

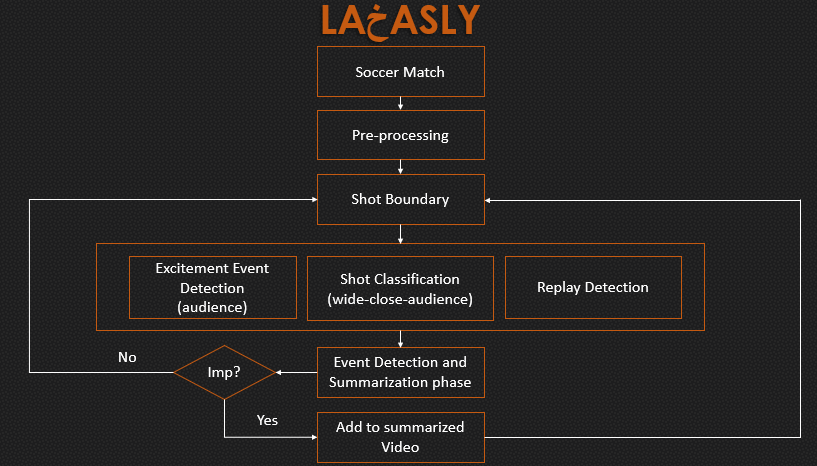
**A Multi-Modal System for Soccer Video Summarization**

**Graduation Project** **Progress Report 2**

Team Members:Ahmed Maher  
Ahmed Salama  
Moamen Hassan  
Mohamed Talaat

Supervised by:Dr.Magda Fayek

**The Proposed System**



In progress report-1 we discussed the pre-processing phase and shot boundary algorithm and its results.

In this report we will be discussing:

* Shot classification phase.
* Excitement Event Detection phase (Audio Processing)
* Excitement Event Detection phase (Goal Mouth Detection)
* Excitement Event Detection phase (Goal detection)
* A finished version of Laخsly using audio and shot boundary only.

1. **Shot Classification**

**1.1 Introduction**

Production crews use different shot types in broadcasting a soccer match, which is  
be used for high-level video analysis in a particular domain.

Cinematographers classify a shot into one of four categories Wide, medium, close-up and audience (out-of-field) shot classes, the definitions of which are usually domain-dependent.  
In the following, we define these four classes for sports videos:

**Wide shot:** A long shot displays the global view of the field; wide shots almost always display some part of the stadium, which decreases the dominant colored pixel ratio.

**Medium (In-field) shot:** A medium shot, where a whole human body is usually visible, is a zoomed-in view of a specific part of the field.

**Close-up:** A close-up shot usually shows the above-waist view of a player or referee.

**Audience (Out-of-field) Shot:** The audience, coach, and other shots are denoted as out of field shots.

The sequence occurrence of a close-up shots and audience (out-of-field) indicates an important event such as (goal, goal attempts …etc) during the match.

**

**1.2 Shot Classification Approaches**

The above definitions consider shot-types as functions of field region. Because field  
region information is available after **Grass dominant ratio extraction** (discussed before in progress report-1) but using only the grass dominant ratio didn’t yield great results in shot classification.

In the following section we discuss three approaches to classification and comparing between them:

* Shot classification using Grass dominant ratio extraction
* Shot Classification using Face Detection
* Shot classification using Deep learning
* Shot classification using image processing techniques

**1.2.1 shot classification using Grass dominant ratio extraction**  
The proposed algorithm is based on a specific threshold (range) for grass ratio (G); which was developed by observing different matches in many lighting and weather conditions to appropriately define a range that will cover the four shot types.

Classification of a shot into one of the discussed three classes is based on spatial features.  
Therefore, shot class can be determined from a single key frame or from a set of frames  
selected according to a certain criterion.

thinking intuitively, G would be almost zero for out-of-field frames, a low G value in a frame corresponds to a close-up, while high G value indicates that the frame is a long view, and in between, a medium view is selected.

Frame Type

***For each shot***

***Choose a set of key frames in the shot***

***Compute grass ratio G of these frames***

***Classify frames***

***Determine the majority type of the frames and assign it to the shot***

Due to the computational simplicity of the proposed algorithm, computing the grass ratio of all frames in the shot would not be a big overhead.

Although the accuracy of the above simple algorithm is sufficient but it does not meet the desired outcome.

**1.2.2 shot classification using Face detection**

In the observation phase mentioned above it can be said that the Wide shot does not contain any clear faces (large enough to be recognized by a human or a computer), in the medium shot a face can or cannot be recognized depending on the angle of the camera filming the shot, in the close shot a face can be clearly recognized as it covers a large area of the frame.

With that said, combining face detection model and the Ratio G to obtain better results in classification.

Frame type

***For each shot***

***Choose a set of key frames in the shot***

***Compute grass ratio G of these frames***

***Determine whether the frames contain face or not and if it does get the***

***Area of the bounding box around the face***

***Classify frames***

***Determine the majority type of the frames and assign it to the shot***

The obtained results are better than using grass ratio only but the computational time and complexity is much worse because the face detection model is very complex and time consuming.

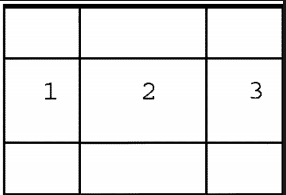
**1.2.3 Shot Classification using deep learning**

In this approach, Frames are Extracted from Matches and filtered (assigned labels) Manually to construct a data set, Then A model **(CNN)** is used to train on the constructed data.

This approach is much better than the previous techniques and has a great classification accuracy 86%.

**1.2.4 Shot Classification using image processing**

In this approach, an image, after a dominant color mask is applied i.e. green is white and any other color is black, is divided into 3:5:3 grid.



The image classes are determined using the green pixels in 1, 2, 3 regions and the absolute difference between 1,2 and 2,3 to estimate is there a close shot or not.

The close out is pretty simple, if green ratio is tiny then it is close-out.

But this approach has a good accuracy, actually near 60%, but compared to the above its subpar.

The advantages of this technique are that its computationally inexpensive (no models) and its implemented from scratch

***3.*Excitement Event Detection phase (Goal Mouth Detection)**

For soccer video, the goal-mouth scenarios can be selected as the highlighted  
candidates, for the reason that most of the exciting events occurs in the goal mouth area  
such as the goal, shooting, penalty, direct free kick, etc

On the other hand, the non-goal-mouth scenarios often consist of the dull passes in the  
midfield, defense and offense or some other shots to the audiences or coaches, etc, which  
are not considered as exciting as the former.



**Goal mouth detection algorithm:**

***For each frame***

***Convert image form RGB to BGR then to gray***

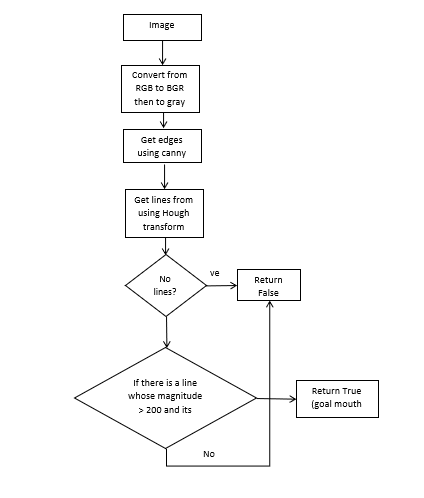
***Perform edge detection using canny***

***Get lines from detected edges using Hough transform***

***Skip lines whose magnitude is less than 200***

***Check if a line is parallel to 2 other lines then there is a potential goal mouth in the image and return true***

**Flow chart of the algorithm:**

****

Accuracy is about 70% - 80%.

**4.Excitement Event Detection phase (Goal detection)**

The score board is a caption region (usually at the top left) distinguished from the surrounding region,  
which provides the information about the score of the game and the match time.  
the score board dimensions and position is always constant in a certain league so detection of the score board itself is not needed



**4.1 Goal detection using OCR**

The goals are detected by the help of scoreboard with optical character recognition



**La5asly** uses **Tesseract** engine for **OCR**, it is free software, developed by **Google**

***Algorithm:  
1: while reading the input video.  
2: for each 5 seconds do.  
 2.1: apply the mask to get the scoreboard.  
 2.2: run the tesseract engine to check if the results changed or not.  
 2.3: if changed and stable:  
 2.3.1: save the results (error free).  
 2.4: if changed:  
 2.4.1: save the results locally and go to 2.***

***4.2 Goal Detection with* structural similarity image index (SSIM)**

Goals are detected with the structural similarity image index (**SSIM**), It is an enhanced way to detect if the image changed or not. It uses the difference between the scoreboard and get the mean, variance and illuminance between the two scoreboards and get if the scoreboard changes or not. If scoreboard changes then it is an event (goal or substitution).

