Applications

Automated Visual Inspection of Solder Paste Deposition on Surface Mount Technology PCBs

James Mahon, Neville Harris
and David Vernon
Department of Computer Science, University of Dublin, Trinity College, Dublin, Ireland

The development of surface mount printed circuit board technology has generated a need for new inspection systems to deal with its problems and peculiarities. This paper describes a system which was designed to locate and measure the area of the solder paste deposited on surface mount printed circuit boards. It measures the distance by which the solder paste is displaced from the solder pads that it is supposed to be covering and reports this to a host system. It simultaneously measures the solder coverage for each solder pad that it inspects. The board is inspected one device at a time using the layout information obtained from CAD data. The inspection route is optimized using a "simulated annealing" technique, as the original CAD layout data is not organized in a suitable manner.

Keywords: Automatic visual inspection, Flexible inspection system, Solder paste, Machine vision, Surface mount process, Printed circuit board, Ternary image; Simulated annealing, Travelling salesman.

James Mahon has been a lecturer in the department of computer science in Trinity College Dublin since 1987. Prior to that he worked as a software engineer in Sord Computers Ltd, and as a doctoral student in T.C.D. He completed his BAI degree in 1982 at the University of Dublin in software. Since 1985, he has been working on automatic visual inspection at T.C.D. He is a member of the Institute of Engineers of Ireland (IEI) and the Irish Computer Society (ICS).

Elsevier Science Publishers B.V.
Computers in Industry 12 (1989) 31–42

0166-3615/89/$3.50

1. Introduction

Computer vision, the processing and analysis of image-based information by computer, has been employed in many disparate areas spanning optical character recognition, remote sensing of geophysical data captured by satellite, enhancement of medical images, and industrial automation. Current research in computer vision is directed predominantly toward the understanding of 3D scenes, so-called image understanding, such as

Neville R. Harris received his BA and BAI degrees from the University of Dublin (Trinity College) in Civil Engineering, his MSc from the University of Illinois in Structural Engineering, and his PhD from the University of Dublin (Trinity College) in Computer Science. He is now Head of the Department of Computer Science at the University of Dublin (Trinity College). His research and teaching are in the area of Distributed Systems and the application of computer techniques in industrial related research projects.

David Vernon has been a lecturer in the Department of Computer Science in Trinity College, Dublin, since 1983. For the two years prior to that, he worked with Westinghouse Electric Inc. as a software engineer. He holds the degrees of BA, BAI, MA, and PhD all awarded by the University of Dublin, Trinity College. He is involved in several research projects in the area of computer vision, including projects funded by the European Strategic Program for Research and Development in Information Technology (ESPRIT) and the National Board for Science and Technology. He is also is a joint director of the Vision and Sensor Research Unit in the Irish Advanced Manufacturing Technology program. Dr. Vernon is a member of the Institute of Electrical and Electronic Engineers (IEEE) and the Society for Optical Engineers (SPIE). He is also a senior member of the Machine Vision Association of the Society of Manufacturing Engineers (SME).

might be encountered by an autonomous automatically guided vehicle. A distinct and important branch of this multidisciplinary science corners the use of image analysis in manufacturing and incorporates both vision for guiding robot manipulators and for visual inspection of manufactured parts. Robot vision and visual inspection differ significantly from the developing subject of image understanding in that they depend on the exploitation of a controlled (visual) environment rather than attempts to deal with the problems caused by shadows, occlusion, movement, and poor illumination.

Inspection is an essential part of the industrial process as a consequence of the importance of product reliability: 100% inspection of the finished parts or subassemblies is often attempted. As a result, the inspection process can be one of the slowest and most expensive phases of manufacturing. There are two aspects to industrial quality control: functional integrity and cosmetic integrity. It may be possible to functionally test a product using electronic probes, but cosmetic quality must be visually tested. This is usually done by human inspectors, but their performance is often inadequate: the human visual system is extremely powerful and flexible, but the human brain becomes distracted from tedious and exacting tasks, with the resulting degradation in the consistency of inspection. Additionally, with the advent of computer integrated manufacturing (CIM), and more automated manufacturing generally, it is desirable to computerize the inspection stage in order to automate the feedback loop in the manufacturing process.

The advantages of automated visual inspection stem from the replacement of lasbour with capital, the automation of the manufacturing process, enhancing consistency of production, removing the need to work in hazardous environments, while producing quantitative measurements and facilitating the integration with other aspects of automated manufacturing.

There are, however, several disadvantages to automatic visual inspection. Commercial systems are quite expensive, starting at about £20000. For most applications, off-the-shelf systems do not exist and special packages must be designed or modified, further increasing the total cost. Special fixturings may be needed for loading and unloading objects and special lighting must be used. More skilled technician and engineering staff may be needed to operate and maintain the systems; integration will probably require more sophisticated software and will increase the complexity of current systems. Despite this, the applications of automated visual inspection are legion [1,2] and there is a definite trend toward the deployment of automatic visual inspection systems as a special-purpose real-time (in the sense that they do not cause delays in the overall process) computer vision systems with control of their own environment.

2. The Inspection of Printed Circuit Boards

A printed circuit board (PCB) is a board on which an electrical circuit is assembled. It has two functions: to hold the electronic components in place and to transmit the electrical signals between the devices on the board. The board itself is usually made of fibreglass and is called a substrate. The electrical signals are transmitted by a copper pattern etched on the surface of the substrate from component to component. The devices have traditionally been placed in holes and soldered into place for mechanical stability and to make electrical connections. The holes are also used to carry the electrical signals from one side of the board to the other. Some boards have inner conductor layers, and are called multilayer boards. Two, four, and six layers are common.

Printed circuit boards were among the first objects to be automatically inspected. There are several reasons why this was so, including the fact that PCB's are flat and thus can be inspected using 2D techniques and modelled with a binary image. The defects are easily classified and inspecting the boards manually is very tedious: operators are not very effective and are expensive. Additionally, the electronics industry was innovative, it was already highly automated, and could recognize and exploit the advantages of using computer vision.

There are several difficulties with inspecting PCBs. There is a large amount of information to be processed due to the size of the features that must be inspected compared with the size of the boards. Further, PCBs may stretch and deform while still being able to function correctly, and the conductor patterns are large and complex.

2.1. Errors in PCB boards

There are five main classes of PCB errors: breaks, blobs, hairline cracks, necks, and whiskers. Breaks are complete gaps in the conductor and cause immediate loss of signal. A large increase in the size of the conductor is called a blob which may or may not cause two conductors to short circuit. Hairline cracks are small cracks in the conductor and may or may not be breaks. Necks are small thin sections in the conductor, which could potentially become breaks. Whiskers are small conductor "hairs" which may or may not be shorts. The main problem associated with blobs, hairline cracks, necks, and whiskers is that they are potential faults: they might not cause a change in the electrical characteristics of the PCB at the time of manufacture but faults might emerge over time, causing failure after shipment of the assembly.

2.2. PCB Inspection Methods

Printed circuit patterns are large, and cover the whole circuit board. They are finely detailed, with tracks as small as 0.254 mm, while boards may be as large as 500×400 mm. To resolve a board 500×400 mm at 0.08 mm resolution would require an image with 6250×5000 pixels, i.e. 31,250,000 pixels. Most frame-buffers can only hold 512×512 or 262,144 pixels. The whole image would thus require over 120 times the storage of an average framebuffer, and hence image acquisition and processing must be done in sections. The track patterns are complex, and joined together, or joined to holes, and the conductors themselves are shiney curved metal tracks that are difficult to illuminate correctly; see [3,4]. A combination of on-axis and side lighting must be used to light the whole of the conductor surface evenly.

Two basic techniques were originally used for PCB inspection: reference comparison and analysis of generic properties.

Reference comparison involves comparing a good reference image with the test image and rejecting the test image if the difference between the two images is too great. Reference comparison is fast and simple and the comparisons may easily be implemented in hardware. It can detect the larger errors easily, e.g. errors larger than the registration accuracy of the two boards. Other than the reference image, no other information is needed about the board, and it is therefore very easy to set up the system to inspect new boards. Any pattern whatever can be inspected, simple or complex, as long as the smallest details are greater than the registration accuracy. The disadvantages are that reference comparison requires very accurate calibration and alignment of the boards and cameras. If one of the boards is misaligned either rotationally or translationally, the system will detect false errors. Similarly, the cameras have to have matched responses. If the conductor pattern has stretched or deformed, a reference comparison system will start to detect errors when the board can still function correctly. It is sensitive to noise, and if it is made less sensitive to noise and minor miscalibration, it will be unable to detect small errors.

The generic property method of PCB inspection relies on detecting and measuring local properties of the image and comparing them with the local properties of good boards. No other information about the board is required and the system does not need to be trained with a good board. There are two principal techniques. The first technique models the conductor patterns as large smooth features, and defects as small angular features. This allows the detection of small errors such as whiskers and cracks, but relies on other inspection methods (e.g. manual inspection) for the larger errors, which is its greatest drawback. The second technique is based on verifying design rules, such as conductor width, feature spacing, pad location, and size. In both techniques, the properties are based on local properties of the image. They may miss large errors, or errors that match their rules, and thus reliability can be low. The generic property technique is also called the "singularity detection method".

Recently, several hybrid techniques have been developed which claim to solve the problems of registration and reliability. Systems such as Hara's PCB inspection system [3] use a reference comparison operator to detect large errors and a fine line detector to detect small errors. Doyle has described a system which can generate a wiring list by using a ternary image to segment out conductors, substrate, and holes [5]. The wiring list specifies the interconnections of the holes. The list is then compared to a reference wiring list. By comparing information at a much higher level, the

correctness of the conductor pattern can be checked in a much more reliable fashion, without too much emphasis on details such as the exact location of the conductor patterns.

Solder joint inspection can be expected to become increasingly important as chip sizes reduce and as it becomes increasingly difficult to inspect the joints manually. Unfortunately, it is an extremely difficult task: solder joints are 3D entities and in this way they differ from all previous inspection tasks that have been examined. Solder is shiney, giving rise to specular reflections, and so it must be illuminated so as to reduce the highlights. There are many kinds of both bad and good solder joints and a classification system must be able to detect both types. The bulk of data to be processed also poses problems: a board might have up to two hundred devices on it with, for example, four thousand joints to be inspected.

3. A System for Visual Inspection of Surface Mount Technology Printed Circuit Boards

In conventional through-hole printed circuit board technology, the devices are connected to the board by soldering pins that pass through holes in the PCB. This process has several disadvantages, mainly due to the size of the pins and holes. The standard distance between pins is 2.54 mm resulting in unnecessarily large devices and reducing the possible board density. Through-hole technology is now being replaced by a new mounting technology, called surface mount. With this approach, the devices are bonded to conductor pads on the surface of the board. No holes are needed and so the pin spacing can be reduced. This gives several advantages over the previous process. Firstly, it is possible to reduce the pin spacing to 1.27 or 0.635 mm, and secondly it is possible to mount devices on both sides of the board. This provides density improvements of at least threefold, or up to sixfold if both sides are used, in addition to which there are shorter signal propogation delays due to the smaller signal paths.

There are six principal surface mount device package types: resistors, capacitors, transistors, dual in line (DIL) ICs, J-lead rectangular ICs, and Gull-wing rectangular ICs. In all cases the leads lie flat on the bonding pads of the PCB. Resistors and capacitors are small and rectangular in shape with contacts at either end. Transistors have a third contact in the center at one side, while DIL IC's have two rows of pads which splay outwards to lie on the pad surfaces. Gull-wing devices have four rows of pads which splay outwards to lie on the pads, while J-leads curl back underneath the devices, giving a smaller footprint.

The solder is used in the form of a paste of small (0.025–0.06 mm diameter) balls of solder in a viscous medium; at room temperature it has the consistency of fine damp sand.

Surface mount devices are designed to be placed on the board automatically and not hand soldered. In general, the whole process is automated, although the boards may be moved manually from process to process. There are effectively four processes involved in the assembly of SMT PCBs: firstly, a bare PCB is placed in a screening press and the solder paste is stencilled onto the board in a process much like a silk screening. Secondly, the components are placed onto the device footprints which should have a covering of solder paste. The required accuracy (0.08 mm) is such that this is usually done automatically. This process is called "onplacement". Thirdly, the board is passed through an infrared oven to slowly heat the board and the components on it to the reflow temperature. This is done to avoid thermal shock at the next stage. Finally, the board is heated above the melting point of the solder. This may take place either in an infrared oven or a vapour phase reflow bath, although an infrared oven may not heat the bright solder pads and pins very effectively. Alternatively, a vapour phase reflow may heat the device too quickly and cause thermal shock which would crack the chips. It is for this reason that the infra red preheat is used.

Inspection may take place after each of the three major stages: silk screening, onplacement, and solder reflow. This paper is concerned with the first of these three inspection tasks. Several problems can occur when depositing the solder paste. The silk screen may deform so that the solder paste is deposited in the wrong place. The silk screen may clog or tear resulting in incorrect amounts of solder being deposited on the pads. This will deposit the wrong amounts of solder on the pads. Additionally, the layer of solder may be too thick or too thin. If too much solder is deposited, it may form bridges, and if too little solder is deposited, the joints may not be sound

and may fail at some time in future. These problems represent errors in location, coverage, and volume of deposited solder paste, respectively. The location and coverage errors are inherently two-dimensional in nature while the volumetric error is three-dimensional. It is the location and coverage errors which are dealt with here. In particular, the purpose of the system is to monitor the location and coverage performance of the screen printer and to feed back error measures to it, should it drift off alignment or if any of the pad templates becomes clogged. In addition, it computes comprehensive statistics on the deposition process for subsequent analysis. As the board layout and the footprint information for the individual chips are available from CAD data, the system is initialized by processing this CAD data, rather than training with a perfect board. The functional requirements of the system where that it should be accurate to ± 0.08 mm and to within 10% of true paste coverage. Furthermore, it had to be capable of processing one device every second: each device can contain from 8 to 84 pins at 1.27 mm pitch and the inspection of both dual in-line and square device was to be faciliatated.

3.1. Hardware Specification

The system is based on Imaging Technology IP612 image processing boards (AP512 framegrabber module and FB512 framebuffer module) and a μVAX-I processor. Images are acquired by the IP512 system and transferred to the μVAX. The μVAX was also used as the development environment, using VAX11-C under VMS. An X–Y table was required to position the SMD devices beneath the camera. The boards were up to 500×400 mm in size, and the required definition was 0.08 mm, and thus the camera had to digitize an area of 40×30 mm to provide sufficient definition with a 512×512 pixel frame buffer, resulting in a field of view just large enough to contain one typical square SMD.

The video camera which is used is a General Electric GE2506 CCD camera which generates a CCIR image with an effective resolution of 380×280 pixels. The lens is an Ashai Pentax 50 mm f2.8 Macro Lens. A Unimatic IF-1 X-Y table is employed under control of the μVAX-1. Four high-frequency fluorescent tubes in a square diffusing box are used to illuminate the boards.

3.2. Assumptions Regarding the Organization of the Surface Mount PCB

A board consists of several devices, each having location, type, and orientation. Each device in turn is composed of several rows of pads, and each row is composed of a regularly spaced identical contact pads. In order to facilitate the inspection process, the following assumptions were made about the system.

- All devices are either parallel or perpendicular to the major axis of the board.
- The board can be located to within ± 1 mm by the X-Y table.
- The angular deviation of the board is less than 1 degree.
- The locations of the devices and their types are known from CAD data.
- The lighting is controllable and constant (in the short term).
- The substrate is green, the pads are shiney silver, and the paste is dull grey or silver.

Thus, there are three distinct features: pads, paste, and substrate. In addition, there are spurious conductors which are not of interest and are ignored. Suitable lighting causes each feature to exhibit a significantly different grey level and pixels can be classified as either pad, paste, or substrate on the basis of brightness alone, allowing very fast segmentation.

Each pad is rectangular or oval in shape and is of a given size (1.93×0.762 mm). They are, however, joined to other features on the board by conductor tracks, forming part of a larger, more complicated pattern. Additionally, the pads are covered by solder paste to varying degrees. A correctly covered pad should be completely covered by paste. This occlusion complicates the segmentation process and will be discussed later in more detail.

3.2.1. Inspection Algorithms

The task of processing a single device comprises five distinct components:

Step 1. Acquire an image.
Step 2. Label every pixel in the image so that it belongs to one of the three categories: substrate, paste, or pads.
Step 3. Locate the rows of pads.

Step 4. Measure the location and coverage of the paste for each pad.

Step 5. Aggregate the results for each row, and then for the whole device. This is used to produce a general measure for the chip. Pass or fail the chip by comparing the figures with predetermined values.

3.3. Image Acquisition

Image acquisition is quite straightforward. An image is grabbed by the IP512 frame buffer and the pixels are read in one at a time from the data register in the IP512 to the μVAX main memory in row order. This process is slow as each pixel has to be read in individually. It takes 1.5 s to transfer a 512×512 byte image to the μVAX-I. For this reason, only those regions (or windows) which are known to contain pads are transferred; the position and size of these windows are established using the CAD data. A window typically encompasses a complete row of pads and is slightly larger to allow for misalignments.

3.4. Segmentation by Automatic Thresholding

The labelling process, known as segmentation, is accomplished by "thresholding" the image, i.e., comparing the image values to three predetermined ranges: the pixel is labelled according to the range in which it falls.

The procedure for automatic thresholding is based on the assumption that the grey level histogram of the image (or the relevent window) has a trimodal structure, e.g. three large peaks. Usually the raw histogram is quite noisy and has more than three peaks, so it has to be smoothed by local averaging to reduce it to three peaks. It is also

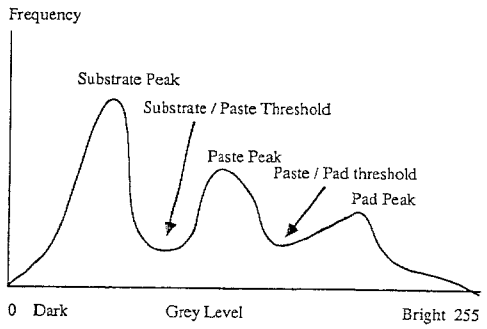

Fig. 1. Grey level histogram.

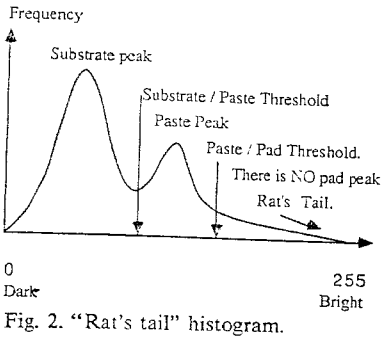

Fig. 2. "Rat's tail" histogram.

assumed that the substrate peak (the darkest one) will be the largest. This can be ensured by adjusting the size of the window so that it contains more substrate than anything else.

The algorithm for automatic thresholding proceeds by finding the highest mode in the histogram. This is the substrate peak (see Fig. 1). Secondly, the next peak is found by searching in the "bright" direction (assuming that the substrate is dark) beginning sixteen grey-levels from the substrate peak. If this peak is strong enough and far enough away from the substrate peak, this is labelled the paste peak. The minimum point between these two peaks is substrate/paste threshold. Thirdly, the next peak is found by searching in the "bright" direction, beginning sixteen grey-levels from the paste peak. If this is strong enough and far enough away from the paste peak, this is labelled the pad peak. The minimum point between these two peaks is the paste/pad threshold. If it has not been possible to find the pad peak, the paste/pad threshold is found using measurements of histogram slope. This type of histogram has been dubbed a "rat's tail" histogram as the segment of the histogram which corresponds to the pad look like a rat's tail: the slope measurements are computed in this rat's tail section (see Fig. 2). If it has not been possible to find any new peaks, an error is flagged, and the set of thresholds found in the previous thresholding selection is used as a default . These defaults are also used if either of the new thresholds is too disimilar from the previous ones.

3.5. Finding the Vertical Bounds of a Row of Pads: The "Cutter" Algorithm

The "cutter" algorithm determines the vertical extent of the row of pads (assuming that the pads

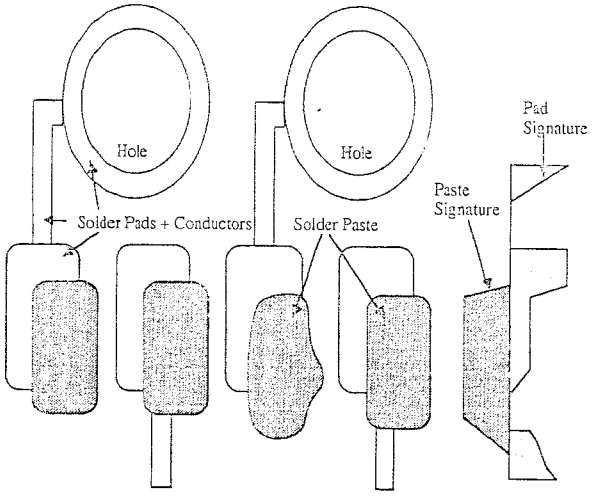

Fig. 3. Location of the vertical bounds on a row of pads—the "cutter" algorithm.

are distributed in a horizontal line). The algorithm derives its name from the need to "cut" the pads off from the attached conductors. It is assumed that the exact height of a pad is known, that the board is aligned to within 1 degree of the camera axis (see Fig. 3) and that the pads can be located ± 1 mm in X- and Y-directions by the X-Y table. The image is divided (conceptually) into four rectangular arrays; each is processed in turn. It is assumed that each array contains a row of pads running along its major axis, with solder tracks, holes, and other irrelevant information above and below the pads. The 2D array is summed along its minor axis to produce two 1D signatures of the summed intensities, one corresponding to the paste and one to the pads. These signatures are thresholded to identify four types of signature subsections corresponding to pad only, paste only, both pad and paste, and either pad or paste (or both). Each of these sections are considered in turn, giving priority to the pad sections, and to pad and paste sections, to find which section best fits the known pad size. The bounds of the section which is chosen correspond then to the location of the top and bottom of the row of pads.

3.6. Detecting the Individual Pads: The "Chopper" Algorithm

To find the individual pads within this row of pads, the 2D array is summed along its major axis to produce another 1D signature (see Fig. 4).

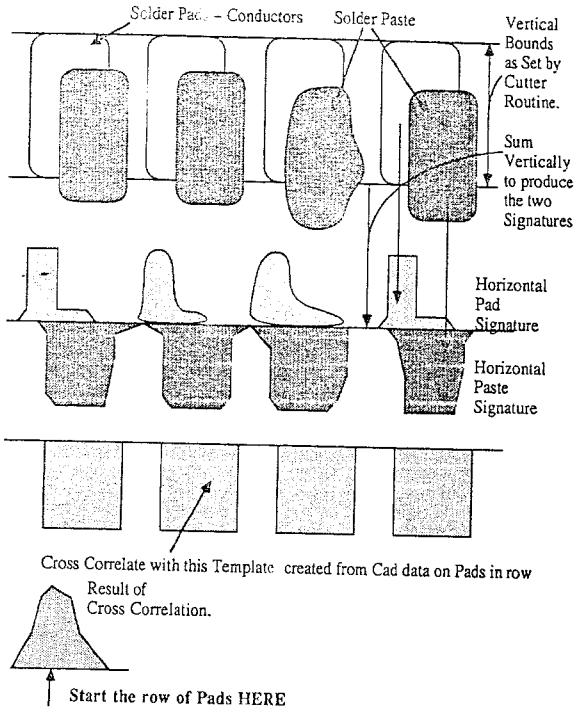

Fig. 4. Detecting the individual pads—the "chopper" algorithm.

Because we are searching for the position of the pads, pixels which correspond to pads are weighted more heavily in the computation of the signature. The position of each individual pad is computed by matching a template signature of an ideal row of pads with the signature computed from the image, and identifying the position at which the template and image signature match best. This allows one to model the pads as a series of perfect rectangles which are then subjected to the subsequent analysis phase, on an individual basis. The "chopper" algorithm is extremely robust and functions correctly, even with very poor image quality, and with serious occlusion by paste.

3.7. Measurement of Location and Paste Coverage

The pads, which have been isolated using the cutter and chopper algorithms, and delineated by perfect rectangles, are comprised of pixels labelled as either paste or pad. The position of the centroid of the paste region is next computed. The relative displacement of the centroid from the centre of the rectangle provides an initial measure of the displacement of the paste and, hence, the position-

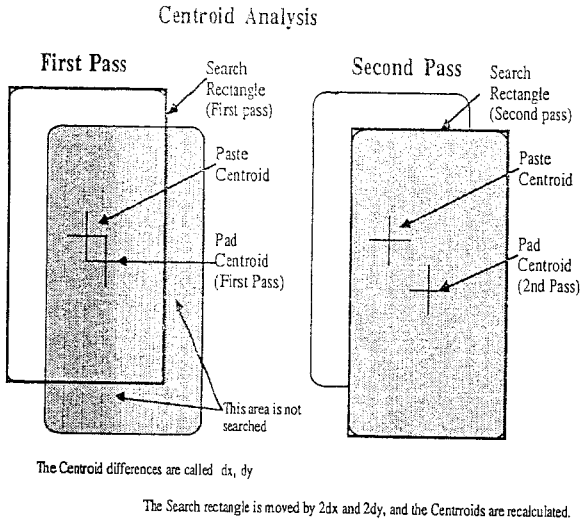

Fig. 5. Two-pass centroid analysis.

ing of the silk screen. However, since we have restricted our analysis to the estimated pad window, some paste may be neglected. To account for this, the window is shifted in both the X- and Y-directions by twice the difference between the centroid positions. New paste area and centroid values are computed for this shifted rectangle. The true displacement is given by these updated centroid positions (See Fig. 5).

3.8. Computation of Statistics on Paste Location and Coverage

Area, X-error, and Y-error figures were generated directly for each pad, as described above. The mean, maximum, and minimum of each figure were calculated for all four rows of pads. The mean and worst-case X- and Y-displacement errors, and paste coverage were calculated for each device, and again for all the devices on the board. These were tested against a set of threshold values. A device could be rejected if the maximum or minimum coverages were above below the threshold values, mainly to detect a blocked part of the stencil or a pad with very little paste on it for whatever reason. It could also be rejected if the mean coverage value had deviated from another set of much stricter limits. This was to detect the process drifting from its correct levels. Similarly, the X- and Y-displacement errors were tested for mean and maximum deviation. A mean deviation would imply that the process was drifting, while a maximum deviation error would imply that a single pad was seriously out, probably due to low paste coverage. The entire board would fail if a single device failed any one of the tests.

3.9. Routing Optimization

The list of inspection points that is generated from the CAD layout data must be reordered to minimize the time spent visiting each point. This task is the "Travelling Salesman in a Plane" (TSP) problem [6]. The problem is NP complete: a system containing N points can only be solved in $(N-1)!$ operations by simple methods. As boards may contain up to 1000 components, this is a serious problem.

Suboptimal solutions are available, however, and the problem can be solved in polynomial time. The solutions may not be perfect, but are adequate for most purposes. The technique chosen was the "Simulated Annealing" (SA) algorithm, invented by Kirkpatrick et al. [7]. This is a "probabilistic hill climbing" algorithm which may be used to optimize functions of very many variables. It is an extension of the "iterative improvement" technique, but the performance is much better. In iterative improvement, the system is said to have a cost C_i at each state. The system is disturbed (in the case of the TSP by swapping two points chosen at random) and if the new cost (the total distance around the points) is lower than the old, the swap is accepted. The process continues until no swaps have been made in a given number of attempts. The problem is that the system may get stuck in a local minimum, rather than the global minimum. For this reason, the process is usually run several times, and the best result is accepted. If the cost reduction process may be likened to descending a hill, the system has got stuck on a small side peak. The essence of simulated annealing is to allow the system to "climb" out of the local minimums and reach the global minimum. To the iterative improvement algorithm we add a control variable, usually called temperature (T). The system is disturbed in the same random way as before: all cost reductions are accepted, but some cost increases may also be accepted, particularly when the system is "hot". (The control variable "temperature" (T) has a high value.) The system is initially heated to "melting point", at which stage it is completely random, and has a

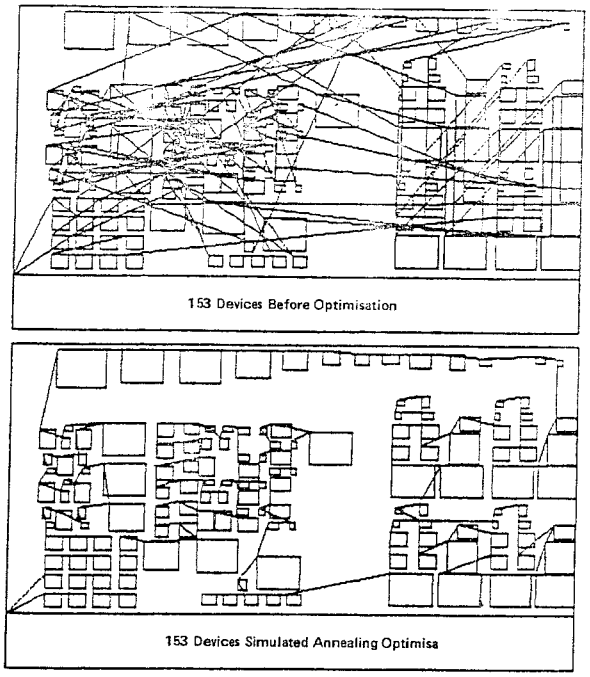

Fig. 6. Simulated annealing optimization.

very high cost. It is then slowly cooled down in the manner of annealing a metal, until it is said to be frozen. At this stage, only cost reductions are accepted. The process stops when no changes are accepted after a given number of iterations in the same manner as the iterative improvement algorithm. Because increases in energy are allowed, the system can "climb out of" a local minimum, and reach the global minimum, or close to it. It is important to "cool" the system slowly enough, as if this is not done, the system will "quench" too quickly, and may not reach the global minimum. The degenerate case of this is the iterative improvement algorithm, where the system is quenched immediately: at a temperature of zero, no increases in energy can be accepted, and the system is likely to get stuck in a local minimum. These techniques are due to the papers of Brady [8], Romeo [9], and White [10]. See Fig. 6 for an example of the results achievable with the simulated annealing optimizer.

4. Discussion

4.1. Tuning System Performance

The system was initially tested on a VAX 11-730 and took 13 s to inspect a 68-pin device. The target speed was one second, and so attempts were made to speed the system up. Four techniques used to increase the system performance: windowing to exclude unnecessary parts of the image, sampling to reduce the number of points processed, straight line coding of loops to reduce the looping overhead on simple actions, and judicious use of register variables and pointers to optimize processor use. The windowing and sampling are general techniques, and would work with any processor, while the straight line and register variable oprimizations are more processor specific. This tuning resulted in an inspection rate of 2.6 s/device.

4.2. Board Illumination

The lighting was critical to the segmentation process as it defined the relative intensity of the light reflected from the various parts of the devices, and this intensity was used to segment the image. Many lighting schemes were tried, including room lighting, viewing at an angle (10 degrees off vertical), and spotlighting, with little success. It was then decided to use a lighting system similar to that used by Hara [3] with a combination of side and on-axis lighting which he considered to be necessary to view the curved specular shapes of printed circuit patterns. The light was provided by a circular fluorescent tube running from a 240 V 50 Hz power supply. This was further improved by hanging a skirt of paper reflectors around the light to reflect a more diffuse light onto the board, and light the board from more oblique angles. Various viewing apertures were used: the best results were found from a 25 mm aperture (significantly better than a 75 mm aperture). This was because the light was being reflected at almost 90 degrees to the shiney surface of the solder pad and straight back up into the lens.

Unfortunately, this setup also proved to be unsatisfactory as waves of alternating bright and dark bands were seen to move up the image at less than 1 Hz. These caused some pixels to be alternately classified as pad or paste, depending on the position of the wave, and caused an unacceptable degree of variability in the image, and the performance of the system. It was considered that the 50 Hz lighting was interfering with the sensor, which was scanned at 25 Hz, and so it was decided to use a higher frequency light source. This was

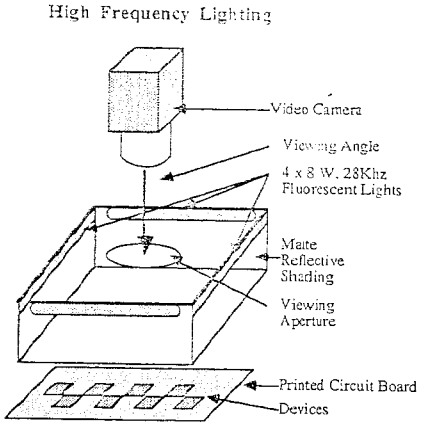

Fig. 7. High-frequency lighting.

achieved by using four 28 kHz high-frequency fluorescent bulbs arranged in a square. A new lighting box was constructed out of sheet aluminium with a matte white interior to diffuse the light. The new box had different reflectance properties, and was not as bright as the old 22 W circular bulb, so a reflecting paper bottom was made for the box, with an aperture of about one fifth the total base area of the box. This reflected most of the light back into the box, while illuminating the devices satisfactorily. Again a 25 mm viewing aperture was found to give the best results, and this light box was used in the final version (See Fig. 7).

To increase the reliability of thresholding at the substrate/paste boundary, a red filter was used to darken the substrate. This worked very well but required an increase in camera gain of 6 dB to compensate for the loss of image intensity, and an increase in the lens aperture. An infrared blocking filter was also used to avoid washout due to the sensitivity of CCD sensors to radiation in the near infrared range of the spectrum. This had no noticable effect on the system, but it was retained for safety sake. The lighting, sensor, and filtration subsystem was critical to the feasibility and accuracy of the system: it is an essential feature of AVI that the user has control of the environment and sensing apparatus and should do all that is possible to "provide a simple image" [11].

4.5. Results

The system was designed, implemented, and tested in a Digital Equipment Corporation factory in Clonmel, Ireland. The prototype system is capable of inspecting all 1.27 mm spacing surface mount devices with 8 to 84 pins. The system can inspect a board containing 153 devices in 10 minutes, an average of 4 s per device. In benchmark tests, when the X-Y table was not used, the system could inspect a 68-pin device in 2.6 s.

Tests were carried out to determine the accuracy of the system by comparing the results to results obtained manually. The test was carried out with an early version of the system which had primitive pad location and analysis algorithms. The results were quite promising; the paste coverage error averaged 5%, and the paste location errors were of the order of 0.0381 mm. This analysis was only applied to one 68-pin device on one board as it took so long (8 hours) to manually classify all 68 pads on the board.

A program was written to analyze the correspondence between the errors detected on several inspections of the same board. The program generated a histogram of errors on the various devices, and produced a figure of merit from this. The histogram was set up with the devices on the abcissa, and the number of errors per device on the ordinate. The values were squared and summed, and divided by the normalized sum of the errors squared. The figure of merit ranged from 0 to 100%. Anything over 95% was considered adequate, but the best result achieved was 92%. This is not very good, but probably better than a human operator. This repeatability problem is being addressed in a subsequent version and it is thought that a better camera will ease the problem and increase the repeatability of the system.

The device list optimization was very successful. The overall inspection path could be reduced to 20% of the original unoptimized length, in a reasonable amount of time. A board with 153 devices could be optimized in 20 minutes on a μVAX-I. The optimized routes were not perfect, but were within 1 or 2% of the optimial distance, and this level of optimization was entirely adequate for the task.

4.4. Future Directions and Related Problems

The main source of concern with the system is the lack of repeatability and the inability to detect the same errors when inspecting the same board

several times. This is due to the quality of the images and the necessity to use TV standard signals. In particular the resolution of the camera is probably inadequate and the sensitivity is uneven. An improvement would be gained by using a higher-resolution camera such as the National Panasonic WVCD-50, which provides a vertical resolution of 450 lines, and better contrast control. The unevenness could be alleviated by photometric decalibaration but this was not implemented in the original system, due to time constraints, or by the use of a better camera, such as the WVCD-50.

The problem of inspecting the next generation of SMT, which use 0.635 mm spacing devices, remains a difficult one. Initially, it was assumed that the devices would have a maximum of 84 pins, but the first devices encountered had 132 pins, and were the same size as the 68-pin devices which had filled the image area. It was only possible to resolve ± 0.08 mm at this magnification, or perhaps less, due to the resolution of the camera. A higher-resolution camera might prove adequate and would improve image quality for all device sizes. Alternately, the devices could be viewed at higher magnification, and inspected one row, or less at a time. This however creates the problem of mounting a second camera, and positioning the devices beneath it. This system would be slower, as more inspections would be necessary per device. However, there are usually only a few of the very large devices on a board as they are usually microprocessors and custom devices, rather than memory chips or resistors. A third solution is to measure coverage only, and then check for pad connectivity using a maze runner algorithm to verify that a pair of pads were not bridged by solder paste. These applications would need less resolution than the pad analysis modules, and might ultimately prove satisfactory.

Future applications include solder depth, device presence, and device orientation inspection as well as post-reflow solder joint inspection. The solder depth problem would require a 3D approach, possibly using structured light, "shape from shading" techniques, or stereo. The device presence checker is relatively simple, but the device orientation and type verification problems are more complex. The device type recognition problem would require an optical character recognition module, or a powerful pattern matcher, to analyze images of the tops of the chips. The solder joint problem is the most difficult of all the problems, as the joints are 3D entities, mostly hidden by the pads, which are very close together. This is currently being attempted using X-rays and thermographic techniques [12], but is at the experimental stage, and is expected to be very expensive ($500,000 per unit).

Any further development should revolve around improving the reliability of the system, increasing the speed and creating a menu-driven interface for the whole system. Finally, to complete the solder paste inspection system, solder depth measurement should be attempted.

5. Conclusions

The objective was to design a system to enable a fairly slow (0.4 MIP) processor to inspect and analyze an image of a 68-pin device in one second. This involved locating the pads and solder paste, calculating their areas, and relative displacement. This was achieved by optimizing the lighting and restricting the orientation of the board, which made the segmentation of the image much easier and allowed high-speed heuristic algorithms to be developed.

Five main algorithms were developed: automatic thresholding, vertical pad location, horizontal pad location, two pass centroid analysis, and the inspection list routing optimizer. In addition, the algorithms were integrated into a system that an engineer or operator could use in a normal production environment.

The automatic thresholding was not used in every inspection as anticipated, but was relegated to a setup function. The algorithm was too sensitive to the proportions of pad, paste, and substrate in the imago to permit it to be used for each device. The system was found to be more reliable if the thresholds were set once per board, or set of boards, as the lighting did not vary significantly over such a small timescale. Despite this, histogram display and analysis proved a valuable tool for assessing the quality and suitability of the lighting.

The pad location routines were refined and developed until they worked very well. The most important feature here was the ability to search for a row of pads one dimension at a time. This approach was inspired by Vanderheydt's Lavim

system [13] which used 1D signatures of the image for faster processing. The effect of compressing a 2D image to a 1D signature was to smooth out the noise, without loosing any detail. In the early versions of these algorithms, the scales of the devices were not fixed precisely, which prevented template matching algorithms from being used, but it was finally decided to fix the scales, allowing the development of simpler, more robust algorithms. This improved the accuracy greatly, especially with noisy images, or when the paste was badly smeared, and the algorithms were considered to be most satisfactory.

The two-pass centroid analysis algorithm also worked well, and was adaptable to pads of any shape, but the overall results could have been better. It was difficult to improve the algorithm, short of providing a better image, and more analysis of its operation would be required for a final system.

The routing optimizer was a great success, as it was fast and reliable, and provided a high degree of optimization. It would have been impossible to complete the project without it, although the need for it was not forseen initially.

The system setup, CAD processing, and integration were much more complex and took longer to implement than was anticipated, but were essential to the use of the project in an industrial environment, and the systems worked satisfactorily. The target speed of 68 pins per second was not reached, but probably could have been if a μVAX-II and a faster *X-Y* table had been available. A camera with improved definition and more even sensitivity would have helped to solve the reliability and accuracy problems.

Overall, the project has demonstrated the feasibility of designing and implementing a working surface mount inspection system with limited resources. Many of the potential problems were solved or avoided with suitable lighting and image registration. The restricted environment allowed novel heuristic algorithms to be developed to analyze the images at maximum speed, making the system perform at an acceptable rate. The remaining problems could be solved with hardware which has become available since the inception of the project, without modifying the algorithms. The successful design and implementation of the system provides a firm basis to attempt the more complex inspection tasks of the future.

Acknowledgment

This work was supported by a research grant provided by Digital Equipment Corporation, Clonmel, Ireland.

References

[1] R. Chin and C. Harlow, Automated Visual Inspection: A survey, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-4, No. 6 (1982).

[2] R. Kruger and W. Thompson, A Technical and Economic Assessment of Computer Vision and Industrial Inspection and Robotic Assembly, *Proc. IEEE*, Vol. 69, No. 12 (1981).

[3] Y. Hara, N. Akiyama and K. Karasaki, Automatic Inspection for Printed Circuit Boards, *IEEE Trans. Pattern Matching and Machine Intelligence*, Vol. PAMI-5, No. 6 (1983).

[4] J. Mundy and G. Porter, Visual Inspection of Metal Surfaces, in: *Proceedings of the 5th International Conference on Pattern Recognition*, FL (1980) pp. 232–237.

[5] K. Doyle and M. Icough, Automatic Inspection of Printed Circuit Boards, in: *Proceedings of the Technical Program*, Internepcon/UK84, Brighton (1984) pp. 145–150.

[6] R. Parry, The Infamous Traveling-Salesman Problem, A Practical Approach, *Byte* (July 1981).

[7] S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, Optimisation by Simulated Annealing, *Science*, pp. 671–680 (May 1983).

[8] M. Brady, A Wire Wrap Design Aid, Written in Prolog, *Comput. Aided Design*, Vol. 16, No. 5, pp. 253–263 (1984).

[9] F. Romeo, Research on Simulated Annealing at Berkely, *Proceedings IEEE International Conference on Computer Design*, Port Chester, NY, 1984.

[10] S.R. White, Concepts of Scale in Simulated Annealing, in: *Proceedings IEEE International Conference on Computer Design*, Port Chester, NY, 1984.

[11] M. Yachida and S. Tsuji, Industrial Computer Vision in Japan, *Computer* (May 1980).

[12] S. Jones, Flexible Inspection Systems In Surface-Mount Device Production, *Robotics Engineering*, pp. 5–9 (1986).

[13] P. Vanderheydt, A. Oosterlinck and H. van den Berghe, Application of Pattern recognition and Image Processing, *Sensor Rev.*, pp. 78–80 (1982).