ANATOMY OF AN APPLICATION:

AUTOMATIC ALIGNMENT IN THE SEMICONDUCTOR INDUSTRY.

Perry West — Automated Vision Systems, Inc.

imagenation
vision system specialists

Table of Contents

Introduction

Automatic alignment, used extensively in the semiconductor industry, includes both "front end" wafer fabrication and "back end" test, assembly, and packaging (TAP) processes. This white paper presents two such automatic alignment application case studies — one from the back end test area, the other from the front end of the process — and shows how each application evolved, lessons learned along the way, and the resulting impact on today's best practices. Of course, automatic alignment is not limited to the semiconductor industry, but is widespread throughout all manufacturing industries using automatic assembly equipment. Thus, the lessons learned from the two case studies detailed here can be applied to a wide spectrum of potential applications.

Generic Alignment Procedure

The alignment of wafers is a superset of the alignment of die and includes 1) prealignment to correct for coarse angular and translational mispositioning and 2) die alignment that provides precision location keyed to specific features on the individual semiconductor die. Although this paper focuses on die alignment, a brief look at wafer prealignment is included below.

Prealignment

The most common method of prealignment is to transfer the wafer onto a chuck where it is rotated while the radial distance from wafer edge to center of the chuck is measured. The sequence of these radial measurements is used to determine the centering of the wafer on the chuck, its translational position, and the location of flats or notches on its periphery which defines its rotational orientation. The movement of the chuck assures rotational alignment, and the mechanism (e.g., robot) that then transfers the wafer to the workspace compensates for its translational misalignment. In this way, prealigners typically are able to align the wafer to within a degree of rotation and a few thousandths of an inch in translation. Ultimately, prealignment accuracy in knowing where die are located relies on the precision of the prealigner, the precision of the wafer transport mechanism and the precision with which die are placed on the wafer. With modern photolithography, device locations can repeat from wafer to wafer to within a few thousandths of an inch.

History

This section reviews the most prevalent means of achieving alignment in the past, including 1) manual methods, 2) mechanically scanned point sensors, 3) binary correlation and 4) projections.

Manual Methods

Originally, alignment was a manual process, with operators looking through a microscope, verifying correct die positioning, and then making any needed adjustments of die or wafer by hand. This was a slow, fatigue-inducing operation with a high yield loss due to misalignment, even among the most conscientious operators.

The main challenges with manual alignment were low speed and yield loss associated with occasional inaccurate alignment. Although the equipment using manual alignment was much less expensive than the same equipment using automatic alignment, the low speed of the process required more equipment, more operators, and more floor space. Furthermore, even though alignment-intensive operations such as wire bonding were relocated to areas with low labor and real estate costs, it was ultimately the yield loss incurred with manual methods that made automatic alignment an economic necessity.

Even so, for many years, companies resisted replacing manual equipment with equipment enabling automatic alignment. They realized that operators, once they gained experience, were adaptable, whereas the available systems were not. This adaptability was evident in both training for a new part and in problem solving. Experienced operators could adapt to a new part using only written instructions, and could, in effect, be trained off-line. Similarly, if a problem arose with the pattern used for alignment — part of it was missing or occluded, for example — the operator was able to adapt and use the remaining portion of the pattern to continue working. These same attributes of adaptability have emerged only recently in automatic alignment techniques.

> *This technique, relying on manually operated equipment, requires much less in terms of capital expense compared to automatic alignment. Even so, the savings in yield loss alone makes automated equipment a profitable investment, even in areas with low labor costs.*

Mechanically Scanned Point Sensors

Two early attempts at automatic alignment relied on point sensors, with the wafer mechanically translated below the sensor. In one instance, a Japanese manufacturer of wafer probers used a laser-based point sensor to detect the streets between die on the wafer. Mounted on a stage, the wafer was mechanically translated under the sensor; since streets at that time were generally clear of any pattern, the sensor's signal could be analyzed to find repetitive patterns correlating to the streets.

> *Although low cost, this technique was also slow, requiring repeated mechanical movement of the wafer. As testing costs rose, the process became more expensive because of the time needed to align the wafer. Also, this approach required relatively wide streets free of patterns. However, as technology progressed, streets became narrower and included test patterns which used to take up space of die, rendering the point scanning approach unreliable as well as time-consuming. What was needed and is used today: an alignment technique using features on the die themselves, and able to recognize features unique to each individual die design.*

In another instance in the late 1970s, an equipment manufacturer used a fiber optic sensor to automatically align individual die for testing. The die required a particular fiducial pattern that allowed for detection by mechanical scanning.

> *While this approach proved workable for a very specific purpose, the need for mechanical scanning meant slow output. Also, the fiducial pattern was programmed into the equipment and could not be changed by the user, precluding its continued use as products evolved. Today, of course, semiconductor real estate is too expensive to allow space for a dedicated fiducial target. As in the previous example, what was needed and is used today: an alignment system able to recognize features inherent in the design of the individual devices without requiring a fiducial mark or other predefined feature.*

Binary Correlation

The first commercially viable automatic alignment, developed in 1978, used a technique called binary correlation, whereby an image template is compared with the contents of a search region. The template must be no larger than the search region and normally is significantly smaller. For each possible position of the template within the search region, the number of pixels which agree in the two regions are totaled. The result is an array of values in which a peak indicates a best match. Mathematically this is represented as:

$$C_{x,y} = \sum_{j=0}^{j=J-1} \sum_{i=0}^{i=I-1} \overline{T_{i,y} \oplus S_{x+i,y+j}} \quad \forall 0 \le x \le X - I + 1, 0 \le y \le Y - J + 1$$

where:

$C_{x,y}$ is the correlation result total at location x,y
$T_{i,j}$ is the template image data at coordinate i,j
$S_{x,y}$is the search region at coordinate x,y
I is the number of columns in the template (i.e., 0=I=I-1)
J is the number of rows in the template (i.e., 0=j=J-1)
X is the number of columns in the search region (i.e., 0=x=X-1)
Y is the number of rows in the search region (i.e., 0=y=Y-1)

In the 1978 implementation, the template was fixed at 64 x 64 pixels, and the search region was fixed at 192 x 192 pixels. To perform the correlation, the number of operations (compares and accumulates) is given as:

$$N = I \cdot J \cdot (X - I + 1) \cdot (Y - J + 1)$$

where:
N is the number of comparisons required.

For the template and search region sizes given above, 68,161,536 operations are needed to perform correlation — a significant computational burden for microprocessors available back in 1978.

Two techniques were designed to expedite the process: use of an 8 x 8 hardware correlator that significantly improved correlation speed, and use of a coarse correlation to identify candidate regions followed by a fine correlation over the areas surrounding the candidate locations. Even with these enhancements, the process took eight camera field times, or about 150 milliseconds.

In this particular implementation, a coarse template was formed by taking every fourth pixel horizontally and vertically, thus reducing the 64 x 64 pixel template to 16 x 16 pixels. The coarse template was moved across the search region skipping every other pixel in both the horizontal and vertical directions, reducing the number of operations for the coarse search to 1,081,600 or only about 1.6% of the operations required for a high-resolution comparison. The locations of the two highest peaks from the coarse correlation were identified as candidate locations. Since correlation peaks are usually not crisp and sharp, but tend to have significant noise, this implementation qualified the peak value in the 7x7 region, and then computed the centroid of the 49 values. A full-resolution correlation using the 64 x 64 template was performed over regions that were 70 x 70 pixels centered about the candidate locations (+/- 3 pixels in each direction about the candidate location). This required an additional 200,704 comparisons. In total, the two-stage correlation proved to be 53 times faster than the exhaustive high-resolution correlation.

In 1978, choice was limited in solid-state area cameras, and those that were available were very expensive, so the implementation used a vidicon-based CCTV camera. To compensate for the drift inherent in the video camera's signal, one side of the vidicon's target area was covered. Thus, the beginning of each scan line was known to be dark. The dark level from the camera was sampled along with the average video level for each scan line, becoming the threshold for the next scan line. In this way, the system compensated for changes in dark level and for brightness changes in the parts being sampled. Lighting was normally incandescent, supplied by a bifurcated fiber optic bundle, with some installations using a small fiber optic ring light.

This technique recognized the need for a trainable pattern recognition system and provided it. This implementation also recognized and addressed the issue of variance in the video signal from part to part and over time. The technique for sampling the vidicon's dark level and compensating for it has been replaced in most solid-state cameras by a more sophisticated technique called double-correlated sampling. The attempt at making the threshold dynamic helped compensate for light level changes and part reflectivity; unfortunately, in some situations, this technique proved inadequate. Later techniques would use gray-scale values to provide an increased opportunity to address these variables.

Vidicon cameras, common at the time, also posed problems. An electromagnetically swept electron beam determined the image target area; however, changes in voltage or temperature or external electric or magnetic fields could change the size and shape of the sensed area. If a camera drifted or was somehow affected, the alignment might give an inaccurate answer or might fail altogether. Later versions of this approach relied on solid-state cameras; today's modern solid-state cameras, used in all machine vision applications, are able to solve these problems.

> *There were two shortcomings to this technique. First, the template was a fixed size (64 x 64 pixels in the above instance). Secondly, the operator selected the pattern for the template, which could be sub-optimal. Current techniques address these problem areas by allowing the operator to designate the template area and by searching to find the most unique pattern in the region to use as a template for matching.*

Projections

In 1980, a major manufacturer of wire bonders developed a successful approach to automatic alignment using projections and relied on machine vision techniques to aid the user to recognize patterns.

A projection is the sum of gray-scale values in all columns or all rows of an image; more generally, it can be the sum of gray-scale values perpendicular to a path in any direction. Thus, a projection is a signature of the image or portion of the image from which it is formed. Although it is not possible to reconstruct the original image from one or even several of its projections, computed tomagraphy uses a large number of X-ray images, each of which is a projection, taken at regular angles around an object or patient, to construct an image of the cross section of that object or patient. In machine vision, it is recognized that a pair of orthogonal projections, while not sufficient to reconstruct the original image, is sufficiently unique to be used for pattern matching.

In this particular implementation, an image from a vidicon-based CCTV camera was digitized into a computer's memory. The image size was 240 rows by 320 columns, and there were 16 possible gray values (4 bits) per pixel. The active pattern search region was limited to a square region 240 x 240 pixels. The target region was nominally 64 x 64 pixels. Centered about this target region, six projections were acquired: three horizontal — 64 x 240, 64 x 200, and 64 x 160 pixels — and three vertical — 240 x 64, 200 x 64, and 160 x 64.

During a search, the horizontal and vertical projection for the 240 x 240 search region were formed. In a sliding 64 row or 64 column window over each projection, the target and search projections were compared using the 64 x 240 and 240 x 64 projections obtained from the target. The absolute differences of the projections were summed. Mathematically, this is represented by:

$$Q_j = \sum_{i=0}^{63} \left| S_{j+i} - T_i \right|$$

where:

Q_j is the error associated with the j^{th} position in the projection of the search region
S_{j+I} is the i^{th} element of the j^{th} position in the search region projection
T_i is the i^{th} element of the target projection

In principal, the minimum value of Q corresponds to the X or the Y coordinate of the best match. However, recognizing that both the reflectivity of the die varies from die to die and from lot to lot and that the light source intensity also varies over time, the approach normalized the projections to compensate for this variation. This is accomplished by multiplying the search projection values by the average of the target projection and the target projection values by the average of the search projection, or mathematically:

$$\overline{S} = \frac{1}{64} \sum_{i=0}^{63} S_{j+i}$$

$$\overline{T} = \frac{1}{64} \sum_{i=0}^{63} T_i$$

$$Q_j = \sum \left| \overline{T} S_{j+i} - \overline{S} T_i \right|$$

The minimum Q for the horizontal projection identifies the best fit Y coordinate. Likewise, the minimum Q for the vertical projection identifies the best fit X coordinate. To improve the search accuracy, the process is repeated in a 200 x 200 pixel sub region centered about the best fit X and Y coordinate using the 64 x 200 and 200 x 64 projections from the target. In a similar manner, a final search region of 160 x 160 pixels is searched to give the best accuracy.

During training a pattern, the operator positions the die under the vision system at a strategic position, usually using a bond pad in one corner of the die. The vision system searches the central 160 x 160 pixel region of the camera's field-of-view to find a target pattern of 64 x 64 pixels that is most dissimilar to any other pattern in the region, thus insuring maximum alignment reliability.

> *This technique, like binary correlation discussed above, recognized the need to mitigate the effects of lighting and reflectance changes. Rather than using an adaptable threshold, it used gray-scale images and normalization. It was the first approach to attempt to help the operator optimize the target for best reliability. The approach used a vidicon camera, not out of necessity, but because solid-state cameras were very expensive in 1980. One of its biggest drawbacks was that although projections were usually unique, there was no guarantee of uniqueness. Therefore, there could be die that would not be handled reliably by this method.*

These four alignment methods demonstrate the history of trial and error. A continual evolution of techniques has led to new solutions in the search for optimum automatic alignment.

Current Applications

Two case studies detailed below -- Wafer Probing and Film Thickness Measurement -- demonstrate the recent application of automatic alignment in semiconductor equipment and illustrate the difference in lighting, cameras and algorithms required by each.

Wafer Probing

A major manufacturer of wafer probers, with a product based on a proprietary hardware processing platform becoming obsolete, is now designing its next-generation product. The key performance criteria: a wafer prober able to align the typically 40µm square pads on the die with an array of probes, all providing contact with the centers of each pad. Alignment accuracy must be within 2µ to insure quality probing.

This company's wafer probers are used for temperature testing. In this process, the chuck that supports the wafer is heated or cooled so that devices can be tested at temperature extremes, thus subjecting the wafer to thermal expansion. In wafer probing where all wafers are at room temperature, the die on a wafer are a known distance apart. However, when thermal expansion is factored in, the difference in spacing from one side of the wafer to the other requires that each die be automatically aligned.

Lighting. In this application, lighting is provided by an LED ring, providing several advantages over conventional incandescence. LED lighting is lower power, producing less heat. An incandescent lamp uses between 50 and 150 watts plus power for cooling fans; an array of 20 LEDs requires approximately 3 watts of power, including the power supply. Also, an LED offers vastly longer life expectancy than incandescent lamps: an average 100,000 hours vs. an incandescent's range of 50 to 1,000 hours. One characteristic of LEDs that may be a benefit or a problem, depending on the application, is that they are essentially monochromatic. However, in the case of a wafer prober where the requirement is to locate the metal pads, monochromatic light is not a handicap.

In the company's previous product version, the operator manually set the light level; thus, there was no way to be certain that a level was optimum for a given product or that it was the same level previously used — a variable which could degrade alignment accuracy. In the new product version, the light level is set by the software, either automatically through image analysis or from a stored recipe for each type of device.

Camera. The camera being used for automatic alignment is a standard monochrome solid-state CCTV camera, with a second color CCTV camera used to provide the operator with an image of the die. The monochrome camera has a very small field-of-view (fixed at 12.8mm x 9.6mm) to give the best alignment accuracy, and is not suitable for operator viewing to monitor the probing process. Because of that limited field-of-view, the wafer prober's automatic alignment system locates one corner of a die, the wafer is moved until the diagonally opposite corner of the die is under the camera, and that corner is located. This procedure provides accurate position correction for both translation and rotation.

The approach uses a frame grabber providing an image resolution of 640 x 480 pixels. Since an ordinary CCTV camera does not supply or accept a pixel clock, both the frame grabber and the camera have their own clocks. Using the synchronizing signals from the camera's composite video, the frame grabber synchronizes itself to the camera. However, in most cases, the pixel clocks in the camera and the frame grabber are different frequencies; even on the same frequency, there is inevitably a phase difference between the two. Therefore, as the frame grabber resamples and digitizes the image information from the camera, there will not be a one-to-one correspondence between photosites (pixels) in the camera and pixels digitized in the frame grabber in the horizontal direction. Because the frame grabber is synchronous on a line-by-line basis, this uncertainty is not present in the vertical direction. The net result is that for any given digitized pixel in the frame grabber, there is a one pixel uncertainty in its true horizontal position. The actual percentage of pixels that are off true position depends on the relationship between the two clocks, but is usually low. In addition, the algorithms used in automatic alignment have the effect of averaging the data so that an occasional positional error of one pixel in a scan line contributes only a very small error to the overall result.

In those applications where precision is critical, the system can be designed with a digital camera. These cameras use a single pixel clock for reading out the sensing array, digitizing the image data and storing the image data, thus eliminating the possibility of pixel position error. In other applications however, cost precludes use of digital cameras, which average two to three times the price of CCTV cameras.

An OEM who chooses a digital camera it reaps a mixed blessing with its 1000 x 1000 pixel high resolution capabilities. The advantages of high-resolution cameras: the ability to achieve higher precision with a given field-of-view or comparable precision as CCTV cameras with a larger field-of-view. But the disadvantages include both increased camera cost and the burden of processing more data. In alignment for photolithography where positional accuracy of a small fraction of a micron must be obtained, the added cost of a high-resolution camera is justified. However, in equipment such as wafer probers, where requirements do not mandate high-resolution, the designer must weigh the tradeoff of a more expensive camera against a less costly prealigner and wafer transport device.

Algorithm. Software in this application is a commercial package that offers normalized gray-scale correlation running on a personal computer. Normally, a board with dedicated hardware processing would be faster than a similar software algorithm running on a PC, but with advances in PC speed, the software-based approach offers a two times speed improvement over the previously used board. Of course, the special processing board's technology is several years old.

Normalized gray-scale correlation, introduced to machine vision in 1987, is a technique that is expressed mathematically as:

$$C_{x,y} = \frac{\sum_{j=0}^{J-1}\sum_{i=0}^{I-1}\left(I_{x+i,y+j} - \bar{I}_{x,y}\right)\left(T_{i,j} - \bar{T}\right)}{\sqrt{\sum_{j=0}^{J-1}\sum_{i=0}^{I-1}\left(I_{x+i,y+j} - \bar{I}_{x,y}\right)^2 \cdot \sum_{j=0}^{J-1}\sum_{i=0}^{I-1}\left(T_{i,j} - \bar{T}\right)^2}}$$

where:

$C_{x,y}$ is the correlation result corresponding to image coordinate x,y

$I_{x,y}$ is the pixel value from the image at coordinate x,y

$T_{i,j}$ is the pixel value from the template at coordinate i,j

I is the number of columns in the template

J is the number of rows in the template

$\bar{I}_{x,y}$ is the average of the image pixel values over an area I x J pixels starting at coordinate x,y

$\bar{T}$ is the average of the template pixel values

Basic correlation is given by:

$$C_{x,y} = \sum_{j}\sum_{i} I_{x+i,y+j} \cdot T_{i,j}$$

By subtracting the average levels from this equation:

$$C_{x,y} = \sum_{j}\sum_{i} \left(I_{x+i,y+j} - \bar{I}\right)\left(T_{i,j} - \bar{T}\right)$$

the result is normalized for shifts in brightness but not for changes in contrast. The denominator:

$$\sqrt{\sum_{j}\sum_{i} \left(I_{x+i,y+j} - \bar{I}_{x,y}\right)^2 \cdot \sum_{j}\sum_{i} \left(T_{i,j} - \bar{T}\right)^2}$$

normalizes the correlation result for contrast changes.

Clearly, correlation, whether binary or gray-scale, requires the same number of operations. In binary correlation, the operation is compare-accumulate; in gray-scale correlation it is multiply-accumulate. As the normalization factors are added, computation for each gray-scale operation increases dramatically. Most gray-scale correlation algorithms are designed for efficiency and precompute as many of the factors having to do with the reference image as practical. Other approaches to decreasing computational burden involve a) using image pyramids to search first at coarser resolution such as was done with the binary correlation discussed above, b) squaring the result, effectively eliminating the need to compute the square root in the denominator and c) using look-up tables and similar techniques to avoid multiplications. Other implementations of normalized gray-scale correlation improve speed by using only a fraction of the pixels in the reference image. In some strategy, the pixels are picked at random; in other strategies, the pixels at edges, especially in the vicinity of corners, are chosen. In many cases, using less than 10% of the number of pixels in the reference image, picked by software, can give very satisfactory results.

While normalized correlation is robust in finding patterns that are translated or that differ in brightness and contrast, it is very limited in finding patterns that are rotated or have changed size. In the case of the wafer prober, wafers are prealigned to an accuracy of a degree or two, and there is no appreciable size change. Normalized correlation is generally accepted to be adequate with rotations up to seven degrees. Of course, with rotation comes a decrease in accuracy and reliability.

> *This implementation of automatic alignment illustrates a contemporary approach. It uses LED illumination which has become widely available in the last four years, solid-state CCTV cameras, which have been the dominant choice since the mid-1980s, and normalized gray-scale correlation. It shows that increases in computing power and speed, even in a personal computer, have overtaken dedicated hardware or processor approaches developed only a half decade ago. It also shows that normalized gray-scale correlation, introduced to machine vision 13 years ago, is still a powerful tool under the right conditions.*

Film Thickness Measurement

During wafer fabrication, many thin films of differing materials are grown or deposited on a silicon wafer. To maintain tight process control, it is necessary to measure the thickness of these films immediately after they are created. In addition, the wafer must be very flat before going through a lithography step. So, the wafer is “leveled” by a chemical-mechanical planarization (CMP) step. The systems that automatically measure film thickness must find the areas on the wafer designated for measurement that are typically on the order of one-thousandth of an inch square.

Complicating the alignment for film thickness is the fact that a stack of thin films becomes an optical thin-film filter. As the wafer progresses through the process, the characteristics of the light it reflects changes; when combined with the effects of CMP, the pattern can exhibit contrast reversal and shading variations. Since contemporary techniques such as normalized gray-scale correlation cannot cope with processes displaying these characteristics, it is imperative that this application be provided a new generation of automatic alignment techniques.

Lighting and Camera . While LEDs offer long life, they are unsuitable for this application because they are monochromatic. Since the film stack acts as an optical filter, the pattern might or might not return sufficient light for an exposure if illuminated with monochromatic light. Therefore, in this case, traditional incandescent illumination delivered through a fiber optic ring light is most effective.(Another possibility is the use of white LEDs, which are actually blue LEDs coated with a phosphor.) Since accuracy is vital to this application, a digital camera is used along with a frame grabber.

Algorithm. The alignment algorithm is shape based. In contrast to correlation which deals with area, shape based algorithms deal with edges. The algorithm consists of these steps:

- Edge point extraction
- Edge point filtering
- Feature detection
- Affine searching

The edge point extraction can use any of the known edge operators. The only requirement for this function is that it provides both magnitude and direction. For that reason, the Sobel edge operator is favored, using convolution to look for horizontal and vertical edge components. The convolving masks for the Sobel operator are:

1	2	1
0	0	0
-1	-2	-1
1	0	-1
2	0	-2
1	0	-1

By combining the magnitudes of each convolution, the magnitude of the edge is obtained; by taking the inverse arctangent of the ratio of the two convolution results, the edge direction is obtained. Software crafted for highest speed can gain significant speed advantages by coding this algorithm specifically rather than using general convolution.

Any edge operator will return both valid and invalid edge points, so an edge point filter is required to prune extraneous edge points. An example of an edge filter is the Canny edge detector, which uses two thresholds. Any edge point with a magnitude above the upper threshold is taken as a confirmed edge point; likewise, any edge point with a magnitude below the lower threshold is discarded. Edge points in between the two thresholds are taken to be candidate edge points, and changed to confirmed edge points if there is a path through other candidate edge points to a confirmed edge point. Other filters can be applied which use edge direction as one of the parameters.

Feature detectors identify characteristics such as straight line segments, arcs, corners, etc., which become features of the object(s). Again, filters can be used to select only relevant features — a straight line must be a minimum length to be significant, for example.

Finally, the features extracted from an image can be compared against those learned from a model. If every feature on a model were unique (e.g., only one corner, one straight line segment, one arc, etc.), if the object in the image were not occluded, distorted, or flawed, and if the object was at the same scale, location, and rotation as the model, then matching would be trivial. However, if the object is subject to one or more of these effects, a more sophisticated search is needed. By identifying a few features that may correspond between the model and the object and then transforming all features of the model with an affine transform to make them correspond to anticipated features in the object being searched, a possible match can be checked. The affine transform is expressed mathematically as:

$$x' = Ax + By + C$$
$$y' = Dx + Ey + F$$

where:

x and y are the original coordinates
x’ and y’ are the transformed coordinates
A performs scaling or stretching along the X axis
B performs rotation or skewing along the X axis
C performs translation along the X axis
D perform scaling or stretching along the Y axis
E performs rotation or skewing along the Y axis
F performs translation along the Y axis

In actual use, the user specifies bounds on each of these parameters. For example, if, as in the case of wafers that have been prealigned, rotation can be limited to a few degrees, the coefficients B and E would be bounded to be only small numbers relating to the maximum rotation allowed. By limiting the range of the parameters, the search is made faster.

> *This study illustrates that the latest technology is not always the best choice for every application. In this case, the low-power, long-life benefits of monochromatic LEDs can not be leveraged; instead, traditional incandescent illumination delivered via fiber optics is a more effective method. The digital camera, though more costly, offers a common clock between it and the frame grabber, providing optimum accuracy in the X direction. Shape-based recognition is used primarily because it can handle contrast reversal which is common on wafers that have been CMP processed.*

Remaining Challenges

Certain trends in the semiconductor industry will continue to affect the demands on automatic alignment. First and foremost, in recognition of the value of silicon real estate, is the trend toward ever- smaller features. Bond pads, where the leads attach to the IC, have shrunk from 125µm square to 40µm square, and are likely to get even smaller. This demands increased accuracy from automatic alignment which will likely be assured through use of higher-resolution (e.g., 1000 x 1000 pixel or greater) cameras.

The semiconductor industry leads all other sectors in using cost-of-ownership as a key criterion in making purchasing decisions. Not only does this require that equipment be cost-effective to acquire, but also places a premium on reliability and speed. Machine vision for alignment will have to contribute to lower overall system prices either by being more affordable while maintaining productivity levels, or by offering performance improvements that provide cost reductions in other parts of the system. For example, use of a high-resolution camera to cover a larger field-of-view than that offered by a CCTV camera can allow the equipment designer to settle for less accurate and less expensive prealigner and wafer transport equipment. Thus, although the camera cost may be higher, the overall system cost may be lower.

Another aspect of cost-of-ownership is reliability. Incandescent lamps, traditionally used in illumination for machine vision, have lifetimes in the 50 to 1,000 hour range. LED light sources, with lifetimes rated at around 100,000 hours, are now very common. Even where white light is needed, such as in the film thickness measurement system discussed above, white LEDs — actually blue LEDs coated with a phosphor — may become more prevalent because of their longer life.

Clearly, one of the biggest challenges remaining is to have fully automated training of the template or model. While current versions of both correlation and shape-based algorithms aid the operator in selecting the template or model and optimizing search parameters, operator involvement is still required. The training must not only be fast, but must provide a template or model that is reliable in use. The current trend toward edge-based pattern recognition will allow off-line training of the vision system from CAD files, and subsequently, should enhance techniques to find the optimum alignment target on a device. This off-line training and alignment target optimization will increase the time equipment is available for use and decrease the cost of ownership.

Conclusion

As demonstrated by the above case studies, as well as by industry trends, we see a move toward increased use of LED lighting and greater reliance on digital cameras. The use of these high-resolution cameras will almost certainly become commonplace as their cost declines and they allow system designers to relax constraints on other components, thereby reducing overall system costs. Processing, which used to require special hardware or processors to meet speed objectives, now can be performed with personal computers with sufficient speed to meet equipment requirements. Processing speeds in common microprocessors will continue to increase, allowing use of even the most computationally burdensome algorithms. Soon, cameras that plug directly into a PC without requiring a frame grabber will be available; however, these cameras appear, at least near-term, to be aimed at desktop conferencing. Cameras for technical and industrial applications such as automatic alignment will continue, for the most part, to utilize frame grabbers. These frame grabbers will continue to evolve to allow the use of high-resolution cameras at the same acquisition speeds now enjoyed by lower-resolution cameras.

There are, of course, other algorithmic techniques which can and have been used for automatic alignment and which are not covered in this paper. Two notable techniques are blob analysis and the Hough transform. Blob analysis, also called the SRI algorithm, isolates regions and computes attributes of these regions such as area, centroid, and moments. The centroid gives the location and the moments can help in recognition of the pattern and determination of its orientation. Drawbacks to blob analysis include the fact that the features to be found in the image must be isolated; also, it is a binary technique that does not compensate for light level or reflectance changes, and it is not tolerant of degraded (e.g., occluded) features.

The Hough transform is a voting process in which extracted edge points “vote” for all possible occurrences of the pattern being sought in all orientations. If a sufficiently strong peak occurs in this voting space, the presence of the pattern is confirmed, and its location and orientation can be determined from the peak. The Hough transform is very time-consuming in its processing demands and has not found wide adoption.

Human vision, as best we understand it, relies primarily on edges for shape recognition. Edge-based shape recognition mimics, better than its predecessors, attributes of human vision. During the reign of manual alignment, the most valuable characteristics of human operators were ease of training and the ability to deal with problems such as degraded targets. Today, automatic alignment is making dramatic progress in these areas in its approach to simulating human vision. While this is exciting, we must not lose sight of our primary objectives — cost effectiveness and reliability — as we work to add to the tool box of hardware components and software techniques for automatic alignment.

Acknowledgments

A great many people in universities, research organizations and factories have contributed in ways great and small to the technology we have today for automatic alignment. It would be impossible to know them all, let alone list them all. A number of people contributed to this white paper through discussions and data input; they are not, however, responsible for the contents of this paper. I would like to thank and acknowledge George Berkin, Bill Cornwall, Frank Evans, Richard Hubach, Bob New, Steve Opatnry and Shanon Stevenson.

Glossary

automatic alignment – any technique by which a system recognizes a feature on a part, and from that feature determines the parts location and pose.

back end – in the semiconductor industry, it refers to the activities of test, assembly, and packaging.

CCTV – an acronym for closed-circuit television.

CMP – an acronym for chemical-mechanical planarization in which raised areas on a silicon wafer are lowered to be level with the recessed features of the wafer.

convolution – a technique for signal or data processing where a value and its neighboring values are multiplied by coefficients and the products summed to make a new value for a corresponding location in a new signal or data stream.

correlated double sampling – a technique for reducing the noise associated with the charge detection process by subtracting a first output sample taken just after reset from a second sample taken with charge present.

correlation – a technique by which a system determines the quality of match between a reference signal or image and an incoming signal or image for all possible locations of the reference.

die – that portion of a semiconductor wafer containing a complete device.

flat – on a semiconductor wafer, a portion of the wafer's periphery which is ground flat or very slightly concave to indicate the crystalline orientation of the wafer.

front end – in the semiconductor industry, that part of the manufacturing process which creates the circuits on the wafer.

image pyramid – an image processing technique in which an original image is converted into a series of images of the same scene, but with decreasing image resolution (e.g., if the original image is 640 x 480 pixels, than the image pyramid might contain the original image plus replicas at resolutions of 320 x 240, 160 x 120, and 80 x 60 pixels).

LED – an acronym for light-emitting diode.

notch – on a semiconductor wafer, a small concave indentation on the wafer's periphery to indicate the wafer's crystalline orientation.

photolithography – in wafer fabrication, the steps that consist of coating the wafer with a photoresist, exposing the photoresist to a pattern, developing the photoresist, and etching the layer through the openings developed in the photoresist.

realignment – the step of bringing a wafer into rough alignment. It almost always precedes automatic alignment.

projection – the summation of pixel values in a series of regularly spaced parallel directions all of which are perpendicular to a particular direction.

ring light – a light source in which the light is directed onto the scene from 360 degrees around the vertical.

search region – in correlation, the portion of the image which is searched for one or more instances of features that match a template.

street – on a semiconductor wafer, the spaces separating the semiconductor die.

TAP – in the semiconductor industry, an acronym for test, assembly, and packaging.

template – in correlation or pattern matching, an image which contains the objects or features which are to be found in subsequent images.

vidicon – an image-sensing device consisting of a target in vacuum tube that is scanned by an electrostatically deflected electron beam.

wafer – in the semiconductor industry, a slice of semiconductor material onto which circuits are fabricated. The wafer is single crystal semiconductor material such as silicon or germanium.

wafer prober – a device for making electrical connection to one or more die on a semiconductor wafer for the purpose of facilitating electrical or functional testing.

wire bonder – a piece of equipment which attaches fine gold or aluminum wires from bond pads on the semiconductor die to the leads on the package.

Perry C. West, President
Automated Vision Systems, Inc.
127 Belcrest Dr.
Los Gatos, CA 95032 U.S.A.
(408) 356-2171
www.autovis.com

Imagenation Corp.
10220 SW Nimbus, Suite K5
Portland, OR 97223 U.S.A.
(503) 495-2200
(800) 366-9131
fax: (503) 495-2201
email: info@imagenation.com
www.imagenation.com