Methodology

3. METHODOLOGY

The term tamper refers to any post-processing operations that have been performed on an image. Image splicing is one of the most common and most challenging tampering forgeries which are detected using camera statistical features and one-class classifier. Four types of features, namely, color features, image quality features, wavelet features and bicoherence features are used during tamper detection. Two one-class classifiers are considered, namely, Support Vector Machine (SVM) based classifier and Radial Basis Function (RBF) based classifier. The tampering method considered in this thesis is the splicing forgery.

The proposed automatic splice tamper detection algorithm consists of three major modules, namely,

(i) Feature Selection
(ii) Feature Extraction
(iii) One-class Classification

First, feature statistics that best represents the property of a camera are extracted from the images taken by that camera. These feature statistics are used for training a one-class classifier (SVM and RBF) in order to get the feature pattern of the given camera. Then, a sliding segmentation is performed on the testing image and the same set of feature statistics are extracted from the segmented blocks. These features are fed as input into the trained one-class classifier to match the feature pattern of the given camera. The images with low percentage of matched blocks are classified as tampered ones. The process is pictorially represented in Figure 3.1. Each of these steps is explained in the following sections.

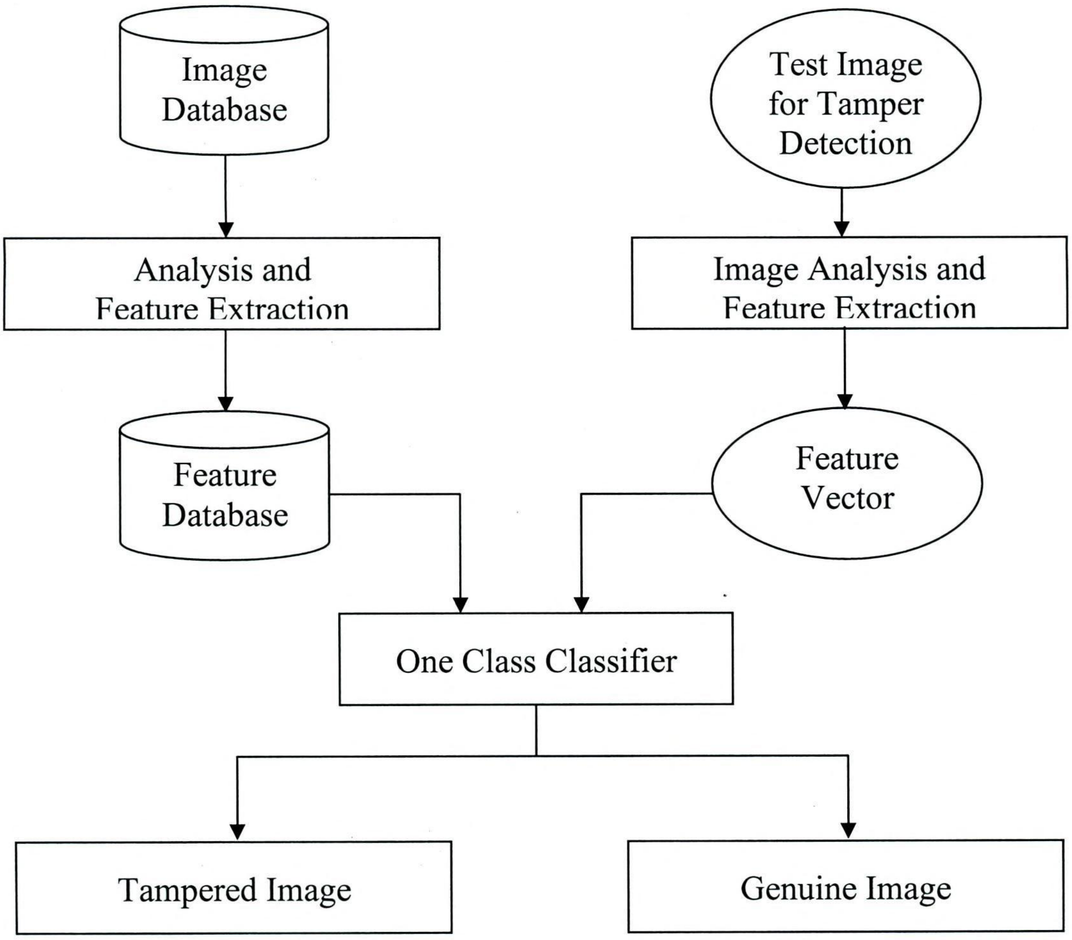


Figure 3.1 : Proposed Tamper Detection Framework

3.1. FEATURE SELECTION AND EXTRACTION

As mentioned before, the system extracts four types of features, namely, color features, quality features, wavelet features and bicoherence features. The method of extraction is described in this section.

3.1.1. Color Features

Cameras of different modes use various CFA interpolation algorithms, transaction modes and transform processes. Color features could show these differences in the aspect of the color property of the whole image (Kharrazi *et al.*, 2004). Four color features namely, mean pixel value, correlation of

RGB color pairs, the neighbor distributing of centroid and the energy ratio of RGB colors are extracted. The energy ratio is defined as using Equation (3.1).

$$E1 = \frac{|G|^2}{|B|^2}, E2 = \frac{|G|^2}{|R|^2}, E3 = \frac{|G|^2}{|R|^2} \tag{3.1}$$

3.1.2. Image Quality Features

Different cameras have different image quality features due to the difference in hardware and software. For example, photographing the same scene, the image of different cameras would show dissimilar brightness or texture definitions. Therefore, quality features would make contribution to distinguishing the images taken by different cameras (Avcibas , 2001). Image quality features used in this work belong to three categories,

(i) features based on pixel difference
(ii) features based on correlation and
(iii) features based on frequency spectrum.

Under the first category, Mean Square Error (MSE), Mean Absolute Error (MAE) and Corrected Infinite Norm (CIN) are extracted. Under the correlation based category, features like image fidelity, normed cross correlation and mean angle similarity are extracted. Under frequency spectrum, the spectrum amplitude error, spectrum phase error, spectrum phase-amplitude error, image block spectrum phase error, image block spectrum phase-amplitude error, are extracted.

3.1.3. Wavelet Features

Apart from the above mentioned features, another frequently used image feature is wavelet feature. These features describe an image in a more detail manner at different resolutions using the characteristics of time-frequency. Besides, various imaging processes and post-processing in different cameras have distinct influences on correlations in the wavelet domain (Lyu and Farid, 2005). Therefore, a Quadrature Mirror Filter separation (QMFs) is performed on each color channel of the images from which, the mean, variance, skewness, kurtosis and the optimal linear prediction error of coefficient amplitude in each level and direction (horizontal, vertical and diagonal) is extracted. These features are extracted for each wavelet subband.

This research work uses the statistic features of wavelet coefficients' prediction error. As the prediction error is the residual of subtracting the content-related portion from the original image, this can be used to improve the accuracy in classifying random image samples by reducing the correlation between statistic of prediction error and image content.

The post-processing operations such as enhancement and gamma correction bring about the nonlinear distortion between the image brightness and pixel value (Figure 3.2) and the nonlinear distortion could lead to the high-order dependence in frequency domain, so the bicoherence features could be used to estimate the geometric distortion and brightness distortion in digital images (Farid, 2001; Farid and Popescu, 2001a; 2001b). Besides, as the bicoherence features extract the information that has no direct relation with the content of image, they could reduce the limitations for the image content. In this approach, two bicoherence features, namely, mean bicoherence amplitude and bicoherence phase entropy are extracted. They could be the features for identifying cameras due to their sensitivity to nonlinear distortion.

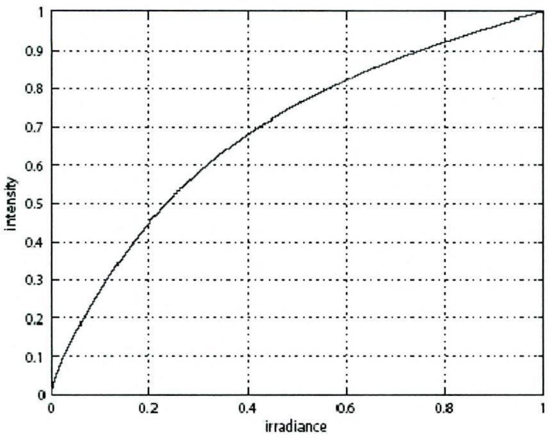


Figure 3.2 : Mean of Camera Responsible Function

In order to make the variation of bicoherence estimation at each frequency pair (ω_1, ω_2) independent of $P(\omega_1)$, $P(\omega_2)$ and $P(\omega_1 + \omega_2)$ the bicoherance is normalized. For single dimension signal f(x), it is defined as :

$$\hat{b}(\omega_1+\omega_2)=\frac{\frac{1}{N}\sum_{k}F_k(\omega_1)F_k(\omega_2)F_k^*(\omega_1+\omega_2)}{\sqrt{\frac{1}{N}\sum_{k}|F_k(\omega_1)F_k(\omega_2)|^2\frac{1}{N}\sum_{k}|F_k^*(\omega_1+\omega_2)|^2}}$$

$$=|\hat{b}(\omega_1+\omega_2)|e^{j\varphi(\hat{b}(\omega_1,\omega_2))} \qquad (3.2)$$

where Fk(.) stands for the kth segment of Fourier transform. Here, in order to limit estimated error, the Fourier transform for each segment is calculated and the mean bicoherance amplitude and bicoherance phase entropy of the frequency is taken as features to estimate correlation using Equation (3.3).

$$M=|W|^{-1}\sum_{w}|\hat{b}(\omega_1,\omega_2)| \qquad (3.3)$$

where

$$W = \{(\omega_1, \omega_2)|\ \omega_1 = 2\pi n_1 / N,\ \omega_2 = 2\pi n_2 / N,\ n_1, n_2 = 0, \ldots, N-1\}$$
$$P = \Sigma_n p(\varphi_n) \log p(\varphi_n)$$

where $p(\varphi_n) = |W|^{-1} \Sigma_w I(\phi\, \hat{b}(\omega_1+\omega_2)) \in \varphi_n)$ and I(.) is indicator function and $\varphi_n = \{\phi| -\pi + (2\pi n)/N \le \phi < -\pi+2\pi(n+1)/N\}$ with n = 0, …, N-1. Thus, the feature vector is formed as a concatenated version of the four features (color, quality, wavelet and bicoherance) to form feature database and feature vector.

3.1.4. Feature Selection

In order to improve the performance of the detection system and reduce the feature dimensions, the features which are more effective for the detection step are selected. In this research, Fisher Classifier is used in for this purpose. Fisher Classifier is a linear Classifier (Vapnik, 1995). If the samples could be classified well using a single feature under linear classifier, it means the samples in the feature space of the used feature are not interlaced, that is: this feature has the great efficiency in classifying these samples. Thus, using Fisher classifier will find the most efficient features for detection system, and consequently, it is possible to obtain the most effective feature set. The features are ordered according to the result of Fisher classifier. Then one feature is added each time according to that order. Figure 3.3 shows the relation between selected feature dimensions and the corresponding accuracy of classification. From the figure, it can be seen that after features exceed 87 dimensions, the accuracy improves little as the features go on increasing. So, only the first 87 features are selected for detection. These selected features are used to train the two one-class classifiers selected in this study.

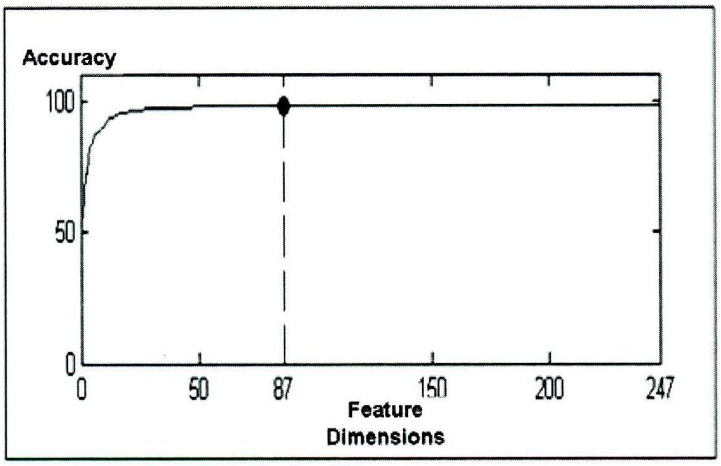


Figure 3.3 : Relation between Selected Feature Dimension and Accuracy of Classification

3.2. ONE-CLASS CLASSIFICATION

Conventional classification techniques mainly focus on two-class or multi-class classification problems. However, in many real-world problems, like tamper detection, there is only one class sample and the other class is severely undersampled. For these situations, one class classification (OCC) algorithms can be designed to describe only one class of target objects and discriminate them from all other possible patterns. Two most frequently used OCC are Radial Basis Function Neural Networks and Support Vector Machines. This section presents description on the same.

A. Radial Basis Function Classifier

The RBF (Lyu and Farid, 2005) first maps training data into a feature space non-linearly by its hidden layer. Then the linear least squares regression can be implemented in its output layer to realize classification. For a binary classification problem, assume there exists l training samples which form the training set $S = \{(x_1, v_1)., (x_2, v_2), \ldots, (x_l, v_l)\}$, where $x_i \in R^d$, $v_i \in \{-1, +1\}$ and v_i represents the class label of the point x_i, $I = 1, 2, \ldots, l$.

A basic RBF consists of three layers, one input, one hidden and one output layer (Figure 3.4), which depicts a 2-class classification model. This can be easily converted to multi-class model by extending the output layer to contain multi-neurons. The input layer passes the input values to hidden layer and the first layer connections are not weighted. Subsequently, the hidden layer, which contains the radial basis function units, transfers the input vector non-linearly to a feature space. The second layer connections are weighted by using a multiple linear regression, then the output neurons are simple summations. The transfer function of neurons in the hidden layer generally adopts the Gaussian Kernel function

$$K(x, c_i) = \exp\left(-\frac{1}{2\sigma_i^2} \| x - c_i \|^2 \right) \qquad (3.4)$$

where $K(x, c_i)$ is the activation value of the ith neuron in the hidden layer for the input x, the c_i denotes the center of this unit and the variable σ_I defines the width or radius of the bell-shape for the Gaussian kernel function.

The method used to find the center c_i is K-means clustering which assumes m neurons in the hidden layer. While adapting this RBF model to OCC, the target dataset $D = \{x_i \in R^n, i = 1, 2, \ldots, N\}$ is mapped into a feature space by the hidden layer of RBF. The corresponding dataset of D in the feature space can be represented as $D' = \{y_i \in R^m, i = 1, 2, \ldots, N\}$, where m is the number of neurons in the hidden layer of the RBF and $y_i = (K(x_i, c_1), K(x_i, c_2), \ldots, K(x_i, c_m))^T$. Then the linear kernel is applied in the feature space for one class classifier.

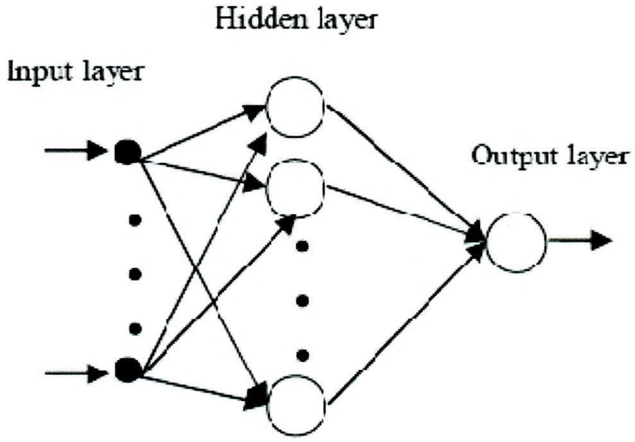


Figure 3.4 : A Basic RBF Structure

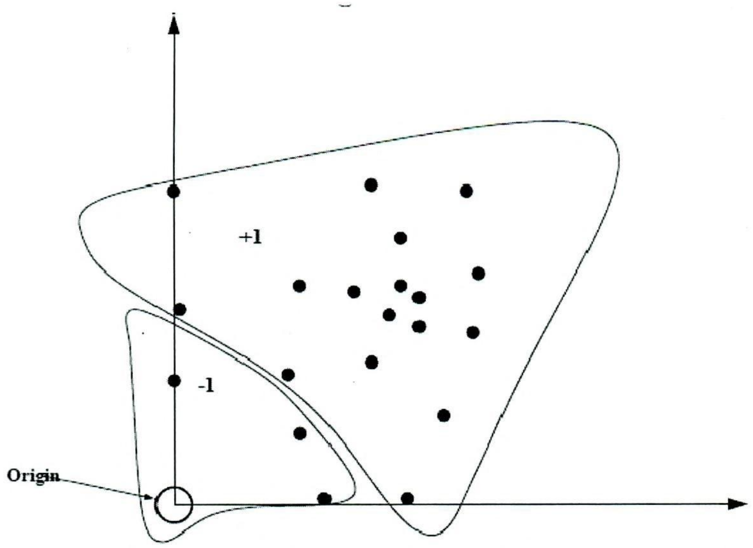
B. One-Class SVM

A one-class SVM classifier proposed by Scholkopf *et al.* (1999) is used in this study. In this method, essentially, after transforming the feature via a kernel, the system treats the origin as the only member of the second class. Then using "relaxation parameters" the image of the one class from the origin is separated. Then the standard two-class SVM techniques are employed. The problem is framed in the following way:

Suppose that a dataset has a probability distribution P in the feature space. Find a "simple" subset S of the feature space such that the probability that a test point from P lies outside S is bounded by some a priori specified value. Supposing that there is a dataset drawn from an underlying probability distribution P, one needs to estimate a "simple" subset S of the input space such that the probability that a test point from P lies outside of S is bounded by some a prior specified $v \in (0, 1)$. The solution for this problem is obtained by estimating a function f which is positive on S and negative on the complement S. In other words, Scholkopf *et al.* (1999) developed an algorithm which returns a function f that takes the value +1 in a "small" region capturing most of the data vectors, and -1 elsewhere.

The algorithm can be summarized as mapping the data into a feature space H using an appropriate kernel function, and then trying to separate the mapped vectors from the origin with maximum margin (Figure 3.5).

$$f(x) = \begin{cases} +1 & \text{if } x \in S \\ -1 & \text{if } x \in \overline{S} \end{cases} \tag{3.5}$$



One-Class SVM Classifier. The origin is the only original member of the second class.

Figure 3.5 : One Class SVM

In the present context, let $x_1, x_2, \ldots, x_l$ be training examples belonging to one class X, where X is a compact subset of R^N. Let $\Phi : X \rightarrow H$ be a kernel map which transforms the training examples to another space. Then, to separate the data set from the origin, one needs to solve the following quadratic programming problem:

$$\min \frac{1}{2} \| w \|^2 + \frac{1}{vl} \Sigma \xi_i - \rho \qquad (3.6)$$

subject to

$$(w. \Phi(x_i)) \geq \rho - \xi_i \quad \text{where } i = 1, 2, \ldots, l \text{ and } \xi_i \geq 0.$$

If w and ρ solve this problem, then the decision function

$$f(x) = \text{sign}((w \cdot \Phi(x)) - \rho) \qquad (3.7)$$

will be positive for most examples xi contained in the training set. In this research the LIBSVM (version 2.0) is used. This is an integrated tool for support vector classification and regression which can handle one-class SVM using the Sholkopf algorithms. The LIBSVM 2.0 is available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

3.3. TAMPER DETECTION

The detection process consists of two steps, namely, training and testing. In the training step, the training images are segmented into blocks of the same size (128x128). This size is chosen because less size would not represent camera features correctly. The next step is feature selection. The one-class classifier is then trained to get the feature pattern of the given camera in a fixed size. Either the radial basis function (RBF) or Support Vecter Machine is chosen as the kernel function of one-class classifier. Finally, the same set of feature statistics extracted from testing blocks are fed into trained one-class classifier to get the matching result. The principle in training and matching is as follows:

Because the input of one-class classifier is statistics, which has the property of fluctuation, a false rejection ratio of 1% is set to one-class classifier, that is: the classifier is allowed to reject one sample by error from each 100 samples. By doing this, a tolerance for the matching step is set to adapt to the fluctuation of statistics. In the matching step, if the percentage of matched blocked is lower than 99% of the whole blocks in a testing image, this image could be seen as a tampered one. For most of tampering operations, the goal is to tamper the kernel objects whose areas are normally much bigger than 1% of the whole image. So the matching result of a tampered image is always much lower than 99%. Thus, the tampered images can be clearly distinguished in the detection system.

Different cameras have dissimilar properties due to the different hardware and software used in the cameras. The features used in the proposed methods represent these properties in various aspects. The tampering operation or any post processing operation such as blurring, retouching and gamma correction would inevitably change these features. These changes would make the tampered blocks hardly match the original pattern features. Thus, the methodology used has practicability in tampered image detection. The results obtained during experimentation are presented in the next chapter, Results and Discussion.