J Intell Manuf (2012) 23:2025–2036
DOI 10.1007/s10845-011-0529-1

A framework for variation visualization and understanding in complex manufacturing systems

Lee J. Wells · Fadel M. Megahed · Jaime A. Camelio · William H. Woodall

Received: 3 November 2010 / Accepted: 17 March 2011 / Published online: 10 April 2011

Abstract This paper provides a framework that allows industrial practitioners to visualize the most significant variation patterns within their process using three-dimensional animation software. In essence, this framework complements Phase I statistical monitoring methods by enabling users to: (1) acquire detailed understanding of common-cause variability (especially in complex manufacturing systems); (2) quickly and easily visualize the effects of common-cause variability in a process with respect to the final product; and (3) utilize the new insights regarding the process variability to identify opportunities for process improvement. The framework is illustrated through a case study using actual dimensional data from a US automotive assembly plant.

Keywords Chance-cause variability · Computer-aided design · Principal components · Process variation · Statistical process control · Visual analytics

Introduction

Understanding process or product variation is an indispensable component in the statistical process control (SPC) methodology. Common and special causes of variation are often discussed in the SPC literature. Common causes (sometimes referred to as chance causes) of variation are considered to be inherent in the manufacturing process and therefore, cannot be changed without changing the process itself. On the other hand, special causes (often called assignable causes) are unusual disruptions in the process, and should be dealt with accordingly. Traditionally, practitioners have used control charting methods to distinguish between common and special causes to ensure an appropriate and timely response to changes in the process. Woodall (2000) noted that the distinction between these two variation types is usually context dependent as current common causes can be special causes in the future. In addition, the use of control charting techniques prevents over and under-reaction to process behavior. Over-reaction is also known as "process tampering" in SPC literature.

Control charts can be used for either retrospective or prospective process analysis. When a control chart is used retrospectively on a historical data set, the goal is to check whether the process has been in statistical control (Phase I methods). On the other hand, if the control chart is used prospectively with samples taken over time to detect changes in the in-control process; this control charting method is referred to as a Phase II method. The use of control charts in Phase I is usually iterative. As highlighted in Woodall (2000), "much work, process understanding, and process improvement is often required in the transition from Phase I to Phase II. (…) The scope of SPC needs to be broadened to include an understanding of the transmission of variation throughout the manufacturing process. This will require more sophisticated modeling and the incorporation of more engineering knowledge of the processes under study."

The method proposed in this paper aids in bridging the gap between Phase I and Phase II methods by integrating two common tools used in manufacturing, principal component analysis (PCA) and computer aided design (CAD). Currently, a significantly large amount of multivariate dimensional data is being collected by modern measurement technologies (coordinate measurement machines (CMMs), machine vision systems (Megahed and Camelio 2010), and/or laser scanners). The goal in combining PCA and CAD tools

L. J. Wells (✉) · F. M. Megahed · J. A. Camelio · W. H. Woodall
Virginia Tech, Blacksburg, VA 24061, USA
e-mail: leejay@vt.edu

together is to present multivariate variation patterns in a manner that can aid human intuition and reasoning to more easily understand the process. Overcoming this hurdle can greatly reduce the amount of time required for determining the cause of process variability, and therefore, it shortens system launch times, leads to faster implementation of process improvements, and results in higher quality products. In addition, it is often argued that the success of control charting techniques as a quality control tool relies heavily on the ability to show operators, in a simple way, how the product varies over time (e.g. Montgomery 2008, p. 157).

In this paper, it is assumed that the reader may not be familiar with the concepts of PCA. Therefore, a brief introduction to PCA is given in the following section. This is followed by a survey of the relevant literature on process understanding. Then, an overview of the visualization method is provided, which is followed by a discussion and analysis of applying the proposed framework in an industrial case study. Finally, our conclusions are provided.

Related SPC techniques and principal component analysis

A major goal of SPC is to detect the occurrence of unexpected process shifts as quickly as possible. This allows the investigation of the root-causes behind the occurrence of these shifts so that corrective action can be taken before non-conforming items are produced. Control charts are a real-time process monitoring technique that is used for that purpose. It should be noted that a control chart is a plot of values of a statistic, which summarizes measurements of important part/process quality characteristics, as a function of time. Control limits are then used to determine if the variability of the statistic over time can be attributed only to inherent randomness in the production process or if an assignable cause for that variation is present. For more details on this topic, the reader is referred to Wheeler and Chambers (1992), Woodall and Adams (1998), or Montgomery (2008).

Profile monitoring is a control charting method used when the quality of a process or product can be characterized by a functional relationship between a response variable and one (or more) explanatory variable(s). It should also be noted that profile monitoring is one of the fastest growing areas of research in SPC as it represents cases that seem to be increasingly common in practical applications. In this section, it is not our goal to review profile monitoring techniques, but to show how that concept can be extended to include other complementary methods, which will make the variation understanding process more intuitive for data-intensive manufacturing environments. For a more comprehensive overview on profile monitoring techniques, the reader is referred to the reviews by Woodall et al. (2004) and Woodall (2007).

Woodall et al. (2004) showed that it is extremely important to distinguish between Phase I and Phase II applications when monitoring profiles. In Phase I, a historical set of process data is analyzed to understand the process variation, evaluate the stability of the process, and model the in-control process performance. It is therefore important to remove any samples associated with special causes of variation, especially if the sources of variation can be determined and prevented, so that the in-control process parameters can be estimated correctly. As mentioned earlier, in our paper CAD techniques are being employed as an ad-hoc method to standard Phase I profile monitoring techniques to ensure that the process variation can be understood and visualized by the operators and engineers on the shop floor. In the SPC framework, Phase I methods are compared by assessing the probability that the process is declared out of control. It is believed that utilizing engineering and process knowledge as a part of the decision process will increase the probability that the correct decision is made regarding process stability.

Woodall et al. (2004) highlighted that PCA can be very useful in understanding the process variation and determining the stability of Phase I profiles. In addition, when dealing with dimensional data, such as those gathered in a manufacturing process, PCA can be used to describe unique geometric variation modes occurring in the data. This use of PCA has seen significant success toward identifying sources of dimensional variation using body-in-white (sheet metal assembly stage) data attained during automotive assembly, as can be seen in the works of Hu and Wu (1992), Wu et al. (1994), Yang (1996), and Lee and Jang (2001). In these papers, variation patterns occurring in the For/Aft (x-direction), In/Out (y-direction), and/or High/Low (z-direction) of the vehicle were determined independently, and the resulting variation patterns were mapped onto 2D sketches using arrows to identify the direction and magnitude of the variation.

PCA can also be used as a dimension reduction technique, where a linear transformation is performed on a set of correlated multivariate variables, resulting in a set of orthogonal variables through eigenvector decomposition. Specifically, PCA aims at reducing a large set of variables to a much smaller set, while retaining the majority of the information. This smaller set of variables can then be further analyzed. To shed light on the structure of PCA, consider a multivariate process consisting of p process variables with N observations, $\mathbf{X}\in\mathbf{R}^{N\times p}$. PCA can be performed by solving the eigenvalue decomposition of the sample covariance matrix, $\mathbf{S}$:

$$\mathbf{S} = \mathbf{VLV}^T, \quad \mathbf{S} \in \Re^{p\times p} \tag{1}$$

where $\mathbf{L}$ is a diagonal matrix containing eigenvalues ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$), and $\mathbf{V}$ is an orthonormal matrix whose columns are eigenvectors (loading vectors). In addition, the process variables can be projected onto the eigenvector space to generate the principal components. The reader should note that any principal component (i) has a mean of zero and a variance of λ_i. In addition, its principal component score can be calculated as

$$z_i = \mathbf{v}_i^T (\mathbf{x} - \bar{\mathbf{x}}), \quad \mathbf{x}, \bar{\mathbf{x}} \in \Re^{p \times 1}, \tag{2}$$

where $\mathbf{v}_i$ is the ith column of $\mathbf{V}$, $\mathbf{x}$, and $\bar{\mathbf{x}}$ are vectors of observations of the original variables and their means, respectively. For each observation, $\mathbf{x}$, in Eq. 2, z_i is the principal component score for that observation. In addition, the quantity

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \tag{3}$$

is the proportion of the variability of the original system that can be attributed to the ith principal component. Therefore, the principal components with the smallest variance provide very little information on the behavior of the system, and in fact can be just reflecting statistical noise. This allows for a reduction in the dimensionality of system, as only k variables ($k < p$) are needed to characterize the process behavior, allowing for a model consisting of k dimensions, $\mathbf{V} \in \mathbf{R}^{p \times k}$, to be obtained. For more details on PCA, the reader is referred to Johnson (1998) and Montgomery (2008).

Jones and Rice approach

Jones and Rice (1992) provided a standard technique, based on PCA, which is applicable for Phase I profile monitoring, where each profile is treated as a multivariate vector of response values. From this response data, a set of mutually orthogonal linear combinations of the response variables are determined, such that they explain the majority of the process variation. The strength of this approach is that if these principal components can be easily interpreted, they can be very effective in understanding the process variation. To simplify the interpretation process, Jones and Rice (1992) recommended plotting the observations which have the maximum and minimum principal component scores, with respect to the most significant principal component's direction (loading vector), to further understand the process variation. Woodall et al. (2004) "strongly recommended the use of these plots in Phase I".

As an illustration to show how the Jones and Rice (1992) approach (referred to as JR approach from this point onwards) can be useful in interpreting a set of Phase I profile data, 100 profiles based on the following kernel density estimate are generated:

$$\hat{f}(x) = .02 \sum_{i=1}^{50} 2\varphi \; \{2\,(x - X_i)\}, \tag{4}$$

based on i.i.d observations, $X_1, X_2, \ldots, X_{50}$, sampled from the standard normal distribution (ϕ is the standard normal density). It should also be noted that this example is quite similar to the one used in the Jones and Rice (1992) paper. Figure 1 shows the 100 profiles generated through simulations of Eq. 4. It is clear that Fig. 1 is uninformative since the behavior of individual curves cannot be seen.

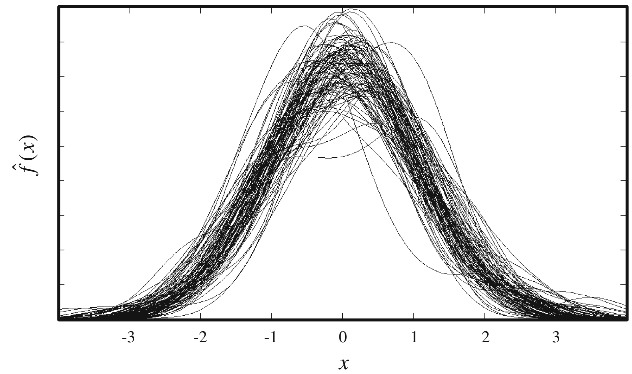

Fig. 1 100 Kernel density estimates

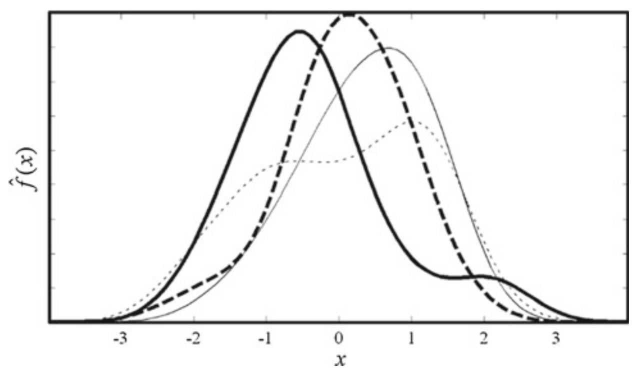

Fig. 2 Profiles corresponding to the maximum (*solid*) and minimum (*dashed*) scores on the first (*darker* and *thicker curves*) and second (*lighter* and *thinner curves*) principal components

Consequently, the JR method suggests plotting only the profiles (observations) having the maximum and minimum principal component scores for the most significant principal component (s). For this example, the majority of the variation can be described by the first two principal components, which account for 43.3 and 34.4% of the variation in the profiles. Following the JR method, the profiles representing the maximum and minimum principal component scores for the first two principal components are shown in Fig. 2. Despite the fact that principal components are often difficult or impossible to interpret, in this example (based on Fig. 2), it is clear that the first principal component corresponds to the location of the curves (i.e. where the mean of the curves are located), while the second corresponds to the spread of the curves (i.e. the standard deviation of the curves).

Despite the usefulness of the Jones and Rice (1992) approach, it is not clear how this approach could be used for dimensional profile data in manufacturing processes since their approach is only applicable for two-dimensional (2D) profiles. In this paper, we extend their approach to the analysis of three-dimensional (3D) profiles so that it can be used for understanding the variation in more complicated manufacturing processes. This is explored in more detail in the following sections.

Extending the JR method to 3D profiles

To understand the difficulties found in applying the JR method to 3D profiles, we first provide a simple simulated example. In this example, 100 simulations are generated for the location of the eight vertices of a standard cube, as shown in Fig. 3. In these simulations, the locations of all points are generated from the same normal distribution in the x, y, and z direction, except for the x-direction of points 5 and 8, which have a larger variance and are correlated with a correlation coefficient of -0.8. The question now arises on whether the JR method can be applied to these simulated data to determine if this variation pattern can be easily interpreted through PCA.

According to the JR methodology, observations corresponding to the maximum and minimum principal component scores for the first principal component are displayed in Fig. 4. Even though the profiles shown in Fig. 4 can be informative in identifying vertices of maximum variation, they do not provide an intuitive description of the variation pattern. This is not surprising since the 3D problem is being reduced to 2D space, resulting in a loss of critical information regarding the product's geometry and the physical relationship between the measurements. Consequently, the JR method should be modified so that it can be used for dimensional profiles where it is important to relate the existing variation patterns to the part geometry. In the paragraphs below, a naïve modification to the JR method is presented, which would enable the practitioner to capture some of the lost 3D information. This modification and its associated complications set the background needed for our visualization method, which is explained in the following sections.

To relate the JR method to the cube's geometry, it is intuitive to consider mapping the observations obtained from the JR method onto a 3D representation of the cube, as this would allow the retention of crucial 3D information, necessary

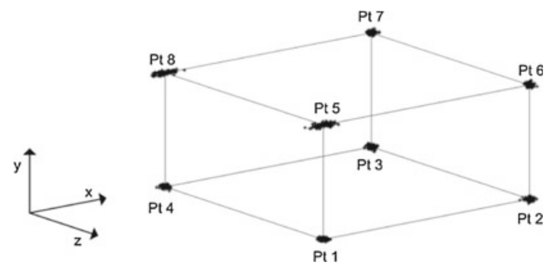

Fig. 3 Simulated data for the cube's vertices

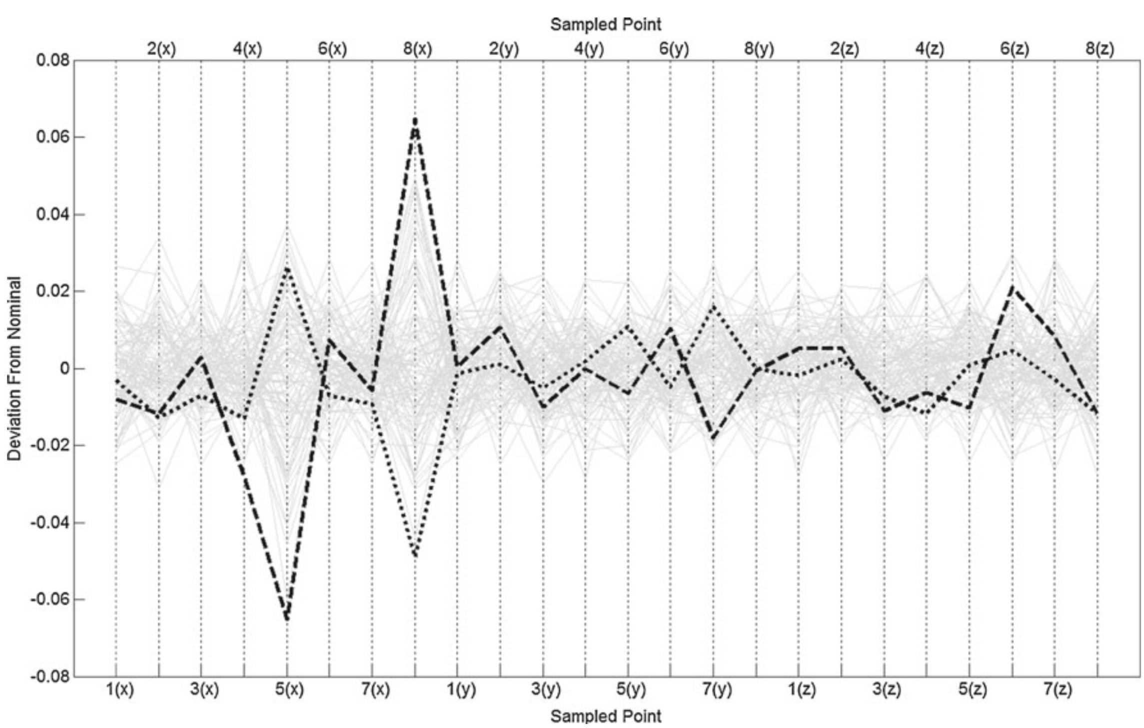

Fig. 4 100 simulated cube vertices (*light* and *thin*) and the maximum (*thick dotted*) and minimum (*thick dashed*) scores for the first principal component

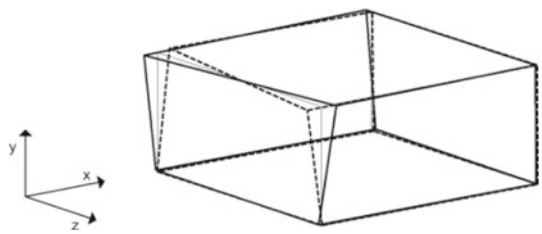

Fig. 5 First variation mode (twisting of the edge defined by vertices 5 and 8 about the y-axis). The *solid* and *dashed profiles* represent the observations with the maximum (*solid*) and minimum (*dashed*) principal component score with respect to the first principal component. It should be noted that the *light* and *thin line* represents the cube's nominal dimensions

for interpreting the variation mode(s). Utilizing the insights from the JR method, only the observations corresponding to the maximum and minimum principal component scores for the most significant principal components (for this example the first principal component) will be used. Based on Fig. 5, it is clear that the most significant variation mode corresponds to a twisting of the edge defined by points 5 and 8 about the y-axis. While this visualization technique is suitable for this example, for more complex cases linearly connecting the sampled points may not adequately represent the part geometry or variation pattern. Therefore, we present a visualization technique, which aims to solve the problem of extending the JR method to 3D datasets present in manufacturing environments by incorporating 3D CAD models. These models will be used as the basis for the mapping the observations back to 3D space, allowing the variation modes to be accurately depicted.

A case study showcasing the importance of variation visualization

To illustrate the difficulty of visualizing common cause variation in complex geometries, consider an industrial example from an automotive assembly plant shown in Fig. 6. The CAD data obtained from this vehicle consist of the geometric representation of the roof and side panels. The dataset for this case study consists of 1,000 observations (vehicles), where during the assembly of this vehicle 100% sampling was performed at 27 locations, resulting in 39 dimensional measurements (deviations from nominal) per observation. It should be noted that due to proprietary issues original information, such as the number and location of sample points (given in Fig. 6 as black spheres), and the sample data have been slightly modified.

JR methodology

To apply the JR methodology on the assembly dataset, the observations corresponding to the maximum and minimum principal component scores for the first principal component (contains the majority of variation for this case study) are displayed in Fig. 7. The interpretation of the profiles in Fig. 7 is very difficult, as the ordering of the data is arbitrary since they are on a nominal scale. It should also be noted that setting a logical order for the sampling locations and directions in the independent axis (prior to the PCA) is difficult, since a logical order would be a function of the variation mode(s) being observed. However, it seems reasonable to first utilize the magnitude of the variation to order these measurements, since this information is not only useful, but it can also be easily extracted from the dataset. Therefore, this method (referred to as JRE) is explored in the following subsection.

JRE methodology

When trying to visualize the results presented in Fig. 7, it would be intuitive to order the measurements in the independent axis from largest variation to smallest. For example, in Fig. 7, it seems obvious that points $11(y)$ and $27(z)$ have the largest and smallest variation, respectively. If the largest deviations are grouped, then it would be possible to identify regions of interest. Consequently, root causes for these regions can be identified using process knowledge in a subsequent process improvement project. Accordingly, these measurements were sorted by their variation magnitude (descending), and are shown in Fig. 8.

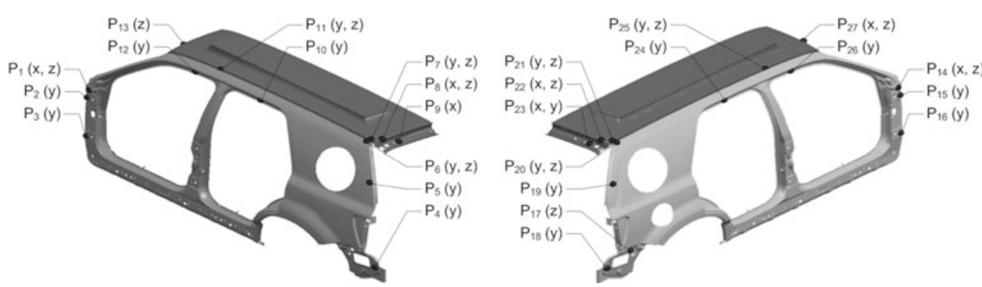

Fig. 6 Case study geometry with sample locations

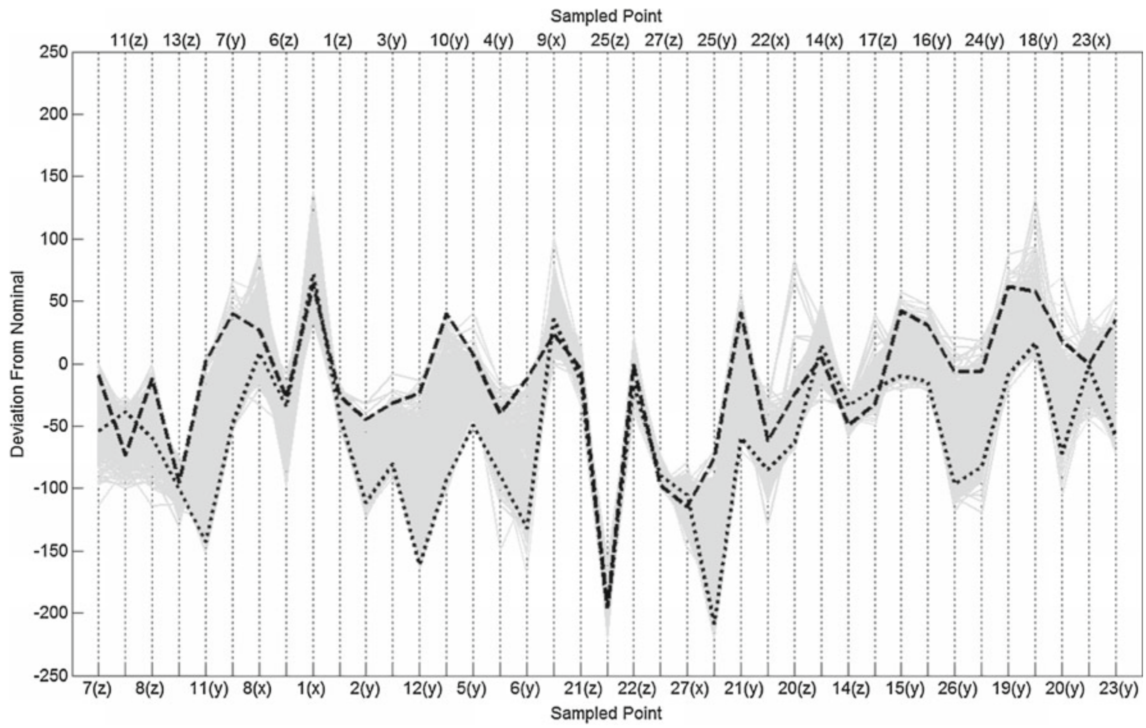

Fig. 7 Observations corresponding to the maximum (*dashed*) and minimum (*dotted*) principal component scores for the first principal component. The remaining profiles are represented by *light* and *thin lines*

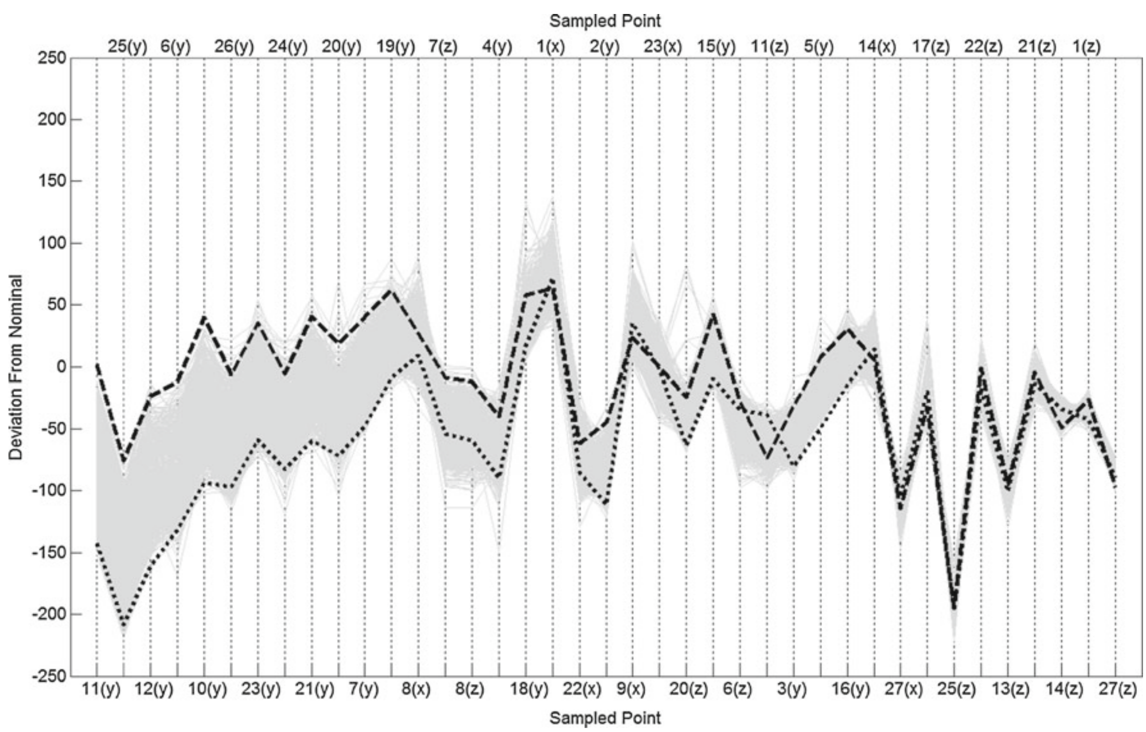

Fig. 8 Rearranged profiles from Fig. 7. It should be noted that the re-arrangement is based on the (*descending*) magnitude of the variation of the measurement points

From Fig. 7, several interesting deductions about the process/product can be made. First, the maximum and minimum scored observations are almost identical for the last 10 measurements, from 14(x) to 27(x), which means that the contribution of these points to the variation mode represented by the first principal component is negligible. On the other hand, the maximum and minimum principal component scored observations for the first principal component are consistent from 11(y) to 4(y). Accordingly, it can be argued that these 16 measurement points are the most significant to represent the first variation mode. It is important to note that the first 12 of these 16 measurement points all measure deviations in the y-direction and are all located where the roof meets the side panels of the vehicle (refer to Fig. 6), which cannot be easily ascertained from this figure. However, using the knowledge gained from this analysis, and comparing that with the known locations of the measurement points, it can be inferred that the variation pattern is that the tops of both the side panels (where they meet the roof) are tilting together in the y-direction.

In addition to ordering the measurements by variation, a similar approach can be done by ordering the measurements by the difference between the measurements in the observations having maximum and minimum principal component scores, as shown in Fig. 9. The results of Fig. 8 are quite similar to those of Fig. 9 in the fact that the majority of the variation across the first principal component can be attributed to measurements in the y-direction. However, what is most important to understand from this figure is the consistent pattern seen between the observations having the maximum and minimum principal component scores. The pattern strongly indicates the presence of a fairly rigid deformation pattern occurring during the assembly of this SUV. A goal of our paper is to make this diagnosis more intuitive through integrating the JR method with CAD tools, which is presented in the following section.

Integrating the JR method with CAD

The framework for the JR/CAD integration scheme presented in our paper builds upon the observations noted in the previous sections. Therefore, the framework utilizes the strengths of PCA (as highlighted by the JR method), in addition to the need for further understanding through visualization tools such as VRML. Specifically, the proposed method consists of the following three steps: (1) Data preparation and reduction, (2) Data analysis and variation identification, and (3) Variation understanding through animated CAD visualization, as summarized in Fig. 10. In addition, a Pre-Processing step is performed in order to develop an adequate 3D model used to graphically map the geometric variation pattern. Each of these steps is discussed in more detail in the subsections below.

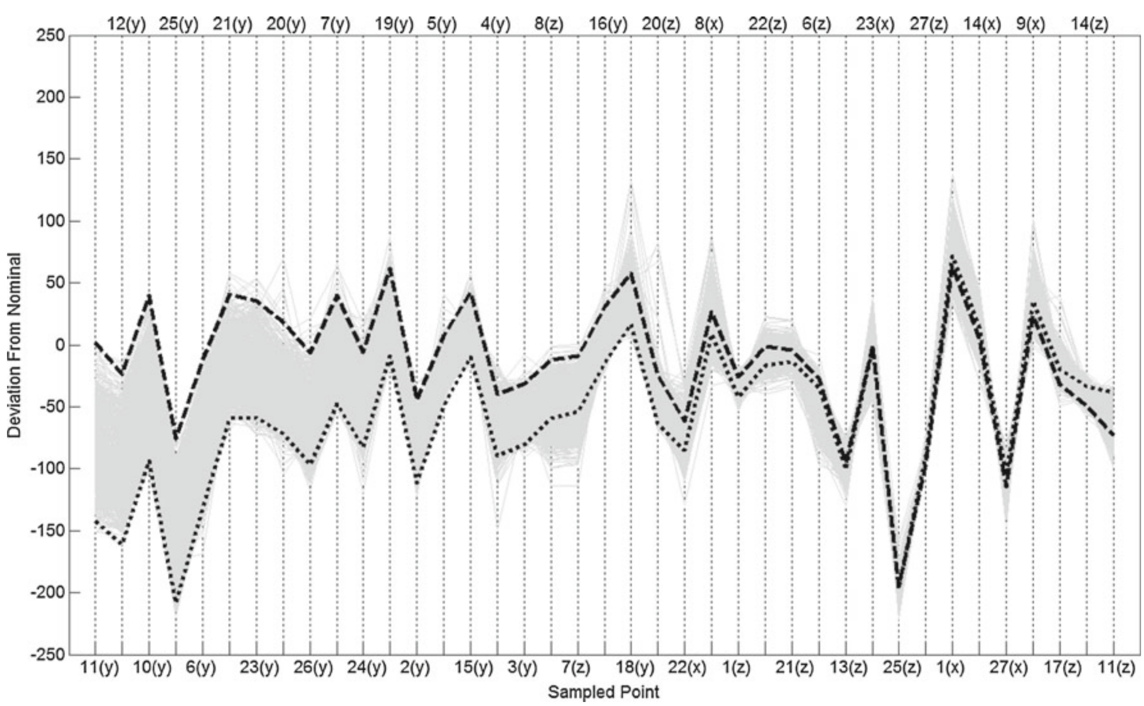

Fig. 9 Rearranged profiles from Fig. 7. It should be noted that the re-arrangement is based on the (*descending*) difference between the measurements in the observations having the maximum and minimum principal component scores

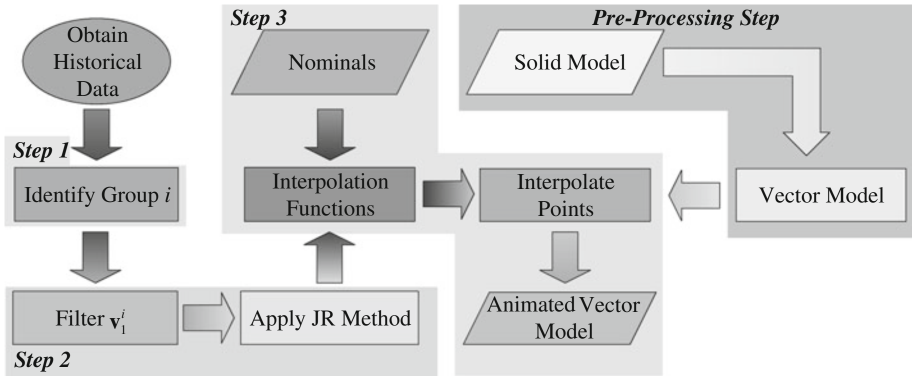

Fig. 10 JR/CAD integration steps

Data preparation and reduction

The effectiveness of PCA to represent a variation pattern can be significantly improved by only including variables that significantly contribute to the principal component loading vector being considered. Therefore, we recommend that these insignificant variables be filtered out before performing PCA. These variables are determined through the simultaneous evaluation of the following two criteria: (1) Identification of the sample points that have the highest contribution to the system variation, which happen to be the sample points with the largest variance; and (2) Grouping highly correlated sample points together.

The purpose of utilizing these criteria is three-fold. First, points that exhibit small variation are likely reflecting statistical noise and do not contribute significantly to a variation pattern. In order to produce the best possible visualization of the significant variation pattern(s), these points must be removed. Second, if a point has no correlation between other sample points, it either contains no contribution to the deformation pattern or it is solely responsible for the variation pattern within the system. If the former case is true, then the point should be removed to produce the best possible visualization. If the latter is true, then PCA will not provide any meaningful insight into the system's variation pattern, which may result in misleading visualization of the variation pattern. Third, if multiple distinct variation patterns with varying severity are occurring simultaneously within the system, it becomes possible to produce multiple visualizations for each distinct pattern.

The implementation of these two criteria in an industrial setting will typically consist of several sequential steps. First, the measurement points are sorted in descending order with respect to their variances. Next, sets of sample points with high variability are identified as regions of interest. For this identification, we suggest selecting points whose variance is larger than a predetermined threshold, σ_t^2. From here, the correlation matrix including all the measurement points can be calculated. Points with the highest absolute correlation are then selected as points of interest. High correlation is defined using a predetermined threshold absolute correlation value, ρ_t, which is typically around 0.7–0.8. It should be noted that the determination of the exact value for this threshold should depend on the process, the part being monitored, and the objective behind the monitoring. The authors believe that industrial practitioners should investigate the effects of using different values for this threshold to ensure that they can meet their objective.

Data analysis and variation identification

Once the significant sampling points are obtained, PCA is performed for each identified group of correlated measurements. This analysis results in the identification of loading vectors of the data most responsible for the variation pattern of interest. Typically for manufacturing systems, the first two to three loading vectors are all that is required to adequately model the variability within a system, as discussed by Kourti and MacGregor (1996). Therefore, it is reasonably safe to assume that when performing PCA on grouped data, only the first loading vector of the ith group ($\mathbf{v}_1^i$) is needed to represent the majority of the variability. Once $\mathbf{v}_1^i$ has been identified, the JR method is performed using the original observations and $\mathbf{v}_1^i$ to calculate principal component scores, resulting in the observations corresponding to the minimum and maximum principal component scores for $\mathbf{v}_1^i$. It should be noted that due to grouping, the dimensionality of $\mathbf{v}_1^i$ may be less than that of the original observations. In this case, the elements within the observations corresponding to the removed variables from $\mathbf{v}_1^i$ are also removed.

It should be noted that it is possible to forego the data preparation and reduction step, and perform PCA directly to the original data set. While this approach is simpler than the suggested approach described above, it also has some drawbacks. For instance, consider a production vehicle similar to the one described earlier where two independent variation patterns exist. Performing PCA on this data would most likely result in two dominant principal components, one for each variation pattern. However, due to the nature of PCA, both of these principal components would contain some information regarding the other variation pattern. When this occurs, it becomes difficult to visually interpret the variation pattern and link this variation pattern to the system. This problem can easily be seen in Fig. 9. In this figure, measurements $7(z)$ and $8(z)$ exhibit significant variation across the 1st principal component. However, these two points are not actually involved in the major variation pattern seen in the vehicle, as this variation pattern is known to be attributed to variations in the y-axis only. By implementing the grouping step described above, this problem can be eliminated through identifying groups of correlated points before PCA is performed.

Variation understanding through animated CAD visualization

Current solid modeling programs involve data intensive binary file structures to store a large amount of data regarding a given model, such as surface spline data, modeling parameters, modeling history, and so forth. In order to facilitate the JR/CAD integration, Virtual Reality Markup Language (VRML) is employed as the CAD format to be used. Originally developed to display web-based content in a 3D environment, VRML is a file format for representing interactive 3D vector graphics. VRML incorporates vertices and edges of 3D polygons in an *American Standard Code for Information Interchange* (ASCII) based text file. This simple structure gives a significant amount of freedom in being able to alter the original geometry. Given that a 3D model of the manufactured part exists, this step merely consists of translating the solid model into VRML, which can be easily performed since most commercially available CAD software packages can export their solid models into VRML format. The reader is referred to Ames et al. (1997) for a detailed introduction on VRML coding and to Krishnamurthy (2001) on some of its applications in manufacturing systems.

The most important aspect of VRML toward the success of JR/CAD integration is that animations can be developed within the language structure. Simple point-to-point interpolations of vertices and edges can be animated quite easily. Any polygon associated with a point that is being animated will inherit this motion, and the entire product can thus be animated to represent a given geometric variation pattern with respect to the nominal geometry. The ability to animate the geometric variation pattern is crucial for the success of the JR/CAD integration as a visualization tool. The final advantage of using a VRML model to represent the geometric variation pattern is that once the model has been generated, and the algorithms have been put into place, no expensive software is required for the visualization of the variation patterns, since a variety of free stand alone and web browser plug-ins are available for reading and viewing VRML files.

Depending on the complexity of the 3D model, the number of vertices within the VRML could range anywhere from hundreds to tens of thousands. However, typical dimensional data measurements recorded during a manufacturing process have significantly fewer points. In addition, there is no guarantee that the sample points are associated with any of the points within the 3D model. Therefore, in order to successfully project the geometric variation modes onto the 3D model, some form of interpolation is required. We suggest the use of a simple Kriging approximation model with a Gaussian covariance function (Stein 1999).

For the JR/CAD integration, the key to a successful visual result is to include all of the sample points when performing the spatial interpolation. During the previous two steps certain sampling points, which had very little contribution to the total variation, were neglected. However, for the spatial interpolation, these points should be set to zero and included along with $\mathbf{v}_1^i$ as known spatial points. This approach will drastically reduce the possibility of poor interpolation results.

JR/CAD integration case study

The automotive assembly case study presented earlier in this paper is used to demonstrate the effectiveness of the proposed JR/CAD integration. Using commercially available CAD software, the geometric model was easily exported to VRML format for the purpose of mapping the geometric variation patterns onto the 3D model. This is the only manual step in the proposed integration process, since the remainder of the process follows the methodology outlined in Fig. 10.

Data preparation and reduction

Despite the fact that 100% sampling was employed in the assembly process for this case study, some measurements are missing, as is often the case in manufacturing systems. To begin the JR/CAD integration process, these missing measurements must be accounted for appropriately. Here, if a measurement is missing, the corresponding vehicle was removed, which resulted in the number of vehicles being reduced from 1,000 to 962. It should be noted that the authors have opted to remove these vehicles for convenience purposes. However, a more structured approach utilizing data mining (see Han and Kamber 2006; Choudhary et al. 2009) techniques can be followed, especially if the number of incomplete observations is significantly high.

Table 1 Case study: correlation matrix of the 15 sample points with the largest variation (in descending order)

	11(y)	25(y)	12(y)	6(y)	10(y)	26(y)	23(y)	24(y)	21(y)	20(y)	7(y)	19(y)	8(x)	7(z)	8(z)
VAR	*0.89*	*0.87*	*0.78*	*0.75*	*0.73*	*0.66*	*0.63*	*0.55*	*0.48*	*0.44*	*0.43*	*0.39*	*0.37*	*0.36*	*0.30*
11(y)	**1.00**	**−0.97**	**0.94**	**0.76**	**0.93**	**−0.82**	**−0.75**	**−0.76**	**−0.75**	**−0.77**	0.69	**−0.73**	−0.37	0.11	−0.43
25(y)	**−0.97**	**1.00**	**−0.91**	**−0.80**	**−0.88**	**0.89**	**0.76**	**0.84**	**0.75**	**0.79**	−0.68	**0.75**	0.39	−0.12	0.46
12(y)	**0.94**	**−0.91**	**1.00**	**0.80**	**0.95**	**−0.80**	**−0.74**	**−0.75**	−0.66	**−0.76**	0.57	**−0.71**	−0.43	0.22	−0.51
6(y)	**0.76**	**−0.80**	**0.80**	**1.00**	**0.81**	**−0.77**	**−0.87**	**−0.77**	**−0.71**	**−0.84**	0.55	**−0.86**	−0.22	0.08	−0.33
10(y)	**0.93**	**−0.88**	**0.95**	**0.81**	**1.00**	**−0.74**	**−0.80**	**−0.71**	**−0.73**	**−0.79**	0.64	**−0.77**	−0.31	0.13	−0.38
26(y)	**−0.82**	**0.89**	**−0.80**	**−0.77**	**−0.74**	**1.00**	**0.76**	**0.92**	0.63	**0.77**	−0.57	**0.73**	0.47	−0.15	0.51
23(y)	**−0.75**	**0.76**	**−0.74**	**−0.87**	**−0.80**	**0.76**	**1.00**	**0.77**	**0.76**	**0.89**	−0.69	**0.94**	0.14	0.02	0.22
24(y)	**−0.76**	**0.84**	**−0.75**	**−0.77**	**−0.71**	**0.92**	**0.77**	**1.00**	0.64	**0.78**	−0.58	**0.75**	0.35	−0.08	0.41
21(y)	**−0.75**	**0.75**	−0.66	**−0.71**	**−0.73**	0.63	**0.76**	0.64	**1.00**	**0.71**	**−0.96**	**0.76**	0.10	0.25	0.18
20(y)	**−0.77**	**0.79**	**−0.76**	−0.84	**−0.79**	**0.77**	**0.89**	**0.78**	**0.71**	**1.00**	−0.66	**0.88**	0.20	0.02	0.27
7(y)	0.69	−0.68	0.57	0.55	0.64	−0.57	−0.69	−0.58	**−0.96**	−0.66	**1.00**	−0.69	−0.06	−0.31	−0.12
19(y)	**−0.73**	**0.75**	**−0.71**	**−0.86**	**−0.77**	**0.73**	**0.94**	**0.75**	**0.76**	**0.88**	−0.69	**1.00**	0.11	0.10	0.18
8(x)	−0.37	0.39	−0.43	−0.22	−0.31	0.47	0.14	0.35	0.10	0.20	−0.06	0.11	**1.00**	−0.38	0.70
7(z)	0.11	−0.12	0.22	0.08	0.13	−0.15	0.02	−0.08	0.25	0.02	−0.31	0.10	−0.38	**1.00**	−0.41
8(z)	−0.43	0.46	−0.51	−0.33	−0.38	0.51	0.22	0.41	0.18	0.27	−0.12	0.18	0.70	−0.41	**1.00**

Once the missing data points are accounted for, the next step in the JR/CAD integration is to isolate groups of highly correlated sample points (locations and directions). This was performed by sorting the sample locations according to their degree of variability and then locating groups of sample points where the absolute correlation between each data source is larger than a predefined criterion, in this case 0.7. Upon analyzing the data for potential groups, a group of 12 sample points was identified. These 12 sample points correspond to the first 12 sample points identified in Fig. 8. Table 1 shows a portion of the grouping process performed on the correlation matrix. Here, the 15 sample points with the largest variation (in descending order) are presented. The first row of this table gives (values in italics) the variance of the corresponding sample point, while the remainder of the table contains the correlation matrix. The (absolute) correlation values larger than the predefined criterion of 0.7 are indicated by bold font. From this table it is quite easy to see that the first 12 sample points are indeed moving together. The only questionable measurement location is 7(y), which only satisfies the predefined criterion once. However, given that 7(y) is close to meeting the criterion, that it exhibits a fairly large variation (with respect to the remaining sample points), and that it has a significant correlation with 21(y) (which is obviously part of the group), it is also included within the group. The remainder of the correlation matrix has sporadic occurrences of values which satisfy the criterion; however, no patterns exist from which a group can be detected. These occurrences are most likely due to the close proximity of the sample points, which in essence measure the same phenomenon.

Data analysis and variation identification

Once the group of highly correlated and high variance variables has been obtained, the next step for the JR/CAD is to perform PCA on the group. The result of the PCA for the identified group produces $\mathbf{v}^1$, which accounts for more than 80% of the variability within the group. Next, the JR method is performed using the original data to obtain the vehicles that have the maximum and minimum principal component scores with respect to $\mathbf{v}^1$. These vehicles are displayed in a skeletal diagram, similar to the one presented in Fig. 5 for the cube, as shown in Fig. 11. From this figure, it can be seen that the variation pattern being observed is that the tops of both the side panels (where the side panels meet the roof) are tilting together in the y-direction.

Despite the fact that Fig. 11 is very useful in identifying the variation pattern, a few problems exist with implementing this as a visualization tool. First of all, if a large number of sample points are being taken, this visualization technique can become chaotic. Second, the selection of which sample points should be connected together to best visualize their relationships is a function of both the physical geometry of the part and the variation pattern. Third, and most importantly, is the issue of having measurements in multiple directions. For instance, the variation pattern shown in Fig. 11 is purely in the y-direction. However, consider a sample point measuring deviations in the z-direction that is located in the area most affected by the variation pattern (i.e. where one of the side panels meets the roof). If this measurement is significant with respect to the principal component loading vector, this visualization

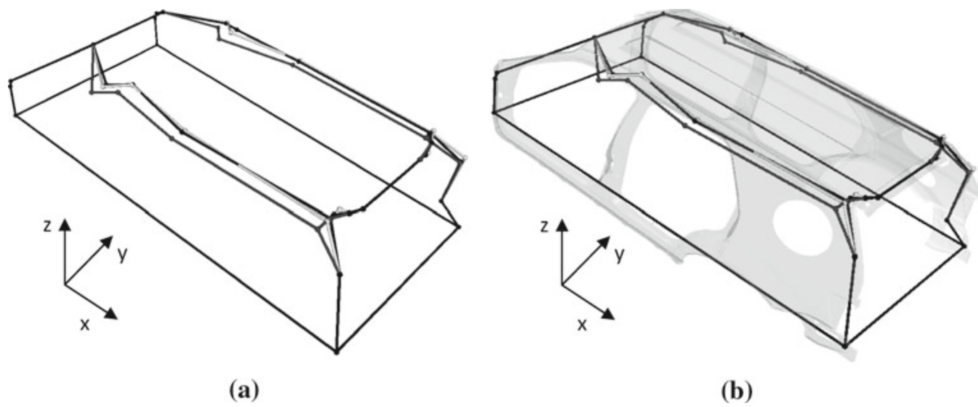

Fig. 11 Skeletal diagram of the observations with the maximum (*gray*) and minimum (*white*) principal component score with respect to the first principal component for the identified group. The *black lines* represent the measurement's nominal value, where the nominal CAD geometry is also given in (**b**)

technique begins to unravel. For the visualization, this point should show movement only in the z-direction; however, it is located at a spot where there should also be movement in the y-direction.

Map loading vector onto CAD geometry

In order to alleviate the aforementioned issues, we propose to map $\mathbf{v}^1$ onto the nominal CAD geometry using a Kriging model. In order to create the model, the nominal values for all data poins are used (including those that do not reside in the identified group). Since the group being analyzed only contains data regarding variation in the y-direction, the Kriging prediction model, results in a 1-dimensional estimation for the geometric variation mode, where the other 2-dimensions stay at their nominal values. The resulting CAD results can be seen in Figs. 12 and 13. By visually combining these two figures, one can easily identify the range of the most significant variation pattern obtained from the data. Figure 12 shows that the variation pattern starts with just a slight tilt in the side panels in the negative y-direction, while Fig. 13 shows that this variation pattern ends with a much more significant tilt in the side panels, still in the negative y-direction, which is a common variation pattern in the automotive industry, known as "matchboxing". For more information regarding "matchboxing", the reader is referred to Olson and Mortimer (1973).

As mentioned previously, the goal of our paper is to develop a 3D rendering of geometric variation patterns, and more specifically present them in the form of an animation. For obvious reasons, the animation results are not presented here. However, the readers are encouraged to visit the authors' website[1] for viewing the animated JR/CAD integration in a fully explorable 3D environment. The animation

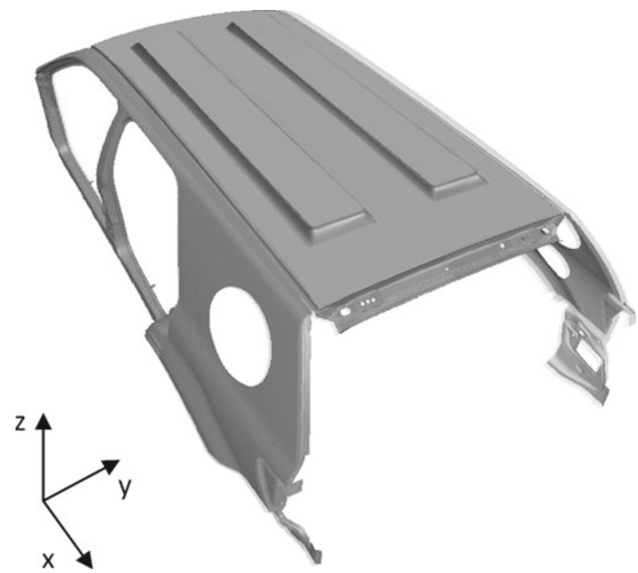

Fig. 12 CAD representation of the observation corresponding to the minimum principal component score with respect to the first principal component for the identified group. In addition the nominal CAD geometry is represented by a transparent model

allows the reader to visualize the variation pattern of a 3D profile through the JR method. Engineering personnel and shop floor operators can then utilize their process understanding to suggest improvement strategies for this (among other) variation patterns, which can be visualized through our proposed method. This can result in significant improvements in the overall quality of the final product.

Conclusions

The proposed JR/CAD demonstrates the benefits of applying PCA to easily visualize and interpret dimensional variation patterns occurring in manufacturing systems. This allows engineers and operators to use their knowledge of the

[1] For more information and results, visit: www.imas.ise.vt.edu/Research/Data_Mining/Projects/Current/pca.html.

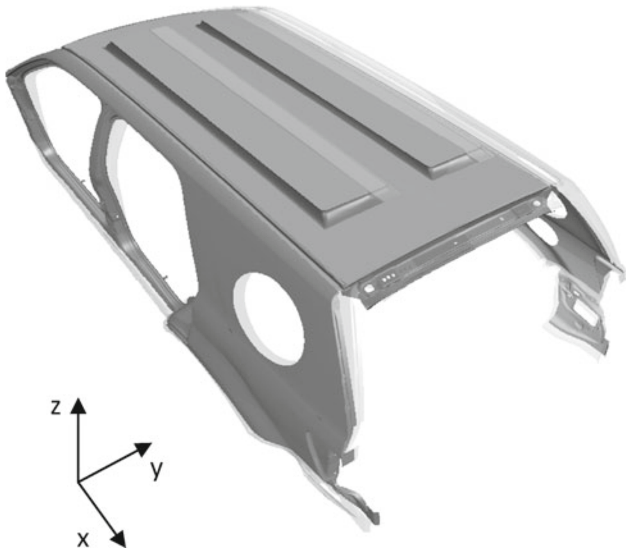

Fig. 13 CAD representation of the observation corresponding to the maximum principal component score with respect to the first principal component for the identified group. In addition the nominal CAD geometry is represented by a transparent model

process to quickly understand and identify common-cause variability. The result of which is faster identification of opportunities for process improvement, and an overall increase in product quality. In addition, as more and more data are being collected in modern manufacturing systems, tools such as the method proposed within our paper are becoming increasingly important. Without tools that allow operators and engineers to easily filter through this immense amount of data and quickly turn it into useful knowledge, manufacturing systems will soon be overflowing with data, causing the information in the data to never be fully utilized.

Acknowledgments This material is based upon work supported by the National Science Foundation under Grant No. CMMI-0927323. The authors would also like to thank Cory Niziolek and Christopher Norris, undergraduate researchers at Virginia Tech, for their contributions to the animations.

References

Ames, A. L., Nadeau, D. R., Moreland, J. L., & Ames, A. L. (1997). *The VRML 2.0 sourcebook* (2nd ed.). New York: Wiley.

Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2009). Data mining in manufacturing, a review based on the kind of knowledge. *Journal of Intelligent Manufacturing, 20*, 501–521.

Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). San Francisco, CA: Morgan Kaufmann Publishers.

Hu, S. J., & Wu, S. M. (1992). Identifying sources of variation in automobile body assembly using principal component analysis. *Transactions of NAMRI, 20*, 311–316.

Johnson, D. E. (1998). *Applied multivariate methods for data analysts*. Pacific Grove, California: Duxbury Press.

Jones, M. C., & Rice, J. A. (1992). Displaying the important features of large collections of similar curves. *The American Statistician, 46*, 140–145.

Kourti, T., & MacGregor, J. F. (1996). Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology, 28*, 409–428.

Krishnamurthy, K. R. (2001). *Application of HTML/VRML to manufacturing systems engineering*. Unpublished Master's Thesis, Virginia Polytechnic Institute and State University, Blacksburg. Retrieved May 5, 2010, from Dissertations and Theses Database.

Lee, S., & Jang, K. (2001). Identification of the deviation from the mean and multimodal distribution in the automotive assembly process: Correspondence analysis approach. *International Journal of Production Research, 39*, 2105–2120.

Megahed, F. M., & Camelio, J. A. (2010). Real-time fault detection in manufacturing environments using face recognition techniques. *Journal of Intelligent Manufacturing*. doi:10.1007/s10845-010-0378-3.

Montgomery, D. C. (2008). *Introduction to statistical quality control* (6th ed.). Hoboken, N.J.: Wiley.

Olson, P. L., & Mortimer, R. G. (1973) *Investigation of some factors affecting the aim of headlamps*. Technical Report No. UM-HSRI-HF-73-13, Highway Safety Research Institute, University of Michigan.

Stein, M. L. (1999). *Statistical interpolation of spatial data, some theory for kriging*. New York: Springer.

Wheeler, D. J., & Chambers, D. S. (1992). *Understanding statistical process control*. Knoxville, TN: SPC Press.

Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology, 32*, 341–350.

Woodall, W. H. (2007). Current research on profile monitoring. *Produção, 17*, 420–425.

Woodall, W. H., & Adams, B. M. (1998). Statistical process control. In H. M. Wadsworth (Ed.), *Handbook of statistical methods for engineers and scientists* (2nd ed.). New York: McGraw-Hill.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., & Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *Journal of Quality Technology, 36*, 309–320.

Wu, S., Hu, S. J., & Wu, S. M. (1994). A fault identification and classification scheme for an automobile door assembly process. *International Journal of Flexible Manufacturing Systems, 9*, 261–285.

Yang, K. (1996). Improving automotive dimensional quality by using principal component analysis. *Quality and Reliability Engineering International, 12*, 401–409.