*International Journal of Engineering & Technology*, 8 (1.2) (2019) 186-193

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

*Research paper*

# Visual Object Categorization based on Gabor Filter Generalization Via K-Means Clustering

**Hayder Ayad[1], Siti Norul Huda Sheikh Abdullah [2], Azizi Abdullah[3]**

*[1,3] Centre for Artificial Intelligence and Technology (CAIT)*
*[2]Digital Forensic Research Group, Center for Cyber Security*
*Faculty of Information Science and Technology,*
*Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia*
**Corresponding author E-mail: hayder.alsultany@ieee.org*

## Abstract

Content based object recognition systems need informative image properties to obtain good performance results. Filter bank such as Gabor filters is believed to be one of the most popular methods for complete characterization of images by having some important properties such as selectivity to orientation, scale, frequency and smooth parameters. Furthermore, such properties are very effective for compact image description and analysis. However, these functions show a strong dependence on a certain number of different parameter values. Hence, the different filter parameters values used to construct the functions may give different filter responds or properties. Besides, the large number of these filters leads to expensive computation to create maps for feature extraction, thus it is necessary to reduce the number of candidates and identify subset of effective and discriminative filters to avoid overfitting and hinder generalization performance. In this paper, we first compute Gabor filters using a set of different values for filter parameters. After that, the k-means clustering algorithm is used to group these filter responds into k different clusters. Next the k different clusters are used to convolve images and the edge histogram then apply to the filter outputs for image description. After that, we combine all outputs of image descriptors using SVMs. Experiment results on 20 and 101 classes of the Caltect-101 object database show that the method significantly outperforms using the standard Gabor filter approach.

***Keywords***: *gabor filter, naïve combination approach, SVM classifier, VOC technique.*

## 1. Introduction

In new technology, image investigation in several areas [1, 2] including computer vision and its diverse techniques (i.e., visual object categorization (VOC) and content-based image retrieval (CBIR)) make an alert to manage these techniques. The VOC technique is the main focus of current research and remains an open research area. The main problem for this research is the categorization problem (semantic gap). Based on state-of-the-art-research, the VOC technique commonly uses the texture feature, as it provides meaningful information that can achieve high-level features and improve the categorization problem [1, 3, 4, 5, 6].

However, much of the research in the area has used a filter to describe the object [4, 7, 8]. Based on the literature, filters can be divided into two kinds: linear filters, such as the Sobel detector, and non-linear filters (filter banks), such as the Gabor filter (GF). The former uses a single orientation and fixed mask size when constructing feature maps, which may inaccurately capture the important salient features, particularly when applying orientation descriptors. Furthermore, it is sensitive to orientation. It thus tends to respond to non-edge orientation, while the most discriminative edge filters respond to edges in a narrow range orientation [9].

The second one (non-linear filters) can efficiently overcome the limitations of the former in capturing the most discriminative edges and salient features based on a different scale, orientation and filter size [10]. Using an orientation descriptor based on this type of filter can provide meaningful and distinctive features for correctly describing an object [1, 4]. More advanced techniques have frequently used the GF to describe objects. [7] showed the effectiveness of the GF when it is used in object detection, recognition and tracking. [11] proposed using the GF with an SVM classifier for object categorization, and their experiment showed that the proposed method performed well. Many researchers also used the GF for 3-D object representation and recognition [12]. These studies all showed the benefit of using the GF to detect the salient feature that can improve system performance.

Conversely, many descriptors have been introduced to extract texture features. To be selected, a descriptor should have a certain ability in human vision to produce the most discriminative and important features. Constructing edge features is one of the primary features of the human vision system. It is thus possible to describe objects efficiently using the edge orientation information. Orientation descriptors, including the edge histogram, have become popular and widely used in object categorization systems [3].

The literature presents the main issue of using the GF and orientation descriptors, including the edge histogram descriptor (EHD), to describe objects. Clearly, [3] used the EHD to extract the texture feature to describe objects. This descriptor used a single orientation to extract the texture feature, which may not be sufficient in describing objects. Applying an advanced filter, such as the GF, to construct diverse feature maps based on a different scale, orientation and filter size will allow this descriptor produce excellent results and a distinctive texture feature. Conversely, the main drawback of the GF is that it constructs redundant and incompact filters that may affect system recognition accuracy and decrease system performance [10, 13]. [10] proposed a principle component analysis (PCA) to reduce the GF dimensionality and produced optimal filters for a face detection technique. The Fast Fourier Transform algorithm (FFT) has also been used to reduce the GF dimensionality in the frequency domain .

In this work, we present a novel filter-based descriptor and classifier denoted by the generalized Gabor filter (GGF) to improve the categorization problem. The GF parameters are selected. We generalize the GF method using an unsupervised machine learning algorithm, denoted by the k-means clustering technique. This step ensures the removal of redundant filters, which eventually leads to a more compact and optimized filter set. We also use an orientation-based descriptor, denoted by edge histograms, to extract the texture feature. We have used single classifier and combination features, denoted by a naïve approach, to construct feature vectors and classify them using a support vector machine (SVM). To evaluate the performance of the proposed GGF method, we implement the VOC technique framework based on the proposed method in the spatial domain for object categorization and compare its result to the standard GF method. Furthermore, we evaluate our proposed method using the first 20 classes and all classes from the Caltech 101 dataset. We extend the experiment by implementing the VOC technique framework based on the GF and proposed GGF methods in the frequency domain and compare it to the proposed VOC technique in the spatial domain.

The rest of this paper is organized as follows. In Section 2, we briefly review the state-of-the-art methods used in the VOC technique. In Section 3, we review the clustering and k-means clustering techniques. In Section 4, we present our novel filter-based descriptor and classifier, denoted by the GGF method. In Section 5, we discuss our experimental results, including performance evaluation and benchmarking. Finally, we summarize our findings in section 6.

## 2. Related work

First, we briefly review the GF method computation. We next present an orientation descriptor, denoted by an edge histogram descriptor (EHD). We then discuss the naïve approach and support vector machine (SVM) methods.

### 2.1. Gabor filter (GF)

The The filter banks are ubiquitous in extracting feature textures. The responses of these filters are mainly based on the distribution, joint filter response distribution and the scale and orientation used to create these filters. These filter banks can be used for segmentation, classification and synthesis [10, 14].

Recent texture classification has widely used the filter banks. The biological plausibility and the hypothesis in various scale and orientation must extract sufficient features and perform accurate classification [15]. Examples of filter banks include the Gabor, Structure Tensor and Steerable Filters, among others [8].

The GF has become important in several applications due to its characteristics, such as providing classification schema and a multi-resolution schema for feature texture. This filter also supplies meaningful information by using diverse magnitude and orientation. Furthermore, several salient features can be captured using this filter, including spatial frequency characteristics, spatial locations and orientation selections. The mathematical equations below describe the GF [10]:

$$\psi k(z) = \frac{\|k\|^2}{\sigma^2} e^{-\|k|^2 z|^2/2\sigma^2}\left[e^{ikz} - e^{\sigma^2/2}\right] \quad (1)$$

The parameters are defined thus: $k$ refers to the wavelength and orientation of kernel $\psi k(z)$ in the image. This equation has two parts: the oscillation refers to the term in brackets, and the dc comprises is the reset. σ refers to the standard deviation of the Gaussian function. $(k)$ is defined thus:

$$k(\mu\,,\upsilon\,) = K_{u^{e^{i\emptyset\mu}}}\,, \quad (2)$$

where parameters μ and υ refer to the orientation and scaling of the Gabor kernel, respectively.

$$k_u = k_{max}/f\upsilon \text{and} \emptyset_\mu = \pi\mu/8 \quad (3)$$

Here, parameter $k_{max}$ refers to the maximum frequency and $f$ refers to the space factor between kernels in the frequency domain.

The GF convolves the image before constructing the image primitives. If $I(x)$ represents the grey level of the image and $\psi k(x)$ the GF kernels, the following equation can yield the image primitive:

$$\boldsymbol{O}k(x) = I(x) * \psi k(x), \quad (4)$$

where $\boldsymbol{O}k(x)$ refers to the convolution result at $k$ and parameter

* refers to the convolution operator. Fig. 1 shows an example of the cosine part (even) of GF.

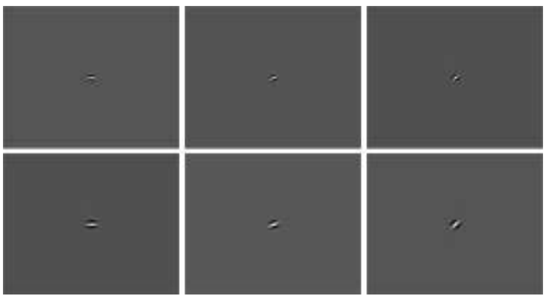
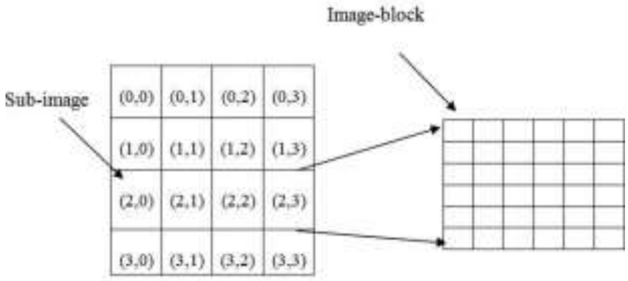
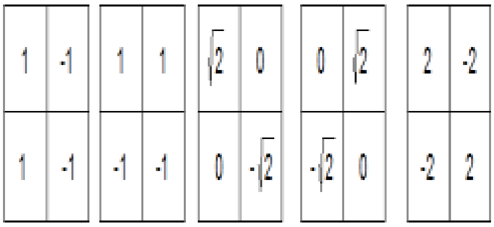


**Fig. 1:** The GFs with parameters values ($\boldsymbol{\upsilon}$ = 2, $\boldsymbol{\mu}$ = 3, σ = 2.5, and filter size= $\mathbf{128} \times \mathbf{128}$).




(a)




(b)

**Fig. 2:** (a) an image divided into sub-images and image-blocks. (b) Edge histogram masks for texture feature extraction: (i) ver_edge_filter, (ii) hor_edge_filter, (iii) dia45_edge_filter, (iv) dia135_edge_filter, (v) nond_edge_filter.

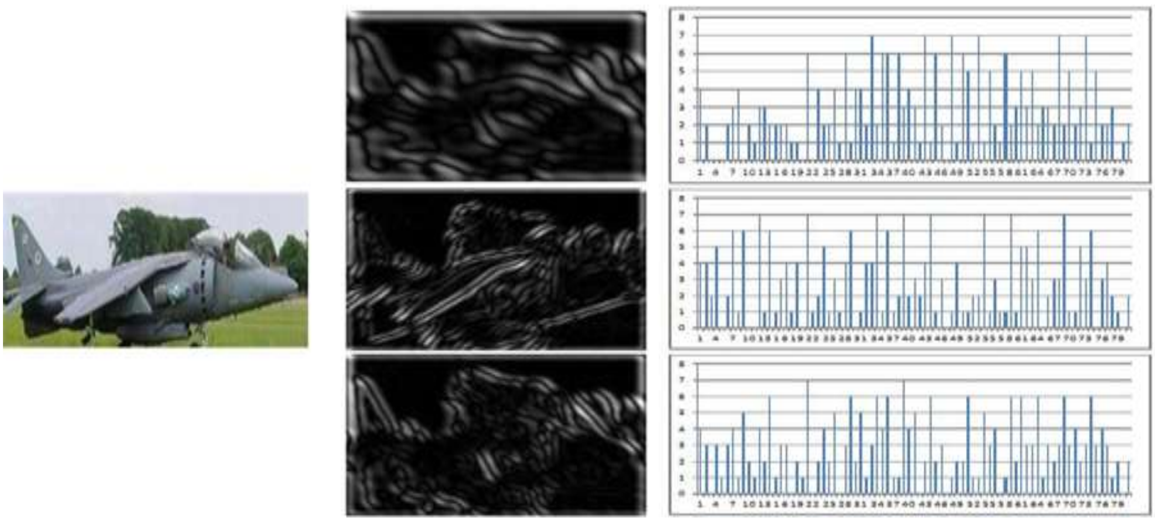
## 2.2. Edge histogram

The edge feature is an important feature for describing image content, especially considering the object shape and texture [16]. We thus use the edge histogram, which represents the edge distribution in an image. At the same time, the edge histogram represents the frequency and directionality of an edge in an image. It also captures the spatial distribution of the edge (four directional edges and one non-directional edge) and describes and extracts the non-homogenous texture [17]. Figure 2 part (A) illustrates sub-images and blocks inside the sub-images.

Extracting the edge histogram first requires partitioning a specified image into 4x4 non-overlapping blocks (sub-images). Each sub-image has a group of image-blocks. The ED method then applies five different edge detectors to perform texture extraction in different directions, including horizontal, vertical, diagonal 45, diagonal 135 and non-directional [17]. In each sub-image, the image blocks are convolved with the filters coefficient, which represents various edge detectors, as in Figure 2 part (B). Only the maximum value among these edge strengths is compared with the edge threshold: if it is greater than the threshold, the image block considers having the corresponding edges. Finally, the edge histogram produces 80 bins of edge features that represent the image [17]. The following equation shows how to obtain the edge strength among different 5 five filters.

$$edg - stg(i,j) = \left|\sum_{k=0}^{n} A_k(i,j) * edge - filter(k)\right| \quad (5)$$

Here, **n** refers to the sub-block number in each image block, $\mathbf{A_k(i,j)}$ refers to the image block, and $\mathbf{edg - stg(i,j)}$ indicates any of the five edge detectors. However, Fig. 3 shows the spatial distribution of the edge histogram.



(a) (b) (c)

**Fig. 3.** The edge histogram applied to real-world images, (a) original airplane class image from the Caltech 101 dataset, (b) image primitives (feature maps) constructed by convolving the original image with the GF, (c) spatial distribution of the edge histogram.

### 2.3. Naïve approach

Combining features outperforms using a single descriptor when representing an image in visual object categorization (VOC) [3, 18]. The naive approach considers a combination method, where it combines several feature vectors in a feature vector and feeds them to the classifier. Similarly, it places several feature vectors produced by various sources directly into an input vector and enters it into the classifier for learning and testing. However, the feature vector produced by the naive approach leads to large dimensional feature vectors. Nonetheless, the naive approach outperforms a single feature vector; however, using the naive approach with combined features increases dimensional feature vectors, produces over-fitting and affects performance [3, 19].

### 2.4. Support vector machine (SVM)

Recent research has shown that the SVM algorithm outperforms state-of-the-art techniques in different applications. The main point of this algorithm is the optimal hyperplane that separates the given data into two classes (categories) {+1, -1} with the maximal margin in the higher dimensional feature space. SVM aims to construct a model based on the given training data to predict target values of the test data attributes [20].

For the classification schema, if X is the input value, SVM will classify this input into two classesy∈{-1,+1}, according to the following formula:

$$y = sign(f(x)) \tag{6}$$

Function f(x) comes from combining the Kernel function between input X and each training data point thus:

$$f(x) = \sum \alpha_m y_m K(X_m, X) + \mathrm{b} \tag{7}$$

where X is the input vector, $X_m$ represents the training data and the Kernel is $K(X_m, X)$. In equation 7, the decision function is computed using the training labels $y_{m\,\in\{+1,-1\}}$ and coefficient $\alpha_m \geq 0$ , while parameter $b$ is determined during training on the labeled dataset. When given a training set of instance label pairs $(x_i,\ y_i)$, i = 1… n, where $X_i \in R$ and the SVMs include the solution for the optimization problem [7].

$$min_{w,b,\xi} \frac{1}{2} w^T + C \sum_{i=1}^{l} \xi_i X_1, \dots, X_n \tag{8}$$

Subjected to

$$y_i(w^T \emptyset(x_i) + b) \geq 1 - \xi_i \tag{9}$$

$\xi\_i \geq 0$.

To find the optimal hyperplane, function $\emptyset(.)$ projects the input pattern Xi onto the higher dimensional space. Among several hyperplanes, SVM finds the optimal one with the maximal margin for the binary problem in this space. In getting a unique hyperplane, weight vector W exists in the terms of a training pattern subset that lies on the margin.

The following equation shows how to construct W:

$$W = \sum_{i=1}^{n} \alpha_i y_i \emptyset(X_i) + b \tag{10}$$

where $n$ is the number of support vectors, $X_i$ is the support vector $i$ and $y_i$ is the label {-1, 1} of Xi . This paper has adopted libSVM with the RBF kernel to classify the data.

## 3. Clustering

The clustering technique is often used to group data into similar groups (cluster). Each group has members that are similar among themselves and dissimilar to those in other groups. In the last decade, researchers have proposed different clustering approaches, namely partitioning, hierarchical, density-based, grid-based and model-based clustering. Partition clustering, which includes the K-means clustering technique and hierarchical clustering, is the most popular technique used in different applications. The main goal of clustering is to maximize intra-cluster and minimize inter-cluster similarity [21].

### 3.1. K-means clustering

K-means clustering is one of the most popular clustering techniques and used in several applications. This algorithm performs clustering by assigning data into different clusters based on the minimum distance between the cluster centroid and the data. Each cluster is then represented by its centroid, i.e., the mean of cluster members [22]. Several advantages make this technique popular and frequently used. First, this clustering technique can work with any standard norm, and it is fast and easy to implement. Second, it has explicit parallelization and is insensitive to data ordering. Conversely, some limitations appear in this clustering technique, as the result is mainly based on the initial cluster centroid. The cluster number is also unknown; the K-cluster must thus be optimized for each case [22].

To explain the k-means clustering technique, assume that we have these observations $\{\boldsymbol{x_i} : \boldsymbol{i} = \boldsymbol{1}, \dots, \boldsymbol{L}\}$, and measure the distance between these data points using order $\boldsymbol{p}$. K-means seeks to divide the observations into different k groups and represent each group with its own centroid (mean) $\{\widehat{\boldsymbol{x1}}, \widehat{\boldsymbol{x2}} \dots, \widehat{\boldsymbol{xk}}, \}$:

$$KCL = \sum_{i=1}^{L} min_{1 \leq j \leq k} (x_i - \hat{x}_j\ )^p \tag{11}$$

The K-means technique should follow these steps: (1) Compute the distance between the centroid and observation data points. (2) Find the smallest distance between the centroid and points and assign a point to the corresponding cluster. The K-means technique converges until the error between the new and old cluster centroids is below a fixed threshold.

# 4. A novel generalization of the Gabor (GGF) method

Over the last decade, researchers have focused on using filters to describe objects. Among several non-linear filters, the GF gives impressive results in different places.

The literature also indicates that the main demerit of the GF is that it constructs redundant and incompact filters, which may decrease system accuracy [10, 11, 13]. An effective method thus requires a generalized and compact GF that can efficiently describe an object. By exploiting a vector quantization method, such as a clustering algorithm, the compact and informative filters can be achieved. This algorithm is considered efficient for constructing compact and generalized filters by grouping similar patterns into similar clusters or groups.

This paper uses the unsupervised machine learning algorithm denoted by the K-means clustering algorithm, to generalize the GF. This algorithm seeks to assign a point to a cluster centroid based on the minimum distance. The main issue of this algorithm is the number of clusters. This parameter should thus be selected to achieve higher accuracy. However, the main goal of data clustering is to provide a precise characterization of unseen samples generated from the same probability distribution. The flowchart below (Figure 4) shows the steps used to generalize the GF:

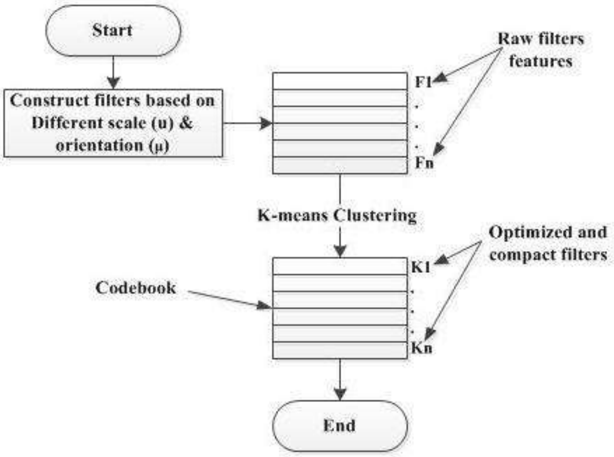



**Fig. 4.** Flowchart of the proposed GGF method, $F1 \ldots, Fn$ represents the raw filters features, and $K1 \ldots, Kn$ represents the cluster centroid *(Generalized filters).*of the edge histogram.

# 5. Experimental setup

This paper uses the first 20 categories of the Caltech 101 dataset with 600 images for training and testing. All categories of the Caltech 101 dataset, containing 3030 images, have been used to test VOC performance in both the spatial and frequency domains.

## 5.1. Caltech 101 dataset

The Caltech 101 dataset contents 101 categories with 9146 images, and each category contains between 40 and 800 images, on average. The images in this dataset are similar in size and resolution. The medium resolution of each image is about 300×200 pixels, and all images are in the JPEG format. Moreover, objects in the images have low occlusion and clutter levels.

This research uses the first 20 categories of the Caltech 101 dataset to demonstrate system accuracy. Furthermore, there are 30 images for each category and the total number of images is 600. To train and test the system, each category contains 15 randomly chosen images each for training and testing [19]. We extend the experiment by using all dataset classes, which provides 3030 images. Figure 5 gives an example of some classes in the Caltech 101 dataset.

**Fig. 5.** Random examples from the Caltech101 dataset categories: (starting from top left) accordion, bonsai, cannon, cell phone, dolphin, elephant, Faces _easy, gerenuk, hedgehog, Joshua _tree, mayfly and yin _yang.

### 5.2. SVM classifier

We used libSVM with the RBF kernel and the one vs. one approach to perform classification. The feature vector is normalized between {+1, -1} to avoid numerical difficulties during the calculation and ensure that larger values do not dominate smaller ones. The normalization equation is thus:

$$\boldsymbol{x} = \frac{\mathbf{2}(\boldsymbol{x}-\boldsymbol{min})}{(\boldsymbol{max}-\boldsymbol{min})} \tag{12}$$

The RBF kernel requires the two best parameters $\mathbf{C}$ **and** $\boldsymbol{\gamma}$ to produce accurate classification. The libSVM grid-search has been used as a straightforward search on the training data to find the best $\mathbf{C}$ **and** $\boldsymbol{\gamma}$ parameters. The search spaces used in this paper are $\{\mathbf{2^{-5}}, \mathbf{2^{-3}}, \dots, \mathbf{2^{15}}\}$ and $\{\mathbf{2^{-15}}, \mathbf{2^{-13}}, \dots, \mathbf{2^{3}}\}$ for $\mathbf{C}$ **and** $\boldsymbol{\gamma}$ respectively. Finally, the K-fold cross-validation has been used to learn the classifier when K=10.

### 5.3. Selecting the best parameters for the Gabor filter (GF) and K-means clustering

The GF has four parameters: scale, orientation, sigma, and filter size. These values and the number of clusters used in the k-means clustering technique differ among applications. These parameters have thus been selected to improve VOC technique performance.

### 5.4. Best Gabor filter (GF) parameters and K-means clustering

Based on different runs, the experiment shows that the best parameters entered into the Gabor filter for object categorization are ($\boldsymbol{\upsilon}$ = 5, $\boldsymbol{\mu}$ = 8, σ = 2.5, and filter size=$\mathbf{128 \times 128}$).

The K-means clustering technique, however, comprises an important part of performing the proposed method. During the experiment, we found a significant difference in accuracy when performing the proposed method in the spatial and frequency domains. The best number of clusters in the spatial domain thus differs from the frequency domain.

For the spatial domain, the experiment started with 4 clusters and ran until reaching 15 clusters in step 1. It stopped at 15 clusters because there is no significant improvement in accuracy in clusters 14 and 15. The 11 clusters provide higher classification accuracy, with a 67.33% average rate based on the naïve approach. Generalization using the Gabor filter is thus considered the best choice.

For the frequency domain, we start with 4 clusters and ran until reaching 20 clusters in step 1. We stop at 20 clusters because there is no significant improvement in accuracy at cluster 20. The best K-cluster is 18, as it gives higher accuracy than the other clusters, i.e., a 65% average rate based on using combined features with the naïve approach. One advantage of using the frequency domain is that it speeds up the system, while some basic information is lost from the raw data when converting it from the spatial to the frequency domains.

## 6. Result and discussion

In this part, we briefly present VOC recognition accuracy based on the standard GF and proposed GGF in both the spatial and frequency domains. We also compare performance at the end of this section.

### 6.1. VOC technique results in the spatial and frequency domains

This section presents the full details of the VOC technique results based on the GF and proposed GGF methods, using the first 20 categories from the Caltech 101 dataset. Table 1 shows the VOC technique results in the spatial domain, based on the naïve approach and single classifier. After different 10 runs, the VOC technique, based on the proposed GGF method, outperforms the standard GF method in average rate and standard deviation, with values of 67.3% and ±0.69299, respectively. The time was reduced by 25%, using 11 generalized filters instead of 40. Referring to Table 1, it shows the accuracy average of the VOC technique in the frequency domain with and without the proposed GGF method. Conversely, Table 1 shows the VOC technique with and without the proposed GGF method in the frequency domain. The GGF performs worse than the VOC technique in the spatial domain, because some raw data is lost when transformation the data from the spatial to the frequency domain using the FFT method. Using the FFT algorithm, however, has the benefit of speeding up the system and reducing time.

**Table 1:** Categorization result based on the standard GF and proposed GGF methods in the (a) spatial and (b) frequency domains.

(a)

| | **Spatial Domain** | | | |
|---|---|---|---|---|
| | Standard GF | | Proposed GGF | |
| | Single Classifier | Naïve Approach | Single Classifier | Naïve Approach |
| Accuracy (%) | 47.345 | 65.1667 | 52.9061 | 67.3 |
| STDEV | ±1.13245 | ±2.16168 | ±0.91206 | ±0.69299 |

(b)

| | **Frequency Domain** | | | |
|---|---|---|---|---|
| | Standard GF | | Proposed GGF | |
| | Single Classifier | Naïve Approach | Single Classifier | Naïve Approach |
| Accuracy (%) | 41.9958 | 63.2667 | 44.5056 | 62 |
| STDEV | ±1.08954 | ±1.9539 | ±0.86007 | ±1.75119 |

### 6.2. Performance comparison

It is interesting to compare the classification rates of the proposed method into two different spaces. This paper uses the t-test approach to measure the significant difference of the proposed method. The p-value (probability-value) of the t-test is 0.05, and the t-test result is compared to this value: if the t-test value is less than 0.05, the proposed algorithm is considered significant. The t-test result of the proposed method decreases as the algorithm is considered more significant.

Based on the t-test result, the proposed GGF method-based naïve approach is considered statistically significant in the spatial domain and outperforms the standard GF method-based naïve approach ($p < 0.05, p = 0.0001$ ). The proposed GGF method-based single classifier is considered extremely statistically significant and outperforms the standard GF method-based single classifier(with $p = 0.0082$). The time has been reduced by 25%, using 11 generalized filters instead of 40. The t-test results for the VOC technique with the proposed GFF method in the frequency domain is not quite statistically significant and performs a bit worse than the naïve approach and single classifier, when compared to the VOC technique-based standard GF method. This result occurs because the number of combined filters is not enough to describe the object compared to the original 40 filters and raw information losses.

Based on the t-test results, the proposed method based naïve approach and single classifier is considered extremely statistically significant in the spatial domain and outperforms the proposed method in the frequency domain ($with$ $p = 0.0001$ ). The t-test results for the standard Gabor filter in the spatial domain show that the naïve approach and single classifier method is considered statistically significant and outperforms the standard Gabor filter in the frequency domain ($with$ $p = 0.0402$ for the naïve approach and $p = 0.0486$ for the single classifier).

In summary, the VOC technique-based standard GF and proposed GGF methods in the spatial domain outperform those in the frequency domain. The VOC technique in the frequency domain performs a bit worse because it loses some basic information when transforming the raw data from the spatial to the frequency domain.

### 6.3. Perform VOC technique with all dataset categories

From the results obtained using the GGF method in the spatial domain, based on 20 classes from the Caltech 101 dataset, it is interesting to perform the VOC technique based on the proposed GGF with the entire Caltech 101 dataset to see how the proposed method performs. Table 2 shows the accuracy rate of the VOC technique-based single classifier and naïve combination approach.

**Table 2:** Classification accuracy of all Caltech 101 categories based on the proposed GGF method in the spatial domain.

| | Single Classifier | Naïve Approach |
|---|---|---|
| Average Accuracy (%) | 40.4207 | 57.0784 |
| STDEV | 0.47436 | 0.98742 |

The VOC technique based on the proposed GGF method in the spatial domain performs well even with all Caltech 101 dataset classes. The standard deviation shows the proposed method is stable with diverse objects.

## 7. Conclusion

This paper has presented the visual object categorization technique based on the filter-based descriptor and a classifier denoted by the standard Gabor filter and proposed GGF methods. The results verify the superiority of the proposed method in describing an object, as it provides different and distinctive feature maps based on a different scale and orientation for the objects.

The VOC technique, based on both the standard GF and proposed GGF methods, has been tested in two different spaces: the spatial and frequency domains. Based on the t-test results, the VOC technique in the spatial domain gives better results than implementing it in the frequency domain for both the standard GF and proposed GGF methods. This occurs for several reasons. One reason is that transforming raw data from the spatial domain to the frequency domain is inefficient and may lose some important information. In other words, the representation of the FFT algorithm to the data is not as efficient as the actual raw data.

Conversely, the VOC technique, based on the proposed GGF method, gives a higher accuracy rate and outperforms the standard GF method. The proposed filter-based descriptor and classifier, denoted by the GGF method, has significantly improved the object categorization problem and increased system performance.

## Acknowledgement

The authors would like to thank the Ministry of Higher Education and University Kebangsaan Malaysia, for providing facilities and financial support under Arus Perdana Research Grant AP-2017-005/2 and PRGS/1/2016/ICT02/UKM/02/1. Furthermore, the authors would like to especially thank Assoc. Prof. Dr. Loay E. George, Dr. Ahmed Talib and Mr. Ahmad Jabir for their important contributions to this research.

## References

[1] Abdullah SNHS, Khalid M, Yusof R & Omar K (2007), Comparison of feature extractors in license plate recognition. *First Asia International Conference on Modelling & Simulation*, 502-506.

[2] Mizher MAA, Ang MC, Abdullah Siti Norul Huda Sheikh & Ng Kok Weng (2017), Action Key Frames Extraction Using L1-Norm and Accumulative Optical Flow for Compact Video Shot Summarisation. *International Visual Informatics Conference*, 364-375

[3] Abdullah, *A Supervised Learning Algorithms for Visual Object Categorization*, Institute of Information and Computing Sciences, Utrecht University, (2010), 1-151.

[4] Abdullah A & Wiering MA (2007), CIREC : Cluster Correlogram Image Retrieval and Categorization using MPEG-7 Descriptors. *IEEE Symposium on Computational Intelligence in Image and Signal Processing*, 431-437.

[5] Liua Y, Zhanga D, Lua G & Mab WY (2007), A survey of content-based image retrieval with high-level semantics. *Journal of Pattern Recognition*, 40: 262– 282.

[6] Wang JZ, Li J & Wiederhold G (2001), SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture LIbraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9): 947-963.

[7] Choi WP, Tse SH, Wang KW & Lam KM (2008), Simplified Gabor wavelets for human face recognition. *Journal of Pattern Recognition*. 41(3): 1186-1199.

[8] Long F, Zhang H & Feng DD, *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, New York, Springer, (2003), 1-26.

[9] Nadernejad E, Sharifzadeh S & Hassanpour H (2008), Edge Detection Techniques: Evaluations and Comparisons. *Applied Mathematical Sciences*, 2: 1507 - 1520.

[10] Zhu J, Vai MI, & Mak PU (2004), A New Enhanced Nearest Feature Space (ENFS) Classifier for Gabor Wavelets Features-Based Face Recognition. *Lecture Notes in Computer Science 3072, Springer*, 124 - 131.

[11] Shen LL & Ji Z (2009), Gabor Wavelet Selection and SVM Classification for Object Recognition. *Acta Automatica Sinica*, 35(4): 350-355.

[12] Xing W & Bir B (1997), Gabor wavelet representation for 3-D object recognition. *IEEE Transactions on Image Processing*, 6(1): 47-64.

[13] Li W, Mao K, Zhang H & Chai T (2010). Designing compact Gabor filter banks for efficient texture feature extraction. *11th International Conference on Control Automation Robotics & Vision*, 1193 – 1197.

[14] Jones JP & Palmer A (1987), An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6): 1233-1258.

[15] Varma M & Zisserman A (2003), Texture classification: are filter banks necessary?. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2:691-698.

[16] Mignotte M (2012), An Energy-Based Model for the Image Edge-Histogram Specification Problem. *IEEE Transactions on Image Processing*, 21(1): 379-386.

[17] Manjunath BS, Salembier P & Sikora T, *Introduction to MPEG-7: Multimedia Content Description Interface*, John Wiley & Sons, (2002).

[18] Alemu Y, Koh JB, Ikram M & Kim DK (2009). Image Retrieval in Multimedia Databases: A Survey. *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 681 – 689.

[19] Mohammed Salameh, Abdullah A, Shahnorbanun Sahran, (2018), Multiple Descriptors for Visual Odometry Trajectory Estimation, *International Journal on Advanced Science, Engineering and Information Technology*. vol. (8), issue 4-2, pp 1423-1430.

[20] Cortes C & Vapnik V (1995), Support-vector networks. *Machine Learning*, 20(3): 273-297.

[21] Yi Y, Dong X, Feiping N, Shuicheng Y & Yueting Z (2010). Image Clustering Using Local Discriminant Models and Global Integration. *IEEE Transactions on Image Processing*, 19(10): 2761-2773.

[22] Jain AK, Murty MN & Flynn PJ (1999), Data clustering: a review. *ACM Computing Surveys*, 31(3): 264–323.