UCF Sports Action Recognition

To identify the sports action, two main steps are:

1. Action Localization (identification of spatio-temporal location)
   It deals with background clutter and spatial complexity of the scene.
2. Action Recognition (assigning videos to a particular action class).

Feature Extraction:

Good features should be invariant to changes in scale, rotation, affine transformations, illumination, and viewpoint.

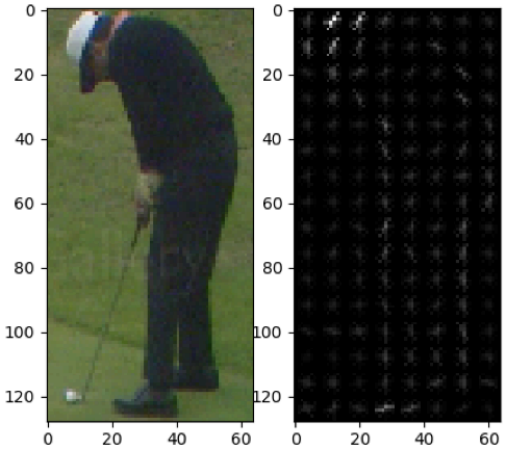
Low Level(Local) Features:

1. **Histogram of Gradients (HoG)** for all three channels - RGB

It is invariant to geometric and photometric transformations, as it uses only the occurrences of gradient orientation in localized portions of an image.

Image is first preprocessed using gt folder values (bounding box) and then cropped into 64*128 colored image.

I used a (2,2) bounding block of (8,8) cells for all three channels (RGB) with 9 orientations which generates a feature vector of length 11340 (3780 * 3). HoG is variant to object rotation, so I also used color values and Histogram for feature extraction.



2. **Color Values (Spatial Features)**

We can change the feature vector length by choosing the appropriate image size, so I used
32 * 32 (1024) image size for all three channels, so feature vector length is 3072 (1024 * 3) as HoG feature vector is very lengthy. Also, color values can easily segregate the background and foreground object for outdoor games like golf and horse riding (grass-green and person), diving (water-blue and person).

If the person wears same color clothes with the background, then it can give the wrong results.

3. **Histogram – RGB**

It clusters the intensity values into specific bins, and it's mostly robust to occlusion and viewpoint changes. I used 16 bins, 3 channels, so feature vector length for this histogram is 48. We can also use the bi-modal histogram to clearly separate the background and foreground and use the intensities of foreground object.

High Level(Holistic) Features:

Shape models and Spatio-Temporal Structures can be used as a feature extraction for human sports action classification. It is taking a lot of time, so I didn't use it while training the data.

Active Contour Model

It is robust to human body part movement variations as it maintains the structural information of actions. Also, it is compact as compared to Low Level Features, so training time is less, but extracting high level feature is computationally expensive. So, if we store the features somewhere and use that while training, then it might be helpful.

I used the 400 boundary points and gt values as center for the boundary circle and stacked the x and y-coordinate, so feature vector length is 800.

UCF Sports Action Recognition

Classifier Design:

SVC

Support Vector Machine uses maximum margin concept as it maximizes the distance to the nearest points for all classes. It is a geometric model processed on the dependent data and used basically for image recognition.
Parameters: Kernel, gamma, C

Kernel: Linear (Faster and less complicated)
Kernel: Radial Basis Function rbf (Complex and computationally expensive)

C – It controls the tradeoff between smooth decision boundary and classifying training points accurately.
Low C - Smooth Boundary between classes (simple)
High C – Wiggly Boundary between classes (complicated as more training points are correct)

Gamma – It defines how far the influence of a single training example reaches.
Low Gamma Value – Influence on Far Points, so smooth boundary
High Gamma Value – Influence on Close Points, so wiggly boundary
I used C = 1.0, kernel = “rbf” and gamma = “auto”

Random Forest

It takes less time to train the data if it is done in parallel. It is robust and it don’t require much tuning of hyper parameters. Its goal is to maximize Information gain.
Entropy controls how a Decision Tree decides where to split the data. Entropy measures the impurity in a bunch of examples.
Information Gain = Entropy[parent] – (Weighted Average) * Entropy[Children]

Max_depth - The maximum depth of the tree
If max depth is 0, then nodes are expanded until all leaves are pure.
I used max_depth value as 8

Deep Neural Net

I tried NVIDIA Model which has 5 convolutional layers and 4 dense layers. I just passed the images, not the extracted features in deep neural network.

UCF Sports Action Recognition

Evaluation:

Leave One Out

Image Histogram alone doesn't give good results with either Support Vector Machine and Random Forest. Spatial Features and Image Histogram combined gives better result than Image Histogram alone, but not up to the mark. Accuracy is best for either Deep Neural Network or with all three features – Histogram of Gradients, Image Histogram and Spatial Features.

Time taken by SVM is higher than Random Decision Forest, but accuracy by SVM is better than Decision Forest.

Swing and Diving class has better results than any other class because the number of test videos for these action classes is bigger than other classes.

Lifting Action class has 0 accuracy because this action class doesn't have gt folder, so I skipped this action class for SVM and Random Decision Forest.

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

Actions	Random Decision Forest			Support Vector Machine		Deep Neural Network
	Image Histogram	Spatial and Image Histogram	Spatial, Image Histogram and HOG	Image Histogram	Spatial, Image Histogram and HOG	
Lifting	0	0	0	0	0	0.38
Golf-Swing-Back	0.24	0.48	0.51	0.23	0.65	0.67
Kicking-Side	0.3	0.34	0.36	0.2	0.46	0.52
SkateBoarding-Front	0.34	0.58	0.61	0.32	0.76	0.74
Walk-Front	0.25	0.34	0.35	0.21	0.53	0.64
Golf-Swing-Front	0.17	0.32	0.44	0.16	0.57	0.61
Diving-side	0.83	0.84	0.88	0.79	0.94	0.83
kicking-front	0.1	0.26	0.32	0.04	0.41	0.45
golf-swing-side	0.38	0.28	0.39	0.24	0.52	0.45
riding-horse	0.44	0.49	0.57	0.42	0.71	0.67
run-side	0.39	0.53	0.58	0.49	0.66	0.73
swing-bench	0.78	0.89	0.91	0.78	0.89	0.95
swing-sideAngle	0.86	0.85	0.88	0.74	0.95	0.98

Actions	Random Decision Forest		Support Vector Machine	
	Sensitivity	Specificity	Sensitivity	Specificity
Lifting	0	0	0	0
Golf-Swing-Back	0.62	0.54	0.64	0.71
Kicking-Side	0.33	0.43	0.42	0.34
SkateBoarding-Front	0.67	0.63	0.72	0.81
Walk-Front	0.36	0.45	0.41	0.48
Golf-Swing-Front	0.46	0.43	0.59	0.52
Diving-side	0.82	0.88	0.92	0.91
kicking-front	0.29	0.38	0.39	0.53
golf-swing-side	0.32	0.54	0.61	0.68
riding-horse	0.52	0.57	0.74	0.79
run-side	0.59	0.61	0.64	0.63
swing-bench	0.85	0.87	0.93	0.83
swing-sideAngle	0.84	0.89	0.96	0.91

A good classification gives almost similar values for all three accuracy, sensitivity and specificity and all having high values. However, if there is high sensitivity and low specificity, then poor candidates are re-evaluated to eliminate the false positives. Also, if there is low sensitivity and high specificity, then good candidates are re-evaluated to eliminate false negatives.