Automated defect recognition of C-SAM images in IC packaging using Support Vector Machines

Y.L. Zhang · N. Guo · H. Du · W.H. Li

Abstract Ultrasonic C-mode scanning acoustic microscopy (C-SAM) is widely used in the semiconductor industry for reliability testing and product inspection due to its ability to non-destructively detect defects in IC packaging. However, image interpretation and defect identification depend largely on the experience of operators, and there is no defect recognition system; this is partly due to current recognition systems, which are based on computer vision algorithms and are not robust for C-SAM images. A new robust defect recognition system and its application to C-SAM images are described in this paper. The iconic domain of two-dimensional C-SAM grey-level image analysis based on the non-linear Mumford-Shah model is used, and defect recognition is achieved through the use of Support Vector Machines (SVMs). The system is verified through experiments on a sequence of C-SAM images corrupted by synthetically generated noise, bias and different shape. The remarkable defect recognition rates achieved indicate that Support Vector Machines (SVMs) are suitable for IC package defect identification.

Keywords Automatic test equipment · C-SAM image analysis · IC package inspection · Support Vector Machine

1 Introduction

Ultrasonic C-mode scanning acoustic microscopy (C-SAM) is widely used in the semiconductor industry for reliability testing and product inspection due to its ability to non-destructively detect defects in IC packaging [1]. However, image interpretation and defect identification depend largely on the experience of operators, and there is no defect recognition system; this is partly due to current recognition systems, which are based on computer vision algorithms and are not robust for C-SAM images, which are susceptible to image noise.

In semiconductor manufacturing, machine vision plays an important role in the automatic inspection of IC chips [2]. The main automated inspection process in IC manufacturing includes mask and reticle inspection, in-process pattern inspection and final chip inspection for quality control [3]. Zhou and Kassim [4] presented a fast method for the detection of die extrusion defects in IC packages using optimal filters for the detection of linear features and other feature enhancement techniques. Dinesh and Teoh [5] developed a software package to carry out the extraction of defects that may exist on the IC leads and the surface of the IC chip, and defect recognition is achieved by comparing the image data with a pre-defined threshold value.

Although there is an abundance of proposed computer vision algorithms for object recognition, there have been few systems that achieve good performance for practical applications, and most such systems do not adapt to changing environments [6]. The main difficulties typically associated with the systems are the complexity due to environments such as noise, pose and lighting conditions. Furthermore, the fixed set of parameters used in various vision algorithms often leads to ungraceful degradation in performance.

Recently, the Support Vector Machine (SVM) learning method has been promulgated to be very effective for general purpose pattern recognition [7]. Intuitively, given a set of points which belong to either of two classes, SVM finds the hyperplane leaving the largest possible fraction of points of the same class on the same side while maximising the distance of either class from the hyperplane. Furthermore, given a fixed but unknown probability distribution, this hyperplane – called optimal separating hyperplane (OSH) – minimises the risk of misclassifying not only the examples in the training set but also the yet-to-be-seen examples of the test [7].

Yun et al. [8] proposed a method for inspecting solder joints using Support Vector Machines (SVMs) and a tiered circular illumination technique. The illumination technique provides visual information that allows the 3D shape of the solder joint surface to be determined. The extracted features are used to classify the solder joint using an SVM classifier.

In this paper, we propose a robust defect detection system using ultrasonic C-SAM images of IC packaging. First, the
Mumford-Shah model segment C-SAM images of the IC package are obtained, and the segmented images are expressed by Hu’s invariant moments; finally, a decision as to whether the defects exist is then made by the SVM classifier. The rest of the is organised as follows. first, the basic theory of SVM is presented. Then the implementation of an automatic defect recognition system for IC packaging is described. Finally, experimental results using original C-SAM images are illustrated. The final section presents our conclusions.

2 SVM learning method in pattern recognition

2.1 Optimal separating hyperplanes

The basic theory of SVM used in this application is first presented; details can be found elsewhere [7]. Let \((x_i, y_i)_{1 \leq i \leq N}\) be a set of training examples; each example \(x_i \in \mathbb{R}^d\), with \(d\) being the dimension of the input space, belongs to a class labelled by \(x_i \in \{-1, 1\}\). The aim is to define a hyperplane that divides the set of examples such that all the points with the same label are on the same side of the hyperplane. This amounts to finding \(w\) and \(b\) such that

\[
y_i(w \cdot x_i + b) > 0, \quad i = 1, \ldots, N.
\]  

(1)

If there exists a hyperplane satisfying Eq. 1, the set is said to be linearly separable. In this case, it is always possible to rescale \(w\) and \(b\) such that

\[
\min_{1 \leq i \leq N} y_i(w \cdot x_i + b) > 1, \quad i = 1, \ldots, N,
\]

i.e., so that the distance between the closest point and the hyperplane is \(1/\|w\|\). Then Eq. 1 becomes

\[
y_i(w \cdot x_i + b) > 1.
\]  

(2)

Among the separating hyperplanes, the one for which the distance to the closest point is maximal is called the optimal separating hyperplane (OSH). Since the distance to the closest point is maximal is called the optimal separation problem Eq. 3 has been found, the OSH \((w_0, b_0)\) has the following expansion:

\[
w_0 = \sum_{i=1}^{N} a_i^0 y_i x_i = 0,
\]  

(4)

where the support vectors are the points for which satisfy \(a_i^0 > 0\) satisfy Eq. 2 with equality.

Considering the expansion Eq. 4 of \(w_0\), the hyperplane decision function can thus be written as

\[
f(x) = \text{sign} \left( \sum_{i=1}^{N} a_i^0 y_i x_i x + b_0 \right) = 0.
\]  

(5)

2.2 Linearly non-separable case

When the data are not linearly separable, we introduce slack variables \((\xi_1, \ldots, \xi_N)\) with \(\xi_i \geq 0\) to allow the possibility of examples that violate Eq. 2:

\[
y_i(w \cdot x_i + b) > 1 - \xi_i, \quad i = 1, \ldots, N.
\]  

(6)

The purpose of the variables \(\xi_i\) is to allow misclassified points which have their corresponding \(\xi_i > 1\). Therefore, \(\sum \xi_i\) is an upper bound on the number of training errors. The generalised OSH is then regarded as the solution of the following problem: minimize

\[
\frac{1}{2} w \cdot w + C \sum_{i=1}^{N} \xi_i
\]

subjected to constraints Eq. 6 and \(\xi_i > 0\).

The first term in Eq. 7 is minimised to control the learning capacity as in the separable case, and the purpose of the second term is to control the number of misclassified points. The parameter \(C\) is chosen by the user, a larger \(C\) corresponding to assigning a higher penalty to errors.

SVM training requires fixing \(C\) in Eq. 7, the penalty term for misclassifications. When dealing with images, the dimension of the input space is generally large (\(>= 1000\)) compared with the size of the training set, so that the training data are generally linearly separable. Consequently, the value of has in this case little impact on performance. This result has been proved by our experimental results.

2.3 Non-linear Support Vector Machines

The input data are mapped into a high-dimensional feature space through some non-linear mapping chosen a priori. The OSH is constructed in this feature space. If we replace \(\phi(x)\) by its mapping in the feature space, Eq. 3 becomes

\[
w(a) = \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i,j=1}^{N} a_i a_j y_i y_j x_i x_j \Phi(x_i) \Phi(x_j).
\]  

(8)