिम परत के साथ-साथ पूरी तरह से जुड़ी हुई परत और उसके बाद सॉफ्टमैक्स परत को क्रिकेट शॉट पहचान कार्य के लिए प्रशिक्षित किया जाता है। प्रस्तावित आर्किटेक्चर द्वारा प्राप्त

सर्वोत्तम सटीकता 96.2% है जबकि बेसलाइन आर्किटेक्चर के लिए, सर्वोत्तम केस सटीकता केवल 86% है।

थीसिस का प्रमुख योगदान काम के माध्यम से है जहां हम मापने योग्य तरीके से क्रिकेट शॉट के लिए बल्लेबाज जोड़ों के प्रभाव का प्रतिनिधित्व करने की संभावना तलाशते हैं। क्रिकेट शॉट्स के दृश्य विश्लेषण से यह स्पष्ट है कि कुछ शॉट्स के लिए कुछ संयुक्त मूवमेंट का अधिक प्रभाव हो सकता है। इसलिए विभिन्न शॉट्स के लिए संयुक्त प्रभाव परिवर्तनशीलता के संदर्भ में बल्लेबाज कार्रवाई की व्याख्या करने के लिए, हम एक आर्किटेक्चर का प्रस्ताव करते हैं जो इस जानकारी को मापने योग्य तरीके से कैप्चर करने के लिए सशर्त अनुपात-अस्थायी ग्राफ कनवल्शन नेटवर्क के साथ बल्लेबाज पोज़ के लिए इस निर्देशित ग्राफ प्रतिनिधित्व का उपयोग करता है।

कंडिशनल स्पैशियो-टेम्पोरल ग्राफ कनवल्शन नेटवर्क [78] पर आधारित प्रस्तावित आर्किटेक्चर जोड़ों के बीच सशर्त कनेक्शन को प्रशिक्षित करने में मदद करता है, जैसे कि अलग-अलग पोज़ विभिन्न गैर-स्थानीय निर्भरता का फायदा उठाने के लिए उपयुक्त कनेक्शन अपना सकते हैं। अंतिम परत का दृश्य विशेष शॉट के लिए जोड़ों की सशर्त निर्भरता में अंतर्दृष्टि प्रदान करता है। अंतिम परत मैट्रिक्स का अनुभवजन्य मूल्यांकन शॉट एक्शन पर जोड़ों के अंतर्निहित प्रभाव को समझने में मदद करता है और संयुक्त प्रभाव मूल्यांकन के लिए विशिष्ट मार्कर प्रदान करता है।

संक्षेप में, हम इष्टतम बल्लेबाज मुद्रा प्रतिनिधित्व के लिए तंत्र और ढांचे का प्रस्ताव करते हैं। काम की श्रृंखला के माध्यम से, हमने क्रिकेट शॉट्स के लिए जोड़ों के अंतर्निहित प्रभाव को समझने के लिए शॉट पहचान और बल्लेबाज मुद्रा का अध्ययन करने के लिए आर्किटेक्चर प्रस्तावित किया। इस समाधान को विभिन्न प्रकार के अनुप्रयोगों जैसे पोज़ सुधार, शॉट तुलना आदि के लिए आसानी से अनुकूलित किया जा सकता है। यह शॉट के समग्र विश्लेषण के लिए जटिल समाधान का हिस्सा भी हो सकता है जहाँ अन्य कारकों जैसे गेंद प्रक्षेपवक्र, गति आदि पर भी विचार किया जाता है।

Table of Contents

Certificate	i
Acknowledgments	ii
Abstract.....	iv
सारांश	vii
List of Figures.....	xiv
List of Tables	xvii
Abbreviations	xviii
1. Introduction.....	1
<i>1.1 Sports Video Analytics</i>	<i>3</i>
1.1.1 Cricket Video Analytics.....	5
<i>1.2 Brief Summary of Literature Survey</i>	<i>6</i>
1.2.1 2D Single Person Pose Estimation.....	7
1.2.2 Human Action recognition.....	9
1.2.2.1 Pose based Human Action recognition	9
1.2.2.2 Action recognition in Sports	11
<i>1.3 Scope of Thesis.....</i>	<i>13</i>
<i>1.4 Thesis Outline</i>	<i>14</i>

2. Literature Survey: Cricket Analytics Using Computer Vision/Image-processing techniques	17
<i>2.1 Bowling Action Analytics.....</i>	<i>18</i>
<i>2.2 Batting Analytics.....</i>	<i>19</i>
<i>2.3 Event Summarization/Highlights Detection.....</i>	<i>26</i>
<i>2.4 Conclusion</i>	<i>28</i>
3. Cricket Temporal Pose Dataset	29
<i>3.1 Introduction.....</i>	<i>29</i>
<i>3.2 Human 2D Pose representation.....</i>	<i>30</i>
3.2.1 MSCOCO Format	30
3.2.1.1 MSCOCO Human Keypoint Annotation Format	31
3.2.2 Batsman Keypoint Annotation Format	34
<i>3.3 Cricket temporal Pose Dataset</i>	<i>38</i>
3.3.1 Cricket temporal Pose Dataset Labelling.....	42
3.3.2 Dataset Normalization	45
<i>3.4 Evaluation Metrics.....</i>	<i>47</i>
3.4.1 2D pose estimation Metrics	47
3.4.1.1 PCP (Percentage of Correct Key-points)	47
3.4.1.2 PCK (Percentage of Correct Key-points)	47
3.4.1 Shot classification Metrics	48
3.4.1.1 Accuracy	48
3.4.1.2 Precision.....	48
3.4.1.3 Recall	49
<i>3.5 Experiment Conclusion.....</i>	<i>50</i>

4. Batsman Pose Estimation	52
4.1 Introduction.....	52
4.2 OpenPose	53
4.3 2D Batsman Pose Estimation.....	55
4.3.1 Training Results	57
4.4 Experiment Conclusion.....	59
5. Latest Pose Estimation based Cricket Shots action Recognition	60
5.1 Introduction.....	60
5.2 Important concepts.....	61
5.2.1 Long - Short Term Memory Networks	61
5.2.2 Fully connected Networks	65
5.2.3 Softmax Function.....	66
5.3 Proposed Architecture	67
5.3.1 Batsman Pose Estimation.....	70
5.3.2 Latent Feature Transformation	70
5.3.3 Action Recognition Framework.....	72
5.4 Baseline Architecture for comparison	73
5.4.1 Spatio-temporal Graph Convolution Networks	73
5.4.1.1 ST-GCN baseline results collection strategy	76
5.5 Proposed Architecture Training and Final Results	77
5.6 Experiment Conclusion.....	79
6. Cricket Shot Analysis using Conditional Directed Spatio-Temporal Graph networks	
81	
6.1 Introduction.....	81

6.2 <i>Important concepts</i>	83
6.2.1 Human Pose Graph Representation	83
6.2.1.1 Bio Vision Hierarchy (BVH) Format	83
6.2.1.2 Batsman Pose Graph Representation	84
6.2.2 <i>Conditional Spatial Temporal Directed Graph Convolution</i>	85
6.3 <i>Proposed Framework</i>	90
6.4 <i>Training and Results</i>	93
6.4.1 Shot Recognition Results	93
6.4.2 Empirical Analysis of Joints conditional dependency pattern for Shots	94
6.4.2.1 Shots on opposite sides of field	96
6.4.2.2 Shots in near-by region in field.....	97
6.5 <i>Experiment Conclusion</i>	98
7. Conclusion and Future Directions	100
7.1 <i>Conclusion</i>	100
7.2 <i>Future Work</i>	103
Bibliography	106
Work Published Based on this Thesis	122
Technical Biography of Author	123

List of Figures

Figure. 1. Sports Video content analytics Pyramid.....	4
Figure 2. A. MSCOCO keypoints B. MSCOCO keypoints annotation JSON format C.....	33
Sample human annotation.....	33
Figure 3. VIP-HARPET dataset details: A. annotated keypoints B. Sample figures [25].....	35
Figure. 4. A. Original MSCOCO keypoints B. Our dataset keypoints	3737
Figure 5. Sample annotated shots	4040
Figure 6. Sample annotated shot sequence – 15 frames	4141
Figure. 7 A. Labelling of dataset based on ground region the ball hits and popular shots. B. class distribution in dataset	4242
Figure. 8. Sample shots from dataset	4343
Figure. 9 A. Labelling of dataset based on ground region the ball hits and popular shots. B. class distribution in dataset	4444
Figure. 10. Statics for dataset A. cricket match format distribution. B. uniform color distribution in dataset.....	4444
Figure. 11. 2D Batch Normalization [82]	4545
Figure 12: OpenPose architecture*	5454
Figure 13: Successful Predictions	5858

Figure 14: Some Failure cases with bad prediction for bat keypoints.....	5959
Figure. 15. Standard LSTM cell [89].....	6363
Figure. 16. Fully connected network [91].....	6666
Figure. 17. Proposed Action Recognition Framework.	6868
Figure. 18. Sample Cricket Shot sequence	6969
Figure. 19 A. Latent Feature Transformer, B. angles used in latent features.	7171
Figure 20: Action Recognition framework architecture	7272
Figure 21. Spatio Temporal-Graph Convolution Network (ST-GCN) architecture	7373
Figure 22: The Bio Vision Hierarchy file format [79] (a) concrete representation of the Human hierarchical anatomical structure. (b) Directed Graph format to represent the articulated human.....	8484
Figure. 23 A. 2D keypoints representation for Human Skeleton B. Proposed Batsman keypoints.....	8585
Figure 24. Basic building block for Conditional ST graph [79] network.....	8787
Figure 25: Conditional directed graph convolution (Cond DG Conv) [79]. (a) Input directed graph with predefined connections as per BVH format along with conditional connections. Part (b), (c) and (d) explain three graph updation steps of the Cond DG Conv.	8989
Figure. 26. Proposed Framework CondST-GCN based Architecture.....	9292
Figure 28: Visualization of the predicted conditional connection for A. ON Shot B. Sample OFF shot.....	9797

Figure 29: Visualization of the predicted conditional connection matrices for A. OFF shot B.

Cover Drive shot9898

List of Tables

Table 1 A: Summary of Literature Survey.....	23
Table 1 B: Summary of Literature Survey continued.....	24
Table 1 C: Summary of Literature Survey continued.....	25
Table 2: List of annotated key-points for each frame.	36
Table 3: OpenPose [27] detector Accuracy results after fine-tuning.....	57
Table 4. Class wise comparison results.....	78
Table 5. Baseline comparison Results.....	79
Table 6. Class wise comparison results; baseline comparison with our prior work- Latent pose features-based action recognition (chapter 4) for cricket shots	94

Abbreviations

2D	Two dimensional
3D	Three Dimensional
BRS	Bowler run-up sequences
BVH	Bio Vision Hierarchical Format
CNN	Convolutional Neural Networks
Cond	Conditional
CPM	Convolutional Pose Machines
DG	Directed Graph
GCN	Graph Convolution Network
GRIP	Global Roads Recovery Project
HOG	Histogram Of Gradients
H36M	Human 3.6 Million dataset
HARPET	Hockey Action Recognition Pose Estimation
HRU	Hierarchal Residual Units
KNN	k-nearest neighbour algorithm
LSTM	Long-Short Term Memory Network
MMDetection	Multimedia Detection Toolbox

MSCOCO	Microsoft Common Objects in Context
NTU-RGB-D	Nanyang Technology University's Red Green Blue and Depth Information
ODI	One Day International Cricket
ORB	Oriented fast and rotated brief method
PAF	Part Affinity Fields
RNN	Recurrent Neural Network
ST	Spatio-temporal
SVM	Support Vector Machine
T20	Twenty20 International Cricket
VGG	Visual Geometry Group