Visual Object Categorization based on Orientation Descriptor

Hayder Ayad\(^1\), Siti Norul Huda Sheikh Abdullah\(^2\), Azizi Abdullah\(^3\)
hayder.alsultany@ieee.org\(^1\), mimi@ftsm.ukm.my\(^2\), azizi@ftsm.ukm.my\(^3\)
\(^{1,2,3}\) Pattern Recognition Research Group, Center of Artificial Intelligence Technology, Faculty of Information Systems and Technology, 43650 Bandar Baru Bangi, Universiti Kebangsaan Malaysia

Abstract—The demand of new fast technology and image investigation in many applications has made managing visual object categorization techniques extremely important. The main problem of visual object categorization is the semantic gap (categorization problem). Currently, several researches show that using a texture feature could improve the categorization problem especially when using orientation descriptors. Mainly, in this research the edge histogram descriptor has been selected to extract the texture feature. Obviously, the main demerit of using this kind of texture descriptor is it uses single orientation to extract the texture feature. Therefore, the Gabor filter has been proposed to improve the performance of this descriptor by constructing different feature maps based on different scale and orientation. To demonstrate the performance of the proposed method, the first 20 classes of the Caltech 101 dataset have been used. Moreover, we compared the performance recognition of the proposed method in two different domains, namely spatial and frequency domains. Finally, the result shows that the proposed method in the spatial domain outperforms the proposed method in the frequency domain. This is because of losing some of the basic raw data through using Fast Fourier Transform algorithm in converting the system to the frequency domain.

Keywords- VOC technique; Gabor filter; Edge histogram descriptor; naïve approach.

I. INTRODUCTION

At present, due to the increase in new technology and the investigation of images in many applications, efficient mechanisms to manage these images are required. In fact, searching for the target image is considered the main issue in many studies [1].

Based on the literature, the semantic gap (categorization problem) is considered the main problem facing the visual object categorization technique. This means the gap between the low-level features and the richness concept in the human mind. Clearly, much of the research performed in the area has used color and texture features in different ways to improve the categorization problem [2]. Basically, the color feature is considered as a popular feature used in object categorization system due to easiest extract from the image [3].

Apart from that, the texture feature is another part of low-level features used in the visual object categorization technique. This kind of feature shown in different applications is an important feature used to obtain the high-level concept. Several texture feature algorithms have presented for different purposes such as object categorization, segmentation and image retrieval system [4]. Examples of these algorithms are co-occurrence matrices, shift-invariant principal component analysis (SPCA), the Tamura feature, Wavelet transform, and the fractal mode [5].

Among these texture descriptors, MPEG-7 introduced edge histogram for extracting local texture features. This descriptor captures the spatial distortion of the edge and is invariant to most image effects, such as transformation [6]. In fact, this descriptor is considered as an orientation descriptor and it has a benefit in capturing the most important features that are used to describe the object [6]. Basically, this kind of orientation descriptors used single orientation in extracting the texture features. This may be inefficient in describing the object correctly [7].

Besides, using filter bank such as Gabor filter provides meaningful information and captures the salient features. Basically, the Gabor filter is used in many applications and has been shown to be robust under certain conditions such as illumination and view point [8].

Therefore, in this research paper, we present a filter-based descriptor and classifier denoted by the Gabor filter to improve the performance of the edge histogram by supplying different feature maps based on different scale and orientation. Furthermore, a naïve approach is used to combine different filters. This operation promising to produce features is robust, distinctive and each feature complements the other in representing the object [9]. The SVM classifier has been presented for classification purposes. Moreover, the FFT algorithm has been used to convert the proposed VOC technique to the frequency domain. In fact, using FFT algorithm has a benefit in speeding up the system performance. Lastly, we evaluate our proposed method with the first 20 categories of the Caltech 101 dataset.

This paper is organized as follows: In section 2, we present in brief the state-of-the art method used in visual object categorization technique. We introduce the proposed VOC technique in both spatial and frequency domains in section 3. In section 4, the experimental results are presented. Lastly, we conclude the fact-finding in section 5.

II. RELATED WORK

Firstly, we introduce the orientation texture descriptor, namely edge histogram. Next, we review the Gabor filter method computation. Then, we discuss in brief the naïve approach and SVM classifier subsequently.
A. Edge histogram descriptor

Edge histogram descriptor was introduced by MPEG-7 under texture descriptor group. This descriptor is considered as an orientation descriptor and it has been used frequently by several researches for different purposes [6, 7]. Basically, this descriptor can describe the non-homogenous texture. In addition, the main point of introducing this descriptor is to describe the distribution of four directional edges and one non-directional edge.

However, this descriptor extracts the texture feature through dividing the image into 16 sub-images and applying 5 fixed filters to extract five kinds of edge namely horizontal, vertical, diagonal 45, diagonal 135 and non-directional. The following figures show the edge histogram mask and the image dividing into sub-blocks and subsequently [6].

![Fig 1. Shows the Edge histogram mask](image)

\[
\begin{bmatrix}
1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 \\
\sqrt{2} & 0 & 0 & -\sqrt{2} \\
0 & \sqrt{2} & -\sqrt{2} & 0 \\
2 & -2 & -2 & 2
\end{bmatrix}
\]

Fig 1. Shows the Edge histogram mask

![Fig 2. Shows the image partition into non-overlapping sub-images](image)

B. Gabor filter

The Gabor filter was introduced in 1946 by Dennis Gabor. Later, this filter become commonly used in different areas of image processing such object categorization, face detection, and variety applications of computer vision such as image retrieval systems, finger print and character recognition [8].

In fact, one of the most important characteristics of this filter is capturing the salient features such as spatial frequency characteristic, spatial location and orientation selection [10]. Furthermore, the Gabor function gets the best time-frequency resolution for signal analysis. Added to that, it provides a multi-resolution for texture feature and classification schema. Lastly, the most important characteristic is the optimal combining between the frequency and the spatial locality. This combination leads to capture the salient feature in different scale and orientation. The following equations below describe the Gabor filter computation [11].

\[g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x^2+y^2\gamma^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x}{\lambda} + \psi\right)\right)\] (1)

Real parts of Gabor filter:

\[g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x^2+y^2\gamma^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x}{\lambda} + \psi\right)\] (2)

Imaginary parts of Gabor filter:

\[g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x^2+y^2\gamma^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x}{\lambda} + \psi\right)\] (3)

Where \(\lambda = x \cos \theta + y \sin \theta\) and \(\gamma = x \sin \theta + y \cos \theta\). In addition, the wavelength of the sinusoidal function represents the \(\lambda\) parameter, \(\theta\) the orientation that entered the Gabor filter. Added to that, \(\psi\) the phase offset represented by the \(\psi\) parameters, \(\sigma\) sigma value of Gaussian envelope and lastly, the \(\gamma\) represent spatial aspect ratio. The figure below shows an example to the Gabor filter.

![Fig 3. The Gabor filters based on \(\mu=3\), \(v=2\), and \(\sigma=6.28\), and filter size =128 × 128.](image)

C. Fast Fourier Transform

This algorithm is considered as an efficient algorithm that is used to compute the discrete Fourier transform. Clearly, this research implements this algorithm for two reasons, first, to convert the system into frequency domain, and second, to speed up the system’s performance. The mathematical equation below describes the forward discrete Fourier transform [12].

\[X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{k}{N} n}\] where \(k = 0, ..., N - 1\) (4)

The complex numbers \(X_k\) represent the amplitude and phase of the different sinusoidal components of the input \(x_n\). The mathematical equation below describes the inverse discrete Fourier transform.

\[X_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{j2\pi \frac{k}{N} n}\] (5)

Equation (5) shows how to compute the \(X_n\) as a sum of sinusoidal components \(\frac{1}{N} X_k e^{j2\pi \frac{k}{N} n}\) with frequency \(K/N\) cycles per sample.

D. Naive approach

Currently, using combination features in representing the image outperforms the single descriptor especially in visual object categorization technique [9]. Naive approach consider as an easier approach used to combine different descriptors.
Basically, it combines different feature vectors in one feature vector regardless to the source of these feature vectors. Obviously, the main issue in using naïve approach in combining diverse feature vectors is the high dimensionality feature vector that is constructed and this could lead to over-fitting problem [7].

E. Support Vector Machine (SVM)

Recently, SVM algorithm in different applications showed that it has outperformed the state-of-the-art techniques. Basically, this algorithm was introduced for binary classification, but it has been shown that it can be easily extend to multi-classes classification. Furthermore, the main point in this algorithm is the optimal hyper-plane that separates the given data into two classes (categories) \{+1, -1\} with the maximal margin in the higher dimensional feature space. Apparently, the SVM goal is to construct a model based on the given training data to predict the target values of the test data attributes [13].

In this research paper, the libSVM has been used to perform the SVM classifier. Additionally, the SVM classifier with RBF kernel used the one vs. one to make the classification, and the classifier output is the probability that is used to make the decision [14]. This kernel requires two parameters $C$ and $\gamma$, and the best values of these parameters will provide a good classification. Therefore, the libSVM grid-search has been used as a straightforward search on the train data to find the best parameters. Added to that, the grid search space values that are used in this research are $\{2^{-5}, 2^{-3}, ..., 2^{15}\}$ and $\{2^{-15}, 2^{-13}, ..., 2^{3}\}$ for $C$ and $\gamma$ respectively. Additionally, the K-fold cross-validation has been used to learn the classifier with $K=10$ [13].

III. PROPOSED VOC TECHNIQUE IN BOTH SPATIAL AND FREQUENCY DOMAINS

The proposed VOC technique implements into two different spaces namely spatial and frequency domain. Figure 4 shows the steps that used to implement the VOC technique for both single classifier and combination features denoted by naïve approach.

On the other hand, Figure 5 shows the VOC technique implementation steps in the frequency domain. The implementation steps in the frequency domain seem to be similar as the spatial domain except the implementation of the Fast Fourier Transform algorithm (FFT) to speed up the system performance.

IV. EXPERIMENT RESULT

The Gabor filter considered as an advanced filter because it takes different scale ($\omega$), orientation ($\mu$), sigma ($\sigma$), and filter size. Therefore, we perform detailed experiments to select the best parameters in Gabor filter. Firstly, we investigate the
number of filters in correlation to accuracy versus sigma values and secondly we test on various filter sizes. The section below shows the best μ, υ, σ, and filter size values.

A. Select optimal scale, orientation, sigma and filter size
Based on different runs, the best parameters entered to the Gabor filter function are μ= 8, υ=5, and σ=2.5, and filter size =128 × 128. The figures below show the optimal parameters that entered the Gabor filter for categorization purpose.

Based on the obtained result in the table above, it becomes clear that the VOC technique in the spatial domain based single classifier and naïve approach outperforms the VOC technique in the frequency domain. To verify our result, the t-test has been used to measure the significant difference of the proposed VOC technique in the spatial domain compared to the frequency domain. The t-test results for the VOC technique based Gabor filter in the spatial domain for both naïve approach and single classifier are considered statistically significant and outperform the VOC technique based on the Gabor filter in the frequency domain (with p= 0.04202) for naïve and (p=0.0486) for single classifier.

Table 1

<table>
<thead>
<tr>
<th>Number of Runs</th>
<th>Spatial domain</th>
<th>Frequency domain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single classifier</td>
<td>Naïve approach</td>
</tr>
<tr>
<td>1</td>
<td>48.3715</td>
<td>66.6667</td>
</tr>
<tr>
<td>2</td>
<td>46.7133</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>46.422</td>
<td>65.3333</td>
</tr>
<tr>
<td>4</td>
<td>48.138</td>
<td>67</td>
</tr>
<tr>
<td>5</td>
<td>45.472</td>
<td>64.3333</td>
</tr>
<tr>
<td>6</td>
<td>48.021</td>
<td>67</td>
</tr>
<tr>
<td>7</td>
<td>47.9725</td>
<td>63.3333</td>
</tr>
<tr>
<td>8</td>
<td>48.6635</td>
<td>67.3333</td>
</tr>
<tr>
<td>9</td>
<td>47.8375</td>
<td>66.6667</td>
</tr>
<tr>
<td>10</td>
<td>45.8388</td>
<td>61</td>
</tr>
<tr>
<td>Average</td>
<td>47.345</td>
<td>65.1667</td>
</tr>
<tr>
<td>STDEV</td>
<td>±1.13245</td>
<td>±2.16168</td>
</tr>
</tbody>
</table>

Based on the obtained result in the table above, it becomes clear that the VOC technique in the spatial domain based single classifier and naïve approach outperforms the VOC technique in the frequency domain. To verify our result, the t-test has been used to measure the significant difference of the proposed VOC technique in the spatial domain compared to the frequency domain. The t-test results for the VOC technique based Gabor filter in the spatial domain for both naïve approach and single classifier are considered statistically significant and outperform the VOC technique based on the Gabor filter in the frequency domain (with p= 0.04202) for naïve and (p=0.0486) for single classifier.

V. CONCLUSION
In this research paper, we present the VOC technique based on orientation descriptor namely edge histogram descriptor. The Gabor filter has been adopted to improve the performance of the edge histogram through providing different feature maps based on different scale and orientation. Next, the SVM classifier has been applied to classify single classifier and combination feature denoted by naïve approach. Furthermore, the first 20 classes of Caltech 101 dataset have used to demonstrate the performance recognition of the proposed VOC technique in both frequency and spatial domains. The experiment results show that the proposed method improve the categorization problem. In addition, the proposed method in spatial domain outperforms the proposed method in the frequency domain due to losing some of the basic information from the raw data.

ACKNOWLEDGMENT

The author would like to thank the Faculty of Information Science and Technology, University Kebangsaan Malaysia, for providing facilities and financial support under Exploration Research Grant Scheme No. UKM-ERGS/1/2011/STG/UKM/2/48(TK) and ERGS/1/2011/STG/UKM/2/41. Furthermore, I would like to
present this work to my parents and my aunt because all of them are like the candle burning in order to light up my way.

REFERENCES.


