DOMAIN SPECIFIC VIDEO RETRIEVAL FOR STRENGTH, HYPERTROPHY AND CONDITIONING

Selman Eroğlu, Tobias Husebø, Brage Ramberg

1. APPLICATION

In this section we are going to delve deeper into the application of our proposed multimedia technology. With regards to background, related work, significance and requirements.

1.1. Background

In today's day and age, every aspect of human behaviour and need is getting digitalized. This involves, but is not restricted to, the car industry, writing industry and others. Contradictory, we have the workout industry. An industry which is yet to be digitalized to an extent at which other industries have strived to meet.

The workout industry is still heavily relying on personal trainers, and somewhat of a prior knowledge within the exercise domain. Which is a concern, regarding both cost of personal trainers and accessibility for everyone. This begs the question, what is the next technological advancement within this domain?

Via raising this question, we have arrived at a possible solution for further advancement in this field. An innovative and technological way of detecting your exercise, without the need for a personal trainer. This is done via utilizing reverse image searching with the assistance of classification models and a webcrawler. This technology can presumably mitigate both cost and accessibility obstacles.

1.2. Related work

The research for further development in the workout industry is an ongoing topic, where there exists a large amount of different approaches aiming to solve different aspects. Via using IEEE Xplore we have managed to successfully find two research articles with focus on points correlated to our topic. Specifically exercise and reverse image searching.

The articles in question are Workout Mentor: Improving posture in real-time using computer vision [1] and Image Vision-Based Pull-Up Recognition [2]. Both projects utilize pose estimation algorithms to extract features from workout videos and (or) photos. Where as the latter article have developed an algorithm employing the Xbox Kinects skeletal tracking, and the first MediaPipe [3]. Both aiming to classify a workout.

These articles have helped us in understanding the task at hand. Where as, we intend on employing the use of Google's mediapipe repository to extract pose features from the images, this to further accelerate our image searching algorithm. This repository contains solutions to extract features from images using ML models, which in return will give us domain specific visual content.

1.3. Significance

The proposed technology we are trying to create could mean a paradigm shift within the workout industry. This namely because of two main aspects:

1. cost
2. accessibility

The cost aspect is in regards to the average cost of personal trainers in the fitness industry. The cost of a personal trainer varies from one gym to another but, there is nearly always a fee involved. With our proposed technology, the user will have mitigated this aspect. This is because our application will give him instructions for free. This will lead to more financial freedom for the user.

Lastly, the accessibility aspect. When indulging in exercise, there is a need for a basic understanding of the exercise at hand. This is with respect to the physical health of the person participating in the activity. Doing exercises wrong may lead to injuries, and in severe cases death. This directs us to a requirement for basic understanding before starting to exercise. This requirement may decrease the number of people having a desire to start working out, because they lack the basic understanding. With our suggested technology, we will eliminate this prerequisite.

1.4. Requirements

For our proposed technology we have several requirements. We have divided the requirements into a user specific and performance & content requirements.

1.4.1. User requirements

User Requirements	Description
Uploading a photo / video	The user should be able to upload a photo or video to our application website.
Retrieving a photo	The user should retrieve n pictures, with $1 \leq n \leq 20$, $n \in \mathbb{N}$
Retrieving a video	The user should retrieve an instructional video explaining the workout.
User experience	The user should not have a difficult time navagating the application, where as the UI should be easy-to-use and comprehensible.

Table 1. User requirements for our proposed technology

1.4.2. Performance & content requirements

P.& C. Requirements	Description
Image / video format	The application should support various image and video formats including: JPG, PNG, PDF, MP4 etc.
Time complexity	The performance of the application should be fast and efficient, preferably, $\mathcal{O}(n)$ where n is the number of pics in the database
Performance	Based on the input, a preferable output would be the correct exercise and instructional video. Where as a preferable accuracy would be $\geq 95\%$

Table 2. Performance & content requirements for our proposed technology

2. SOLUTIONS

Our algorithm takes a workout video as an input and utilizes multimedia retrieval to find similar workouts from a dataset.

2.1. Minimum Viable Product

As a first step, our algorithm uses several frames extracted from the video and applies reverse image search against a workout image dataset, to find similar images. This raises the question about the following aspects of the reverse image search pipeline:

2.1.1. Image Preprocessing

We use the following dataset from Kaggle, available at `www.kaggle.com/datasets/hasyimabdillah/workoutexercises-images/data` This is a dataset of 20 different exercises, where each exercise folder contains around 500 images. This means that the images in the dataset are already labeled based on the exercise name. The images in the image workout dataset have been sourced from YouTube videos and Google Images. The images extracted from videos have the resolution 480×360, whereas the images sourced from Google Images remain original, so we need to find a way to standardize image sizes. Furthermore, many images extracted from workout videos contain graphics that have to be taken into account, as they are not part of the original video and may distort the feature extraction.

2.1.2. Feature extraction

We want to try different approaches and combine them in a meaningful way:

- Pose estimation: Analyze the body positions to compare with the poses in our database. Making use of the Mediapipe repository [3].
- Visual similarity assessment: Employ deep learning to find similarities in the colors, shapes, and textures.
- Scene recognition to identify the workout environment, which could be a gym, outdoors, or a home setting
- Another possibility is to utilize a convolutional neural network that can enhance the feature extraction process.

To get the best accuracy, we have to try weighting the feature vector in different ways, such that the right linear combination of feature vectors gives the best accuracy in exercise recognition.

2.1.3. Feature Indexing and Storage

Once the features of the images in the database have been extracted, we need to find a way to index and store the features of each image in a database that allows for efficient image retrieval of similar images.

Furthermore, we have to take the space complexity into account, to find a good balance in the space-time tradeoff that suits our system configuration best.

2.1.4. *Search Algorithm*

In order to ensure a smooth user experience, we need to find a way to efficiently search in the database for the n images that have the closest feature similarities. This includes finding a good similarity metric. One possibility would be to employ a linear search that applies the similarity metric to each image in the database to find the n most similar images.

2.1.5. *Retrieving the result*

We define a number of n (e.g. $n = 5$), to find the n best matching images from the dataset and return them to the user. Furthermore, we group the images based on their exercise label, which allows us to aggregate the results and predict the exercise that is being performed in the input video. This approach not only predicts the exercise that is being performed in the uploaded video, but also allows the user to evaluate a range of similar exercises for comparison. This is particularly beneficial for users that are unfamiliar with the specific exercises such that they don't perform the exercise correctly or combine various exercises and enables them to find the best fitting example.

2.2. Enhanced solution

In our enhanced solution, we make use of the entire video to have a more profound analysis of the workout. This gives us the opportunity to employ a reverse video search against a similar workout video database from Kaggle, available at `https://www.kaggle.com/datasets/hasyimabdillah/workoutfitness-video`. In addition to the frame-based approach, we can make use of the following video features:

- Movement tracking to analyze movement throughout the exercise, providing a detailed understanding of the exercise movement.
- Timing analysis to understand the sequence and duration of exercises, which helps identify complete exercise programs rather than isolated movements.
- Audio analysis can be used to understand the speed and intensity of the exercise

2.3. User Interface Implementation

Our goal is to make the application user friendly by implementing a user interface, which will provide the following functionalities:

- Video Upload: The user can easily upload the input video that should get analysed.
- Displaying Similar Videos: The user gets displayed similar videos from the data set, which are grouped by the exercise label.
- Exercise prediction: Display the final prediction of the algorithm.
- Video Access: The user gets access to an instructional video demonstrating the correct performance of the exercise, giving the user a learning experience to adjust their form and technique

3. PLAN

3.1. Project recap

The main product should classify a workout based on a video from a workout session. We aim to achieve this through reversed image search. Using a database we aim to determine which exercise is done one the video and retrieve another video with instructions on how to do it correctly.

3.2. Task breakdown

- Feature extraction from video frames.
- Database management.
- Majority vote for exercise.
- Create User Interface.
- Return information from input.
- Writing report.

3.3. Project timeline

Week	Tasks
8	Proposal and Feature extraction
9	Machine Learning pose estimation
10	Training and testing
11	User Interface and response generation,
12	Finalize code and report / Prepare presentation

Table 3. Project Tasks Schedule (Week 5 to Week 12, USYD semester 1 timeline)

Table 3 will be our initial project plan, however it may change depending based on different time frames and shifting priorities depending on unforeseen events, such as programming difficulties and debugging. We will however adapt along the way and finish with a viable product within the Project deadline at: 11:59pm, 14 May 2024 (Tue, Week 12).

3.4. Communication Plan

In this project we are going to use the task manager called Trello, to create and delegate tasks. In Trello we plan to create tasks as we go, using "Backlog", "Doing" and "Done" as our main categories for our task. Additionally the task will be created using SMART goals [4], meaning: Specific, Measurable, Achievable, Relevant and Time-Bound. This will create structure, a sense of progress and a way to not be overwhelmed by throughout the project. furthermore SMART goals will help us by avoiding overlapping work and merge conflicts.

For all other communications we will be using WhatsApp, to schedule meetings, asking questions and discussing important matters as they arise. Since we are students with different timetables we will have to plan as we go.

3.5. Version control and Collaboration

We are going to create a project in GitHub. All team members are familiar with git and agree that this is the best option for version control and collaboration.

4. REFLECTION

Our reflection part consists of an initiation subsection, research & analysis subsection and a moving forward subsection. These parts as a whole, convey information about how we have reflected around the project at hand while writing our proposal.

4.1. Project initiation

Our creative process surrounding finding the optimal project started with brainstorming and looking at sample projects in multimedia. We narrowed down the search by selecting fields we found the most interesting. Where the dominating multimedia types where audio, image and video. After discussing internally we found different areas which could be applicable. More specifically sports, like for example soccer or boxing. We created a list and narrowed it down by factors such as difficulty and uniqueness. As section 1.1 summarize we ended up with the exercise domain. Mainly, because this is still a developing field in terms of digitization. We found the use case interesting, because people may benefit in terms of both cost and accessibility.

4.2. Research and Analysis

A crucial part of this proposal was to research similar projects. Where the main idea was to see how beneficial our application would be in the specific domain chosen. We are thinking about using pose estimation as a feature to improve the algorithm. Where as this feature could be crucial in differentiating similar exercises. Our assumption is that image search alone, based on colour features for example, would not provide the best results. Therefore pose estimation can almost guarantee an improved overall accuracy.

4.3. Moving Forward

Moving forward we are eager to see how our algorithm works in terms of time complexity. Our aspirations is to create an exercise analysis app with swift results. We have also brainstormed different ideas on how to improve our database and time complexity. We think that this can be done by storing a complete feature analysis of each labeled photo. Where as pose estimation is one of other features. This is currently just an idea and we will have to look closer into the matter. Depending on how slow or fast our application performs.

Technical Req.	**Desc.**
Reverse Image search	Implement a reverse image searching algorithm to find corresponding exercise photos
Pose classification	Utilize googles mediapose repository to extract poses as a feature to our reverse image search
Majority vote algorithm	A majority vote algorithm to let the retrieved photos determine the exercise at hand
Performance metric	A metric to evaluate the performance of our algorithm
Photo database	A database containing specific photos of exercises
UI	A user interface for efficient user interaction

Table 4. Technical requirements for our proposed technology

5. REFERENCES

[1] Umme Aiman and Tanvir Ahmad, "Workout mentor: Improving posture in real-time using computer vision," in *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2023, pp. 1062–1067.

[2] Mingyan Ke, Haoran Tang, Huanan Zhu, and Mo Li, "Image vision-based pull-up recognition," in *2023 3rd International Conference on Electronic Information Engineering and Computer Science (EIECS)*, 2023, pp. 872–876.

[3] Google, "Mediapipe," GitHub repository, 2019, Accessed: 24th of March, 2024.

[4] Rob Watts Kimberlee Leonard, "The ultimate guide to s.m.a.r.t. goals," 2022, Accessed on march, 25, 2024.