A Fault Identification and Classification Scheme for an Automobile Door Assembly Process

SHING-KUO WU, S. JACK HU, AND S.M. WU
Department of Mechanical Engineering and Applied Mechanics, The University of Michigan, Ann Arbor, MI 48109

Abstract. A process fault identification and classification scheme for an automobile door assembly process is presented based on multivariate in-line dimensional measurements and principal component factor analysis. First, the door assembly process and the dimensional measurement system are briefly introduced. Second, the technique of principal component factor analysis is presented for process fault identification. Process faults are summarized based on off-line identified case studies. Finally a machine classification scheme based on principal components and principal factors is presented and evaluated, using the pattern knowledge obtained off-line. This scheme is shown to be effective in classifying process faults using production data.

Key Words: automobile, assembly, fault classification, principal components.

1. Introduction

An inadequate door fit on the car body will not only affect the aesthetic value of the vehicle, but wil also cause functional problems such as wind noise, water leakage, and excessive door closing effort. Variation in the door fit may arise from four different sources: (1) Dimensional variation of the doors; (2) variation in the body openings; (3) variation in the door hanging and fitting process; and (4) effects of painting and general assembly, such as deformation due to weather strips, etc. Variation in the body opening and door fitting process have been addressed in previous publications by Hu and Wu (1992), and Wu et al. (1994). This paper will only address variation source identification and classification for a door assembly process.

Variation in a complete door results from variation in the panel stamping process and the assembly process. The panel shape variation in the stamping process is in general not correctable in the downstream process (Wu 1991), and needs to be corrected upstream in stamping. However, variation in the door assembly process also results from faults occurring in each assembly station, and can be detected and corrected case by case.

Statistical process control (SPC) using control charts has been used as a tool to monitor and control some assembly processes. To obtain data for process monitoring, traditional checking fixtures were used. With checking fixtures, although measurements are taken on selected points on a door, these checking fixtures can usually provide visual feedback of the spatial relationship between different points checked, which could help operators identify the root cause of variation. But the speeds of these fixtures are too slow to provide a statistically significant sample size. Since the late 1970s, coordinate measuring machines

(CMM) have been introduced into stamping and assembly plants. Because of their accuracy and flexibility, CMMs are gaining wide acceptance. CMMs measure the dimensions of selected points on a door based on sampling. Usually, only two to three doors are measured in an eight-hour shift due to the difficulty of transporting the door from the production line to the CMM room, setting it up for measurements, and transferring back to the production line. Therefore, both fault detection and diagnosis were not very effective because of the limited amount of data. Also, when plotting the data, relationships between measurement points were not explored.

With the availability of in-line measurement, such as the in-line optical coordinate measurement machines (OCMM), every door assembled can be measured, resulting in 100% measurement data (Wang 1991). Also, the in-line OCMM can measure a large number of points on a door, resulting in a high dimensional vector series over time. The abundant process information contained in the multivariate data makes effective process fault identification feasible.

The multivariate approach to quality control was first introduced by Hotelling (1947), who proposed the use of the T^2 control chart as a technique for monitoring two related variables. Later, Jackson (1980) extended Hotelling's procedure for use with principal component analysis. However, these procedures lacked the physical interpretation of the eigenvectors of the principal components.

This article will present a fault identification and classification scheme for a door assembly process using multivariate data from the in-line OCMM. The paper is organized as follows. The door assembly process and the in-line dimensional measurement system is introduced in section 2. Case studies identified off-line from the door assembly processes are presented in section 3. Based on these case studies, a two-level fault classification scheme is developed and evaluated in section 4.

2. The door assembly process and dimensional measurement

To provide the background for process faults identification and classification, the door structure, the assembly process, and the dimensional measurement system are briefly introduced in this section.

2.1. Elements of an automobile door

A typical automobile door is made by assembling the inner panel, the outer panel, the inner and outer belt reinforcements, the crash bar, and the hinge pillar. The inner panel constitutes the major stiffness of the door, while the outer panel constitutes the fit and finish. The crash bar provides the stiffness for side impact. These structural elements are shown in figure 1. Also shown in figure 1 is the coordinate system used to measure doors, described using a car body coordinates Fore/Aft (F/A), High/Low (H/L), and Inboard/Outboard (I/O).
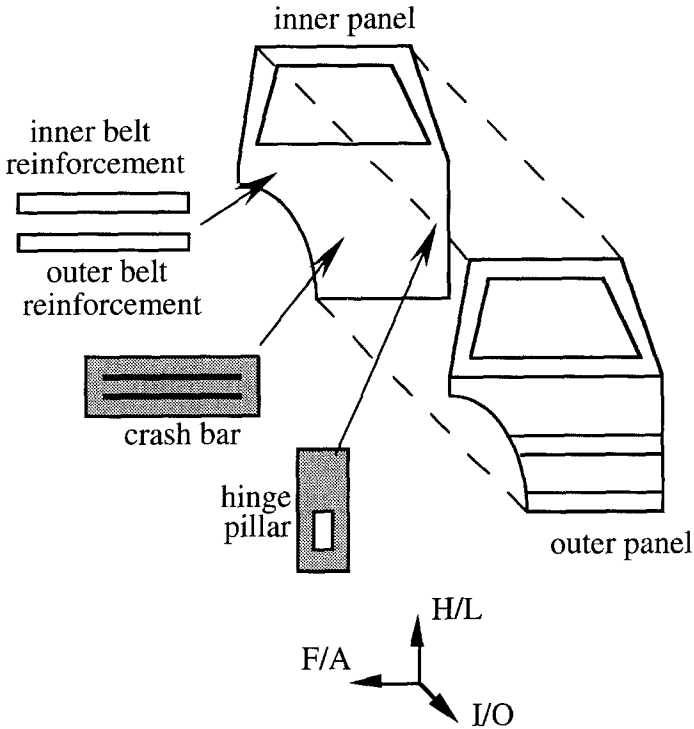
Figure 1. Elements of an automobile door with coordinate convention.

2.2. Door assembly process

The essential parts of a door assembly process include part positioning, clamping, welding, and hemming. The schematic flowchart of a typical door assembly process is shown in figure 2. The operations at each station are tabulated below:

Station 1: Weld inner belt reinforcement and hinge pillar to door inner panel
Station 2: Enhancement welding
Station 3: Weld outer belt reinforcement and crash bar to door inner
Station 4: Provide deadener to outer panel
Station 5: Provide adhesive to outer panel
Station 6: Match inner panel to inside outer panel
Station 7: Position inner relative to outer and weld
Station 8: Hem the outer panel flange to cover the inner panel
Station 9: Heat and solidify the adhesive between inner and outer
Station 10: In-line dimensional measurement by OCMM
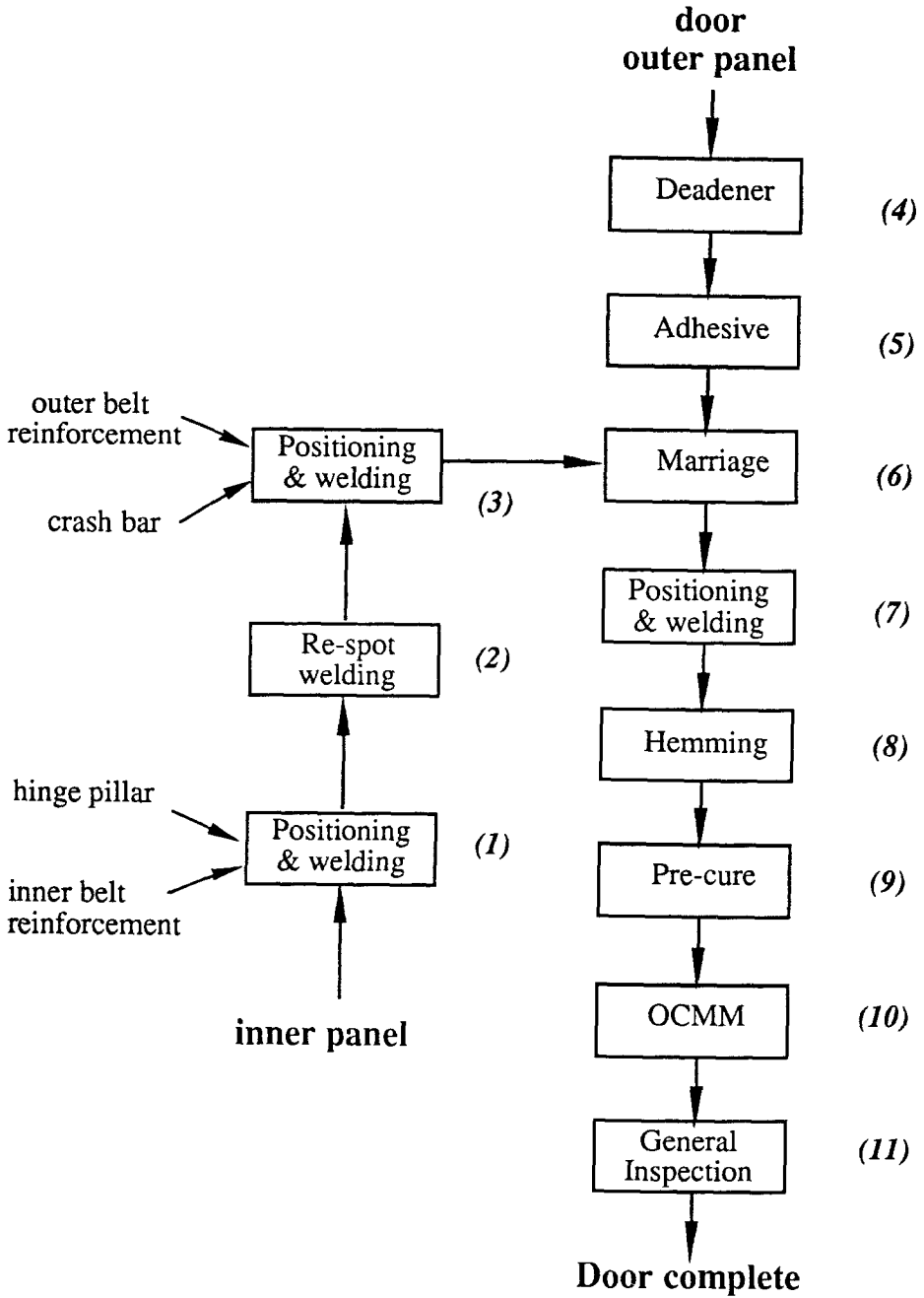Station 11: Surface inspection, general visual inspection

**door
outer panel**

Deadener	*(4)*

Adhesive	*(5)*

outer belt
reinforcement

Positioning & welding	

crash bar

(3)

Marriage	*(6)*

Re-spot welding	*(2)*

Positioning & welding	*(7)*

Hemming	*(8)*

hinge pillar

Positioning & welding	*(1)*

inner belt
reinforcement

Pre-cure	*(9)*

inner panel

OCMM	*(10)*

General Inspection	*(11)*

Door complete

Figure 2. Door assembly process flowchart.

2.3. In-line dimensional measurement

Every door assembled was measured by the in-line OCMM. A typical OCMM for door measurement consists of a frame supporting about 30 vision sensors. Sensors are aimed at selected checking points. A sensor projects a laser line to the part feature to be measured with a camera looking at the projected laser line. The laser line forms an image in the photodetector arrays inside the sensor. The positions of the image in the image space will reflect the feature position in the part coordinate. For a station, say, with 30 sensors checking about 50 dimensions on a door, it generally takes less than 10 s to complete the measurement. Figure 3 shows the typical measurement locations on a door.

Positioning of doors on the OCMM station does not have to be exact. In the case of nonexact positioning, the door is left in the fixture free from clamping. The six degrees of freedom needed for part locating can be replaced by six measurements. Then the dimensional variation of the doors is reported relative to these six measurements.
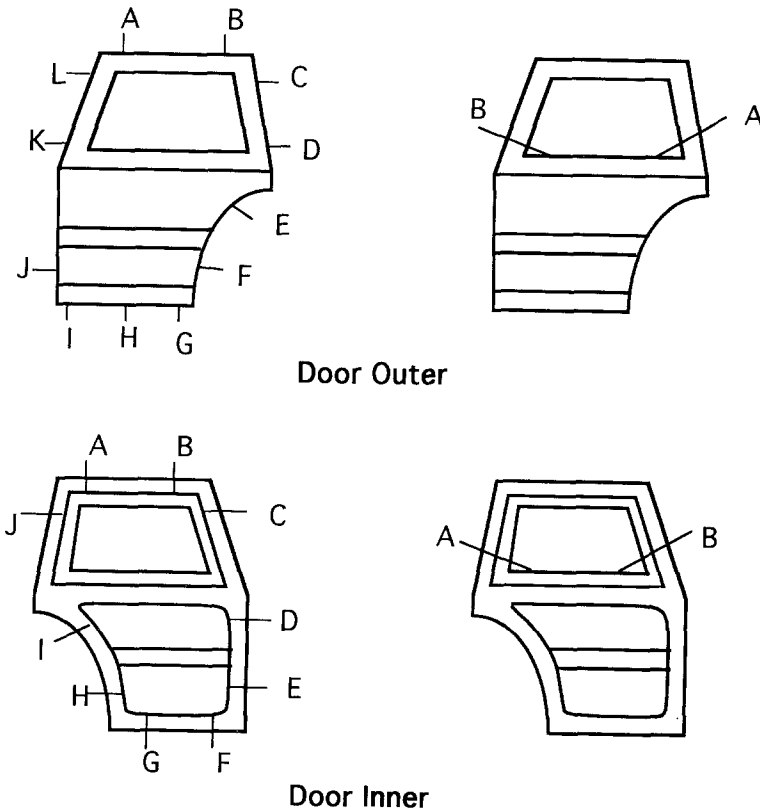
Figure 3. Typical optical sensor aiming locations on a door.

3. Assembly process fault identification

OCMM measurements are made at spatially sampled points. To help identify the sources of dimensional variation, it is important to relate the variation of various points on the door. One technique to accomplish this is multivariable correlation analysis and principal component analysis.

3.1. Principal component and factor analysis

The purpose of principal component analysis is to find the interrelationships between multiple variables by finding combinations of them to produce uncorrelated variables.

3.1.1 Principal component analysis. Principal component analysis is a simple analysis method for multivariate data interpretation. It finds combinations of p variables, x_1, x_2, ..., x_p, to produce indexes, z_1, z_2, ..., z_p, that are uncorrelated. The uncorrelatedness of the p indexes is then utilized to separate the different "modes" of variation in the data.

When conducting a principal component analysis, it is assumed that the variation in the data set can be adequately described by a few ($< p$) new variables, say, z_i. The variances of the rest of the indexes are expeced to be negligible. Since the variation of the p original x-variables can be accounted for by a few z-variables, the data reduction results in a certain degree of economy. For highly correlated variables, principal component analysis will result in a relatively few principal components. The principal components will also facilitate easier interpretation of the major sources of variation in the data. For p original variables, the ith principal component is a linear combination of variables, x_1, x_2, ..., x_p:

$$z_i = a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{ip} x_p$$

subject to the condition that

$$\sum_{j=1}^{p} a_{ij}^2 = 1 \tag{1}$$

The variance of z_i, denoted as $\text{Var}[z_i]$, should be as large as possible given the constraint of equation (1). The a_{ij} can be interpreted as the contribution of each variable x_j on the ith principal component z_i. The indexes z_is are also ordered in such a way that

$$\text{Var} [z_1] \geq \text{Var} [z_2] \geq \ldots \geq \text{Var} [z_p]$$

Principal component analysis can be approached by solving the eigenvalue problem of the sample correlation matrix or covariance matrix (Manly 1986; Therrien 1989), with the variances of the principal components being the eigenvalues of the correlation matrix $\mathbf{R}$, and weight a_{ij}s being the elements the normalized eigenvector corresponding to the ith eigenvalue. The steps in a principal component analysis can be summarized as:

1. Estimating the correlation matrix $\mathbf{R}$
2. Solving the eigenvalue problem of $\mathbf{R}$
3. Interpreting the patterns of variation using eigenvectors with large eigenvalues.

Therefore, the relative magnitude of the eigenvalues, i.e., the variance of principal components, will prioritize actions for variation reduction. The eigenvector for each principal component, showing the relative contribution of each original variable to the component, can be used to reveal any relative movement or deformation pattern of each "mode" of variation.

3.1.1.1. Improvement evaluation. When the process faults are prioritized by the principal components, the improvement by correcting each eigenmode can be evaluated both in the space of the principal component and in the original data space. To clarify the procedure, the eigenvalue problem is briefly reviewed and then the evaluation of improvement is discussed.

For a p-dimensional correlation matrix $\mathbf{R}$, the diagonalization of $\mathbf{R}$ can be done by:

$$\hat{\mathbf{R}} = \mathbf{Q}^{-1}\,\mathbf{R}\,\mathbf{Q}$$

where

$\mathbf{R}$ = correlation matrix (or covariance matrix); $\mathbf{Q} = [\mathbf{v}_1\ \mathbf{v}_2 \cdots \mathbf{v}_n]$;

$$\hat{\mathbf{R}} = \begin{bmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & & 0 \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \lambda_n \end{bmatrix}$$

and

$\mathbf{v}_i\mathbf{s}$ = normalized eigenvectors.

The eigenvalue represents the variation association with each principal component. When the ith principal component is removed, the reduction of variation in the principal component space thus can be calculated as follows:

$$\% \text{ variation reduction} = \frac{\lambda_i}{\displaystyle\sum_{i=1}^{n} \lambda_i}. \tag{2}$$

It should be noted that when the covariance matrix is used in the analysis, this percentage of variation reduction is compatible in scale with the variation reduction in the original data space. However, when the correlation matrix is used for principal component analysis, the percentage of variation reduction calculated using equation (2) is not compatible with the variation reduction in the original data space, and the evaluation needs to be done in the original data space.

To evaluate the variation reduction associated with each physical measurement point, variation in each principal component needs to be transformed back to the original variable domain. For example, when the first principal component is eliminated, the first eigenvalue in the diagonalized matrix becomes zero and results in matrix $\hat{\mathbf{R}}_1'$ as in equation (3). The matrix is then transformed back to the original variable domain as $\mathbf{R}_1'$ by equation (4). The diagonal terms in $\mathbf{R}_1'$ can then be evaluated separately as the improvement for each measurement point.

$$\hat{\mathbf{R}}_1 = \begin{bmatrix} 0 & 0 & & 0 \\ 0 & \lambda_2 & & 0 \\ & & \cdot & \\ & & & \cdot \\ 0 & & & \lambda_n \end{bmatrix} \tag{3}$$

$$\mathbf{R}_1' = \mathbf{Q}\,\hat{\mathbf{R}}_1'\,\mathbf{Q}^{-1}. \tag{4}$$

Improvement through elimination of other principal components can be evaluated in a similar way by setting the corresponding eigenvalues to zero and transforming back to the original variable space.

When the covariance matrix is used for principal component analysis, the inverted matrix after removing the ith principal component directly provides the information of variance reduction in the original data space. When the correlation matrix is used in principal component analysis, the diagonal term in the correlation matrix needs to be scaled by the variance of each variable. That is,

$$r_{ii}' = r_{ii} \cdot \text{Var}[X_i].$$

3.1.2. Factor analysis. The orthogonal eigenvectors were shown to be optimal in representing relationships between a set of multiple variables in the sense of minimal mean-square-error (Therrien 1989). Principal component analysis solves the eigenvalue problem for the multivariate correlation matrix and approximates the original data set by smaller dimensions of principal components. The omitted eigenmodes thus become errors associated with representing the original data set with smaller dimensionality of principal components. Although the principal components (eigenvectors with eigenvalues larger than one) are orthogonal to each other, the errors (also called specific factors) are not orthogonal. A rotation, orthogonal or oblique, is frequently applied to the eigenvectors to seek for clearer feature of the factors. The task to search for clearer factors is called factor analysis.

Factor analysis is frequently approached by performing the principal component analysis first (Manly 1986). The orthogonal principal components are then further converted to factors and then subject to rotation if necessary. The p uncorrelated principal components can be represented as:

$$Z_j = b_{j1}X_1 + b_{j2}X_2 + \ldots + b_{jp}X_p, j = 1, \ldots, p.$$

The inverse of the above gives the representation of the original variables using the uncorrelated principal components:

$$X_j = b_{1j}Z_1 + b_{2j}Z_2 + \ldots + b_{pj}Z_p, j = 1, \ldots, p.$$

By retaining only m $(m < p)$ principal components and normalizing each principal component with the square root of their associated eigenvalues, the unrotated factor model results as equation (5). Proper rotation then can be applied to the orthogonal unrotated factor model to obtain the better interpretable factors.

$$X_j = a_{1j}F_1 + a_{2j}F_2 + \ldots + a_{mj}F_m + e_j, j = 1, \ldots, p, \tag{5}$$

where

$$F_j = Z_j/\sqrt{\lambda_j} \text{ and } a_{ij} = \sqrt{\lambda_j}\, b_{ji}.$$

Rotation of the unrotated factor model can be either orthogonal or oblique. Orthogonal rotation retains the uncorrelatedness between factors after rotation. The oblique rotation relaxes the constraints of uncorrelatedness and allows for correlation between rotated factors. The amount of rotation is determined by maximizing or minimizing certain criteria (e.g., variance of square of factor loadings). Different criteria have been developed and compared. The main purpose of the rotation is such that a simple structure can result with the factor loadings being either closer to zero or further away from zero. (Thurstone 1947). Therefore, the resulting simple structure is easier for factor interpretation. Since uncorrelatedness is one characteristic of interest in this research, the orthogonal Varimax criterion is implemented in searching for factors.

3.1.2.1. Improvement evaluation. The original correlation matrix **R** can be approximately reconstructed by the factors as in equation (6). When the first factor is removed from the correlation matrix, the reduction of the variance in the original variable space can be evaluated by setting the associated eigenvalues $\lambda_i' = 0$ in the F^*F^{*T}, and then rotated back to the original variable space as in equation (7). Since λ_i' is nothing but the normalized eigenvalues rescaled by themselves in the factor analysis procedure, they will simply be unity. Thus the reduction in the variance can be evaluated by the diagonal terms in the correlation matrix with proper scaling by their standard deviations, as was discussed in principal component analysis.

$$\hat{R} = [\hat{X}\,\hat{X}^T]_{pxp} = G_{pxm}\,[F^*F^{*T}]_{mxm}\,G_{mxp}^T \tag{6}$$

$$[F^*F^{*T}]_{mxm} = \text{diag}\,[(\lambda_1')^2, (\lambda_2')^2, \ldots, (\lambda_m')^2]_{mxm}$$
$$\approx (G^TG)^{-1}G^T\,R\,G\,(G^TG)^{-1} \tag{7}$$

where $\hat{R}$ = approximated correlation matrix; F^* = rotated factor matrix; and G = factor loading matrix.

3.1.3. Principal components versus rotated factors. It has been stated (Gorsuch 1983) that Varimax orthogonal criterion is inappropriate when there is a general factor existing among the variables. In fact when there is one dominant general factor existing among the variables under study, the eigenvector corresponding to the largest eigenvalue is the best representation of the underlying factor in minimum mean square error sense. However, when there are more than one common factors existing among variables, the eigenvectors, although having largest eigenvalues in order, are in general inappropriate for factor interpretation. An appropriate rotation is usually needed to extract interpretable factors. Therefore, both the eigenvectors and the rotated factors are needed for data interpretation, with the former for one general factor situation and the latter for multiple factors.

As has been stated earlier, the eigenvalues in principal component analysis are so ordered such that the first eigenvalue is larger than the second, and so forth. To discriminate between the single factor case and the multiple factor case, only the second eigenvalue needs to be inspected. The procedure to determine which feature to use for data interpretation is:

1. Perform the principal component analysis.
2. If $\lambda_2 < 1$ (for correlation matrix), then conduct analysis using the eigenvector from principal component analysis.
3. If $\lambda_2 > 1$ (for correlation matrix), then conduct analysis using the Varimax rotated factors.

3.2. Process fault types

Based on principal component and factor analysis, four subgroups of faults can be classified based on data. They are: (1) Whole door shifting (type I shifting); (2) inner panels shifting inside outer panels (type II shifting); (3) increased variations due to improper fixturing; and (4) local errors. The type I shifting usually resulted in large variations in measurements of the door. The type II shifting usually causes variations either in the inner panel or outer panel only. The improper fixturing happens by bending the door header. The inconsistent spring back will cause the header dimensions of both inner and outer panels to vary. When the fault is not any of the above, it is then categorized as local process fault. A detailed description of the four types of process faults is provided in Wu (1991).

A typical local variation is the window opening channel width variation (Wu 1991), which results from the inconsistency in respot welding of the inner belt bar to the door inner panel. Figure 4 shows the measurement location on both the inner panel and the outer panel. By ranking the 6-sigma variation from smallest to largest for all the points measured, points on the outer panel showed very small variation, while points on the inner panel showed large variation. Small variation on the outer was due to the fact that doors were measured relative to four datum points on the outer surface. The patterns
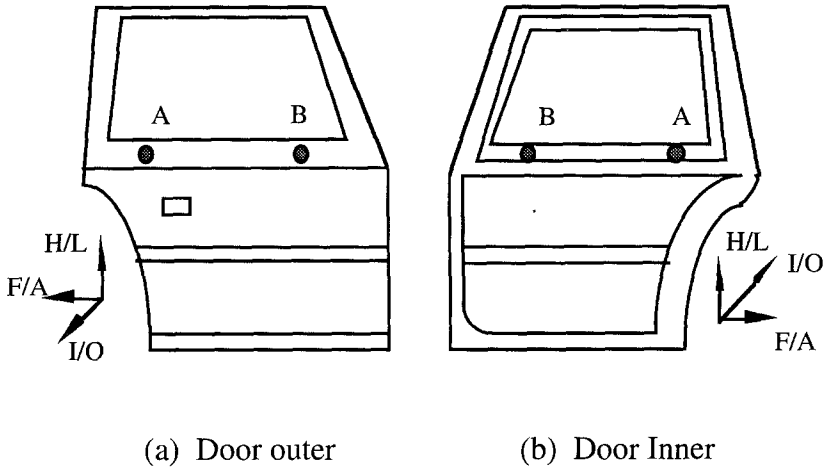
(a) Door outer   (b) Door Inner

Figure 4. Checking points at day light opening in inner and outer panels.

of variation on the inner panel showed to be bimodal as shown in Figure 5. Principal component analysis was performed using data obtained form measuring 80 doors. The relationship between the points, i.e., A(I/o), A(H/L), B(I/O), and B(H/L), are represented in the correlation matrix, eigenvalues, and eigenvectors as shown in tables 1 and 2. It was found from the first mode that the belt bar area was not only being pulled in and out, but also being pulled up and down as shown in figure 6. The bimodal pattern was due to the fact that two robots performed the belt bar welding alternately, and the two robots were not calibrated the same in their nominal position. Figure 7 shows the result after reprogramming the robot in the direction normal to the inner panel.

4. Fault classification scheme

The dimensional quality problem identified and corrected previously can reoccur. With the knowledge of door assembly process faults constructed based on off-line data analysis, a machine fault classifier is proposed using principal components and factors from the multivariate data to quickly classify the faults for efficient fault correction.

4.1. The two-level classification scheme

As was presented in the previous section, principal component factor analysis is effective in locating the process faults. The eigenvectors and rotated factors, which are the multidimensional door variation vectors in the pattern space, are used for machine fault classification in this section.
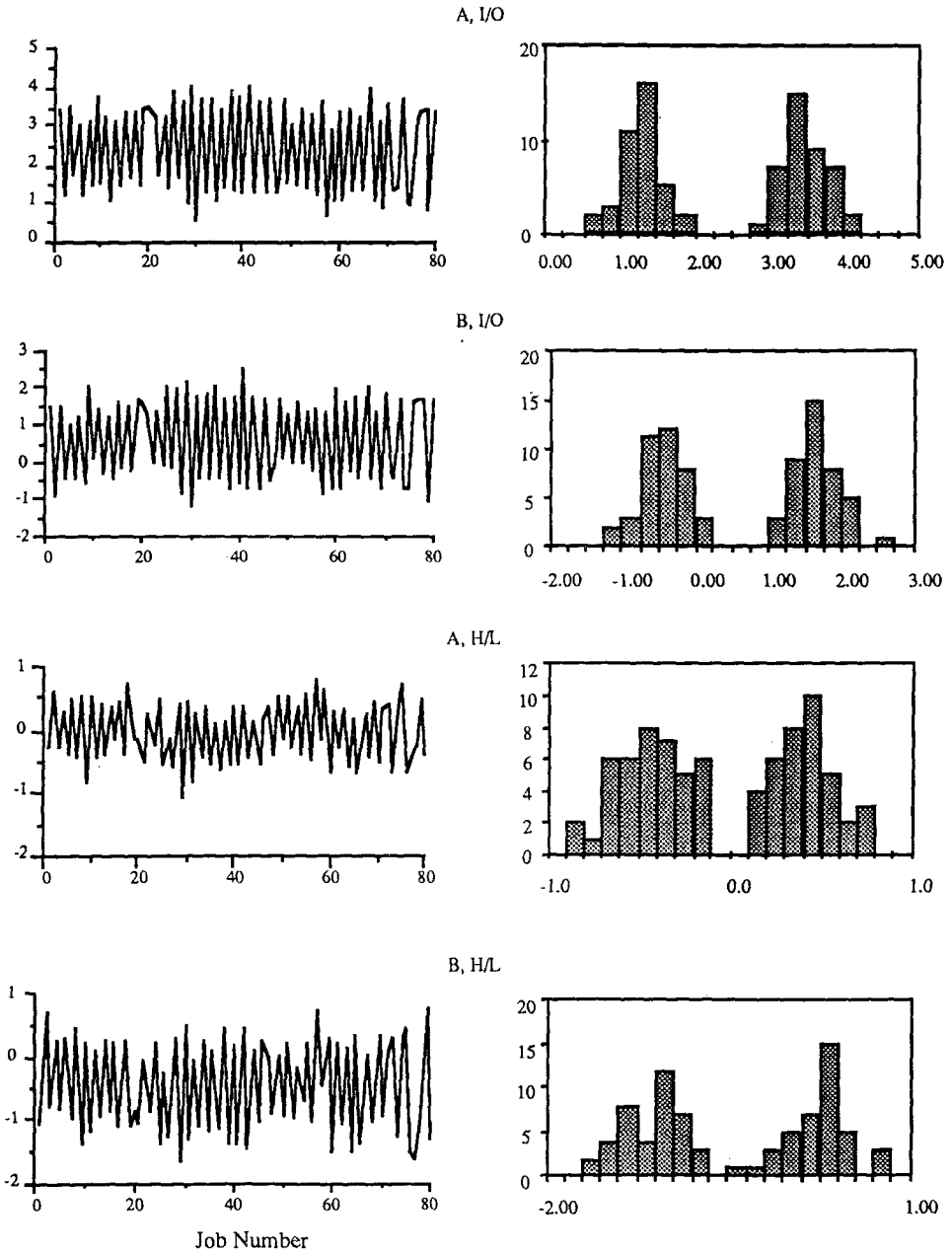
Figure 5. Measurement data from inner panel and their distributions.

Table 1. Correlation matrix for H/L and I/O of sensor A and B.

Point	A, H/L	A, I/O	B, H/L	B, I/O
A, H/L	1.000	−0.930	0.985	−0.956
A, I/O		1.000	−0.914	0.946
B, H/L			1.000	−0.968
B, I/O				1.000

Table 2. Eigenvalues, their energy proportion, and eigenvectors of the correlation matrix.

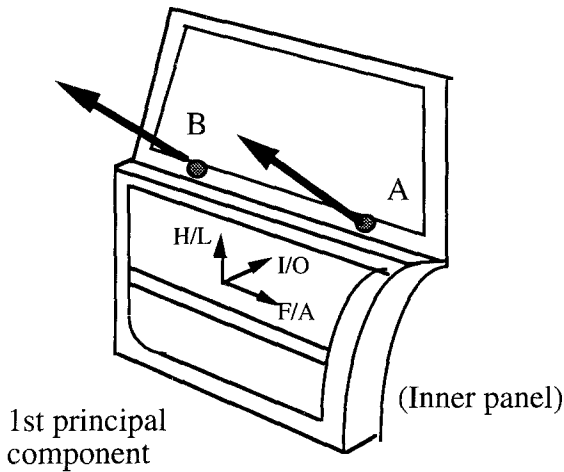| Component	Eigenvalue	Percent	1st Eigenvector	2nd Eigenvector	3rd Eigenvector	4th Eigenvector
1	3.850	96.250	−0.503	−0.345	0.534	−0.585
2	0.100	2.500	0.492	−0.804	−0.271	−0.195
3	0.041	1.025	−0.502	−0.482	0.002	0.718
4	0.009	0.225	0.503	−0.040	0.801	0.323

Figure 6. First principal component of inner belt bar twisting.

A two-level classification scheme is implemented for fault diagnosis for the door assembly process. The first level classification is the *problem classification*, and the second level classification is the *diagnostic classification*. The reason for implementing a two-level classification scheme is to reduce the data dimensionality by focusing only on the specific data points with variations exceeding a certain level, e.g., 3 mm. After the problem zone has been classified by the first level classification, the principal components and rotated factors are calculated and the second classification scheme is used for diagnostic classification. The schematic flowchart for the two-level classification scheme is shown in figure 8. The first level and second level classification subschemes will be discussed in the following sections.
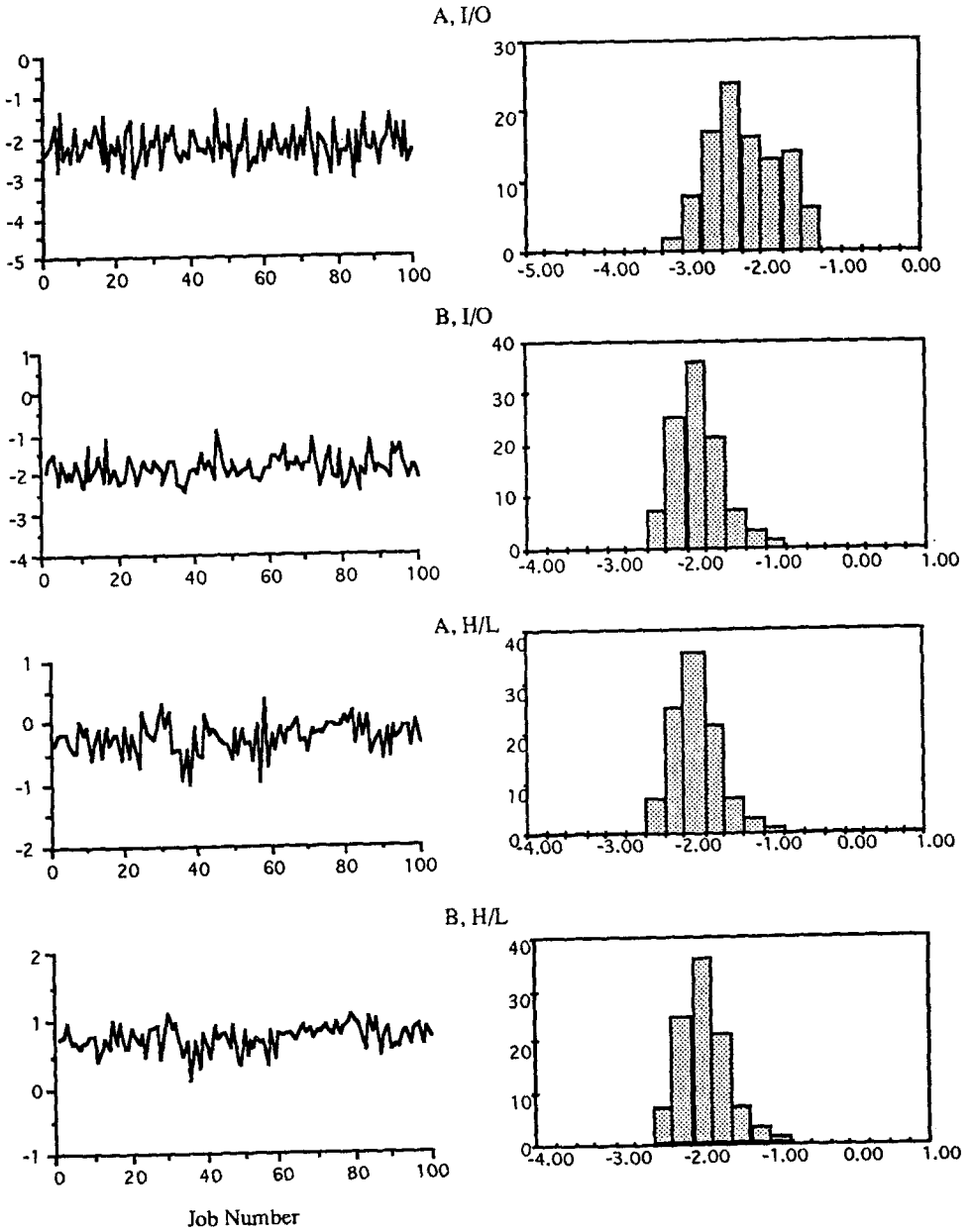
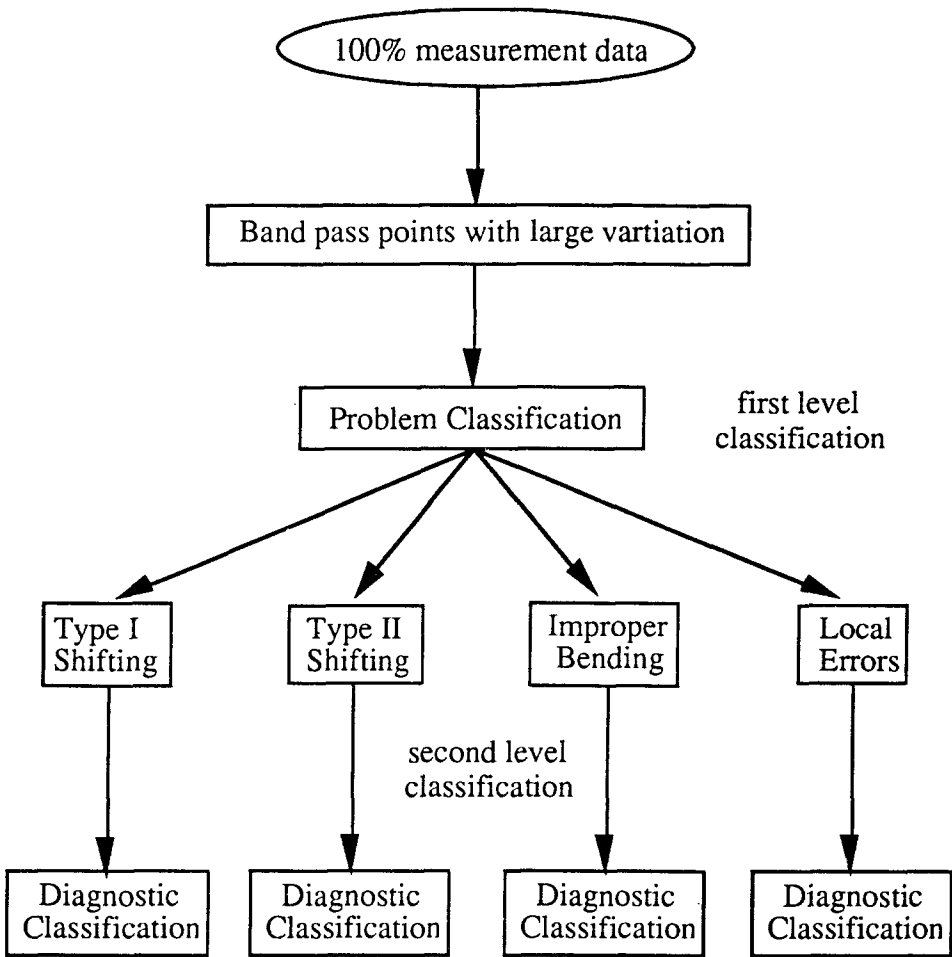*Figure 7.* Measurement data and histograms after correction.

Figure 8. Flowchart for the two-level classification.

4.2. First level classification

The first level classification is to identify if the points with large variation are the same as the points previously identified as having large variation. For a p-dimensional measurement vector, $X = [x(1),x(2), \ldots\ldots, x(p)]^T$, a corresponding p-dimensional class pattern can be formed. The p-dimensional class pattern is so formed for a specific type of problem that the $x(i)$s are set to one if previously found to have large variation, and the rest of the $x(j)$s are set to zero.

As an example, let $x(a),x(a+1), x(a + 2)$, and $x(a + 3)$ be the four DLO measurements. Since the DLO was found to have large variations, a class pattern for first level classification can be formed as figure 9, such that $x(a)$ to $x(a + 3)$ are ones and zeros elsewhere.
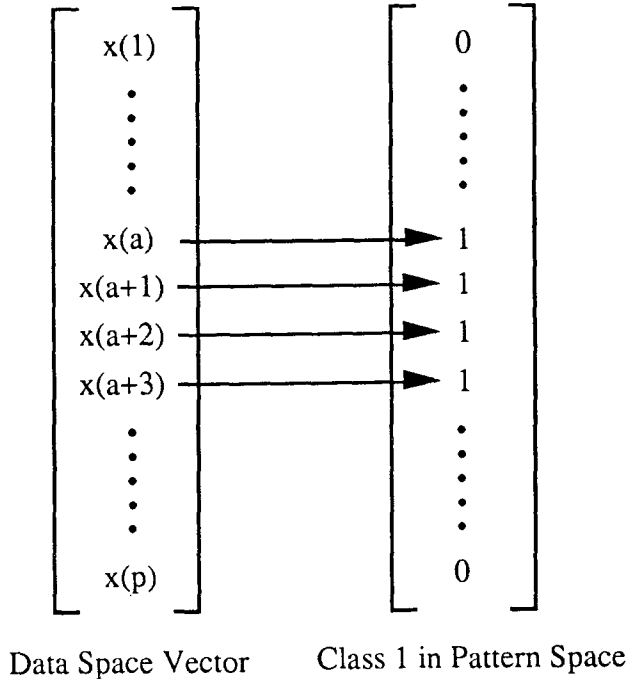
Figure 9. Mapping of data space vector to first level pattern space.

When a new set of data showing large variation in different locations (including DLO) is collected, a new pattern feature (p-dimensional vector in pattern space) can be formed by switching the corresponding elements in the null vector from zero to one for points with large variations as in figure 10.

A simple inner product classifier is implemented for quick problem zone classification. The inner product between the new pattern and the class pattern is calculated. The inner product is then compared with a threshold for decision making. Using the DLO problem as an example, when any of the DLO points have large variation, the DLO variational problem an be suspected to be happening again. The threshold thus can be set to 0 such that a decision can be made for the following hypothesis.

Inner Product [new pattern, class 1 pattern)] = 0, Null hypothesis
Inner Product [new pattern, class 1 pattern)] > 0, DLO problem

The first level classification for the other types of problems can be designed in a similar way.

$$
\begin{bmatrix}
x(1) \\
\vdots \\
x(a) \\
x(a+1) \\
x(a+2) \\
x(a+3) \\
\vdots \\
x(p)
\end{bmatrix}
\longrightarrow
\begin{bmatrix}
0 \text{ or } 1 \\
\vdots \\
0 \text{ or } 1 \\
0 \text{ or } 1 \\
0 \text{ or } 1 \\
0 \text{ or } 1 \\
\vdots \\
0 \text{ or } 1
\end{bmatrix}
$$

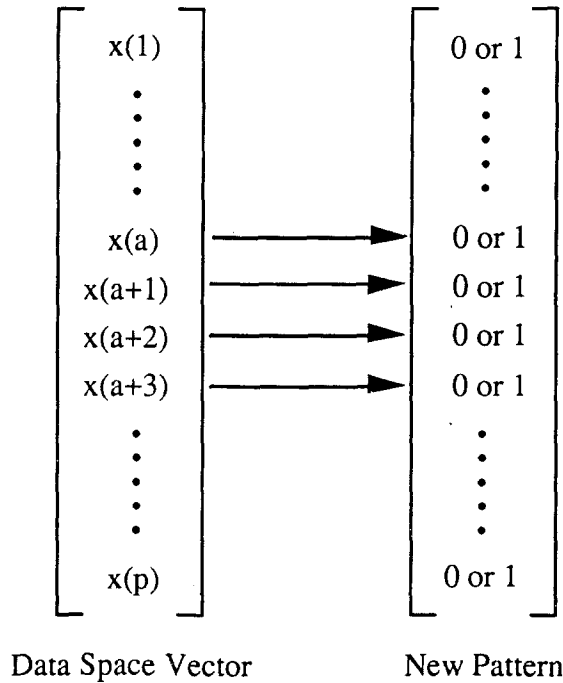Data Space Vector   New Pattern

Figure 10. Mapping of data space vector to first level pattern space for new observation.

4.3. The second level classification

The second level classification is for the purpose of fault diagnosis once the fault zone has been identified by the first level classifier. In the second level classification, the class patterns are the average normalized factors obtained from previously solved problems. For a problem with m principal components, there will be m factors and thus m classes in the pattern space. Take the DLO case study discussed previously as an exmaple.. A 4×1 normalized pattern vector, which is the eigenvector for the case, can be formed for the second level classification.

After the first level classification is completed, a principal component analysis will be conducted for measurements of these points with large variation. The second eigenvalue will then be checked to decide whether the eigenvector or the rotated factors should be used for diagnostic classification. Then the quadratic Mahalanobis distance, as in equation (8), is calculated and the nearest neighbor approach is implemented to select the shortest distance for fault classification.

$$
D_i^2 = (X - \bar{X}_i)^T C_i^{-1} (X - \bar{X}_i) \tag{8}
$$

where D_i^2 = quadratic Mahalanobis distance; X = newly observed pattern; $\bar{X}_i$ = pattern of class i; C_i = variance-covariance matrix of class i.

Since the nearest neighbor classification will always result in a solution, even if the newly observed factor does not belong to any of the existing classes, a confidence checking is necessary to confirm the classified result. To do this, the maximum distance in the training data is stored for result checking. When the newly calculated distance is less than the maximal distance, the classified result is confirmed without vagueness. When the newly calculated distance is larger than the maximal distance in the training set, a heuristic decision has to be made to either confirm the classification and update the class pattern, or add a new class pattern to the classifer.

To summarize, a schematic structure of the second level classification is shown in figure 11.

5. Evaluation of the proposed classification algorithm

To evaluate the classification scheme, the DLO data was collected over a wide span of production time. Three patterns were found to happen more frequently than others. The patterns represent different types of distortion in the DLO, and are stored as class patterns. These three different patterns with their associated variance-covariance matrices are listed in tables 3 (a) and (b).

The Mahalanobis distances of the training data for the three classes are first calculated to show that these class patterns are separable. Figures 12 (a), (b), and (c) tabulate and plot the square roots of the quadratic Mahalanobis distances. As can be seen from the orders of the distances, the three class patterns are well separable in the pattern space and the Mahalanobis distance is shown to be an adequate classification index. The distributions of the distances for factors 1 and 2 around their class patterns are also shown in figures 13 (a) and (b), which can be seen to approximate exponential distribution.

The patterns of the newly collected test data with large variations in DLO are tabulated in table 4. The distances of newly observed patterns to each class are also listed in table 4. For the data used in the evaluation, no misclassification is found for pattern 1 and pattern 2.

The last set of data was classified to be pattern 2 by the nearest neighbor approach. However, it does not pass the confidence threshold (maximal distance in training data), and is concluded to be new. As can be seen in the patterns, the DLO reinforcements are not only pulled in I/O direction, but also twisted especially in the B-sensor side.

Based on the test results, it can be seen that the proposed classification scheme functioned effectively in detecting and diagnosing an already identified variation problem. In our case, it is the door DLO variation. In addition to the implementation of the proposed techniques on production floors, techniques will be developed in the future to facilitate the automatic grouping of measurements into variation case studies.

Due to the necessity of having large samples for conducting the proposed analysis, 100% process measurement is the key for the proposed approach. Recent development and successful implementation of the real-time-in-line optical coordinate measuring machines made this proposed approach feasible.

Principal Component Analysis

Eigenvectors

If $\lambda_2 > 1$

If $\lambda_2 < 1$

Nearest Neighbor
Classification

Rotated Factor

Check
Confidence Threshold

Classified ?

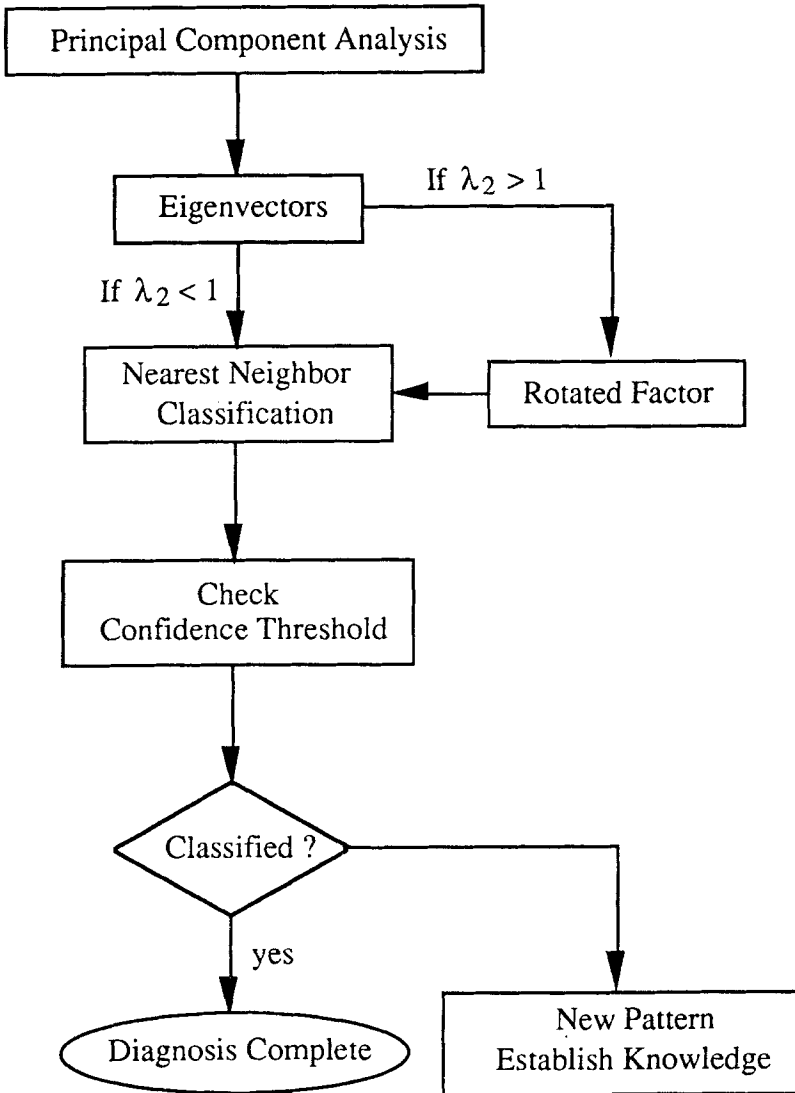yes

Diagnosis Complete

New Pattern
Establish Knowledge

Figure 11. Second level classification procedure.

Table 3(a). Class patterns for DLO.

Sensor, Direction	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$
A, I/O	−0.032	0.698	−0.5021
A, H/L	0.692	0.025	0.4948
B, I/O	0.023	0.685	−0.5020
B, H/L	0.698	−0.028	0.5012

Table 3(b). Variance-covariance matrix for DLO class patterns.

$$C_1 \quad \begin{bmatrix} .015 & 0 & -.001 & .001 \\ 0 & .002 & -.002 & -.001 \\ -.001 & -.002 & .017 & 0 \\ .001 & -.001 & 0 & .001 \end{bmatrix}$$

$$C_2 \quad \begin{bmatrix} .001 & .001 & -.001 & -.001 \\ .001 & .034 & -.002 & -.006 \\ -.001 & -.002 & .001 & 0 \\ -.001 & -.006 & 0 & .01 \end{bmatrix}$$

$$C_3 \quad \frac{1}{10000} \begin{bmatrix} 0.0755 & 0.1168 & -0.0029 & -0.0454 \\ 0.1168 & 0.3136 & 0.0071 & -0.1893 \\ -0.0029 & 0.0071 & 0.0229 & 0.0129 \\ -0.0454 & -0.1893 & 0.0129 & 0.1564 \end{bmatrix}$$

6. Conclusions

The development in sensing and computers made it possible to measure a large number of points on every part produced. Effective utilization of measurement data promoted the development of automated process monitoring and diagnostic systems. A simple and robust two-level classification scheme based on principal components and rotated factors is presented for the monitoring and diagnosis of an automobile door assembly process.

- Principal component analysis was used to investigate the spatial relationship between measurement points. The eigenvectors obtained represent the patterns of variation among the points. Visualization and animation of these patterns will help manufacturing engineers identify the possible locations of root causes of variation.
- Based on off-line identified case studies, a two-level fault classification scheme is proposed using eigenvectors or rotated factors. This scheme is shown to be effective in classifying process faults using production data.
- This classification scheme can be easily expanded upon discovery of new process fault patterns, and is also suitable for other production processes where in-line measurements are available.

Distance of class 1 data to		
class 1	class 2	class 3
1.273	49.836	.1865.8
1.518	55.967	1576.3
2.425	64.710	842.0
1.590	59.398	1107.5
1.282	44.786	2205.9
1.190	48.016	1958.9
3.654	41.299	3016.5
1.209	49.423	1804.1
1.049	47.933	1932.4
1.091	52.922	1597.1
1.871	41.268	2498.5
2.635	48.964	1897.2
2.017	52.767	1858.0
2.267	63.342	1013.1
1.961	56.931	1292.2
1.499	53.607	1543.1
1.583	54.895	1566.3
2.694	50.027	1787.6
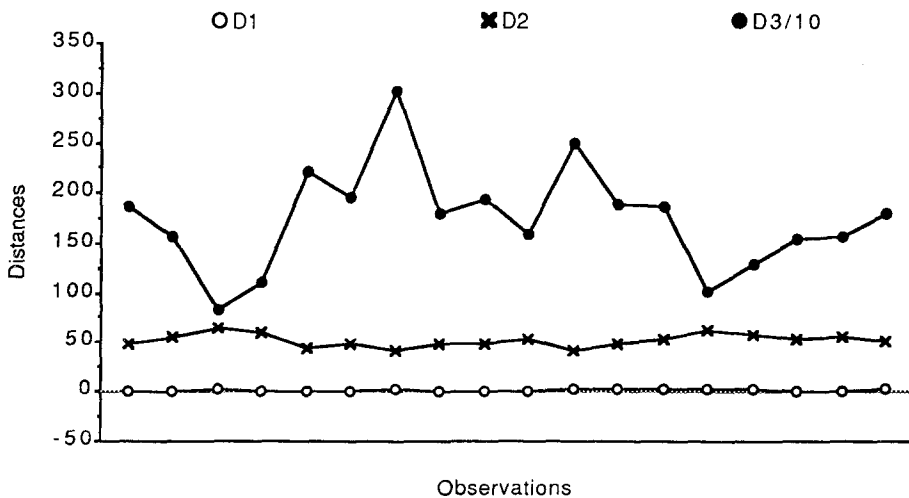
Figure 12(a). Distances of class 1 data to every class.

Distance of class 2 data to		
class 1	class 2	class 3
63.723	1.100	9602
68.662	1.584	9905
82.303	2.457	10655
74.647	1.237	10268
57.264	1.384	9081
61.823	1.145	9370
61.253	3.407	9208
64.406	1.286	9676
61.778	1.740	9528
67.530	1.116	9862
52.620	2.144	8744
58.317	2.640	8956
62.203	3.248	9366
78.997	2.199	10480
74.589	1.726	10218
67.304	1.021	9810
68.935	0.992	9909
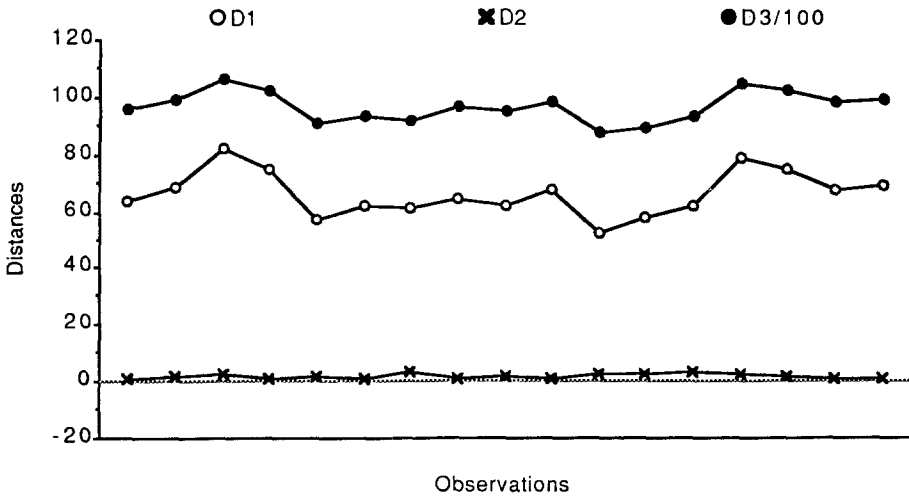
Figure 12(b). Distances of class 2 data to every class.

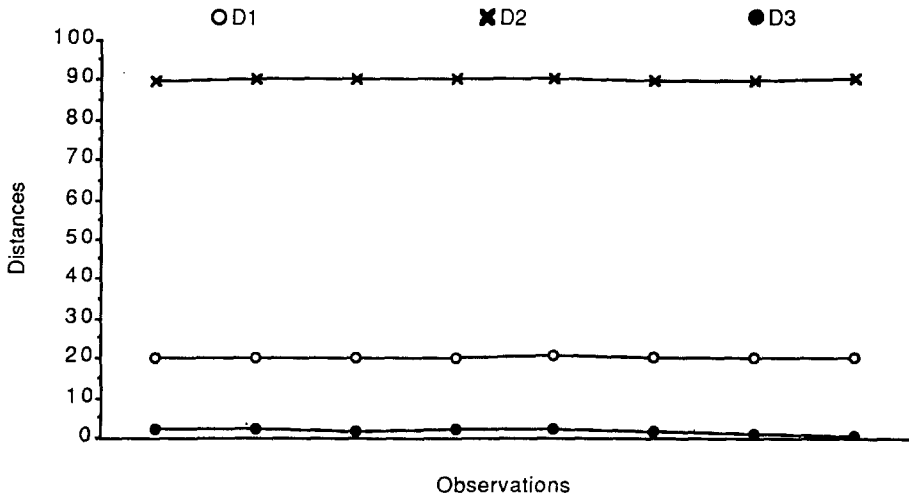| Distance of class 3 data to		
class 1	class 2	class 3
20.1423	89.8269	2.0921
20.3859	89.9945	2.0561
20.3148	89.9717	1.5530
20.3388	90.0197	2.4749
20.5908	90.2306	2.3798
20.2340	89.8635	1.7288
20.3189	89.9359	1.2612
20.3934	90.0141	0.7851

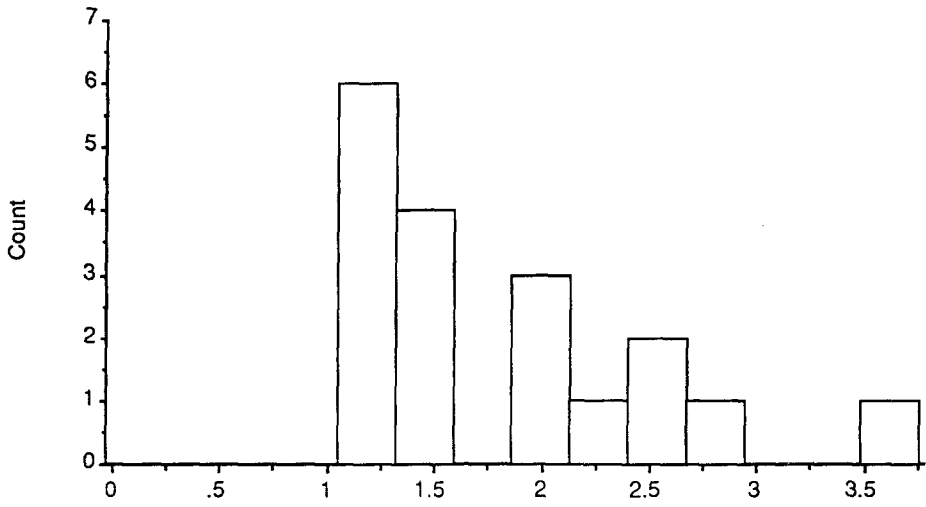*Figure 12(c)*. Distances of class 3 data to every class.

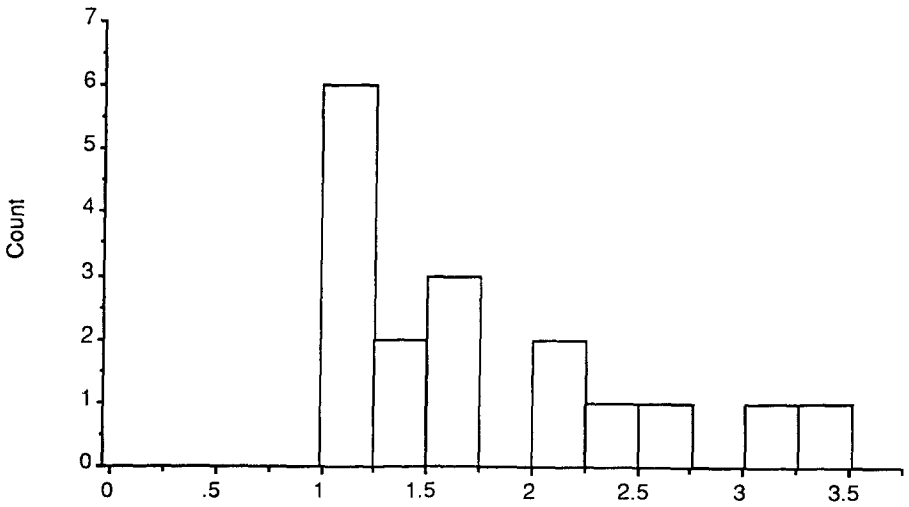*Figure 13(a).* Histogram of distances in pattern 1.

Figure 13(b). Histogram of distances in pattern 2.

Table 4. Classification test results.

Test Pattern				Distances			
A I/O	A H/L	B I/O	B H/L	D_1 max 3.65	D_2 max 3.41	D_3 max 2.48	Resulting Class
−0.124	0.693	−0.096	0.704	**1.288**	59.743	1,117.0	1
−0.100	0.713	0.072	0.691	**1.318**	53.219	1,654.6	1
0.039	0.709	−0.101	0.697	**1.098**	53.117	1,529.0	1
−0.001	0.704	0.001	0.710	**0.988**	51.305	1,692.5	1
0.052	0.709	0.153	0.687	**1.567**	44.102	2,334.7	1
−0.017	0.709	−0.093	0.698	**0.907**	55.161	1,393.7	1
0.013	0.708	−0.102	0.699	**1.008**	54.221	1,450.9	1
0.119	0.699	−0.029	0.705	**1.283**	47.348	1,967.1	1
0.699	−0.112	0.703	−0.072	74.256	**1.135**	10,259	2
0.696	0.063	0.710	−0.085	67.304	**1.021**	9,810	2
0.708	0.039	0.699	−0.095	68.935	**0.992**	9,909	2
0.705	0.049	0.705	−0.049	66.129	**1.023**	9,757	2
0.705	0.038	0.689	0.167	55.770	**2.316**	9,120	2
0.706	0.001	0.698	−0.117	71.684	**1.100**	10,072	2
0.706	0.021	0.698	−0.115	70.722	**1.053**	10,009	2
0.698	0.085	0.711	0.005	61.785	**1.338**	9,497	2
0.687	0.059	0.653	−0.315	79.442	**3.824**	10,293	new
0.658	0.144	0.661	−0.330	76.490	**4.551**	10,035	new
0.696	0.149	0.622	−0.327	76.434	**4.406**	10,005	new
0.654	0.255	0.648	−0.297	70.209	**4.726**	9,575	new
0.640	0.028	0.633	−0.434	86.769	**6.942**	10,534	new
0.670	−0.024	0.674	−0.311	82.613	**3.915**	10,531	new

References

Gorsuch, R.L., *Factor Analysis*, Lawrence Erlbaum Associates, Publishers (1983).

Hotelling, H., "Multivariate Quality Control," in *Techniques of Statistical Analysis*, C. Eisenhart, M. Hastay, and W.A. Wallis (Eds.), New York (1947).

Hu, S.J. and Wu, S.M., "Identifying Root Causes of Variation in Automobile Body Assembly Using Principal Component Analysis," *Transactions of North America Manufacturing Research Institution*, pp. 311–316 (May 1992).

Jackson, J.E., "Principal Component and Factor Analysis: Part I—Principal Components," *Journal of Quality Technology*, Vol. 12, No. 4, pp. 201–213 (October 1980).

Manly, B.F.J., *Multivariate Statistical Methods—A Primer*, Chapman and Hall, New York (1986).

Therrien, C.W., *Decision Estimation and Classification—An Introduction to Pattern Recognition and Related Topics*, John Wiley and Sons, New York, NY (1989).

Thurstone, L.L., *Multiple Factor Analysis*, University of Chicago Press, Chicago, IL (1947).

Wang, W.W., "Accuracy Analysis for In-Line OCMM." Ph.D. Thesis, Department of Mechanical Engineering and Applied Mechanics, The University of Michigan, Ann Arbor, MI (1991).

Wu, S.K., "A Methodology for Optimal Door Fit in Automobile Body Manufacturing." Ph.D. Thesis, Department of Mechanical Engineering and Applied Mechanics, The University of Michigan, Ann Arbor, MI (1991).

Wu, S.K., Hu, S.J., and Wu, S.M., "Optimal Door Fitting with Systematic Fixture Adjustment," *International Journal of Flexible Manufacturing Systems*, Vol. 6, No. 2, (1994).