Journal of Theoretical and Applied Information Technology

30th September 2023. Vol.101. No 18


ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# ACF-GSVM: CASCADE AGGREGATE CHANNEL FEATURE WITH GABOR FILTERS AND SUPPORT VECTOR MACHINE FOR ENHANCED FACE DETECTION

**AGIELA OMAR MOHAMED ALGASSIR[1], SELVAKUMAR MANICKAM[2], MOHAMMED ANBAR[3], ASAMA KUDER NSEAF[4]**

[(1,2&3)] National Advanced IPv6 Center, Universiti Sains Malaysia, Penang, 11800, Malaysia

[4] Faculty of Applied Sciences and Technology, Humber College Blvd, ON M9 W5L7, Canada

E-mail: [1]akheala22@gmail.com, [2]selva@usm.my, [3]anbar@usm.my, [4]asama.nseaf@humber.ca

**ABSTRACT**

Face detection has recently attracted a lot of attention because of the groundbreaking work of Viola and Jones. Despite the efforts to improve the performance using several subsequent enhancements, feature representation for face detection remains unsuitable for handling faces with a complicated appearance in an efficient and effective manner. In order to solve this dilemma, we look into the concept of face detection that elaborates the channel overview in a variety of ways, such as using gradient, oriented gradient, histograms, which are known to be efficient for decoding uncomplicated sound information. Consequently, the researchers suggest a novel hybridized and improved face detection method that addresses the problems associated with face detection schemes by utilizing an aggregate channel feature with Gabor filters and a support vector machine, called the ACF-GSVM model. The procedure for improving the face detection is based on a cascade of an aggregate channel feature with Gabor filters and a support vector machine, where the bounding boxes of the face region are initially detected with the use of aggregate channel features. The key decision is made by using Gabor filters and a support vector machine that identifies which bounding boxes belong to the face region. Experiments on the CASIA-FaceV5 database with the proposed ACF-GSVM face detection scheme showed that the accuracy rate it achieved in face detection was 96.6%. This study demonstrates the effectiveness of the proposed method.

**Keywords:** *Face Detection; ACF; Gabor Filters; SVM; CASIA-FaceV5; Decision Model*

## 1. INTRODUCTION

Because of the great advances in technology, computers have become more genuinely intelligent and are therefore intelligently capable of demonstrating great interaction with people by watching them with cameras, hearing them with microphones, and responding to them kindly. Face detection is one of the most popular forms of human–computer interaction (HCI).

The use of face detection for computer vision communication and access control systems is now of great interest to everyone. As a result of numerous images that seem to vary, including position (front, non-front), occlusion, image orientation, illumination, and facial appearance, face detection is not straightforward [5]. Therefore attention to this aforementioned problem is vitally important.

Over time, many studies have revealed that face detection constitutes a crucial issue in the field of computerized forms of recognition [15, 28, 36] and face clustering [36].

The goal of face detection is to discover if faces are shown in the image or not, as well as, if present, to determine the location and extent of every single face [47]. This makes use of the vision of computers, and is based on various factors such as scale, location, orientation, pose, facial expressions, light, conditions, and other aspects of appearance such as glasses, facial hair, and makeup.

The Viola–Jones (VJ) framework utilizes rectangular and Haar-like qualities and evaluates the hypothesis using the AdaBoost algorithm.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Over the past decade, these Viola–Jones (henceforth VJ) face detection frameworks have been regarded as the most influential [38]. The VJ detector accomplishes real-time face identification in conjunction with an attention cascade structure. Nevertheless, despite VJ's great success, because the face is an unlimited environment showing a great volume of appearance-oriented variations, its performance remains unsatisfactory.

Many VJ framework sub-sequences have been combined to handle faces in the wild, and more complex features [24], [30], [48] and/or more potent learning algorithms [18], [4], [46] are two key performance improvements achieved using these methods. Because the combination of amplification and cascade is considered effective in face detection, a dilemma remains in features representation, as complex features in the adoption of the aforementioned references provide limited performance gains at the expense of high computational costs.

A family of channel features has recently obtained recorded results in several areas of pedestrian detection [10], [9]. To extract features from the extended channels, the channel features first calculate the recorded maps of the original images such as gradients and oriented gradient, and histograms are then computed. The VJ presentation follows the classifier learning process.

This process is guided by the VJ framework pipeline; in this study, many channel features are named aggregate channel features, which are genuinely used by researchers [9]. These aggregate features are then obliquely taken out as pixel values from sub-sampled channels. While simple features can quickly compute guarantees, channel extension offers a broad representation capacity. As a result of these two benefits, aggregate channel features contribute to reducing the bottleneck in the VJ framework and advancing face detection significantly.

The scope of this research focuses on the aggregate channel features (ACF) using a Gabor filter and support vector machine (SVM) for face detection. The open dataset used in this research is the CASIA-Face-Image database, Version 5.0.

Owing to the breakthrough results achieved with them, aggregate channel features were used for face detection in this study. Given the wide range of sizes and placements of faces, the aggregate channel features approach can detect bounding boxes that may have both correct and incorrect bounding boxes of the face region in the image, by exhaustively searching for every conceivable position in the image. To address the issue of incorrect bounding boxes of the face region, the researchers propose an improved face detection method using a cascade of aggregate channel features and Gabor filters together with a support vector machine, the ACF-GSVM model, which locates the face position more precisely.

The following questions are addressed in this study: In the context of these tasks, what is the significance of studying face detection? In an image, how can a face be located? And how can we develop an enhanced face-detection method to achieve peak performance?

Our research work makes the following new contributions to face detection research:

- We develop a system based on machine learning for face detection;

- After detecting the bounding box in the image via aggregate channel features, which represents the initial face region, the features of that region are extracted using Gabor filters, and the support vector machine is then applied to determine the final face region.

The rest of this study is organized as follows. The second part highlights prior research on face detection methods as well as algorithms. Part 3 offers a thorough explanation of the proposed ACF-GSVM face-detection method. Part 4 provides an experimental examination of the proposed method, and part 5 presents the summary and conclusion.

## 2. LITERATURE REVIEW

Facial detection has attracted considerable attention in the initial stages of visual computing. Although several options have been put forth, face detection has seen major advances only in the past few decades, when Viola and Jones [43] proposed their milestone work. Three characteristics distinguish the VJ face detector: fast feature computing via integral picture representation, classifier learning using AdaBoost, and attentional cascade structure are all examples of methods used. The VJ framework has a significant disadvantage because the functionalities have a restricted representational capability, despite the fact that the lineaments pool size is too great to compensate. The amount of Haar-like features in a 24 × 24 detection window is typically 160,000 [43]. Efforts are being made to address this issue in two ways.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Some studies have concentrated on more complex features, such as HoG [48] and SURF [17]. Some attempts have been made to use heuristics to speed up the feature selection [29], [6]. However, this issue has not been completely resolved. Recently, the proposed channel features have demonstrated record performance in pedestrian detection [10], [9]. These features have a greater representational capacity for classification as a result of the channel expansion to other kinds, such as gradients and local histograms. These features, nonetheless, were extracted as rectangular sums at various sizes and places, potentially resulting in a redundant feature pool. Mathias et al. [51] discovered the effectiveness of integral channel features in the domain of face detection independently.

Using fully nonlinear activation networks, we created a face component score map and utilized it to produce face suggestions for further categorization. [16] provides a CNN cascade for accurate face detection. This study was developed further in [31] by collaborative training. [13] provided an end-to-end training version of the detection network that predicted the bounding box and object confidence directly. [11] shows that fine-tuning the CNN model from the ImageNet classification job for face/background classification produces good results. In [7], pose-invariant face detection was implemented using a supervised spatial transformation layer. Popular generic object identification algorithms, such as faster R-CNN [34], R-FCN [8], YOLO [33], and SSD [21], may also be used for face detection directly.

Mukherjee et al. (2017) [25] have described the formulation for both strategies, which include employing hand-crafted features and then training a simple classifier, as well as a completely new method for acquiring information utilizing neural networks. Real-time face detection and tracking have been demonstrated by Ren et al. (2017) [35]. The suggested approach combines tracking with Kalman filter tracking and the detection of convolutional neural networks. The face in a video was identified using a convolutional neural network, which is more precise than the standard detection technique. A face's position can be predicted by using Kalman filter tracking when it is significantly deflected or obscured. They are working to meet the real-time requirements while striving to improve the probability of face recognition. The computation, common state, and actions that alter that state are all represented as dataflow graphs. Luo et al. used facial recognition software to identify potential faces in images (2018) [22] and proposed a localization. Bounding-box regression is used frequently in a process referred to as deep consecutive detection. They used it to boost the overall performance by taking into account the natural correlation between classification and bounding-box regression. They used a three-stage, sequential structure of very deep convolutional networks to anticipate the presence of faces. The large-scale machine learning system TensorFlow can function in a wide range of environments. Dataflow graphs are used by TensorFlow to show the computation, common state, and actions that alter that state. Mapped computing resources include multicore CPUs, and specialized Tensor Processing Units (TPUs) that are ASICs, as well as numerous clustering machines and various computing resources inside a single device. Since shared state management was incorporated into earlier "parameter server" designs, this architecture gives the application developer flexibility. TensorFlow allows programmers to try out different training and optimization strategies. A wide range of uses are supported by TensorFlow, with a focus on deep neural network inference and training. TensorFlow was employed in Google's manufacturing. Abadi et al. in 2016 described the TensorFlow dataflow model and illustrated how well it performed in various practical contexts [1] (Figures 1, 2).

*Figure 1: A Selection of Faces.*

*Figure 2: Identified Faces.*

The proposed ACF-GSVM method for face detection in this study makes use of aggregate channel feature techniques along with Gabor filters and a support vector machine. In summary, the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

initial face region was found using an aggregate channel feature, while the final face region was found using Gabor filters and a support vector machine.

## 3. PROPOSED METHOD

### 3.1 Outline of the Proposed Method

The suggested ACF-GSVM-enhanced face-detection algorithm's structure is discussed in this section. The algorithm is divided into two sub-modules as illustrated in Figure 3: (i) an aggregate channel feature that determines the bounding box of the image's initial face coordinates, and (ii) Gabor filters and a support vector machine that anticipates and determines the face's final bounding box coordinates.

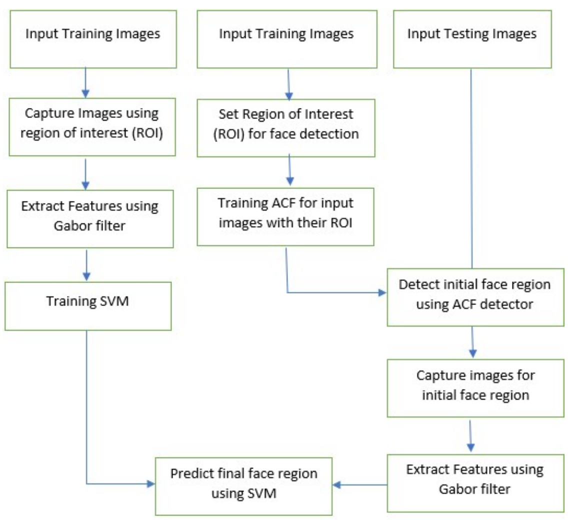



*Figure 3: The Framework of the Proposed ACF-GSVM-Enhanced Face Detection Method.*

### 3.2 Initial Face Area Detection Using an Aggregate Channel Feature

Face regions were detected using images from CASIA-FaceV5 as inputs for the aggregate channel features (ACF) [9]. Figure 4 depicts the aggregate channel features structure used in this example.

This section provides a thorough investigation of the channel features using the same context of face recognition. The channel functionality is described in detail, including its calculation, qualities, and advantages over the standard Haar-like features used in the VJ framework. The ACF detector is proposed in [49] as a combined feature made up of three channels in the LUV color space, a normalized gradient channel, and a six-channel oriented gradient histogram (HOG), which are then set in a boosted tree.

The basis area for the ROI was an input picture with dimensions of 2352 × 1728 pixels (width × height), which made it possible to identify faces more effectively. The ROI was established using the training information collected during the study, taking into account different and variable positions because the image is of a person at a distance. Figure 5(a) shows several actual pictures from the training dataset, and Figure 5(b) shows the corresponding labels. The ROI or the ground truth is shown by the red bounding boxes that serve as annotations.

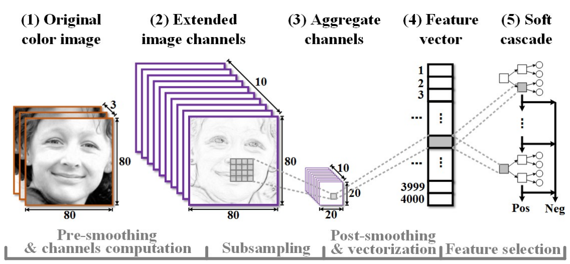



*Figure 4: Structure of Aggregate Channel Features (ACF).*

#### 3.2.1 Structure of ACF

**Channel extension:** A channel is the fundamental structure of the total channel features. Since the emergence of digital images, this channel has been in widespread usage. The color channel should be the most apparent sort of channel in a picture, followed by grayscale and RGB. Numerous other channel types in addition to color channels have been developed to encode various types of data for more challenging issues. For increasingly complicated tasks, other sorts of information are available. Channels are defined as a registered map of the original picture using pixels generated from the matching areas of the original pixels [10]. The novel image can be transformed linearly or nonlinearly to compute the different channels. The transformations must adhere to a limitation that renders them internationally invariant in order to allow sliding-window recognition.

**Feature computation:** The calculation of the aggregate channel features is fairly simple, as stated in the definition of the channel. All defined channels and sub-sampling are carried out in accordance with a predetermined factor in the color image shown in Figure 4. The aggregated pixels from all the sub-sampled channels are vectorized and stored in a pixel look-up table. It should be

Journal of Theoretical and Applied Information Technology
30th September 2023. Vol.101. No 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

noted that an optional smoothing process involving a binary filter was utilized both before and after sub-sampling.

*(a)*

*(b)*

*Figure 5: Illustrations of ROI in Images Input to Aggregate Channel Features (ACF).*

**Classifier learning:** Learning is a straightforward process. Two changes were made in comparison to the VJ framework. Instead of a decision stump, the weak classifier is now a depth-2 decision tree, which is the first modification. When looking for distinctive channel correlations for classification, the more sophisticated weak classifier performed better than the simpler weak classifier [19]. The adoption of a soft-cascade construction is the second distinctive feature [4]. In contrast to the attentional sequential structure employed in the VJ framework, which contains several sequence stages, only one classifier is trained on all the training data. AdaBoost then selects the weakest classifiers, and the threshold is established. Training and detection are more effective as a result of these two changes.

**Overall superiority:** Compared to the conventional aggregate channel features, which are Haar-like key features utilized in the VJ framework, the following benefits and drawbacks are offered. 1) Several types of visual channels are added to store more details, including colors, gradation, and local graphs, providing richer representation capabilities. 2) Rather than using integral images to create rectangular sums with downscaled networks, different position and scale properties are directly retrieved as pixel values. In order to increase the learning, this results in a smaller feature pool size and faster feature calculation. Using the cascade structure, the detection speed was increased even more. 3) Because its structure is compatible with the entire image, the boosted classifier effectively stores hierarchical patterns from huge sets of training information to give a more exact localization of faces in the image.

### 3.3 Using the Decision Model to Detect the Final Face Region

Figure 6 shows how the ACF determines the initial bounding box for the face, which can then be used by the decision model to predict the final face region. To improve the performance of face detection schemes, we used Gabor filters and support vector machines in our decision model, based on reflections. In the ACF module, all areas of the image are taken into account, including irrelevant areas, for precisely locating the face. The ACF is then used in the decision model to determine the final face region based on the bounding box of the initial face.

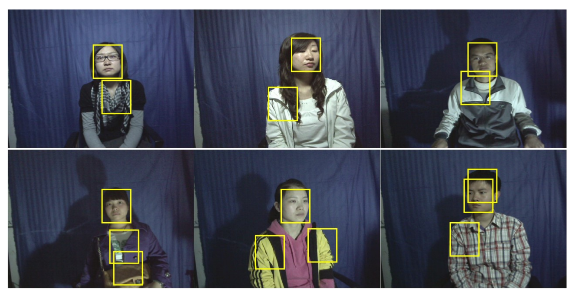


*Figure 6: The Initial Bounding Box Captured.*

Gabor filters were used to extract the characteristics of a face captured using the initial bounding box [12]. The primary benefits of Gabor filters are their variability in translation, magnification, and rotation. Aside from being resistant to noise and changing lighting, they are also resistant to photometric turbulence [14], [20], [23], [37].

The Gabor filter properties were deleted from the gray-level images instantly. In a three-dimensional area, a two-dimensional Gabor filter is a Gaussian kernel function guided by a plane polyhedron sine wave, which may be explained as follows:

$$G(x,y)=\frac{f^2}{\pi\gamma\eta}\exp\left(-\frac{x^2+\gamma^2y^2}{2\sigma^2}\right)\exp\left(j2\pi fx'+\phi\right)$$
$$x'=x\cos\theta+y\sin\theta$$
$$y'=-x\sin\theta+y\cos\theta \qquad (1)$$

where $f$ is the sinusoidal wave occurrence, θ is the standard Gabor function's orientation, the stage of counterbalance is ϕ, the standard deviation of the Gaussian envelope is σ, and γ is the three-dimensional ratio of the aspects that identify the Gabor function support elasticity.

Figure 7 shows 40 Gabor filters in eight distinct orientations and five different scales. The main image of the face area used in this experiment was 100 × 100 pixels in size, and the feature vector had dimensions of 100 × 100 × 40 = 400,000 pixels when using the 40 Gabor filters. Down-sampling the feature image produced by Gabor filters can reduce the redundancy of information because the pixels adjacent to an image are frequently closely linked [20], [37]. The feature vector will ultimately be 1960 bytes in size because of the 8-fold down-sampling that is used. Subsequently, the vectors and unit modifications were adjusted to have a mean of zero.

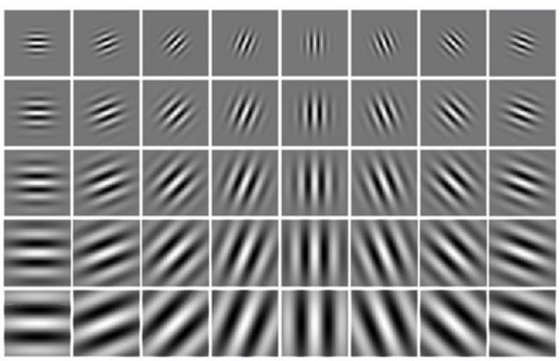


*Figure 7: Gabor Wavelets in Five Different Sizes and Eight Orientations.*

A support model machine was used to forecast the final face area using features taken from the original face area. Vapnik developed the support vector machine (SVM) algorithm in 1963 [41].

Support vector machines (SVMs) are a set of supervised learning methods. SVM is significantly better than other ML techniques in identifying patterns that are specific in large sets of data [3]. SVMs may be used to recognize faces, identify false credit cards, identify speakers, and discern readable handwriting [26]. SVMs are a widely used method for creating labels. This aims to establish a decision boundary between two classes, and such labels can be based on one or more feature vectors [27]. As far as is physically possible, the hyperplane, a decision boundary, is far from the class's nearest data points. The nearest data points are referred to as basic functions in the presence of a labeled training dataset.

$$(x_1, y_1), \ldots, (x_n, y_n), x_i \in R^d \text{ and } y_i \in (-1, +1) \quad (2)$$

where $x_1$ is a feature vector representation and $y_1$ is the positive or negative class label of training compound I, and the ideal hyperplane is provided by

$$wx^T + b = 0 \quad (3)$$

where $w$ denotes the weight vector, $x$ is the input feature vector, and $b$ is the bias.

For all elements of the training set, $w$ and $b$ should satisfy the following inequality:

$$wx_i^T + b \geq +1 \quad if \quad y_i = 1 \quad (4)$$

$$wx_i^T + b \leq -1 \quad if \quad y_i = -1 \quad (5)$$

The objective of training an SVM model is to find $w$ and $b$ such that the hyperplane separates the data and maximizes the margin $1/\|w\|^2$.

Vectors $x_i$ for which $|y_i|(wx_i^T + b) = 1$ will be termed support vectors (Figure 8).

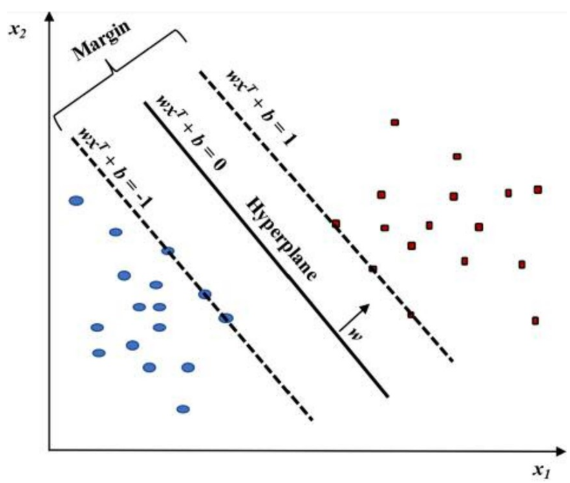



*Figure 8: A Regular SVM Model.*

There are two distinct classes (red and blue). The kernel approach, which is a different use for SVM, enables the modeling of higher-dimensional, nonlinear models [2]. In a nonlinear mode, a kernel function can be employed to boost the number of dimensions in the raw data, resulting in a more advanced spatial dimension (Figure 9). In short, the kernel's function can speed up some computations that would need to be done in a three-dimensional space.

This is outlined as follows:

$$K(x, y) = \langle f(x), f(y) \rangle \quad (6)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

where $K$ is the kernel function, while $x$ and $y$ are $n$-dimensional inputs, and $f$ is utilized to transform the input from an $n$-dimensional space to an $m$-dimensional space. $x$, $y$> indicates the linear combination. Using kernel functions, the scalar product may be determined between two data points in a higher-dimensional space without explicitly computing the mapping from the input space to the higher-dimensional space. Computing the kernel is often straightforward, but computing the inner combination of two feature vectors in a high-dimensional space is complex. Even simple kernels can have huge feature vectors, and the corresponding feature vector for kernels such as the radial basis function (RBF) kernel (KRBF(x, y) = exp (-γ||x - y|| 2) is infinite-dimensional.

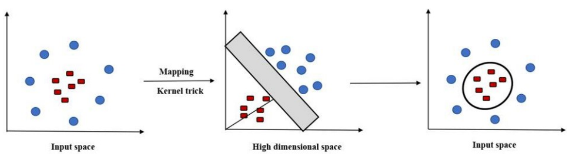



*Figure 9: Kernel Function. Data that Cannot be Transformed or Separated Using a Linear SVM Can be Transformed and Separated Using a Kernel Function.*

The SVM is used to generate values for the face region and other regions, which represent two classes. Following the grading of these two values, we determine the maximum value indicating that the support vector machine (SVM) mode is used for determination (prediction).

## 4. RESULTS AND ANALYSIS OF THE EXPERIMENT

This section provides the empirical findings relating to the proposed ACF-GSVM model for enhanced face detection in terms of accuracy and efficiency. These investigations used an open dataset (CASIA-Face-Image database, Version 5.0) to detect faces in 2,500 UV images of 500 participants with intra-class changes in expression, illumination, imaging distance, glasses, position, and so on [52]. Every face image is a 16-bit color BMP file with 640 × 480 resolution. CASIA-FaceV5 volunteers include graduate students, workers, waiters, and others. Figure 10 shows examples of these images. The images were separated into two data sets for training and testing purposes. The training dataset has 127 images, while the test dataset has the remaining images.

*Figure 10: Examples of Images from the Training Dataset.*

A personal computer with 20 GB of RAM, an NVIDIA GeForce GT 430 1024 MB GPU, and an Intel I5-2400 CPU running at 3.10 GHz were used for the training. On 300 images, all of which were taken from the same training images on an aggregate channel feature with a resolution of 100 × 100 pixels, the classification model was trained (using a Gabor filter and then a support vector machine). Figure 11 shows some samples of these images. The 300 images were separated into two groups: 100 for the "face region" and 200 for the "other region".

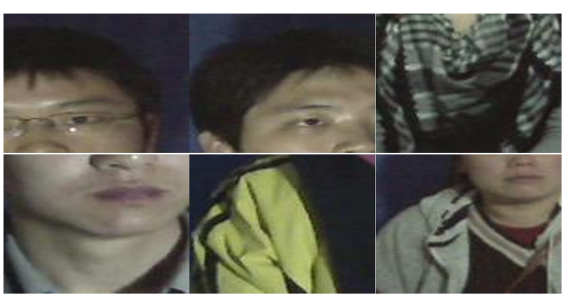
*(a)*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*(b)*

*Figure 11: Images from the Decision Model's Training Dataset Contain (a) Images of the "Face Area" and (b) Images of the "Other Areas".*

After training, our suggested ACF-GSVM model was utilized to detect faces in the test dataset. The accuracy levels of the two models (both before and after using our improved face detection algorithm) were measured throughout the testing phase. The performances of the two models were evaluated both before (with the aggregate channel feature model) and after (with the aggregate channel feature with Gabor filters and support vector machine (ACF-GSVM) model) enhanced face detection. These networks were compared and their accuracy was evaluated for the test dataset. Figure 12(a) displays several incorrect detections using an aggregate channel feature, whereas Figure 12(b) displays the matching correct detection using the ACF-GSVM model.



*(a)*

*(b)*

*Figure 12: Examples of Face Detection Using (a) ACF and (b) the Suggested ACF-GSVM Method.*

To evaluate the object detection quality, the accuracy was analyzed. The accuracy results of face recognition on the CASIA-FaceV5 images are shown in Table 1.

*Table 1: Comparison of the Suggested ACF-GSVM Face Accuracy detection with a State-of-the-Art Approach Using the CASIA-FaceV5 Database.*

| Method | Accuracy Rate |
|---|---|
| Our implemented Viola–Jones algorithm [42] | 68.8% |
| Our implemented ACF [49] | 23% |
| Proposed ACF-GSVM | 96.6% |

Table 1 shows the findings of the proposed ACF-GSVM model compared with those of state-of-the-art algorithms, aggregate channel features (ACF) [49], and the Viola–Jones algorithms [42].

The model of the Viola–Jones algorithm [49] is seen to have a face detection accuracy 45.8% higher than our implemented ACF [42].

Our proposed ACF-GSVM model, however, is more accurate by 27.8% and 73.6% than the Viola–Jones algorithm [42] and ACF [49], respectively.

The ACF-GSVM model is significantly more efficient than existing models of face detection because it uses an aggregate channel feature to determine the basic bounding boxes for the facial image. A determination model of Gabor filters and a support vector machine was used to determine and locate the correct face regions in order to achieve the objectives of the study.

## 5. CONCLUSION

To increase the accuracy of face detection under the aforementioned conditions, an aggregate channel feature with Gabor filters and a support vector machine (ACF-GSVM) model was used. In this study, the improved face detection model was found to be resistant to pose variations, illumination

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

variations, facial expressions, aging conditions, imaging distance, glasses, etc. This proposed method attempts to detect the faces and must be able to predict the face region instantly. Our new contributions have used the aggregate channel feature with Gabor filters and a support vector machine (ACF-GSVM) model to increase the accuracy of face detection under the aforementioned conditions. It should be noted that the proposed method first detects the face region as an area of interest and then trains the ACF to detect and capture the initial face region using this region of interest. The results are then passed through a Gabor filter to extract features before being input into a support vector machine model. We proved the proposed model's high detection rate and superior performance in contrast with earlier detection methods using experiments with the CASIA-FaceV5 database. MATLAB 2019a was used for evaluations, and the presented approach performed well. These experiments demonstrated that the proposed improved model can be successfully used for face detection. To predict the correct face region, the proposed method uses a decision model with an object identification mechanism, namely a Gabor filter as well as a support vector machine.

**REFERENCES:**

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, et al., "Tensorflow: A System for Large-Scale Machine Learning", Paper presented at the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 2016.

[2] M.A. Aizerman, "Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning", *Automation and Remote Control*, Vol. 25, 1964, pp. 821-37.

[3] S. Aruna and S.P. Rajagopalan, "A Novel SVM Based CSSFFS Feature Selection Algorithm for Detecting Breast Cancer", *International Journal of Computer Applications*, Vol. 31, No. 8, 2011, pp. 14-20.

[4] L. Bourdev and J. Brandt, "Robust Object Detection via Soft Cascade", Paper presented at the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005.

[5] P. Brimblecombe, "Face Detection Using Neural Networks", Dissertation, H615–Meng Electronic Engineering, School of Electronics and Physical Sciences, University of Surrey, 2002.

[6] S.C. Brubaker, J. Wu, J. Sun, M.D. Mullin, and J.M. Rehg, "On the Design of Cascades of Boosted Ensembles for Face Detection", *International Journal of Computer Vision*, Vol. 77, No. 1, 2008, pp. 65-86.

[7] D. Chen, G. Hua, F. Wen, and J. Sun, "Supervised Transformer Network for Efficient Face Detection", Paper presented at the European Conference on Computer Vision, 2016.

[8] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object Detection Via Region-Based Fully Convolutional Networks", *Advances in Neural Information Processing Systems*, Vol. 29, 2016.

[9] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast Feature Pyramids for Object Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 8, 2014, pp. 1532-45.

[10] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral Channel Features", 2009.

[11] S.S. Farfade, M.J. Saberian, and L.-J. Li, "Multi-View Face Detection Using Deep Convolutional Neural Networks", Paper presented at the 5th ACM International Conference on Multimedia Retrieval, 2015.

[12] M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb, "CloudID: Trustworthy Cloud-Based and Cross-Enterprise Biometric Identification", *Expert Systems with Applications*, Vol. 42, No. 21, 2015, pp. 7905-16.

[13] L. Huang, Y. Yang, Y. Deng, and Y. Yu, "Densebox: Unifying Landmark Localization with End to End Object Detection", arXiv preprint arXiv:1509.04874, 2015.

[14] Kamarainen, J-K, Ville Kyrki, and Heikki Kalviainen. "Invariance Properties of Gabor Filter-Based Features-Overview and Applications." IEEE Transactions on image processing 15, no. 5, 2006, pp. 1088-99.

[15] Kumar, K Susheel, Vijay Bhaskar Semwal, and Ramesh Chandra Tripathi. "Real Time Face Recognition Using Adaboost Improved Fast Pca Algorithm." arXiv preprint arXiv:1108.1353, 2011.

[16] Li, Haoxiang, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. "A Convolutional Neural Network Cascade for Face Detection." Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[17] Li, Jianguo, and Yimin Zhang. "Learning Surf Cascade for Fast and Accurate Object Detection." Paper presented at the Proceedings

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of the IEEE conference on computer vision and pattern recognition, 2013.

[18] Li, Stan Z, Long Zhu, ZhenQiu Zhang, Andrew Blake, HongJiang Zhang, and Harry Shum. "Statistical Learning of Multi-View Face Detection." Paper presented at the European Conference on Computer Vision, 2002.

[19] Liao, Shengcai, Anil K Jain, and Stan Z Li. "Unconstrained Face Detection." Technical report, 2012.

[20] Liu, Chengjun, and Harry Wechsler. "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition." IEEE Transactions on Image processing 11, no. 4, 2002, pp. 467-76.

[21] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. "Ssd: Single Shot Multibox Detector." Paper presented at the European conference on computer vision, 2016.

[22] Luo, Dazhi, Guihua Wen, Danyang Li, Yang Hu, and Eryang Huan. "Deep-Learning-Based Face Detection Using Iterative Bounding-Box Regression." Multimedia Tools and Applications 77, no. 19, 2018, pp. 24663-80.

[23] Meshgini, Saeed, Ali Aghagolzadeh, and Hadi Seyedarabi. "Face Recognition Using Gabor-Based Direct Linear Discriminant Analysis and Support Vector Machine." Computers & Electrical Engineering 39, no. 3, 2013, pp. 727-45.

[24] Mita, Takeshi, Toshimitsu Kaneko, and Osamu Hori. "Joint Haar-Like Features for Face Detection." Paper presented at the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 2005.

[25] Mukherjee, Subham, Sumalya Saha, Sounak Lahiri, Ayan Das, Ayan Kumar Bhunia, Aishik Konwer, and Arindam Chakraborty. "Convolutional Neural Network Based Face Detection." Paper presented at the 2017 1st International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech), 2017.

[26] Noble, William Stafford. "Support Vector Machine Applications in Computational Biology." Kernel methods in computational biology 71, 2004.

[27] Noble, William S. "What Is a Support Vector Machine?". Nature biotechnology 24, no. 12, 2006, pp. 1565-67.

[28] Parkhi, Omkar M, Andrea Vedaldi, and Andrew Zisserman. "Deep Face Recognition." , 2015.

[29] Pham, Minh-Tri, and Tat-Jen Cham. "Fast Training and Selection of Haar Features Using Statistics in Boosting-Based Face Detection." Paper presented at the 2007 IEEE 11th International Conference on Computer Vision, 2007.

[30] Pham, Minh-Tri, Yang Gao, Viet-Dung D Hoang, and Tat-Jen Cham. "Fast Polygonal Integration and Its Application in Extending Haar-Like Features to Improve Object Detection." Paper presented at the 2010 IEEE computer society conference on computer vision and pattern recognition, 2010.

[31] Qin, Hongwei, Junjie Yan, Xiu Li, and Xiaolin Hu. "Joint Training of Cascaded Cnn for Face Detection." Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[32] Ranjan, Rajeev, Vishal M Patel, and Rama Chellappa. "A Deep Pyramid Deformable Part Model for Face Detection." Paper presented at the 2015 IEEE 7th international conference on biometrics theory, applications and systems (BTAS), 2015.

[33] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection." Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[34] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks." Advances in neural information processing systems 28, 2015.

[35] Ren, Zihan, Shuangyuan Yang, Fuhao Zou, Fan Yang, Chaoyang Luan, and Kai Li. "A Face Tracking Framework Based on Convolutional Neural Networks and Kalman Filter." Paper presented at the 2017 8th IEEE international conference on software engineering and service science (ICSESS), 2017.

[36] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A Unified Embedding for Face Recognition and Clustering." Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[37] Shen, LinLin, Li Bai, and Michael Fairhurst. "Gabor Wavelets and General Discriminant Analysis for Face Identification and Verification." Image and Vision Computing 25, no. 5, 2007, pp. 553-63.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[38] Shen, Xiaohui, Zhe Lin, Jonathan Brandt, and Ying Wu. "Detecting and Aligning Faces by Image Retrieval." Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013.

[39] Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deep Learning Face Representation from Predicting 10,000 Classes." Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, 2014.

[40] Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. "Deepface: Closing the Gap to Human-Level Performance in Face Verification." Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, 2014.

[41] Vapnik, Vladimir. "Pattern Recognition Using Generalized Portrait Method." Automation and remote control 24, 1963, pp. 774-80.

[42] Vikram, K, and S Padmavathi. "Facial Parts Detection Using Viola Jones Algorithm." Paper presented at the 2017 4th international conference on advanced computing and communication systems (ICACCS), 2017.

[43] Viola, Paul, and Michael J Jones. "Robust Real-Time Face Detection." International journal of computer vision 57, no. 2, 2004, pp. 137-54.

[44] Yang, Bin, Junjie Yan, Zhen Lei, and Stan Z Li. "Convolutional Channel Features." Paper presented at the Proceedings of the IEEE international conference on computer vision, 2015.

[45] Yang, Shuo, Ping Luo, Chen-Change Loy, and Xiaoou Tang. "From Facial Parts Responses to Face Detection: A Deep Learning Approach." Paper presented at the Proceedings of the IEEE international conference on computer vision, 2015.

[46] Zhang, Cha, and Paul Viola. "Multiple-Instance Pruning for Learning Efficient Cascade Detectors." Advances in neural information processing systems 20, 2007.

[47] Zhang, Cha, and Zhengyou Zhang. "A Survey of Recent Advances in Face Detection.", 2010.

[48] Zhu, Qiang, Mei-Chen Yeh, Kwang-Ting Cheng, and Shai Avidan. "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients." Paper presented at the 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), 2006.

[49] Dollár, Piotr, Ron Appel, Serge Belongie, and Pietro Perona. "Fast Feature Pyramids for Object Detection." IEEE transactions on pattern analysis and machine intelligence 36, no. 8, 2014, pp. 1532-45.

[50] Yang, Bin, Junjie Yan, Zhen Lei, and Stan Z Li. "Convolutional Channel Features." Paper presented at the Proceedings of the IEEE international conference on computer vision, 2015.

[51] Mathias, Markus, Rodrigo Benenson, Marco Pedersoli, and Luc Van Gool. "Face Detection without Bells and Whistles." Paper presented at the European conference on computer vision, 2014.

[52] CASIA Face Image Database. [Online]. Available: http://english.ia.cas.cn/db/201610/t20161026_169405.html, 1/06/2023.