Human Activity Recognition using kernelized SVMs (Group 1)

Arjit Singh Arora
Student
IIITD
arjit21452@iiitd.ac.in

Barneet Singh
Student
IIITD
barneet23028@iiitd.ac.in

Aman Sharma
Student
IIITD
aman21010@iiitd.ac.in

Keshav Goel
Student
IIITD
keshav20512@iiitd.ac.in

Shubham Kumar Choudhary
Student
IIITD
shubham23093@iiitd.ac.in

Abstract

The Human Activity Recognition(HAR) model differentiates the different activities a human performs. It has very wide applications, such as in the field of healthcare to monitor the physical activity of patients, in the field of sports and fitness to analyze the activities of athletes to improve their performance, in security, and many more. In this study, you can see the pre-processing and EDA done on the HAR data set provided to us, before applying the model to classify the images into different classes. Some of the parts of data pre-processing are - checking for imbalance, checking for the Gaussian distribution, Standardization, Normalization, etc.

1 Introduction

In this study, we work with two essential components of our project: Image data and Associated labels, aiming to classify them from a list of 15 actions. These include sitting, clapping, dancing, using laptop, hugging, sleeping, drinking, calling, cycling, laughing, fighting, eating, listening to music, texting and running.
We have two main resources at our disposal—CSV files, 'training.csv' and 'test.csv,' and the raw image data in the 'training' and 'test' directories.
The **CSV** files, 'training.csv' and 'test.csv,' play a critical role in our project. Each CSV file contains two columns. The first column, labeled '**filenames**,' corresponds to the name of each image in our dataset. The second column, labeled '**labels**,' provides information about the content of each image. These labels define the categories or classes to which each image belongs, helping us understand the action being performed in the image.
The raw **Image data**, stored in the 'training' and 'test' directories, forms the visual foundation of our project. It is here that we find the actual images corresponding to the file names listed in the CSV files. This image data is crucial for training and testing our classification model. We have structured this data by **organizing the images in their respective label subdirectories.**

2 Data Pre-processing Techniques

2.1 Class Imbalance

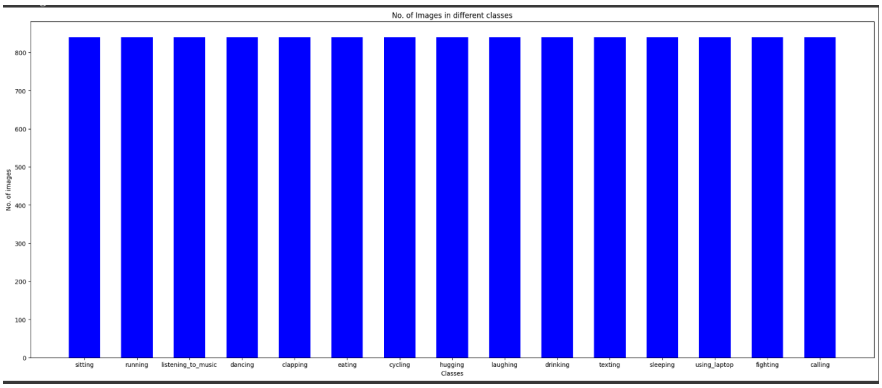

Figure 1: A Bar Plot depicting class balance.

Class imbalance in machine learning refers to the situation where one class has significantly fewer instances than the others, potentially leading to biased model performance.
All the 15 classes in the HAR dataset have the same number of images, hence, there will be no class imbalance based on the no. of images per class. This will ensure balanced training (which can help the model learn the characteristics of all classes effectively) and reduced bias.

2.2 Pie chart and Data Distribution

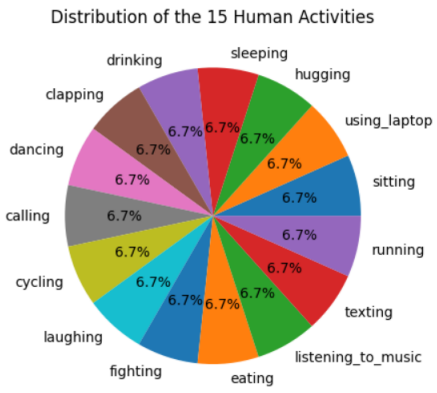

Figure 2: A Pie chart depicting class distribution.

There are a total of 12600 images in the training folder of the dataset, along with an additional 5400 testing images. Each of the 15 classes in the training set accounts for 6.7% of the total images. This means there are 840 images per class.

2.3 Histogram plots for all data and per class based (RGB and Grayscale)

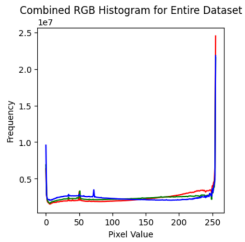

Figure 3: An RGB Histogram wrt the entire dataset

The dataset clearly doesn't follow a gaussian distribution as evident with strange peaks at the beginning and end (These suggest the presence of ouliers and noise which needs to be removed) . Also on an average the colour distribution is even , evident from the closely grouped peaks and modes.

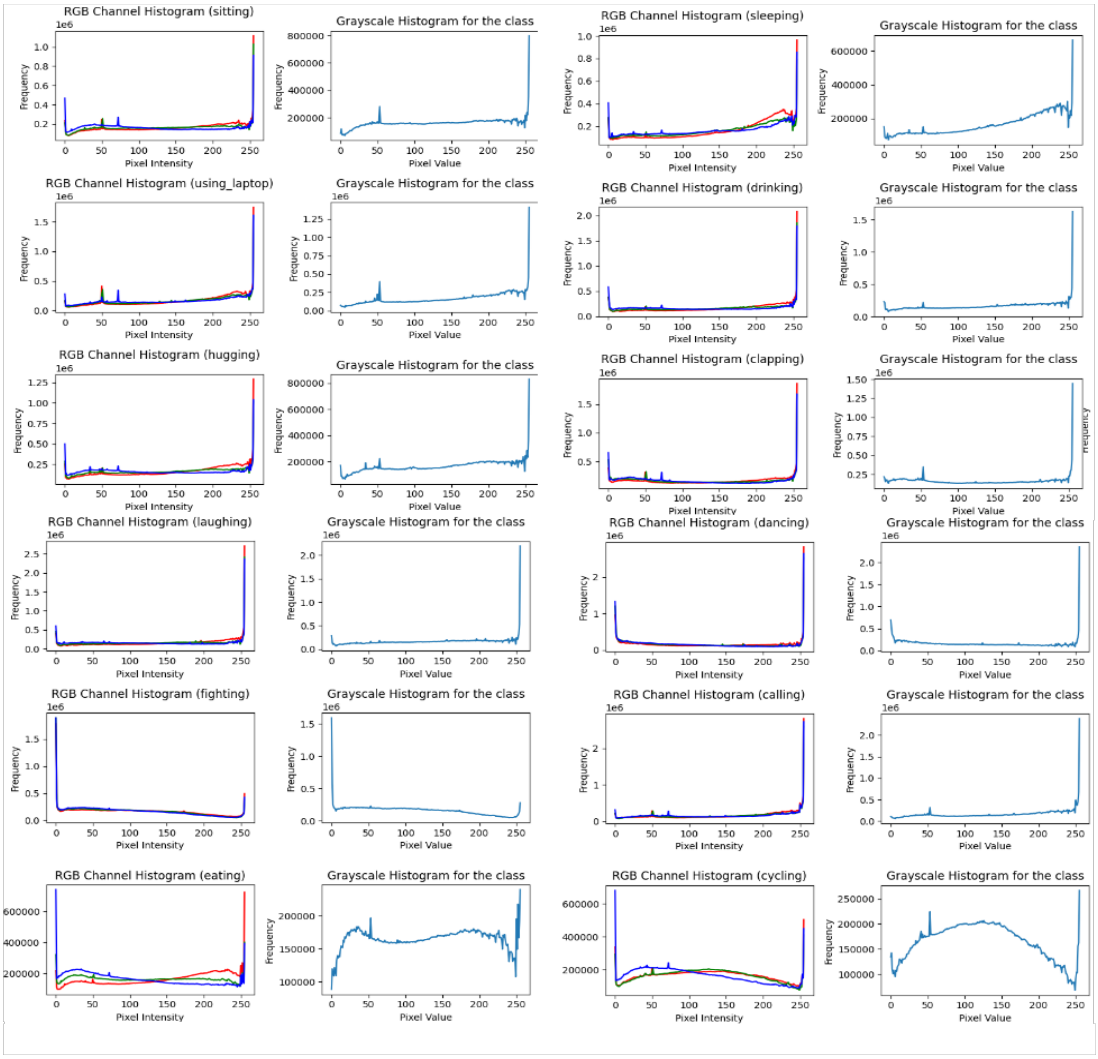

Figure 4: RGB Histograms for all classes

All the classes roughly follow the same distribution as the whole dataset with 2-3 small initial peaks and one large peak at the end.(Except the cycling class and the eating class which has a broader peak in the middle as well).

2.4 Standardization with respect to entire data set

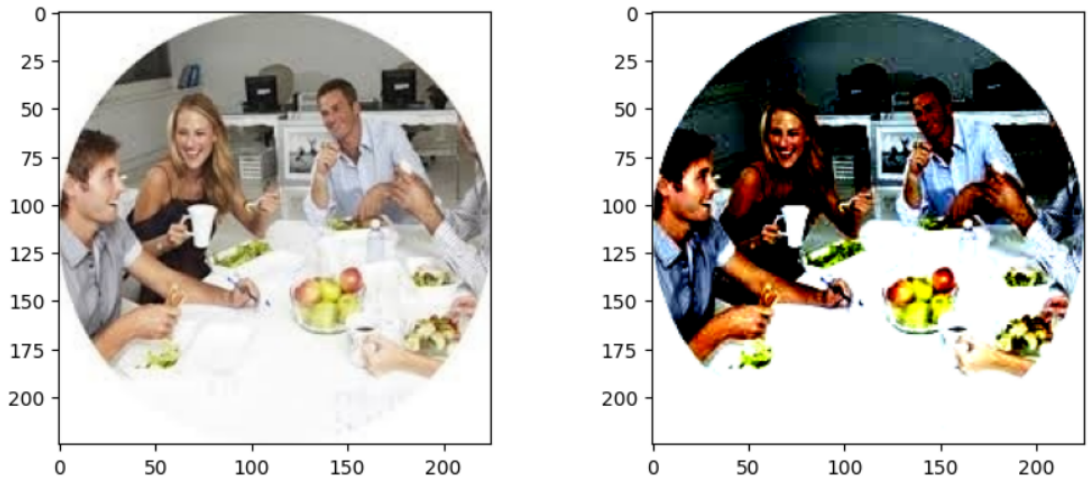

Figure 5: Visual Effect of standardization on the image

The dataset doesn't follow a Gaussian distribution meaning technically we shouldn't standardize. Still standardized data might be helpful and both versions can be used while testing for accuracy and the more accurate one can be chosen. This is less sensitive to outliers and doesn't maintain the shape of the distribution. Standardization scales pixel values within each image and across the entire dataset. This can be essential for machine learning algorithms that are sensitive to the scale of input features. For example, support vector machines (SVMs)

2.5 Normalisation

This is useful when the data distribution is unknown or not Gaussian.Also, this is sensitive to outliers and retains the shape of the original distribution. This is also called min-max scaling and can be beneficial because of following reasons:-

Feature Scaling: Normalization is often used to scale features to a common range, typically between 0 and 1. This ensures that all features have the same scale, which is important for SVMs since they are sensitive to feature scales. Normalized features can prevent certain features from dominating the distance calculations or influencing the model.

No visible change was observed in the images after normalizing.

2.6 Consistent image size

Resizing images distort the aspect ratio and its geometry (if the new width and height are different). Since we are using kernelized svms we might not need it at all since it is a non-linear method that might handle different image sizes.

2.7 Histogram of oriented Gradients (HOG)

Figure 6: Gradient features of the image (The silhouette)

Histogram of Oriented Gradients, also known as HOG, is a feature descriptor. It is used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in the localized portion of an image. For the regions of the image, it generates histograms using the magnitude and orientations of the gradient.

2.8 Edge map using Sobel's filter

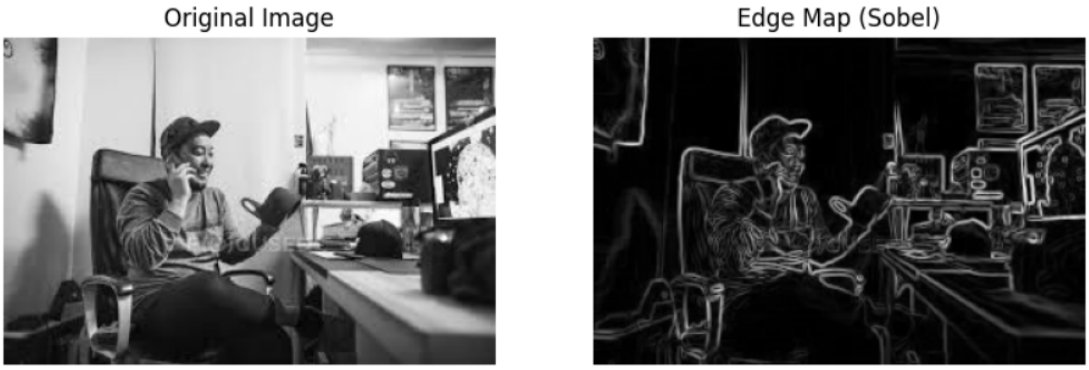

Figure 7: The Edge map of the image

Sobel's Edge detection filter is used to detect edges in the images. These can be converted to statistical features like mean, median, etc. to convert into feature vectors.

2.9 Local Binary Patterns (LBP)

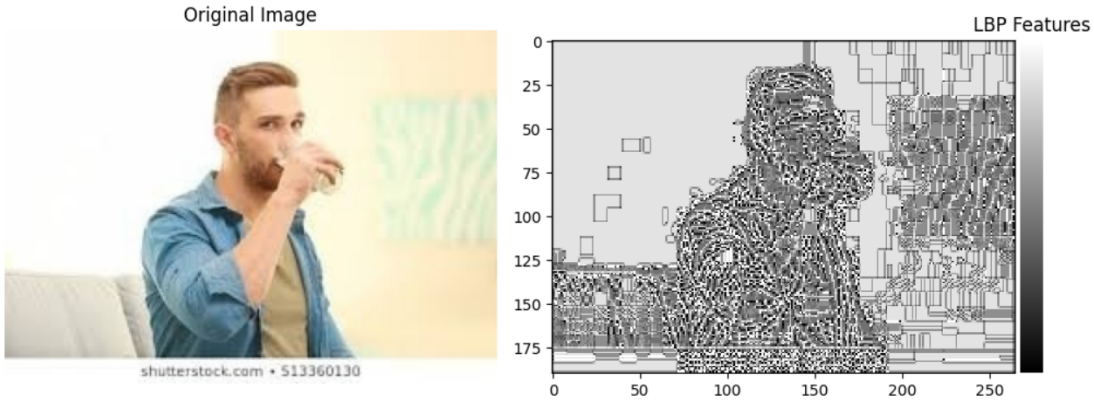

Figure 8: LBP texture Features of the image

LBPs compute a local representation of texture. This local representation is constructed by comparing each pixel with its surrounding neighborhood of pixels. LBP looks at points surrounding a central point and tests whether the surrounding points are greater than or less than the central point (i.e., gives a binary result).

2.10 Gaussian filter to remove random noise

A Gaussian filter is used to remove random noise from the image. If applied with a large sigma, it starts to remove edge information beyond removing noise. (i.e. it blurs). For now, it has been applied to the image and it might increase accuracy (will be verified later)

2.11 Converting to HSB space and finding HSV Histograms

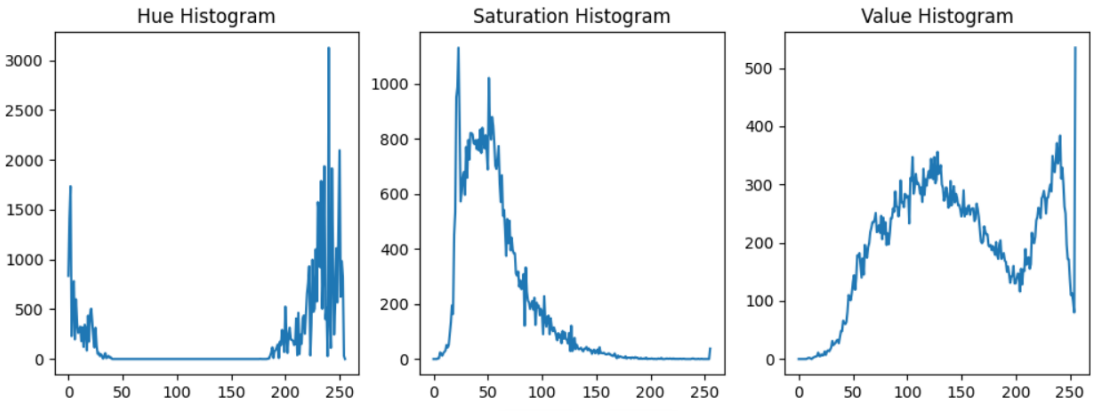

Figure 9: HSV color features of the Image

HSV (Hue, Saturation, Value) is a color model that represents colors based on three components:

Hue: This component represents the type of color, such as red, green, or blue. Hue is measured in degrees (0° to 360°) and represents the position of a color on the color wheel.

Saturation: Saturation refers to the intensity or purity of a color. A high saturation value indicates a vivid, pure color, while a low saturation value results in a more muted or grayish color.

Brightness: Brightness, often referred to as value or luminance, controls the lightness or darkness of the color. It ranges from 0 (black) to 100 (white).

Converting an image to the HSB color space is a common technique in image processing and computer vision for various reasons:

Color Analysis: The HSB color space is well-suited for color analysis and feature extraction. By separating color information into hue, saturation, and brightness components, you can perform tasks like color-based object detection and tracking. This can be done by using colour histograms(HSV histograms) in this form as features

2.12 Random Image from each class generator

A random image for each of the 15 classes can be taken out. This can be used for further analysis of the data.

2.13 Split data into training and testing

The data has been split into training and testing components comprising 75% and 25% of the total data respectively. Care has been taken to ensure that even after the split, all the classes remain balanced. This has been achieved by moving 210 random images from each class to the testing data. So the training now has 630 random images per class

3 Advanced Preprocessing

Advanced preprocessing techniques are crucial for refining raw image data. They enhance image quality, extract relevant features, and eliminate noise and unwanted elements, making images more suitable for higher-level analysis and computer vision tasks.

Going beyond basic transformations, sophisticated preprocessing methods specifically target image optimization, feature extraction, and preparation for subsequent analysis stages. These methods play a pivotal role in ensuring the accuracy and effectiveness of image processing pipelines.

3.1 Canny Filter

Canny edge detection filters, a foundational technique in computer vision, precisely identify edges within an image while reducing noise. These filters employ a multistage algorithm, including blurring, gradient calculation, and thresholding, to extract clear and prominent edges.

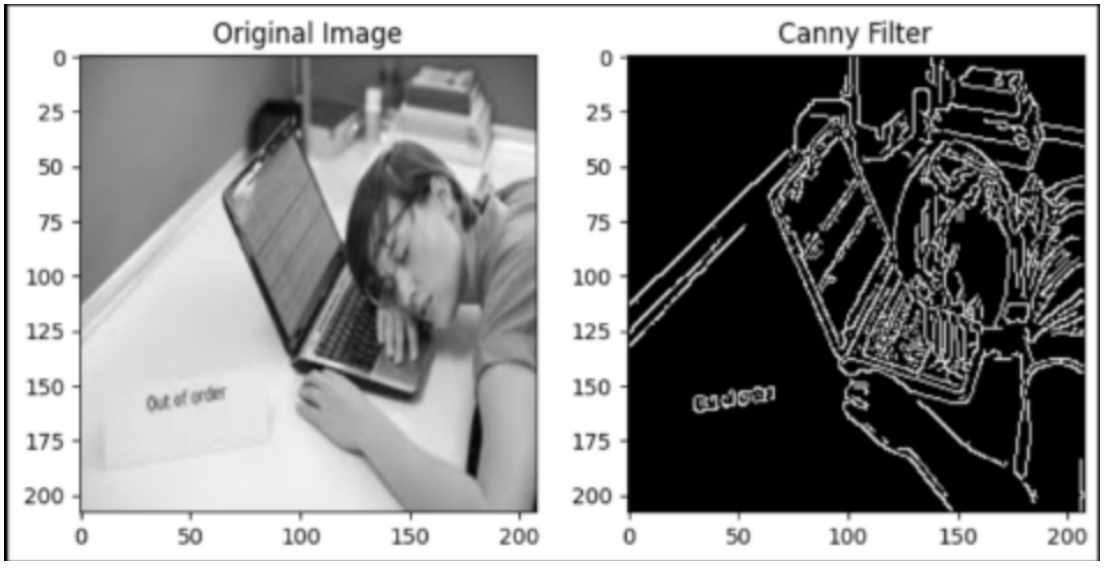

Figure 10: Effect of Canny filter on image

Effects on Image Data:

- **Edge Localization:**
  Canny filters precisely locate edges within an image, enabling accurate identification of object boundaries and shapes.
- **Noise Reduction:**
  By employing Gaussian blurring, these filters effectively minimize noise, enhancing the clarity of detected edges.
- **Feature Extraction:**
  Essential for feature extraction, Canny filters identify salient features critical for subsequent analysis or computer vision tasks.
- **Enhanced Edge Visibility:**
  Canny filters enhance edge visibility, making edges more pronounced and distinct, aiding in object detection and recognition.

Canny edge detection filters have a significant impact on image data by pinpointing edges accurately, reducing noise, and extracting crucial features, essential for various computer vision applications.

3.2 Bilateral Filter

The Bilateral Filter, a good image processing technique, excels in noise reduction while preserving critical image features and edges. It combines spatial proximity and pixel intensity similarity to achieve selective smoothing without compromising edge sharpness. In our case, it worked better than Gaussian for noise because Bilateral filters are often preferred in scenarios where preserving edge details is crucial.

Figure 11: Effect of bilateral filter on image

3.3 SIFT features

SIFT (Scale-Invariant Feature Transform) features are local descriptors that capture the distinctive appearance of an object at specific interest points in an image. These features are remarkable for their invariance to image scale, rotation, and illumination changes, making them highly robust to variations in image conditions. Additionally, SIFT features exhibit resilience to noise and minor viewpoint alterations, ensuring their effectiveness in a wide range of image-processing tasks.

3.4 Other Techniques

- Min Max scaling
- Template matching / Patch similarity
- Contours
- Median filters

4 Existing work and analysis

After going through the papers published earlier, we have found out about the work that has been done and the methods used for feature extraction. Here are a few listed below:

4.1 Paper Title: "Faster Human Activity Recognition with SVM", Year: "2012" (Authors: K. G. Manosha Chathuramali, Ranga Rodrigo) [1]

In this paper, the author employs a process that involves creating a silhouette for feature extraction. This is done through background subtraction, which essentially means isolating the subject from its background. Then, they incorporate motion descriptors, which are numerical representations that capture the movement characteristics of the subject.

In simpler terms, they're using a machine learning technique (SVM) to understand and categorize different activities. To do this, they first create a silhouette, which is like an outline of the moving subject. They do this by removing the background. Then, they add details about how the subject is moving (motion descriptors). This combination of silhouette and motion information forms the basis for classifying activities.

4.2 Paper Title: "Research on human activity recognition based on image classification methods", Year: "2017" (Authors: istė Štulienė, Agnė Paulauskaitė-Tarasevičienė)[2]

This paper focuses on classifying human activities using different machine-learning techniques. These methods include Convolutional Neural Networks (achieving around 90% accuracy), Bag of Features model (69% accuracy), Support Vector Machine (60% accuracy), and K-Nearest Neighbors (41% accuracy).

The paper conducts a comparative analysis of these methods for recognizing human activities. It uses a dataset of images representing five distinct categories of daily life activities. The study demonstrates that incorporating hardware and sensors can significantly enhance the accuracy of activity recognition.

4.3 Paper Title: "A Detailed Review of Feature Extraction in Image Processing Systems" Year: "2014" (Authors: Gaurav Kumar, Pradeep Kumar Bhatia) [3]

In this paper for feature extraction, several techniques are employed:

- Statistical features like zoning, characteristic loci, and distances are derived from the distribution of points in the image.
- Global features such as Fourier transforms and wavelets capture broader characteristics of the image.
- Geometric Features like strokes provide information about the shapes and structures within the image.

These techniques collectively enable the extraction of important information from images, which is valuable in various applications of image processing.

4.4 Paper Title: " Histograms of Oriented Gradients for Human Detection" Year: "2015" (Authors: Navneet Dalal, Bill Triggs) [4]

This paper examines feature sets for robust visual object recognition, using linear SVM-based human detection as a case study. After evaluating various existing edge and gradient-based descriptors, it was demonstrated through experiments that grids of Histograms of Oriented Gradient (HOG) descriptors perform notably better than other feature sets for detecting humans.

4.5 Paper Title: " Towards a deep human activity recognition Approach based on video-to-image transformation with a skeleton data" Year: "2020" (Authors: Ahmed Snoun, Nozha Jlidi) [5]

This paper introduces a new HAR approach based on the extraction of human skeletons from videos. Three feature extraction techniques are proposed in this work. These techniques are- dynamic skeleton,(founded on the concept of dynamic images introduced in the literature), skeleton superposition, and body articulations (which use only the body joints instead of the whole skeleton to recognize the ongoing activity). The Deep learning library- OpenPose is used for Pose Detection here. The obtained images from these three techniques are analyzed and classified using CNNs.

4.6 Paper Title: " Feature selection with kernelized multi-class support vector machine" Year - "2021" (Authors: Yinan Guo, Zirui Zhang, and Fengzhen Tang) [6]

This paper introduces an innovative approach for non-linear feature selection tailored for multi-class classification challenges within the support vector machine framework. The proposed method leverages a kernelized multi-class support vector machine and employs an efficient variant of recursive feature elimination. This technique identifies effective features across all classes by enabling the classifier to construct distinct decision functions that separate each class from the rest simultaneously.

5 New Ideologies and Methodology

In this project, the methodology combines traditional statistical features (mean, variance, etc.) extracted from Sobel's filter edge map with color (HSV) histograms and HOG features to form a comprehensive set of image features. To improve computational efficiency and address the curse of dimensionality, dimensionality reduction techniques will be applied to reduce the feature space. The classification task will be carried out using kernelized support vector machines (SVMs), allowing us to explore various kernel functions (e.g., linear, radial basis function, polynomial) and different preprocessing strategies (standardization, normalization). This innovative approach leverages the power of kernelized SVMs to capture complex relationships in the data, ultimately enhancing the accuracy of human activity recognition from images. Furthermore, the flexibility to experiment with multiple kernels and preprocessing variations promises to unveil new insights and optimize performance in this challenging task.

6 Results

6.1 HOG + HSV + LBP features

On employing a combination of HOG, HSV, and LBP features. HOG features, extracted from grayscale images, captured shape and texture information. HSV features, extracted from color images, provided color information. LBP features, extracted from both grayscale and color images, captured local texture patterns. These extracted features were concatenated to form a comprehensive feature vector for each image, resulting in approximately 4,500 features per image. This feature matrix was then fed into an SVM classifier with a polynomial kernel of degree 6, leading to optimal classification results. Min-Max scaler boosted the accuracy by 4%

Accuracy on Cross-validation: 35%

6.2 Bilateral Filter (for noise removal), Sobel filter (for BG separation) then HOG + HSV + LBP features

On removing noise and highlighting the edges of an image would make it easier to extract HOG features, but it actually made the accuracy worse. This might be because the filter removed some of the important edges that HOG needs to work well.

Accuracy on Cross-validation : 30%

6.3 Ensemble of HOG, HSV

On combining HOG and HSV features and trained two separate SVMs on different scales of the data. During classification, we calculated the confidence score of each SVM and chose the one with the higher confidence for prediction.

Accuracy on Cross-validation : 31%

6.4 SIFT Features

Integrating SIFT, HSV, and LBP features augmented our image analysis pipeline, amalgamating distinctive key points, color nuances from HSV, and local texture patterns from LBP. This fusion enriched the model's comprehension, enabling a more detailed and nuanced interpretation of the visual data and enhancing its ability to discern intricate details, colors, and texture variations within the images.

Accuracy can be improved if we consider some more features of SIFT, as for now we only took the first row of each feature extracted, which is 1 x 128.

Accuracy on Cross-validation : 25%

6.5 HSV+LBP

The fusion of HSV (Hue, Saturation, Value) color space with LBP (Local Binary Patterns) descriptors creates a powerful feature extraction technique for image analysis. HSV captures rich color information, while LBP encodes local texture patterns. This fusion enables a comprehensive image representation, facilitating nuanced analysis and interpretation. The combination leverages both color nuances from HSV and fine-grained texture details from LBP, providing a holistic understanding of the visual content. This fusion finds applications in a wide range of image-processing tasks, including object recognition, image segmentation, and content-based image retrieval.
Accuracy on Cross-validation : 30%

6.6 Performance Evaluation of Various Kernels for Optimal Model Selection

- Linear Kernel: 28%
- RBF Kernel (Gaussian): 34%
- Polynomial Kernel (Deg-6): 35%
- Sigmoid Kernel: 14%

6.7 Other Techniques

- To address the dimensionality issue, we augmented the number of HOG features and employed mean and variance calculations to reduce the feature space. This approach proved effective as PCA, our initial choice, failed to yield the desired results.
- Z-score Normalization

7 Feedback and Conclusion

Noise in the dataset:
Traditional feature extraction methods involve handcrafted operations on the image. While these methods (filters and thresholds) may capture certain local features, they do not perform noise removal in a learned and adaptive way as they are manual. Ties in the dataset added to the woes.

Imperfect feature extraction:
Handcrafted techniques like HOG, HSV, and LBP have limited capacity to represent and generalize complex and high-dimensional patterns inherent in human activities.

Thus, traditional ML techniques and non-deep learning-based feature engineering do not lead to high accuracy in this task. Existing analysis also showed a majority of DL techniques (like OpenPose, CNNs, and RNNs) being used to solve this problem. CNNs are capable of learning hierarchical representations of data and reducing noise inherently due to multiple layers.

Note: Even using CNNs we achieved an accuracy of almost 70%.

References

[1] Chathuramali, K.M. and Rodrigo, R., 2012, December. Faster human activity recognition with SVM. In International conference on advances in ICT for emerging regions (ICTer2012) (pp. 197-203). IEEE.

[2] Štulienė, A. and Paulauskaite-Taraseviciene, A., 2017. Research on human activity recognition based on image classification methods. Comput. Sci.

[3] Kumar, G. and Bhatia, P.K., 2014, February. A detailed review of feature extraction in image processing systems. In 2014 Fourth international conference on advanced computing communication technologies (pp. 5-12). IEEE.

[4] Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). Ieee.

[5] Snoun, A., Jlidi, N., Bouchrika, T., Jemai, O. and Zaied, M., 2021. Towards a deep human activity recognition approach based on video to image transformation with skeleton data. Multimedia Tools and Applications, 80, pp.29675-29698.

[6] Guo, Y., Zhang, Z. and Tang, F., 2021. Feature selection with kernelized multi-class support vector machine. Pattern Recognition, 117, p.107988.

[7] M. M. Hossain Shuvo, N. Ahmed, K. Nouduri and K. Palaniappan, "A Hybrid Approach for Human Activity Recognition with Support Vector Machine and 1D Convolutional Neural Network," 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington DC, DC, USA, 2020, pp. 1-5, doi: 10.1109/AIPR50011.2020.9425332.

[8] Althloothi, S., Mahoor, M.H., Zhang, X. and Voyles, R.M., 2014. Human activity recognition using multi-features and multiple kernel learning. Pattern recognition, 47(5), pp.1800-1812.

[9] Kästner, M., Strickert, M., Villmann, T. and Mittweida, S.G., 2013, April. A sparse kernelized matrix learning vector quantization model for human activity recognition. In ESANN.

[10] Horn, D., Demircioğlu, A., Bischl, B. et al. A comparative study on large scale kernelized support vector machines. Adv Data Anal Classif 12, 867–883 (2018). https://doi.org/10.1007/s11634-016-0265-7

[11] Vrigkas, M., Nikou, C. and Kakadiaris, I.A., 2015. A review of human activity recognition methods. Frontiers in Robotics and AI, 2, p.28.