

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Multi-Classifer Approaches for Post-Placement Surface-Mount Devices Quality Inspection

Stefanos Goumas¹ and Michalis Zervakis²

¹*Technological Educational Institute of Kavala*

²*Technical University of Crete
Greece*

1. Introduction

In the continuing effort to shrink the electronics components and assemblies, the need for streamlined production processes and quality assurance is emerging stronger than ever. Surface Mounted Devices (SMD) is one of the breakthrough techniques that drove printed-circuit board production to a new level, increasing substantially the component density and reducing the size of produced circuits. Quality inspection of SMD is recognized as a critical and complex task in the production process (Bartlet et al., 1988). Speed is also important as to reduce the overall production costs (Lecklider, 2004). Specific SMD defects have been reported in the literature (Loh & Lu, 1999) including component misplacement and absence, component with wrong polarity, solder joint defects and component shifting. Much of the current research efforts are concentrated on detecting solder joint defects. The types of solder joint defects include surplus solder, insufficient solder and lacking solder. Component shifting has also been reported as a special solder joint defect.

Among the methods employed by industrial automatic solder-joint quality inspection systems are laser infrared signatures, digital radiography, laser Doppler vibrometry or laser acoustic microscopy. However, the high cost, low throughput and sampling loss of the above approaches call for research on non-destructive machine vision systems. Several optical inspection systems for solder joints have been reported (Bartlet et al., 1988; Capson & Eng, 1988; T-H. Kim et al., 1999; Ryu & Cho, 1997), using different illumination techniques and defect classification schemes. There have also been efforts to evaluate data fusion to combine data from various sensors for quality inspection of soldering processes (Lacey et al., 1993).

In an earlier work (Zervakis et al., 2004), our effort focused on overcoming the degrading effects of illumination and/or inaccurate measurements by exploiting stochastic modeling of lead displacement and its effects. As part of the aforementioned work, we have provided a novel framework to inspect the placement quality of SMD immediately after they have been placed in wet solder paste on a Printed Circuit Board (PCB). This type of inspection has the advantage that the critical data is available immediately after placement, so no extra time and components are spent on an already faulty PCB. Three measures of quality placement from individual lead images are of general interest, namely overlap, insulation distance and slump gap. Under general geometric conditions and using simple geometric relations, it can be shown that these measures are only affected by the displacement (i.e., shift and rotation)

of the component, relative to its pad region (Goumas et al., 2002; Zervakis et al., 2004). Furthermore, positioning measures can be inferred from quantitative analysis of the inter-lead images. Instead of concentrating in one and every (poorly imaged) lead, we may fuse complementary information from all leads into a Bayesian estimation framework. In this work we attempt to further improve the positioning measurements of individual leads, by means of information fusion. To our knowledge this is the first time higher level (classifier) fusion is applied to the problem of automated solder joint inspection. More specifically, the quantification of positioning measures is viewed as a classification problem, where the lead displacement is inferred from characteristic features associated with image analysis for optical inspection. In order to overcome inaccuracies due to the poor optical quality of the component images, we use a variety of multiple classifiers fusion strategies based on statistical and soft computing methods to improve the performance of the classification task on individual leads.

Hyper classifiers or classifier ensembles have been intensively studied with the aim of overcoming the limitations of primary classifiers (Kittler et al., 1998; Xu et al., 1992). The most often used classifiers fusion approaches include the majority voting (Xu et al., 1992); the weighted combination (weighted averaging) (Kuncheva, 2004); the probabilistic schemes (Kittler et al., 1997; Kittler et al., 1998); various rank-ordered rules, such as the Borda count (Ho et al., 1994; E. Kim et al., 2002); the sum rule (averaging), product-rule, max-rule, min-rule, median rule (Kittler et al., 1998); the Bayesian approach (naïve Bayes combination) (Altincay, 2005; Kuncheva, 2004; Xu et al., 1992); the Dempster-Shafer (D-S) theory of evidence (Denoeux, 1995; Xu et al., 1992); the behavior-knowledge space method (BKS) (Huang & Suen, 1995; Shipp & Kuncheva, 2002); the fuzzy integral (Chi et al., 1996; Kuncheva, 2004; Mirhosseini et al., 1998); fuzzy templates (Kuncheva et al., 1998); decision templates (Kuncheva, 2001, 2004); combination through order statistics (Kang et al., 1997a, 1997b); combination by a neural network (Ceccareli & Petrosino, 1997). In a recent review paper (Oza & Tumer, 2008) a summary of the leading ensemble methods and a discussion of their application to four broad classes of real-world classification problems is provided. In addition, two novel information fusion approaches are presented recently in (Giacinto et al., 2008; Parikh & Polikar, 2007).

From the point of view of training and the form of *input pattern representation*, there are basically two classifier combination scenarios. In the first scenario, all the classifiers use the same representation of the input pattern (*identical pattern representation*). In the second scenario, each classifier uses its own representation of the input pattern (*distinct pattern representation*) (Kittler et al., 1997; Kittler et al., 1998; Mashao & Skosan, 2006; Rodriguez-Linares et al., 2003). In this case, the measurements extracted from the pattern are unique to each classifier, i.e., each individual classifier uses a different set of features. This poses practical encoding and statistical difficulties much in the sense of presenting a number of experts (classifiers) with different representations of the same phenomenon and deriving an unbiased overall decision. In fact one can argue that the primary classifiers that use distinct feature representations are inclined to be more biased than others operating on a common set of features. However it has the added benefit that the classification results tend to be uncorrelated and complementary, thus contributing to improved ensemble accuracy.

The objective of this chapter is to highlight the aspects of classifier fusion in solder-joint inspection and identify under which conditions it improves positioning measurements. Both fusion schemes, using identical and distinct pattern representations are considered. In the former case, the features for classification are obtained directly from the lead images. The

latter scheme uses features that encode “reduced content” of the original images, i.e. the edge structure and the projection profile of leads. Notice that all features are obtained from the same primary source of information, i.e. the lead images. In this form, only a subset of features can be viewed as complementary. In fact, the edge and projection characteristics can be considered as only slightly correlated and, thus complementary, but the set of optical features are highly correlated to the rest. We elaborate on two schemes for distinct pattern representations. In the former scheme we use only reduced dimensionality features (i.e., topological and projection features), whereas the latter enriches the topological and projection features with optical ones, in order to improve the classification rates and robustness across all lead-displacement classes. By combining the power of the individual classifiers through multi-modular architectures we attempt to improve the classification results and enhance the robustness of the overall classification system. Furthermore, through the use of distinct feature representations we test the potential of using and combining reduced-content data towards increasing the speed of inspection.

With respect to the identical pattern representation scheme, we apply four representative schemes for soft fusion of multiple classifiers, ranging from simple majority voting to Bayesian, possibilistic and fuzzy schemes. The hyper classifiers used with the distinct pattern features are non-parametric methods. We use the product, sum, min and max combiners which are among the simplest classifier fusion rules, yet provide adequate results. We should notice here that following the process of classification of individual leads, we can further proceed with the Bayesian estimation approach developed in (Lacey et al., 1993; Zervakis et al., 2004) to accurately estimate component displacements based on the measurements from many individual leads.

The rest of this chapter is organized as follows. Section 2 involves a brief description of the experimental set up. In Section 3 we present all feature extraction schemes from individual lead images, along with the corresponding primary classifiers utilized. The formulation of various classifier combination strategies is described in Section 4. The results of the classifier fusion methods applied on SMD post placement quality inspection are presented in Section 5. Section 6 concludes this comparative study on several classifier combination methodologies with relevant observations and future research directions.

2. Experimental set up

A high dynamic range CMOS camera equipped with a simple LED illumination device and a general purpose processor performs the acquisition of the component image. For the purpose of presenting our results, QFP (Quad Flat Pack) SMD components with 120 leads (30 leads per side) are employed. Following image acquisition, the four regions of interest (ROIs), one at each side of the component, are isolated and all 120 small lead-images are extracted. The density of the CMOS sensor is 1024x1024 pixels, deriving an image resolution of 20x20µm per pixel. To capture the entire area of interest around each lead the size of the lead images is set to 36x56 pixels. Each lead image captures information about four material areas of interest, namely lead, pad, paste and background. Notice that the extraction of lead images can be easily customized to any conventional SMD component. Our approach aims to estimate each lead displacement over its ideal position centered at the pad/paste region. For the purposes of estimating other quality measures, only the displacement along the side of the component is essential (Lacey et al., 1993). Thus, our problem is restated as estimating lead displacement on the direction perpendicular to the lead axis. Essentially, we consider

quantized displacement estimations organized at multiples of a pixel displacement. The displacement classes considered are $\{-6, -3, 0, +3, +6\}$ and $\{-6, -4, -2, 0, +2, +4, +6\}$, in pixel displacements over the lead's central position.

Visual inspection techniques proceed by extracting features from segmented lead images. An example of input lead image and its segmented version is illustrated in Fig. 1.

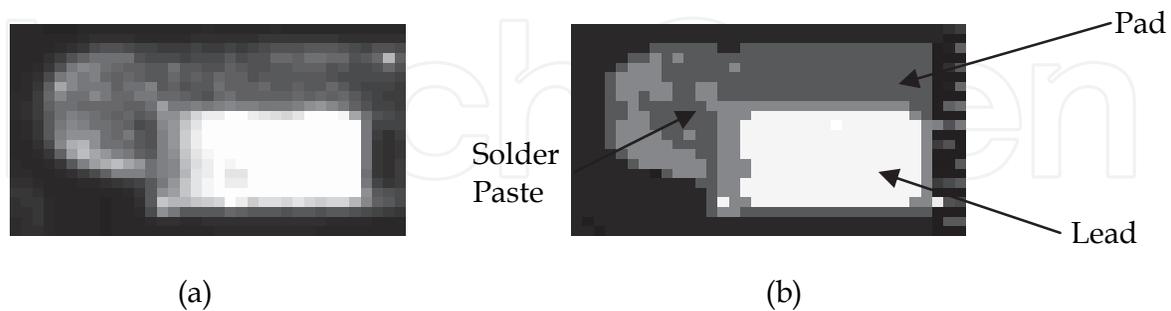


Fig. 1. Segmentation of lead image based on the 4-level Otsu algorithm: (a) original lead image, (b) segmented lead image.

In this case a four-level Otsu algorithm (Otsu, 1979) is applied on each ROI as to segment the lead images that are included in the examined ROI image. The 4-level Otsu-segmented image has been further processed by region growing/merging, labeling and line fitting approaches at the lead sides to arrive at the segmented result e.g. as in Figure 1b. The outcome of the segmentation algorithm is a four-level image that corresponds to the regions of lead, pad, solder paste and background. A labelling algorithm that relies on certain criteria, such as intensity, shape, location, and size, is subsequently applied in order to define (label) the four areas of interest.

3. Feature extraction and primary classifiers

Three sets of features are extracted from each segmented lead image, which encode different characteristics of this image. The first set encodes optical characteristics, by means of simple area measures that sustain the most desirable image attributes. The second set reflects only the edge information, whereas the third set pertains to features derived from the one-dimensional projection profile of the lead image in one direction. Even though all three sets are eventually derived from the same primary source (i.e. the original lead image), each set reflects different attributes of this source. The optical set may be seen as reflecting information from its full representation, whereas the latter two capture attributes of a reduced-attribute representation of the source. In fact, the second set is based on a reduced dynamic range representation of the binary edges, whereas the third set is based on a reduced dimensionality representation of the optical image through its intensity projection on only a single spatial dimension. In this form, the first set may be viewed as a "complete" characterization of the source image, whereas the other two can be interpreted as "incomplete", uncorrelated and complementary attributes of the same source. The last two approaches have been extensively evaluated in (Goumas et al., 2004).

3.1 Optical features

The feature extraction process is geared towards a single lead region. Our method requires only the roughly segmented lead, as in Fig. 2, which is simply enclosed by its bounding

rectangle. For each lead we define two sub-regions presented in Fig. 2, based on the bounding rectangle. One region (L1) concerns the area where the lead is located and the other (L2) spans the pad area in front of the lead outwards the component. The area backwards the lead towards the body of the component is disregarded, since it contains misleading (non-useful) information. The features of each sub-region are appropriately normalized to the length of the corresponding region, in order to make them independent of the axial (u-direction) shift of the lead within the area of its pad.

From lead sub-region-1 (L1) we extract 7 features, which are the area of pad, the area of solder paste, the center-of-gravity distance on v axis between all (non-background) areas and the lead, the center-of-gravity distance on v axis between solder paste and lead, the center-of-gravity distance on v axis between solder paste and pad, the pad mean width on v axis, the pad total length on u axis. Furthermore, from lead sub-region-2 (L2) we extract the following 5 features: the area of pad, the area of solder paste, the center-of-gravity distance on v axis between all non-background regions and the lead, the center-of-gravity distance on v axis between solder paste and pad, and the elongation of pad.

For more details regarding the computation of the aforementioned features, the interested reader is referred to (Jain et al., 1995). The above 12 features constitute a feature vector for pattern classification of each lead. Any classifier can be utilized to perform this task. In our work we use a Bayes classifier, a multilayer Perceptron (MLP) neural network classifier and a learning vector quantization (LVQ) neural network classifier as primary classifiers for optical features.

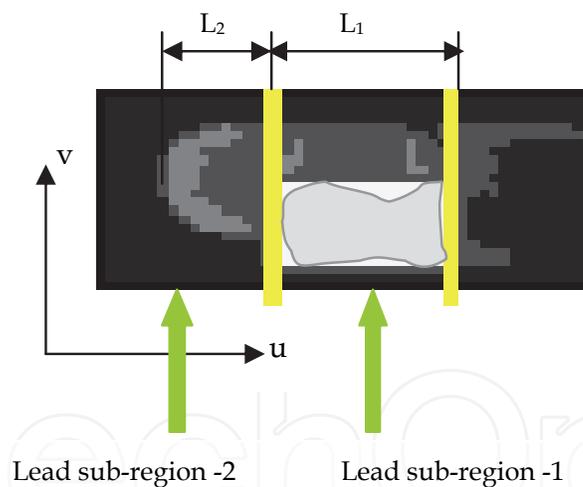


Fig. 2. The lead sub-regions used in the feature extraction process.

3.2 Reduced dynamic-range features

In our first approach related to data-space reduction, we utilize the edge structure extracted from the input lead image for classification purposes. We employ a Laplacian edge detector followed by simple thresholding. In most cases, the derived edge structure is partially deformed or destroyed. Thus, the major task is to relate edge patterns so that we can recall a class assignment for each test pattern that may be presented for classification. We exploit the concept of associative memories (AMs) as stored patterns representing the desirable classes and the Hamming distance for quantifying the distance between the input pattern and each

one of the stored memories (Fausset, 1994). For classification of input patterns we employ the Hamming network, which is used to determine the proximity of an input vector to several exemplar vectors or prototype patterns. An input pattern that partially resembles the stimulus of an association invokes the associated response pattern by means of the shortest Hamming distance. Thus, an associative memory can retrieve a stored pattern given a reasonable subset of information embedded in that pattern. Moreover, an associative memory is error correcting in the sense that it can override inconsistent information in the cues presented to it. The input pattern to the network is a binary edge pattern obtained from the grayscale input lead image through segmentation and edge detection. The stored AM patterns reflect the edge structure of the “typical” edge image representing each class of lead displacements. Thus, the reduced-dimensionality binary edge image is fed into the Hamming network to determine pattern similarities and implement the desirable classifier. The classifier is trained for 5 and 7 classes, corresponding to integer lead displacements from -6 to +6 pixels per 3 and 2 pixel displacements, respectively. Its operation aims at selecting one of the stored patterns (or classes) that is at a minimum Hamming distance (HD) from the binary input vector. The Hamming network consists of two layers. The first layer calculates the M distances between the input vector \mathbf{p}^{probe} and the stored $\mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^M$ fundamental memories in a feed-forward pass. The strongest response of neurons in this layer is indicative of the minimum HD between the input and the fundamental memories. In our implementation the input in Hamming neural network is a binary image $36 \times 56 = 2016$ pixels. Thus, the input vector of Hamming neural network has dimension 2016, i.e., the first layer of Hamming neural network is constituent of 2016 neurons. The second layer of the Hamming network is a winner-take-all network (MAXNET), implemented as a recurrent network. The MAXNET's ε parameter was set to $\varepsilon = 0.0385$. The MAXNET suppresses all of its input values except the one at the maximum node of the first layer.

Given a set of binary prototype (exemplar) vectors $\mathbf{p}_i^k, i = 1, \dots, N$ and $k = 1, \dots, M$, the operation of the Hamming network is summarized as follows.

1. For storing the M prototype vectors, initialize the weights:

$$w_{ij} = \frac{p_i^j}{2}, (i = 1, \dots, N; j = 1, \dots, M)$$

and the bias terms:

$$b_j = \frac{N}{2}, (j = 1, \dots, M)$$

2. For each unknown N -dimensional input vector \mathbf{x} compute the input to each unit Y_j of second layer:

$$Y_j = b_j + \sum_{i=1}^N w_{ij} x_i(t), (j = 1, \dots, M)$$

3. Initialize activations for MAXNET:

$$y_j(0) = Y_j, (j = 1, \dots, M)$$

4. MAXNET iterates to find the best-match exemplar pattern based upon the equation:

$$y_j(t) = f \left(y_j(t-1) - \varepsilon \sum_{k \neq j} y_k(t-1) \right)$$

where f is the activation function : $f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$ and ε is a small parameter $0 < \varepsilon < \frac{1}{M}$.

In our application we set $\varepsilon = \frac{1}{2} \left(\frac{1}{M} \right) = 0.0385$.

Since we exploit the concept of associative memory, the input pattern must have a structure similar to its closest one of fundamental memories. In order to enforce such pattern similarity, we fix the location of the lead in both the test image and the fundamental memories, so that large pattern differences in the comparison of two images can only be attributed to different shifts of the outside boundaries over the fixed lead location.

An important issue of associative memories is the definition of its *fundamental memories*. Each fundamental memory comprises the specific characteristics for discriminating its class. Moreover, the fundamental memories used in lead displacement must assess the standard characteristics of the problem, such as same image size, uniform lead position, etc. To satisfy these requirements, we first select the memory for one displacement (0 pixels) and then construct the memories associated the other classes by shifting the outside edge structure with respect to the fixed structure of the lead. The basic fundamental memory at shift 0 is selected from a number of test images reflecting exactly this specific case through statistical analysis of the mean pattern in this class. The resulting memories are depicted in Fig. 3.

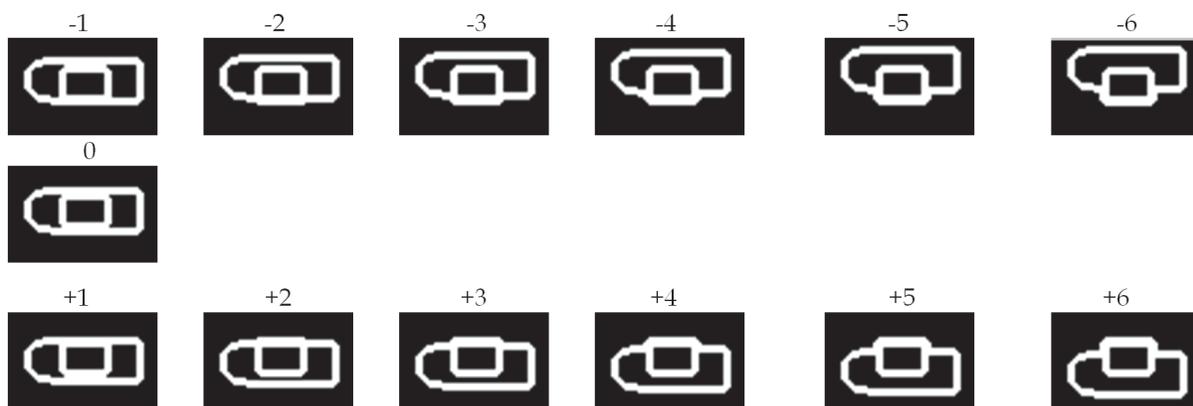


Fig. 3. Fundamental memories after dilation.

Based now on the design of the fundamental memories, we train the network so that it recovers the closest stored pattern in response to each test-input. An operation example of the associative memory in the case of a test image with +3 pixels lead-shift is illustrated in Fig. 4.



Fig. 4. Associative memory operation: (a) test input image, (b) output response.

3.3 Reduced input-dimension features

In this approach we exploit the structure of the lead image profile (projection) along one, the most descriptive direction vertical to the lead axis for extracting meaningful features related to displacement measurements. The important component of this classification scheme is its feature extraction unit. We propose a complete feature extraction and classification approach that consists of three distinct modules. The first module receives the lead projection function at its input and utilizes a nonlinear filter based on a high-order neural network (HONN) for feature extraction. The second module implements feature reduction and de-correlation of the feature space by using the Karhunen-Loeve transform (KLT). The third module comprised by the Bayes classifier serves as a classifier that assigns each feature vector to one of the predetermined classes for lead displacements.

HONNs are fully interconnected single-layer networks, containing high order connections of sigmoid functions in their neurons. If we define as x, y its input and output respectively, with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ the input-output representation of a HONN is given by:

$$y = \mathbf{W}^t \mathbf{S}(x) \quad (1)$$

where \mathbf{W} is a $q \times m$ matrix of adjustable synaptic weights and $\mathbf{S}(x)$ is a q -dimensional vector of sigmoids. For sufficient high order terms, there exist weight values $\hat{\mathbf{W}}$ such that the HONN structure $\hat{\mathbf{W}}^t \mathbf{S}(x)$, can approximate an unknown function $f(x)$ to any degree of accuracy, in a compact domain (Rovithakis et al., 2001).

The KLT is used to de-correlate and reduce the dimensionality of feature vectors, disjoint class spaces in the new (reduced) feature space and aid the classifiers in performing accurate discrimination. The KLT projects the feature vector to the K most important directions. In essence, the KL transform projects feature vectors on the directions that best preserve class properties. Two different forms of the KLT are studied. In the first form only one KLT matrix (1 KLT) is created for the entire data set, whereas in the second form a KLT matrix is created for each class (each displacement). Thus, the first approach computes the most significant directions of the entire problem space and preserves directions where the data set expresses the largest diversion. In the second approach each individual class is represented by its most significant directions. Thus, it encompasses class specific characteristics and uses them to better isolate and discriminate classes by avoiding class mixing in irrelevant directions. The theoretical background of the multiple KLT approach is given in (Cappeli et al., 2001), whereas its application as a general analytic tool is established in (Goumas et al., 2002).

3.3.1 HONN based feature extraction

The HONN based feature extraction module receives as input a normalized projection function of the tested lead image (as in Fig. 5) and updates its weights by stable Lyapunov learning laws as to approximate this input function. Prior to entering, the input function is linearly transformed in the range $[0, 1]$, as to avoid the appearance of destabilizing

mechanisms caused by purely numeric issues, (i.e., large variations in the image projection data). Moreover, for uniformity reasons the rising point of this function is shifted to the origin. The position of the lead on the pad region determines the location of the main lobe in the projection function. Hence, the location of the main lobe and the overall structure of the projection function become the main characteristics that can be exploited for classifying the lead shift.

In the following we study the construction of the feature extraction system through an approximate modeling of the projection function and we rigorously analyze its performance. Let $x \in \mathbb{R}_+$ be the data point on the projection axis, $y \in \mathbb{R}_+$ be the projection value of lead image (\mathbb{R}_+ denotes the set of positive real numbers), and f represent the actual but unknown projection function. Obviously the projection profile is modeled as a function $y = f(x)$. Moreover, let $\hat{y} = \mathbf{W}^t \mathbf{S}(x)$ be a HONN approximation of the actual projection function $f(x)$. Due to the one-dimensional structure of the problem, the HONN is designed for scalar input/output pairs linked at a higher dimension with a weight vector \mathbf{W} . Define the projection approximation error as

$$e = f(x) - \mathbf{W}^t \mathbf{S}(x) = y - \hat{y} \quad (2)$$

Observe that e is directly measured even though $f(\cdot)$ is unknown. The nonlinear adaptive filter

$$\dot{z} = -\alpha z + y - \mathbf{W}^t \mathbf{S}(x), \quad \alpha > 0, z \in \mathbb{R} \quad (3)$$

equipped with the update law

$$\dot{\mathbf{W}} = -\gamma \mathbf{W} + z \mathbf{S}(x), \quad \gamma > 0 \quad (4)$$

guarantees the uniform ultimate boundedness of its output $z \in \mathbb{R}$ with respect to the arbitrarily small set

$$Z = \left\{ z \in \mathbb{R} : |z| \leq \frac{\varepsilon}{2\alpha} + \frac{1}{2} \sqrt{\left(\frac{\varepsilon}{\alpha}\right)^2 + \frac{2\gamma |\hat{\mathbf{W}}|^2}{\alpha}} \right\} \quad (5)$$

as well as the boundedness of the optimal HONN weights $\hat{\mathbf{W}} \forall x \geq 0$ (Rovithakis et al., 2001). In the aforementioned relations α, γ are design constants and $\varepsilon \geq 0$ is an unknown but small bound on the HONN reconstruction error.

After convergence of the HONN, the feature vector \mathbf{F} is formulated by the vector of trained weights \mathbf{W} augmented with the approximation error e . In our approach, we form the feature vector \mathbf{F} to be

$$\mathbf{F} = \begin{bmatrix} \mathbf{W} \\ e \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ e \end{bmatrix}, \quad \text{where } \mathbf{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} \text{ is the HONN weights vector (of specific dimension}$$

$N=12$) and e is the approximation error. This selection allows \mathbf{F} to encode all HONN

variables that characterize the projection function. An obvious feature is the approximation error e . Furthermore, since the HONN possesses a linear-in-the-weights property, the existence of a unique optimal vector \hat{W} different for each different projection function is guaranteed. Thus, the weights vector \hat{W} also serves the purpose of a relevant feature.

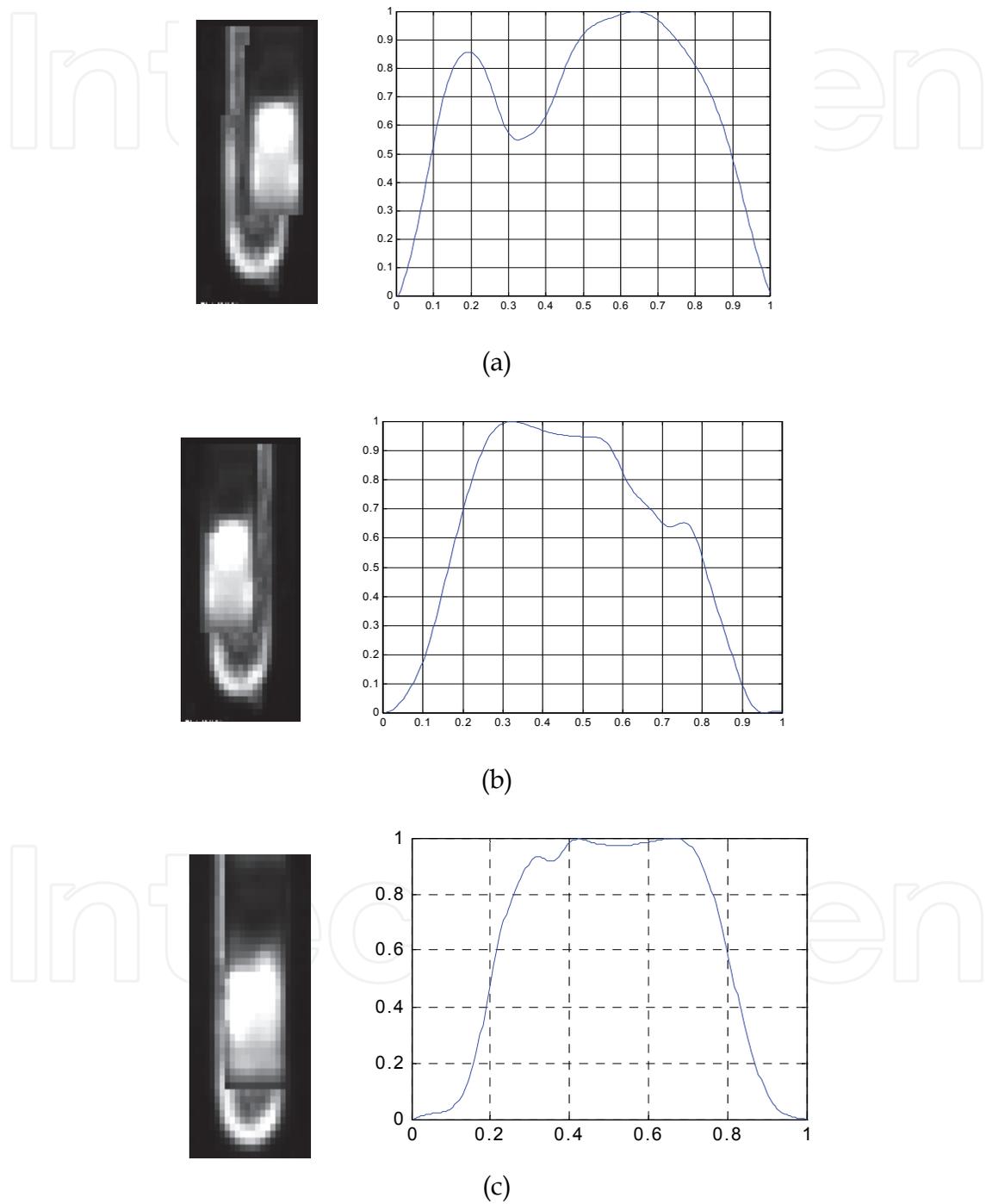


Fig. 5. Original lead images and projection functions for (a) +3 pixels, (b) -3 pixels, (c) zero pixels lead shift.

To approximate the unknown projection function the following HONN structure is used:

$$y = \mathbf{W}^T \mathbf{S}(x) = \sum_{i=1}^3 w_i s_1^i(x) + w_4 s_2^4(x) + \sum_{i=5}^8 w_i s_3^{(i-4)}(x) + \sum_{i=9}^{12} w_i s_4^{(i-8)}(x)$$

with

$$s_1(x) = \frac{0.9571}{1 + e^{-35.703(x-0.076)}} + 0.2245, \quad s_2(x) = \frac{0.3838}{1 + e^{-0.3598(x-1.488)}} - 0.2607$$

$$s_3(x) = \frac{0.9625}{1 + e^{-22.4438(x-0.7927)}} + 0.5625, \quad s_4(x) = \frac{1.2906}{1 + e^{-51.468(x-0.3287)}} - 0.3572.$$

The HONN weights are updated according to:

$$\dot{w}_i = -0.000534w_i + z s_1^i(x), \quad i = 1, 2, 3, \quad \dot{w}_4 = -0.000756w_4 + z s_2^4(x),$$

$$\dot{w}_i = -0.000825w_i + z s_3^{(i-4)}, \quad i = 5, 6, 7, 8, \quad \dot{w}_i = -0.000407w_i + z s_4^{(i-8)}, \quad i = 9, 10, 11, 12.$$

The parameter α that appears in (5) is fixed to $\alpha = 8.0913$ through the use of a genetic algorithm. For the training of HONN, which derives 13 features from projection profiles, the network architecture involves 13 neurons. The use of Karhunen-Loeve transform (KLT) reduces the dimensionality down to 11.

4. Multiple classifier combination methods

4.1 Formulation of the combined classifier problem

In this chapter we assume that a small set of trained classifiers is available operating on the same dataset and we are interested in combining their outputs aiming at the highest possible accuracy. Let $C = \{C_1, C_2, \dots, C_K\}$ be a set of classifiers and $\Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$ be a set of class labels. Each classifier gets as input a feature vector $\mathbf{x} \in \mathbb{R}^n$. The classifier output is an M -dimensional vector $\mathbf{y}_i = C_i(\mathbf{x}) = [c_{i,1}(\mathbf{x}), \dots, c_{i,M}(\mathbf{x})]^T$, where $c_{i,j}(\mathbf{x})$ is the degree of "support" given by classifier $C_i, i = 1, \dots, K$ to the hypothesis that \mathbf{x} comes from class $\omega_j, j = 1, \dots, M$. Without loss of generality we can restrict $c_{i,j}(\mathbf{x})$ within the interval $[0, 1]$ and call them "soft labels", with 0 meaning "no support" and 1 implying "full support", $i = 1, \dots, K, j = 1, \dots, M$ (Shipp & Kuncheva, 2002). Most often $c_{i,j}(\mathbf{x})$ is an estimate of the posterior probability $P(\omega_j | \mathbf{x})$. The process of combining classifiers attempts to combine the K classifier outputs $C_1(\mathbf{x}), \dots, C_K(\mathbf{x})$ as to obtain a soft label for \mathbf{x} , denoted $C(\mathbf{x}) = [\mu_1(\mathbf{x}), \dots, \mu_M(\mathbf{x})]^T$, where $\mu_j(\mathbf{x})$ denotes the overall degree of support for ω_j

given by the ensemble classifier. If a crisp class label of \mathbf{x} is needed, we can use the maximum membership rule, which assigns \mathbf{x} to class ω_s if,

$$c_{i,s}(\mathbf{x}) \geq c_{i,j}(\mathbf{x}) \quad \forall j = 1, \dots, M, \quad i = 1, \dots, K, \quad \text{for individual crisp labels and} \quad (6)$$

$$\mu_s(\mathbf{x}) \geq \mu_l(\mathbf{x}), \quad \forall l = 1, \dots, M, \quad i = 1, \dots, K, \quad \text{for the final crisp label} \quad (7)$$

The minimum-error classifier is recovered from (7) when $\mu_j(\mathbf{x}) = P(\omega_j | \mathbf{x})$. In the following we briefly address more advanced combination methods used in this chapter.

4.2 Non-trainable and probabilistic combination schemes for identical pattern representations

This type of combiners constitutes a group of simple, yet often surprisingly effective, methods for fusing the primary classifiers' soft labels. Their key advantage, apart from speed, is the fact that, having no tunable parameters, they do not impose a second training phase on the model. Additionally they belong to the group of class-conscious hyper classifiers since they utilize only one column of the ensemble's Decision Profile (Kuncheva, 2004). More specifically non-trainable combiners use only soft labels $c_{1,j}(\mathbf{x}), c_{2,j}(\mathbf{x}), \dots, c_{k,j}(\mathbf{x})$ corresponding to class ω_j to estimate the fused maximum support value $\mu_j(\mathbf{x})$ for this class. A combination function F maps the primary to the fused labels (Kuncheva, 2004)

$$\mu_j(\mathbf{x}) = F \left[c_{1,j}(\mathbf{x}), c_{2,j}(\mathbf{x}), \dots, c_{k,j}(\mathbf{x}) \right] \quad (8)$$

Some popular choices for the functional F are the sample mean (average), min, max, median and the product rules. Furthermore, majority voting is a popular and easy to implement method (Xu et al., 1992). The primary classifiers "vote" with their class labels and the class label with most votes is assigned to \mathbf{x} .

Whereas the voting method only considers the result of each classifier, the approach of Bayesian formalism (E. Kim et al., 2002; Kuncheva, 2004) considers the error of each classifier. The "naïve Bayes" scheme assumes that the classifiers are mutually independent given a class label (conditional independence).

Consider the crisp class labels obtained from the K classifiers and let L_1, \dots, L_K be the class labels assigned to \mathbf{x} by classifiers $C_1(\mathbf{x}), \dots, C_K(\mathbf{x})$, respectively and denote by $P(L_j)$ the probability that classifier C_j labels \mathbf{x} in class $L_j \in \Omega$. The conditional independence allows for the following representation

$$P(\mathbf{L} | \omega_k) = P(L_1, \dots, L_K | \omega_k) = \prod_{i=1}^K P(L_i | \omega_k) \quad (9)$$

Then, the posterior probability needed to label \mathbf{x} is given by

$$P(\omega_k | \mathbf{L}) = \frac{P(\omega_k)P(\mathbf{L} | \omega_k)}{P(\mathbf{L})} = \frac{P(\omega_k) \prod_{i=1}^K P(L_i | \omega_k)}{P(\mathbf{L})}, \quad k = 1, \dots, M \quad (10)$$

Since the denominator does not depend on ω_k and can be ignored, the support for class ω_k by the set of classifiers can be computed as

$$\mu_k(\mathbf{x}) \propto P(\omega_k) \prod_{i=1}^K P(L_i | \omega_k) \quad (11)$$

Naïve Bayes fusion is applied as follows. Assuming M classes labeled 1 through M , the error for the i th classifier, $i = 1, \dots, K$, can be represented by a two-dimensional *confusion matrix* as follows:

$$CM^i = \begin{pmatrix} \alpha_{1,1}^i & \dots & \alpha_{1,M}^i \\ \vdots & \ddots & \vdots \\ \alpha_{M,1}^i & \dots & \alpha_{M,M}^i \end{pmatrix} \quad (12)$$

For each classifier C_i , a $M \times M$ confusion matrix CM^i is calculated by applying C_i to the training data set. The (k,L) th entry of this matrix, $\alpha_{k,L}^i$ is the number of elements of the data set whose true class label was ω_k and were assigned to class ω_L by C_i . By N_L we denote the total number of elements of \mathbf{S} that truly belong to class ω_L . Taking $\alpha_{k,L}^i / N_k$ as an estimate of the probability $P(L_i | \omega_k)$, and N_k / N as an estimate of the prior probability of class ω_k , Eq. (11) is equivalently written as:

$$\mu_k(\mathbf{x}) \propto \frac{1}{N_k^{K-1}} \prod_{i=1}^K \alpha_{k,L_i}^i \quad (13)$$

4.3 Multi-classifier combination based on fuzzy Integral

The theory of fuzzy measures and fuzzy integrals was first introduced by Sugeno (Mirhosseini et al., 1998) and has been successfully used in decision fusion (Chi et al., 1996). Contrary to fuzzy sets, in fuzzy measures a value is assigned to crisp subset of the universal set signifying the degree of evidence or belief that a particular element belongs in the subset. In this form, fuzzy measures are used to solve *ambiguity* associated with making a choice between two or more alternative decisions. In classifier fusion, the fuzzy measure relates to a measure of competence of each classifier. The ensemble support $\mu_j(\mathbf{x})$ for class ω_j , $j = 1, \dots, M$ is obtained from the support values of individual classifiers $c_{i,j}(\mathbf{x})$, $i = 1, \dots, K$, but also taking into account the competences of experts expressed

through a *fuzzy measure* denoted by $g_i(\mathbf{x}), i = 1, \dots, K$. This form of fusion is implemented by means of a *fuzzy integral*. The Choquet fuzzy integral is often utilized, which is based on Sugeno's λ -fuzzy measure (Chi et al., 1996; Mirhosseini et al., 1998).

4.4 Multi-classifier combination based on Dempster–Shafer theory of evidence

The Dempster–Shafer theory of evidence (Xu et al., 1992), also known as the theory of belief functions, is a generalization of the Bayesian theory for subjective probability. This theory is more flexible than Bayesian when our knowledge is incomplete and we have to deal with uncertainty and ignorance. Belief functions allow us to assign degrees of belief (or evidence) for one event (i.e. support for a class from one classifier) based on evidence for a related event (i.e. support of this class from another classifier). In the context of measurement-level classifier combination, a method for evidence combination is presented in (Rogova, 1994) and is also adopted here.

For the input vector \mathbf{x} the output of the i th classifier is $\mathbf{y}_i = C_i(\mathbf{x}), i = 1, \dots, K$. Let $\{\mathbf{t}_j\}$ be the training set for class ω_j and $\mathbf{r}_{i,j}$ be the mean output vector of the i th classifier on this training set. The support function for class ω_j by the classifier C_i can be obtained by using the Euclidean distance between $\mathbf{r}_{i,j}$ and \mathbf{y}_i :

$$d_{i,j} = \varphi(\mathbf{r}_{i,j}, \mathbf{y}_i) = \frac{\left(1 + \|\mathbf{r}_{i,j} - \mathbf{y}_i\|^2\right)^{-1}}{\sum_{k=1}^M \left(1 + \|\mathbf{r}_{i,k} - \mathbf{y}_i\|^2\right)^{-1}} \quad (14)$$

and the overall evidence for class j from this classifier is computed as:

$$e_j(\mathbf{y}_i) = \frac{d_{i,j} \prod_{k \neq j} (1 - d_{i,k})}{1 - d_{i,j} \left[1 - \prod_{k \neq j} (1 - d_{i,k})\right]} \quad (15)$$

Finally, evidences for all classifiers may be combined (according to a simplified Dempster–Shafer rule) to obtain a measure of confidence for each class ω_j for the feature vector \mathbf{x} as

$$e_j(\mathbf{x}) \propto \prod_{i=1}^K e_j(\mathbf{y}_i), \text{ so that we may assign class } \omega_k \text{ to the feature vector } \mathbf{x} \text{ if}$$

$$e_k(\mathbf{x}) = \max_{j=1}^M \{e_j(\mathbf{x})\}.$$

4.5 Combination rules for distinct pattern representations

The case of distinct pattern representation poses an additional burden to the design of the combination rules, since the representations of information (features) are quite

inhomogeneous. Nevertheless, based on a Bayesian framework for relating the available information similar simple rules can be derived for classifier combination under distinct pattern representations. Assume that K classifiers are available, each representing the given pattern by a distinct feature vector. We consider K conditionally independent feature subsets (or distinct pattern representations). Each subset defines a part of the feature vector, $\mathbf{x}^{(i)}$, so that $\mathbf{x} = \left[\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(K)} \right]^T$, $\mathbf{x} \in \mathbb{R}^n$. Notice that there is a one-to-one correspondence between each feature vector $\mathbf{x}^{(i)}$ and its underlying classifier C_i , $i = 1, \dots, K$. Using a Bayesian framework, Kuncheva (Kuncheva, 2004) derives the posterior class probability using the entire information from all representations as:

$$P(\omega_j | \mathbf{x}) \propto P^{(1-K)}(\omega_j) \prod_{i=1}^K P(\omega_j | \mathbf{x}^{(i)}) \quad (16)$$

so that we can assign

$$\mu_j(\mathbf{x}) = \hat{P}^{(1-K)}(\omega_j) \prod_{i=1}^K c_{i,j}(\mathbf{x}^{(i)}) \quad (17)$$

Using product expansion, we can further expand this assignment as in (Kittler et al., 1998) to derive the sum combination rule:

$$\mu_j(\mathbf{x}) = \hat{P}(\omega_j) (1 - K) + \sum_{i=1}^K c_{i,j}(\mathbf{x}^{(i)}) \quad (18)$$

Furthermore, for equal prior probabilities, the assignment of Eq. (17) can be reduced to:

$$\mu_j(\mathbf{x}) = \prod_{i=1}^K c_{i,j}(\mathbf{x}^{(i)}) \quad (19)$$

and the sum combination rule (18) can be viewed as the average a posteriori probability for each class over all the classifier outputs (Kittler et al., 1998) :

$$\mu_j(\mathbf{x}) = \frac{1}{K} \sum_{i=1}^K c_{i,j}(\mathbf{x}^{(i)}) \quad (20)$$

The aforementioned combination rules (17) and (18) or their simplified versions (19) and (20) constitute the fundamental schemes for combining classifiers, each representing the given pattern by a distinct feature vector. Some additional non-trainable fusion strategies can be developed from these rules by considering the inequalities:

$$\prod_{i=1}^K c_{i,j}(\mathbf{x}^{(i)}) \leq \min_{i=1}^K \{c_{i,j}(\mathbf{x}^{(i)})\} \leq \frac{1}{K} \sum_{i=1}^K c_{i,j}(\mathbf{x}^{(i)}) \leq \max_{i=1}^K \{c_{i,j}(\mathbf{x}^{(i)})\} \quad (21)$$

The relationship (21) suggests that the product and sum combination rules can be approximated by their upper or lower bounds deriving the max and min combination rules, respectively.

5. Experimental results

5.1 Generation of test data

In this research we present two types of results. The first one deals with simulated data operating in an external leave-one-out validation scheme. The results presented show the average accuracies attained for each lead displacement through this recursive cross validation scheme. The second type of results refers to testing on the real data. The training of classifiers is performed on the entire set of simulated data, whereas testing is performed on the completely independent set of real images. For the generation of simulated data we use the *Monte Carlo simulation* process (Bremaud, 1999; Fishman, 1996; W. Martinez & A. Martinez, 2002; Robert & Casella, 1999), in order to generate lead samples with appropriate size and intensity distributions for training the classifiers. Our purpose is to simulate after placement the four Regions of Interest (ROIs), namely left, right, bottom and top ROI of an entire QFP component consisting of four sides with 30 leads on each side.

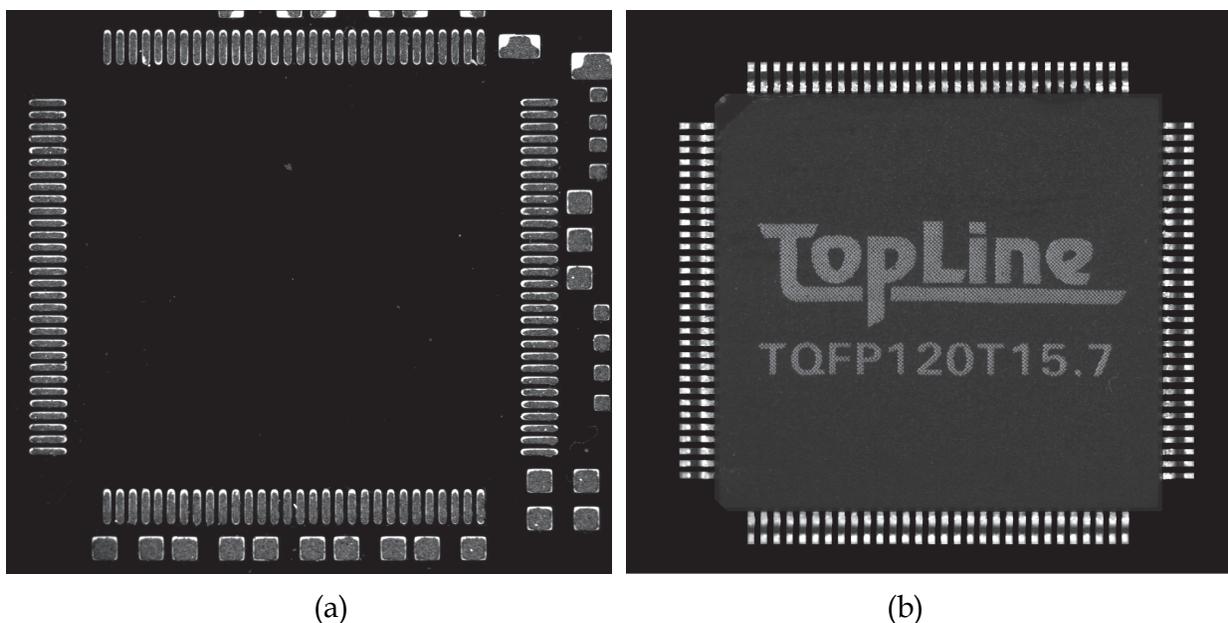


Fig. 6. Pad and QFP component images; (a) pad and smeared solder paste, (b) the component QFP 120.

The data available consist of one set of 4 actual images of pad and smeared solder paste and another set of 5 actual images with only the component QFP 120 in front of a dark background (as in Fig. 6a and b, respectively). Thus, the process of simulation of new images with controlled component translations is based on the constituent images of individual pad regions with solder paste and individual lead regions. By superimposing individual lead images on pad regions (as in Figures 7b and 7a, respectively) we can control both the displacement and varying illumination conditions of the actual placement environment.

It is apparent that the leads inside a ROI may have varying dimensions and intensity levels. Similarly, the pads may have varying intensity levels due to the different distribution of the smeared solder paste. In order to simulate these varying factors of a ROI, we first estimate the distributions of these characteristics. For every lead-image we estimate length, height and mean intensity and for every pad-image we compute mean intensity. The distributions

of these characteristics are used in the Monte-Carlo process that each time extracts a lead and pad image controlled by these distributions. Then, the displacement is implemented by shifting the lead on the pad region accordingly.

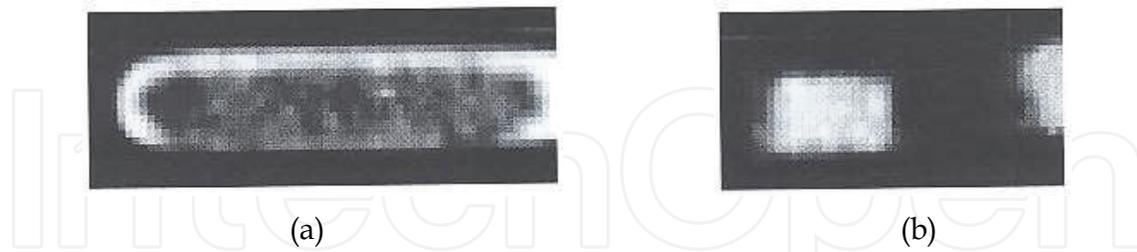


Fig. 7. Individual pad and lead images: (a) individual pad, (b) individual lead.

The Monte Carlo process simulates variable size and illumination conditions for individual leads and implements entire component displacements on the pad regions, which are then employed in training. For class labeling we define 13 classes of component displacements i.e., $\{-6, -5, \dots, 5, 6\}$ pixels and each displacement involves three neighbouring values of displacement in simulations; e.g., class -4 involves displacements $\{-4.2, -4, -3.8\}$. Both directional displacements have been considered, namely horizontal and vertical.

From the total set of 13 classes we form two groups for training and testing the classifiers, composed of 5 and 7 classes, respectively. These two cases study the ability of the classifiers to discriminate classes in the feature space separated by 3 and 2 pixels apart, respectively. We do not consider training on all 13 classes, since all these classes are hardly separable in the feature spaces defined. A larger training sample size might improve class bounds and allow 13-class mapping. By increasing the target displacement step from 1 pixel (13 classes) to 2 pixels (7 classes) to 3 pixels (5 classes), we aim at balancing the classifiers' mapping capability to the overall displacement estimation accuracy. This means that although it is unfeasible to estimate small pad shifts using the given dataset, it is useful to derive larger shift estimates with high confidence. All the above apply for the given resolution level of the current dataset.

For testing with real images, a set of 20 real component images are kindly provided from the actual placement environment of Philips, The Netherlands. Ten actual boards with different shifts are provided, with two images from each case. Each individual case is controlled by the placement machine and conveys the limited accuracy of placement. The sides of each component are located and the individual lead areas are extracted. These images are used for testing of our developed algorithms; the training stage of classifiers is performed with the simulated data.

5.2 Classification results obtained from primary classifiers

For the simulated data, the training set consists of 120 lead images randomly generated from each class. To overcome the problem of statistical significance of the results caused by the rather small data set (for a quite high dimensional feature space) we apply a jack-knifing validation process. Regarding the design of the primary classifiers used in identical feature representation, the LVQ neural network architecture was defined by the feature vector size, training set size and output class mapping. In particular, for use with the 12 geometric (optical) features the LVQ input layer consisted of 12 neurons. In accordance to LVQ theory the hidden competitive layer contained a number of neurons equal to the number of training

set cases. In the output layer for 5 classes (2 pixel shift precision) 5 output neurons were used. Accordingly, discrimination of 7 classes required 7 output neurons. The model was trained for 1000 epochs with a learning parameter $\alpha=0.09$. The MLP neural network was designed with 50 hidden layer neurons and 5 or 7 output neurons depending on the required output classes. The input layer was as above defined by the dimensionality of the feature vector. As stated before, 12 features per lead formulate the feature vector that forms the input to each classifier.

The design of primary classifiers for distinct pattern representation uses the topological features as input vectors to a Hamming neural network. As already mentioned, the first level of Hamming neural network is constituted of 2016 neurons. The second layer of the Hamming network is a winner-take-all network (MAXNET), implemented as a recurrent network. The MAXNET's ϵ parameter was set to $\epsilon=0.0385$. For the case of projection features, which uses 13 features from projection profiles, the training of HONN is performed with a network architecture involving 13 neurons. The use of Karhunen-Loeve transform (KLT) reduces the dimensionality down to 11. These final 11 projection features are used by a non-parametric Bayesian classifier to test the efficiency of this representation.

The classification rates of primary classifiers on 5 classes of simulated lead-images are shown in Table 1a. Table 1b presents the classification rates of individual classifiers on 7 classes. The classification accuracies from the jack-knifing process are presented along with the 95% confidence interval. From the classification results of primary classifiers, we can initially conclude that the Bayes classifier provides better results than the MLP and LVQ classifiers on the 5-classes case. However, competing performances of Bayes and MLP classifiers are observed on the 7-classes case. Furthermore, as can be observed in Tables 1a and 1b, the discrimination between different classes becomes easier as we move to larger displacement intervals; the distinction of 3-pixel difference in Table 1a is more efficient than that of 2-pixel difference in Table 1b. Overall, we observe a large variance of each classifier's performance along the classes of interest.

Features type	Classifier type	- 6 pixels shift	-3 pixels shift	0 pixels shift	+ 3pixels shift	+6 pixels shift
optical	Bayes	97.75 (± 0.73)	94.35 (± 0.88)	94.35 (± 1.35)	93.14 (± 0.86)	98.25 (± 0.75)
optical	MLP	95.32 (± 0.97)	93.87 (± 1.05)	92.28 (± 1.40)	95.83 (± 0.98)	97.63 (± 0.99)
optical	LVQ	93.27 (± 0.60)	90.67 (± 0.76)	78.24 (± 0.74)	95.42 (± 0.80)	94.78 (± 0.32)
topological	Hamming	86.67 (± 1.20)	79.17 (± 1.03)	95.00 (± 1.42)	92.50 (± 1.12)	93.33 (± 0.84)
projection	Bayes	92.50 (± 0.90)	93.33 (± 0.96)	86.70 (± 0.54)	93.33 (± 0.88)	97.63 (± 0.79)

Table 1. a. Classification rates of primary classifiers on 5 classes using Monte Carlo simulated images.

In the sequel we test the primary classifiers on the set of 20 real component images from the actual placement environment. The testing set consists of 120 lead-images obtained from the components of the corresponding class. The classification rates of primary classifiers on 5 classes for the real lead-images are shown in Table 2a, whereas Table 2b presents the classification rates of individual classifiers on 7 classes. As we observe by comparing the results for real and simulated data, there is a small decrease (ranging from 0.30 to 1.30 in different classes) in classification rates for the real data, which are used as an independent test set. Nevertheless, the results on real data are only slightly inferior to those from cross validation, indicating the robustness of developed techniques in realistic operation.

Features type	Classifier type	- 6 pixels shift	- 4 pixels shift	- 2 pixels shift	0 pixels shift	+2 pixels shift	+4 pixels shift	+6 pixels shift
optical	Bayes	93.00 (±0.45)	81.82 (±0.67)	75.72 (±0.75)	85.87 (±0.92)	77.80 (±0.97)	85.87 (±0.70)	91.86 (±0.57)
optical	MLP	91.18 (±0.83)	80.46 (±1.16)	80.36 (±2.94)	79.74 (±1.30)	82.41 (±1.46)	86.83 (±0.90)	88.27 (±0.89)
optical	LVQ	83.47 (±0.56)	76.42 (±1.43)	78.90 (±0.93)	57.06 (±0.13)	80.40 (±0.74)	74.35 (±0.37)	81.16 (±0.30)
topological	Hamming	85.00 (±0.94)	84.17 (±0.74)	77.50 (±0.69)	92.50 (±2.79)	82.50 (±1.85)	82.50 (±1.55)	88.33 (±1.07)
projection	Bayes	76.70 (±0.98)	80.83 (±1.03)	86.07 (±2.01)	68.64 (±0.66)	83.30 (±0.41)	90.00 (±1.91)	93.10 (±1.29)

Table 1. b. Classification rates of primary classifiers on 7 classes using Monte Carlo simulated images.

Features type	Classifier type	- 6 pixels shift	-3 pixels shift	0 pixels shift	+ 3pixels shift	+6 pixels shift
optical	Bayes	96.43	93.61	93.17	92.08	97.33
optical	MLP	94.56	93.19	91.84	94.42	96.37
optical	LVQ	92.73	91.24	79.66	94.77	94.19
topological	Hamming	85.83	78.56	94.13	92.06	92.79
projection	Bayes	91.46	93.76	87.24	92.95	97.18

Table 2. a. Classification rates of primary classifiers on 5 classes using real images.

Features type	Classifier type	- 6 pixels shift	- 4 pixels shift	- 2 pixels shift	0 pixels shift	+2 pixels shift	+4 pixels shift	+6 pixels shift
optical	Bayes	91.87	80.26	74.66	85.26	77.32	85.42	91.03
optical	MLP	90.33	79.74	79.23	78.44	81.92	86.05	87.49
optical	LVQ	83.16	77.28	79.57	56.59	79.63	73.94	80.71
topological	Hamming	84.25	83.79	76.63	91.16	81.59	81.37	87.75
projection	Bayes	75.68	79.96	84.75	67.60	82.45	89.26	92.53

Table 2. b. Classification rates of primary classifiers on 7 classes using real images.

The classifiers designed on reduced dimensionality features achieve lower classification rates than those designed on optical features. There are some exceptions to this general trend, especially associated with the class of zero pixel-shift. Nevertheless, the variance of these classifiers within and among classes is quite large, showing unstable performance. Regarding classification on 7 classes, the Hamming network based on topological (edge) features achieved high rates for the 0 pixel class, but this is rather incidental due to the large variation of results within this class.

On classes corresponding to negative lead shifts we observe slightly worse classification rates than in the positives shifts. This effect is more evident in the cases of reduced dimensionality features. Following extensive experimentation with the data set we conclude that this is attributed to lighting effects during acquisition of the test images, which result in better contrast in one direction.

5.3 Results of combined classifiers using identical pattern representations

The methods of section 4 are used here to combine the three primary classifiers (Bayes, MLP, LVQ) using different methodologies, but operating on the same feature sets (optical). For such a three-classifier combination, majority voting (MV) assigns classification to one class if two or three classifiers produce this same class. Otherwise, the input pattern is rejected. To apply the naïve Bayes (NB) combination, the conditional probabilities $P(L_i | \omega_k)$, $i = 1, 2, 3, k = 1, \dots, 5$ or $k = 1, \dots, 7$ are obtained from the resulting confusion matrices of individual classifiers on the training set. In a same manner, to fuse the results using the Choquet fuzzy integral (CFI), the initial fuzzy densities $g_i, i = 1, 2, 3$, are computed from the resulting confusion matrices of individual classifiers on the training set. The Dempster-Shafer (D-S) fusion is performed in the context of measurement-level classifier combination based on the proposed method in section 4.4.

The classification results obtained from the above four combiners are presented in tables 3a and 3b on 5 and 7 classes, respectively. As we observe from these tables, all combinations of classifiers achieve better performance than any individual classifier used for fusion. Examining deeper their performance, we conclude that the naïve Bayes and the Dempster-Shafer combiners achieve better overall performance than the other schemes, with the naïve Bayes reaching the best performance of all combining classifiers employed. The largest improvement achieved by the combined classifiers over the best individual classifier

performance is also depicted in Tables 3a and 3b for the naive Bayes scheme. In fact, the maximum improvement is achieved by this fusion approach for the class of -3 pixels shift. The advantage of naïve Bayes combiner over the others fusion schemes, along with the advantage of primary Bayes classifier over the others individual classifiers, cannot be generalized. The ranking of classification schemes observed in this application is partially attributed to the stochastic properties of the data set, supporting the assumption that the distribution of our experimental data follows the normal (Gaussian) distribution.

Combiner type	- 6 pixels shift	-3 pixels shift	0 pixels shift	+ 3pixels shift	+6 pixels shift
MV	98.25 (±0.38)	95.13 (±0.98)	94.78 (±1.08)	96.41 (±0.33)	98.43 (±0.82)
NB	99.20 (±0.52) (>1.45)	98.33 (±0.81) (>3.98)	97.21 (±0.64) (>2.86)	98.84 (±0.70) (>3.01)	99.87 (±0.41) (>1.62)
CFI	98.34 (±0.34)	95.89 (±0.49)	95.44 (±1.02)	96.90 (±0.23)	98.79 (±0.67)
D-S	98.67 (±0.22)	97.71 (±0.19)	96.63 (±0.68)	97.56 (±0.95)	99.36 (±0.31)

Table 3. a. Classification rates of combined classifiers on 5 classes using identical (optical) features based on simulated images

Combiner type	- 6 pixels shift	- 4 pixels shift	- 2 pixels shift	0 pixels shift	+2 pixels shift	+4 pixels shift	+6 pixels shift
MV	94.12 (±0.75)	82.53 (±0.95)	81.05 (±0.83)	86.38 (±0.71)	83.44 (±0.53)	87.85 (±0.28)	92.48 (±0.90)
NB	96.82 (±1.13) (>3.82)	85.49 (±0.94) (>3.67)	83.87 (±0.80) (>3.51)	88.79 (±0.46) (>2.92)	85.21 (±0.55) (>2.80)	90.35 (±0.38) (>3.52)	95.53 (±0.61) (>3.67)
CFI	95.27 (±0.40)	83.76 (±0.69)	82.64 (±0.65)	87.30 (±0.98)	84.37 (±1.33)	88.54 (±1.21)	93.86 (±0.77)
D-S	95.79 (±0.89)	84.88 (±0.32)	83.26 (±0.76)	88.14 (±0.83)	84.69 (±0.20)	89.87 (±0.19)	94.64 (±0.34)

Table 3. b. Classification rates of combined classifiers on 7 classes using identical (optical) features based on simulated images

In the sequel we derive the classification results using the four combination schemes based on the classifiers Bayes, MLP and LVQ employing real images, given in Tables 2a and 2b. These results are presented in Tables 4a and 4b on 5 and 7 classes, respectively. By comparing these results with Tables 3a and 3b, we can detect a small decrease (ranging from 0.30 to 1.30 in different classes) in classification rates from the case of testing simulated data, which can be attributed to small differences in the formation of the training and the testing data. Nevertheless, by comparing them with the results of individual classifiers on real

image data (Tables 2a and 2b), we observe a consistent increase of the success rate achieved by any fusion methodology.

Combiner type	- 6 pixels shift	-3 pixels shift	0 pixels shift	+ 3pixels shift	+6 pixels shift
MV	97.16	94.42	93.56	95.73	98.04
NB	97.70 (>1.27)	96.97 (>3.36)	95.56 (>2.39)	97.61 (>2.84)	98.61 (>1.28)
CFI	96.95	94.88	95.00	96.49	98.40
D-S	97.26	95.76	94.63	96.15	98.27

features based on real images.

Table 4. a. Classification rates of combined classifiers on 5 classes using identical (optical)

Combiner type	- 6 pixels shift	- 4 pixels shift	- 2 pixels shift	0 pixels shift	+2 pixels shift	+4 pixels shift	+6 pixels shift
MV	92.65	80.69	79.94	85.90	82.73	86.89	91.61
NB	95.33 (>3.46)	83.50 (>3.24)	82.76 (>3.19)	87.84 (>2.58)	84.29 (>2.37)	89.33 (>3.28)	94.44 (>3.41)
CFI	93.69	81.80	81.43	86.45	83.67	87.53	92.68
D-S	94.23	82.98	80.94	87.30	83.77	88.48	93.61

features based on real images.

Table 4. b. Classification rates of combined classifiers on 7 classes using identical (optical)

The high accuracies of fusion schemes can be partially attributed to the diversity of the three primary classifiers. It is the authors' opinion that an additional improvement can be achieved in the 7-class case by enriching the primary classifier's pool. This would require a very careful choice of additional classifiers that would contribute to the ensemble's diversity, if possible. The 5-class case is less likely to benefit since the obtained accuracies are already nearly maximized. Such a refinement might also render 1-pixel resolution shift estimation (13 classes) manageable. In any case, one has to keep in mind that model complexity should not outweigh possible minimal gains and that results have to be extended to other datasets.

The increased computational complexity of fusion in a real time inspection system was also a factor considered. The overhead in a multiple classification process of this type is additive. This problem is addressed in three ways towards minimizing this overhead. Firstly the number of classes is kept to a minimum required for quality inspection by quantizing the

output displacements. Secondly the features used were chosen so that no intensive image processing or costly transformations are involved in their computation. Thirdly a minimal primary classifier pool is used whilst maintaining a decent misclassification rate. From a different point of view, classifiers in this application area can benefit from certain symmetries and prior knowledge inherent to the problem. Limiting the displacements to one axis (for the corresponding component side) reduces the degrees of freedom in problem specification and classifier design. Additionally, the shape and size of the areas is roughly known or can be easily inferred for any new dataset and thus geometry metrics can be used reliably.

5.4 Results of combined classifiers using distinct pattern representations

In this section, the combination rules of Section 4.5 are used to combine the two primary classifiers (Hamming classifier and Bayesian classifier), using two sets of distinct pattern representations (topological and projection) for individual lead classification. Four different combination rules are tested under the assumption of equal priors and their results are compared. Each combiner uses the outcomes of primary classifiers as estimates of a posterior class probability, in a soft-level combination manner.

The classification results obtained from the above four combiners are presented in tables 5a and 5b for 5 and 7 classes, respectively. As we observe from these tables, the sum combination rule achieves better performance than any individual classifier alone with the exception of the class of -3 pixels shift on the 5 class formulation and the class of 0 pixels shift on the 7 class formulation. The max combination rule follows closely in performance, whereas the worst results are achieved when using the product and min combination rules. These results are in close agreement with the findings of (Kittler et al., 1998), based on a theoretical error sensitivity analysis, where the sum combination rule is found to be much more resilient to estimation errors of the posterior probabilities $P\left(\omega_j | \mathbf{x}^{(i)}\right)$ than the

product combination rule. In particular, the product combiner is oversensitive to classification estimates close to zero. Presence of such estimates from one classifier has the effect of veto on that particular class, regardless the outcome of other classifiers.

We should further emphasize that fusion may not improve the classification results for each and every lead displacement compared to the individual classifiers, but it rather improves the overall classification ability for all lead-shifts examined. Even though fusion increases the classification accuracy for lead shifts where individual classifiers generally lag in performance, there are a few cases where one or the other individual classifier (based on topological or projection features) by chance achieves extremely high accuracy. The results of primary classifiers show a large variance of performance across the lead displacements, as in Tables 1a and 1b or 2a and 2b for simulated and real data, respectively. From these results, we cannot claim that one individual classifier, either Hamming based on topological or Bayes based on projection features, surpasses the other in performance. Each one attains maximum performance by chance at some specific lead displacement. We cannot generalize such results of individual classifiers due to the limited number of available data. Notice that this large variation is reduced by the fusion approaches. Thus, fusion using distinct, reduced-content representations not only boost the overall classification performance, but also makes the overall classification performance more consistent across all lead-displacements examined.

Combination Rule	- 6 pixels shift	-3 pixels shift	0 pixels shift	+ 3pixels shift	+6 pixels shift
Product	83.08 (±0.46)	78.25 (±0.62)	84.23 (±0.70)	88.76 (±0.32)	89.68 (±0.09)
Sum	94.64 (±0.37)	90.68 (±0.78)	95.67 (±0.39)	97.42 (±0.57)	98.73 (±0.87)
Max	90.78 (±0.06)	87.14 (±0.79)	91.62 (±0.94)	93.46 (±0.34)	96.09 (±0.56)
Min	84.33 (±0.45)	79.04 (±0.58)	85.59 (±0.75)	89.94 (±0.83)	90.81 (±0.47)

Table 5. a. Classification rates of combined classifiers on 5 classes using distinct features (topological & projection) based on simulated images

Combination Rule	- 6 pixels shift	- 4 pixels shift	- 2 pixels shift	0 pixels shift	+2 pixels shift	+4 pixels shift	+6 pixels shift
Product	71.24 (±0.49)	75.81 (±0.32)	74.41 (±0.65)	70.93 (±0.97)	76.13 (±0.23)	78.95 (±0.31)	83.76 (±0.94)
Sum	85.08 (±0.46)	86.63 (±0.58)	86.12 (±0.77)	85.91 (±0.24)	87.38 (±1.22)	90.53 (±0.45)	94.97 (±0.75)
Max	81.59 (±0.64)	83.43 (±0.21)	82.36 (±64.)	81.22 (±0.57)	83.87 (±0.59)	87.11 (±0.88)	91.55 (±0.98)
Min	72.09 (±0.44)	76.88 (±0.23)	75.23 (±0.38)	71.64 (±0.55)	77.39 (±0.16)	79.83 (±0.29)	85.03 (±0.72)

Table 5. b. Classification rates of combined classifiers on 7 classes using distinct features (topological & projection) based on simulated images

Considering the classification of real data, the results of these four combination rules are presented in Tables 6a and 6b for the 5 and 7 class formulations, respectively. We recall that the individual classifiers used at first level are the Hamming neural network operating on topological features and the Bayes classifier operating on projection features extracted from the set of real images considered. As can be observed in Tables 6a and 6b, the sum combiner again achieves overall better results, but there is a small decrease (ranging from 0.30 to 1.30 in different classes) in classification rates in comparison with Tables 5a and 5b for the simulated data.

In general, the classification scores achieved using reduced dimensionality features are inferior to those obtained using the optical features. Furthermore, the combination of topological and projection features in a distinct representation fusion scheme also lags in performance to the combination of classifiers trained with optical features alone in Section 5.3. This is expected since all feature sets are obtained from the same primary source (original lead images), so that the information captured by topological and projection features does not add much to the information already conveyed by optical features.

Furthermore, the primary data in reduced content representation (1-bit edge images and 1-D projections) are inter-related, rendering the corresponding features (topological and projection) not quite independent. At this stage we do not perform any feature selection process, since we are focusing on the nature of primary data (edges and projections) and the information conveyed in these forms, rather than the nature of features. Nevertheless, it is worth mentioning that reduced dimensionality features using just a portion of information available can still attain acceptable results, especially through the employment of fusion. Reduced dimensionality features have the benefit of summarizing the required information for adequate shift detection in a compact format that can significantly reduce processing time. It is the authors' opinion that such features should be used when balance is required between speed and effectiveness. In addition, the particular reduced dimensionality features possess conceptual attributes that can instigate further speed-up and improvement in component inspection systems. More specifically, the topological features (edges) may be used for appropriate modeling of the component placement process and can be directly obtained from a number of commercial cameras, eliminating the need of preprocessing. The projection features on the other hand may eventually enable the use of faster and cheaper line sensors instead of area cameras for component inspection.

Combination Rule	- 6 pixels shift	-3 pixels shift	0 pixels shift	+ 3pixels shift	+6 pixels shift
Product	81.89	77.34	85.00	87.93	89.44
Sum	93.33	91.38	94.61	92.77	94.55
Max	89.28	87.94	90.47	94.38	95.35
Min	83.09	78.38	85.79	88.94	90.55

Table 6. a. Classification rates of combined classifiers on 5 classes using distinct features (topological & projection) from real images

Combination Rule	- 6 pixels shift	- 4 pixels shift	- 2 pixels shift	0 pixels shift	+2 pixels shift	+4 pixels shift	+6 pixels shift
Product	69.97	75.14	73.06	69.54	75.54	77.45	82.94
Sum	84.68	85.98	85.12	84.75	86.08	89.63	93.95
Max	80.41	82.77	81.31	86.10	81.67	86.01	91.24
Min	70.85	76.00	73.98	70.04	77.03	77.86	84.93

Table 6. b. Classification rates of combined classifiers on 7 classes using distinct features (topological & projection) from real images

Elaborating on the use of distinct feature representations and its potential in increasing accuracy and robustness for all classes of lead displacements, we further consider a combination of optical, topological and projection features. We define the resulting distinct features set (i.e., *optical* and *topological* and *projection features*) as *distinct features-2*. This set of features captures information from many different aspects of the problem and contains features that are more likely to be independent than the set used before employing only topological and projection features. The quite diverse nature of information handled by each approach justifies the assumption of class conditional independence (at least approximately) for the distinct representations used by the individual classifiers. Motivated by good results of the sum combination rule we also use it as a fusion rule on the Bayes classifier with optical features, Hamming classifier with topological features and Bayes classifier with projection features. The classification rates based upon Monte Carlo simulated and real images are presented in Tables 7a and 7b on 5 and 7 classes, respectively. We observe that the sum combination rule achieves better performance than any individual classifier alone based on distinct features-2, with the exception of the class of 0 pixels shift on the 7 class formulation. It improves the results of the first level Bayes classifier and derives quite uniform results across all classes. Comparing this distinct feature combination with the one in Section 5.3 using identical pattern representation, we can claim that the former achieves comparable and at cases (7-class formulation) even better performance than the latter. This result further supports the potential of the distinct representation scheme, requiring however further investigation on the appropriate selection of distinct features, which is out of the scope of this work.

Sum Combination Rule	- 6 pixels shift	-3 pixels shift	0 pixels shift	+ 3pixels shift	+6 pixels shift
Simulated Images	98.37	97.14	96.27	97.45	99.73
Real Images	97.49	95.53	95.04	95.52	98.92

Table 7. a. Classification rates of Sum Combination Rule on 5 classes using distinct features-2 (optical & topological & projection) on simulated and real images

Sum Combination Rule	-6pixels shift	- 4 pixels shift	- 2 pixels shift	0 pixels shift	+2 pixels shift	+4 pixels shift	+6 pixels shift
Simulated Images	93.65	88.25	86.92	92.70	87.53	91.46	95.19
Real Images	92.45	86.66	85.44	92.28	86.18	90.27	94.12

Table 7. b. Classification rates of Sum Combination Rule on 7 classes using distinct features-2 (optical & topological & projection) on simulated and real images

5.5 System performance

With respect to time requirements, the tested approaches achieve the following performance using a fast Intel Core 2 Duo workstation. The optical feature approach takes about 0.34 sec

for processing an entire QFP chip of 120 leads. The reduced dynamic-range approach requires 0.15 sec, less than half of the computation time of the conventional approach. Finally, the reduced input-dimension processing requires about 0.22 sec for the entire QFP-120 component. In comparison with existing industrial systems for PCB inspection, the proposed fusion system can achieve better throughputs, even though it considers each lead separately. The results of our survey of industrial systems are outlined on Table 8. The performance data for commercial products have been obtained through the vendors' online available product datasheets. In our case, the speed of the fusion algorithms was mapped to throughput (cm^2/sec) by simulating performance on a 120-lead QFP component of roughly $3.3 \times 3.3 \text{cm}$ surface at a sampling resolution of $20 \mu\text{m}/\text{pixel}$. The speed of each algorithm was estimated with respect to the chip's total lead area. The reported times refer to processing alone, without including the board placement/ adjustment times required by the mechanical operation of the production line.

System	speed (cm^2/sec)	resolution ($\mu\text{m}/\text{pixel}$)
optical feature	32.4	20
reduced dyn range	72.6	20
reduced input dim	49.1	20
Agilent Medalist SJ50 3	38.7	16
Orbotech Symbion P36	22-60	20
Viscom S3088	20-40	15

Table 8. Inspection speed comparison

High abstraction features are generally less descriptive than pixel-based features for classification purposes. With respect however to computational complexity, the distinct features in cooperation with a fusion scheme can yield appreciable reduction at the cost without compromising the effectiveness of inspection

6. Conclusion and future work

In this research, we tested several combination methods for soft fusion of the outputs of multiple classifiers. The aim is to improve the performance of primary classifiers used for individual lead-image classification in post-placement quality inspection of components. Two different schemes of classifier fusion are considered. The first one refers to identical feature representations, where the primary classifiers operate on the same feature set. The second scheme uses distinct pattern representation, where each of the primary classifiers operates on a different set of features. Comparing the classification results of the proposed combined classifiers, we can derive that all combiners have better performance than any individual classifier alone. In addition, it is verified that both the naïve Bayes and the Dempster-Shafer combiners on identical feature representations achieve better overall performance, with the naïve Bayes reaching the best performance improvement over the primary classifiers. The combiners based on distinct feature representations present lower overall performance and higher variability of their results. This is expected due to the

reduced content of information exploited. Despite that, their performance is still better than that of most primary classifiers, showing a good potential for accelerating the inspection process when speed needs to be balanced against effectiveness.

According to market studies (Frost & Sullivan, 2005), the PCB inspection field is in need of reliable systems in order to sustain growth as component densities get higher. Use of exhaustive solder paste inspection helps reduce the contribution from the print process to solder joint defects, in-turn saving money by reducing the cost of scrap with minimal cost to rework (i.e. wash boards) and with no penalty in solder joint reliability (Lecklider, 2004). Some companies claim this number to be as high as 80% of their overall defect Pareto chart (Mendez, 2000). Furthermore, the total misclassification cost in an automated optical inspection system is the product of the production volume, cost-per-defective PCB and accuracy. Taking into account the ranges of the first two variables it is evident that even a minor, yet consistent, improvement in classification accuracy is translated to amplified profits.

Overall, classifier fusion can contribute to the visual solder-joint inspection domain by improving accuracy and speed. One of the conditions under which fusion is favorable is the high diversity in features and primary classifier outputs. Evaluation of a number of diversity metrics indicated that using distinct representations (different feature sets) of leads, in most cases leads to a reduction in the correlation between the outputs of individual classifiers. This is attributed to the reduced correlation in the input vectors of distinct information content. Since this is a desirable feature in fusion, a further research is required to establish the effects of combining truly different input representations besides exploiting different attributes of the same primary source of information (as with the use of the same optical images to obtain the different features sets). Fusion at different levels (measurements, features, and outputs) can then be evaluated overall.

7. Acknowledgement

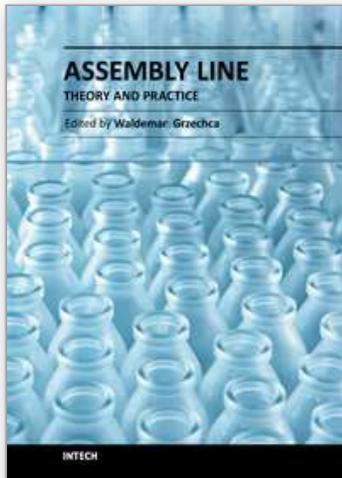
The authors would also like to thank Philips, The Netherlands, for the provision of test images.

8. References

- Altincay, H. (2005). On naïve Bayesian fusion of dependent classifiers. *Pattern Recognition Letters*, 26 (2005) 2463 – 2473.
- Bartlet, S., Besl, P., Cole, C., Jain, R., Mukherjee, D., & Skifstand, K. (1988). Automatic Solder Joint Inspection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 10, No. 1, pp. 31-41.
- Bremaud, P. (1999). *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer Science + Business Media, New York, USA.
- Cappelli, R., Maio, D. & Maltoni, D. (2001). Multispace KL for Pattern Representation and Classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 9, pp. 977-996.
- Capson, D. & Eng, S. (1988). A tiered-colour illumination approach for machine inspection of solder joints, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 10, No. 3, pp. 387-393.
- Ceccareli, M. & Petrosino, A. (1997). Multi-feature adaptive classifiers for SAR image segmentation. *Neurocomputing*, 14 (1997) 345-363.

- Chi, Z., Yan, H. & Pham, T. (1996). *Fuzzy algorithms: With Applications to Image Processing and Pattern Recognition, Advances in Fuzzy Systems – Applications and Theory, Vol. 10*, World Scientific Publishing Co. Pte. Ltd., USA.
- Denoeux, T. (1995). A k-nearest neighbor classification rule based on Dempster- Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics - Part B*, Vol. 25, No. 5, pp. 804-813.
- Fausset, L. (1994). *Fundamentals of Neural Networks Architectures, Algorithms, and Applications*, Englewood Cliffs, NJ: Prentice-Hall, USA.
- Fishman, G. (1996). *Monte Carlo: Concepts, Algorithms and Applications*, Springer-Verlag, New York, USA.
- Frost & Sullivan. (2005). Growth Opportunities for the World SMT Inspection Equipment Markets. Market survey report, May 2005.
- Giacinto, G., Perdisci R., Del Rio, M. & Roli, F. (2008). Intrusion detection in computer networks by a modular ensemble of one-class classifiers. *Information Fusion*, 9 (2008) 69 - 82.
- Goumas, S., Zervakis, M. & Stavrakakis, G. (2002). Classification of Washing Machines Vibration Signals Using Discrete Wavelet Analysis for Feature Extraction. *IEEE Trans. on Instrumentation and Measurement*, Vol. 51, No. 3, (June 2002), pp. 497-508.
- Goumas, S., Rovithakis, G. & Zervakis, M. (2002). A Bayesian Approach to Post Placement Quality Inspection of Components, *Proceedings of IEEE Int. Conf. on Image Processing*, 2002.
- Goumas, S., Zervakis, M. & Rovithakis, G. (2004). Data-Space Reduction in Component Quality Inspection, *Proceedings of ESANN'04*, pp. 275-280.
- Ho, T., Hull, J. & Srihari, S. (1994). Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, pp. 66-75.
- Huang, Y. & Suen, C. (1995). A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 1, (January 1995), pp. 90-94.
- Jain, R., Kasturi, R. & Schunck, B. (1995). *Machine Vision*, McGraw-Hill International Editions, Singapore.
- Kang, H., Kim, K. & Kim, J. (1997a). Optimal approximation of discrete probability distribution with k -order dependancy and its application to combining multiple classifiers. *Pattern Recognition Letters*, 18 (1997) 515-523.
- Kang, H., Kim, K. & Kim, J. (1997b). A Framework for Probabilistic Combination of Multiple Classifiers at an Abstract Level. *Engng Aplic.Artif. Intell.*, Vol. 10, No. 4, pp. 379-385.
- Kim, T-H., Cho, T-H., Moon, Y. & Park, S. (1999). Visual Inspection System for the Classification of Solder Joints. *Pattern Recognition* 32 (1999) 565-575.
- Kim, E., Kim, W., & Lee, Y. (2002). Combination of multiple classifiers for the customer's purchase behavior prediction. *Decision Support Systems*, 34 (2002) 167-175.
- Kittler, J., Hojjatoleslami, A. & Windeatt. (1997). Strategies for combining classifiers employing shared and distinct pattern representations. *Pattern Recognition Letters*, 16 (1997) 1373-1377.
- Kittler, J. Hatef, M., Duin, R. & Matas, J. (1998). On Combining Classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, (March 1998), pp. 226-239.
- Kuncheva, L., Bezdek, J., & Sutton M. (1998). On combining multiple classifiers by fuzzy templates, *Proceedings of NAFIPS'98*, Pensacola, FL, 1998, pp. 193-197.
- Kuncheva, L. (2001). Using measures of similarity and inclusion for multiple classifier fusion by decision templates. *Fuzzy sets and systems*, 122 (2001) 401- 407.

- Kuncheva, L. (2004). *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience publication, New Jersey, USA.
- Lacey, G., Waldron, R., Dinten, J-M. & Lilley, F. (1993). Flexible Multi-Sensor Inspection System for Solder-joint Analysis. *Proceedings of SPIE Machine Vision Applications, Architectures, and Systems Integration {II}*, 1993, No. 2064.
- Lecklider, T. (2004). PCB Inspection Outlook for 2005. *Evaluation Engineering*, (Dec 2004).
- Loh, H-H. & Lu, M-S. (1999). Printed Circuit Board Inspection using Image Analysis. *IEEE Trans. Industry Applications*, Vol. 35, No.2, pp. 426-432.
- Martinez, W. & Martinez, A. (2002). *Computational Statistics Handbook with MATLAB*, Chapman & Hall / CRC, Florida, USA.
- Mashao, D. & Skosan, M. (2006). Combining Classifier Decisions for Robust Speaker Identification. *Pattern Recognition*, 39 (2006) 147-155.
- Mendez, D. (2000). An Integrated Test and Inspection Strategy, *Proceedings of APEX*, 2000.
- Mirhosseini, A., Yan, H., Lam, K-M. & Pham, T. (1998). Human Face Image Recognition: An Evidence Aggregation Approach. *Computer Vision and Image Understanding*, Vol. 71, No. 2, (August 1998), pp. 213-230.
- Ng, G. & Singh, H. (1998). Data equalization with evidence combination for pattern recognition. *Pattern Recognition Letters*, 19 (1998) 227 - 235.
- Otsu, N. (1979). A Threshold Selection Method from Grey-Level Histograms. *IEEE Trans. On Systems, Man and Cybernetics*, Vol. 9, No. 1, pp. 62-66.
- Oza, N. & Tumer, K. (2008). Classifier ensembles: Select real-world applications. *Information Fusion*, 9 (2008) 4-20.
- Parikh, D., & Polikar, R. (2007). An Ensemble-Based Incremental Learning Approach to Data Fusion. *IEEE Trans. On Systems, Man, and Cybernetics – Part B: Cybernetics*. Vol. 37, No. 2, (April 2007), pp. 437-450.
- Robert, C. & Casella, G. (1999) *Monte Carlo Statistical Methods*, Springer-Verlag, New York, USA.
- Rodriguez-Linares, L., Garcia-Mateo, C. & Alba-Castro, J-L. (2003). On Combining Classifiers for Speaker Authentication. *Pattern Recognition*, 36 (2003) 347-359.
- Rogova, G. (1994). Combining the Results of Several Neural Networks Classifiers. *Neural Networks*, Vol. 7, No. 5, pp. 777-781.
- Rovithakis, G., Maniadakis, M., Zervakis, M., Filippidis, G., Zacharakis, G., Katsamouris, A. & Papazoglou, T. (2001). Artificial Neural Networks for Discriminating Pathologic From Normal Peripheral Vascular Tissue. *IEEE Trans. on Biomedical Engineering*, Vol. 48, No. 10, (Oct. 2001), pp. 1088-1097.
- Ryu, Y. & Cho, H. (1997). A Neural Network Approach to Extended Gaussian Image Based Solder Joint Inspection. *Pergamon: Mechatronics*, Vol. 7, No.2, pp. 159-184.
- Shipp, C. & Kuncheva, L. (2002). Relationships between combination methods and measures of diversity in combining classifiers. *Information Fusion*, 3 (2002) 135-148.
- Xu, L., Krzyzak, A. & Suen, C. (1992). Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 22, No. 3, (May/June 1992), pp. 418-435.
- Zervakis, M., Goumas, S., & Rovithakis, G. (2004). "A Bayesian Framework for Multi-lead SMD Post-Placement Quality Inspection", *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, Vol. 34, No. 1, (February 2004), pp. 440-453.



Assembly Line - Theory and Practice

Edited by Prof. Waldemar Grzechca

ISBN 978-953-307-995-0

Hard cover, 250 pages

Publisher InTech

Published online 17, August, 2011

Published in print edition August, 2011

An assembly line is a manufacturing process in which parts are added to a product in a sequential manner using optimally planned logistics to create a finished product in the fastest possible way. It is a flow-oriented production system where the productive units performing the operations, referred to as stations, are aligned in a serial manner. The present edited book is a collection of 12 chapters written by experts and well-known professionals of the field. The volume is organized in three parts according to the last research works in assembly line subject. The first part of the book is devoted to the assembly line balancing problem. It includes chapters dealing with different problems of ALBP. In the second part of the book some optimization problems in assembly line structure are considered. In many situations there are several contradictory goals that have to be satisfied simultaneously. The third part of the book deals with testing problems in assembly line. This section gives an overview on new trends, techniques and methodologies for testing the quality of a product at the end of the assembling line.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Stefanos Goumas and Michalis Zervakis (2011). Multi-Classifer Approaches for Post-Placement Surface-Mount Devices Quality Inspection, *Assembly Line - Theory and Practice*, Prof. Waldemar Grzechca (Ed.), ISBN: 978-953-307-995-0, InTech, Available from: <http://www.intechopen.com/books/assembly-line-theory-and-practice/multi-classifier-approaches-for-post-placement-surface-mount-devices-quality-inspection>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen