Tshepiso Mahoko

Automated Classification of Ballet Moves Using Supervised Learning

Abstract

In recent years, data traffic has surged dramatically. For instance, between 2002 and 2009, data traffic grew 56-fold (Kambatla, et al., 2013), and a decade later, researchers estimate that approximately 402.74 million terabytes of data are generated daily (Jenik, 2021). This exponential growth has given rise to the field of Big Data, which focuses on managing and extracting value from vast amounts of information. Big Data analytics, the application of advanced analytical techniques to large datasets, provides powerful solutions for deriving insights from this data.

One of the significant applications of Big Data analytics is video analysis, which leverages machine learning models to extract meaningful information from video content. This has broad implications, including video summarization, action recognition, video retrieval, visual simultaneous localization and mapping (SLAM), video annotation, and the generation of realistic videos. Within this context, the automated classification of human actions, such as ballet movements, represents a compelling challenge and opportunity for the application of supervised learning techniques (Kim, et al., 2017).

I. Introduction

The intersection of machine learning and video analysis has opened new frontiers in various domains, including entertainment, sports, and education (Kim, et al., 2017) (Brooks, et al., 2009). One of the emerging challenges in video analysis is the automatic classification of dance moves from video footage. This capability has the potential to be leveraged in a wide range of applications, from enhancing virtual reality experiences to offering actionable insights in educational tools for dancers.

The acquired dataset, totalling approximately 10.5 GB, is relatively small compared to the vast amounts of data stored in data centres. However, it is sufficient for our purposes. A supervised machine learning model, such as a neural network, is well-suited for analysing this dataset and extracting key frames necessary for the accurate classification of dance moves (Yan, et al., 2020).

This research focuses on developing a machine learning model to classify ballet moves from video footage using supervised learning techniques. The dataset used consists of annotated videos from solo ballet performances, providing a comprehensive foundation for training and evaluating our model.

II. Problem Background

Classifying ballet moves from video data is a complex task due to the intricate and varied nature of ballet choreography. The dataset we are working with includes 1,020 videos with over 24,600 temporal annotations for 11 different ballet action classes. The challenge lies in accurately recognizing these classes in new video data, considering the variations in lighting, camera angles, and the complexity of the movements.

one of the primary objectives is to achieve an accuracy level close to 93.6%. This target reflects the general precision typically demonstrated by an experienced dancer (Matsuyama, et al., 2021). Reaching this level of accuracy is essential to ensure that the classification model can reliably replicate the nuanced expertise of a professional dancer in identifying and categorizing ballet moves.

The primary challenges include:

- Variability in Video Data: Handling different lighting conditions, camera angles, and occlusions.
- Temporal and Spatial Complexity: Capturing both the spatial posture and the temporal sequence of movements
- Generalization: Ensuring the model performs well on unseen video data.
- Feature extraction: Identifying and extracting relevant features from video data is crucial for effective classification.

III. Literature Review

Annotation

One of the key challenges in big data is managing vast quantities of data, which can often be unstructured. To extract value from this data, machine learning models are commonly employed. However, a significant challenge faced by researchers when developing supervised machine learning models is the annotation of data. In the context of video analysis, for example, a one-minute video sampled at 24 frames per second (fps) consists of 1,440 frames. Manually annotating each of these frames is a time-consuming process, highlighting the complexity and effort required in preparing data for training models (Yan, et al., 2020).

Previous research in action recognition and dance move classification has predominantly utilized deep learning techniques such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to analyse video data (Yan, et al., 2020). Studies using motion capture data have shown promising results, but real-world video data, such as the ballet performances in our dataset, introduces additional challenges due to uncontrolled environments.

As mentioned, capturing temporal data presents significant challenges. However, incorporating a temporal awareness structure has been shown to greatly enhance model performance in video analysis (Matsuyama, et al., 2021).

Preprocessing

In training machine learning models for big data analytics, it's crucial to utilize both hardware and software resources, often taking advantage of parallelism (Kambatla, et al., 2013). Before training, the data must be thoroughly cleaned and analysed to identify and correct any anomalies that could impact accuracy. In a study on ballroom dance recognition, researchers demonstrated that eliminating certain dance movements can be effective. Since movement distance and direction can vary from person to person, this approach reduces complexity and enhances the model's performance (Matsuyama, et al., 2021).

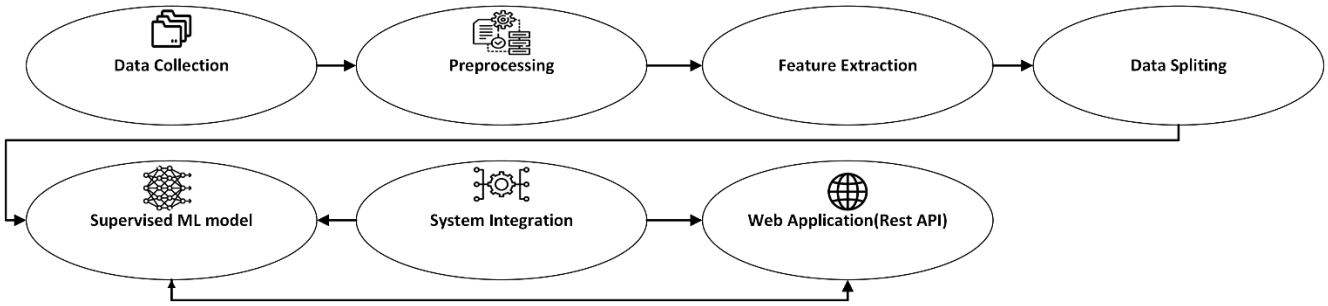
Feature extraction

There are several techniques used to classify data, with some researchers taking the extra step of automatically extracting features and labelling data (Yan, et al., 2020). Techniques such as PCA, LBP, HAAR, SIFT, SURF, BRISK, HOG, Hu moments extraction, and Zernike moments extraction are employed (Gupta & Singh, 2024) to significantly enhance the efficiency of supervised machine learning models, thereby achieving high accuracy.

Technology Review

Key technologies in this area include pose estimation algorithms, such as OpenPose and MediaPipe, which are used to extract key body points from video frames. These points can then be fed into machine learning models for classification. Deep learning models like CNNs have been effective in capturing spatial features, while RNNs and LSTMs are often used for capturing temporal sequences.

IV. Proposed Solution



SUPERVISED ML MODEL PROCESS FLOW 1

This solution proposes a two-method approach to classify the ballet moves, leveraging both pose-based and sequence-based analysis.

Model Choice

The choice of Random Forest and SVM is motivated by their balance between simplicity and effectiveness (Amanpreet, et al., 2016). Random Forest is particularly good at handling high-dimensional data and provides robust performance without requiring extensive hyperparameter tuning. SVM, on the other hand, is effective for classification tasks where the classes are well-separated and provides good generalization to unseen data (Amanpreet, et al., 2016).

1. Pose-Based Classification using Random Forest
   - Preprocessing:
     - Pose Estimation: We will use OpenPose to extract key body points from each video frame. These points include coordinates for major joints such as shoulders, elbows, hips, and knees.
     - Feature Extraction: The extracted key points will be used as features representing the spatial posture of the dancer in each frame.
   - Data Splitting: The dataset will be split into 80% for training and 20% for testing.
   - Model Training: A Random Forest classifier will be trained using the key points as input features. Random Forest is chosen for its robustness and ability to handle high-dimensional data.
   - Data Visualization:
     - Visualize the distribution of key points across different ballet classes.
     - Use t-SNE or PCA to reduce dimensionality and visualize the feature space.
   - Model Evaluation: Performance will be evaluated on the test set using metrics such as accuracy, precision, recall, and F1-score.
2. Sequence-Based Classification using Support Vector Machine (SVM)
   - Preprocessing:
     - Sequence Formation: The key points extracted from each frame will be organized into sequences representing the temporal flow of the ballet moves.
     - Feature Extraction: Temporal features, such as velocity and acceleration of key points, will be computed to capture the dynamics of the movement.
   - Data Splitting: The same 70/30 split will be applied to maintain consistency.

- Model Training: An SVM will be trained to classify the sequences into the respective ballet action classes. SVM is chosen for its effectiveness in classification tasks, especially when the data classes are well-separated.
- Data Visualization:
  - Visualize the temporal trajectories of key points for different ballet classes.
  - Plot confusion matrices to identify any misclassifications.
- Model Evaluation: The model's performance will be evaluated using the same metrics as the Random Forest approach, with an additional focus on temporal accuracy.

3. Increasing Accuracy and Fine-Tuning
   - Hyperparameter Tuning:
     - Random Forest: Fine-tune hyperparameters such as the number of trees, maximum depth, and minimum samples split using GridSearchCV or RandomizedSearchCV to optimize the model's performance.
     - SVM: Adjust hyperparameters like the regularization parameter (C), kernel type, and gamma value to find the optimal settings for the SVM model.
   - Cross-Validation:
     - Implement cross-validation (e.g., k-fold cross-validation) to assess the generalization capability of the models. This will help in identifying any overfitting issues and improving the model's performance on unseen data.
   - Feature Selection and Engineering:
     - Experiment with different feature selection techniques to identify the most relevant features for classification. Consider engineering new features, such as combining or transforming existing key points to capture more complex patterns.
   - Ensemble Methods:
     - Explore the potential of combining the predictions from both the Random Forest and SVM models using ensemble techniques like voting or stacking to improve overall accuracy.
   - Data Augmentation:
     - Perform data augmentation by artificially increasing the dataset size through transformations such as rotation, scaling, or flipping of video frames. This helps in creating more diverse training examples and improving model robustness.
   - Model Evaluation:
     - Reevaluate the fine-tuned models on the test set using the same metrics (accuracy, precision, recall, and F1-score). Compare the results with the initial models to measure the improvement in performance.
4. Integration into an End-to-End System
   - Inference Engine Development: The best-trained model from the fine-tuning phase will be deployed as an inference engine using frameworks such as TensorFlow Serving, TorchServe, Nvidia Triton, or Flask. This engine will be

designed to handle REST or gRPC requests, enabling it to process incoming video samples and return predictions in real-time.

- Web or Mobile Application Development: A user-friendly web or mobile application will be developed to allow users to upload video samples. This application will interface with the inference engine, sending video data for classification and displaying the predicted ballet moves. The app will be designed to handle various file formats and ensure a seamless user experience, making the solution accessible and practical for end-users.
- System Integration and Deployment: The inference engine and user-facing application will be integrated into a cohesive system that can be packaged and distributed. The system will include an executable file for easy deployment, sample data for testing, and comprehensive documentation to facilitate its use in various environments.

The following tools and libraries will be utilized:

- OpenCV: For video processing, including frame extraction and data augmentation.
- OpenPose: For extracting key body points from video frames.
- Scikit-learn: For implementing the Random Forest and SVM models.
- Python libraries: For data visualization, including plotting key point distributions, confusion matrices, and t-SNE/PCA results.
- Python: The primary programming language, chosen for its extensive machine learning and data processing libraries.

Dataset

The dataset used in this project is publicly available and can be accessed at the following link: UJAnnChor - A video dataset consisting of ANNotated CHOReography for temporal action localization in ballet. This dataset includes 1,020 videos with over 24,600 temporal annotations for 11 action classes, providing a robust foundation for training and evaluating the model.

V. Why Supervised Learning?

Supervised learning is suitable for this project due to the availability of a well-annotated dataset. The dataset's labels enable the models to learn to classify ballet moves accurately, making supervised learning an ideal choice.

Significance

The successful classification of ballet moves has broad implications beyond academic interest (Gupta & Singh, 2024). It can significantly enhance user experiences in interactive media, assist in developing educational tools for dancers, and provide detailed performance analysis for professional ballet dancers.

Given the complexity and difficulty of mastering dance, the automated classification of ballet moves can significantly contribute to both learning and performance analysis. Similar approaches have been successfully applied in previous studies, where supervised machine learning models were used to classify ballroom dance moves (Matsuyama, et al., 2021). Moreover, across various platforms, analysing video data has become crucial for numerous industries, including security, healthcare, entertainment, sports, and autonomous systems (Flick, 2014).

Expected Outcomes

Existing solutions often struggle with the complexity of real-world video data and may not effectively capture the temporal dynamics of ballet movements. This project aims to address these gaps by using a combination of pose-based and sequence-based methods in a supervised learning framework, focusing on both accuracy and practicality.

By using both pose-based and sequence-based methods, we expect to achieve a comprehensive solution for classifying ballet moves. We will measure the performance of each model using metrics such as accuracy, precision, recall, and F1-score. The expected outcome is a model that is both accurate and practical for real-world applications, offering a solid foundation for future work in dance move classification.

REFERENCES

Amanpreet, S., Narina, T. & Aakanksha, S., 2016. *A review of supervised machine learning algorithms*. New Delhi, India, IEEE.

Brooks, C., Amundson, K. & Greer, J., 2009. *Detecting Significant Events in Lecture Video Using Supervised Machine Learning*. Brighton, UK, s.n.

Flick, U., 2014. *The handbook of Qualitative Date Analysis*. 1st ed. London: SAGE Publications.

Gupta, S. & Singh, S., 2024. Indian dance classification using machine learning techniques: A survey. *Entertainment Computing,* Volume 50, pp. 1-3.

Jenik, C., 2021. *Statista*. [Online]
Available at: https://www.statista.com/chart/25443/estimated-amount-of-data-created-on-the-internet-in-one-minute/
[Accessed 27 08 2024].

Kambatla, K., Kollias, G. & Kumar, V., 2013. Trends in big data analytics. *J. Parallel Distrib. Comput.,* 74(7), pp. 2-8.

Kim, D., Kim, D.-H. & Kwak, K.-C., 2017. Classification of K-Pop Dance Movements Based on. *Sensors,* 17(6), pp. 1-2.

Matsuyama, H. et al., 2021. Deep Learning for Ballroom Dance Recognition:A Temporal and Trajectory-Aware Classification Model With Three-Dimensional Pose Estimation and Wearable Sensing. *Sensors,* 21(22), pp. 1-4.

Yan, X. et al., 2020. Self-Supervised Learning to Detect Key Frames. *Sensors,* Volume 20, pp. 1-3.