한국 *CAD/CAM* 학회 논문집
제 *1* 권 제 *3* 호 *1996*년 *12*월 *pp. 215-223*

기하 공차의 표현 및 조립성 확인에의 응용

박상호* · 이건우**

Representation of Geometric Tolerances and its Application to Assemblability Checking

Sangho Park* and Kunwoo Lee**

ABSTRACT

Every mechanical part is fabricated with the variations in its size and shape, and the allowable range of the variation is specified by the tolerance in the design stage. Geometric tolerances specify the size or the thickness of each shape entity itself or its relative position and orientation with respect to datums while considering their order of precedence. It would be desirable if the assemblability of parts could be verified in the computer when the tolerances on the parts are stored together with the geometric model of the parts of an assembly and their assembled state. Therefore, a new method is proposed to represent geometric tolerances and to determine the assemblability. This method determines the assemblability by subdividing the ranges of relative motion between parts until there exists the subdivided regions that do not cause the interference.

Key words : Geometric tolerances, Assemblability, Assembly

1. Introduction

In most engineering designs, the final goal is a composition of parts, formed into as an assembly. The simplest method to represent the assembly is to specify relative position such as the location and orientation of each component together with its nominal shape. The relative position can be represented by homogeneous transformation matrix of the coordinate frame attached to each part. In real practice, however, the shape error always accompanies each part because of the inaccuracies of the manufacturing processes and the range of the variation from the nominal geometry is specified by the tolerances. Thus it is necessary to store the tolerances as well to represent a part and an assembly of parts. Therefore, a mathematical representation of the tolerances and a method to add this information to the nominal B-Rep is proposed in this paper.

Geometric tolerances constrain the size or the thickness of each shape entity itself or its relative position and orientation with respect to other shape entities, called datums. Since the range of shape variation can be represented by the variation of the coordinate system attached to the shape, the transformation matrix of the coordinate system would mathematically express the range of shape variation if the interval numbers are inserted for the elements of the transformation matrix. For the shape entity specified by the geometric tolerances with reference to datums, its range of variation can be also derived by concatenating the transformation matrices composed of interval numbers. The components of the transformation matrices in the concatenation depend upon the order of precedence of datums. Thus storing the tolerance information would be equivalent to storing the transformation matrix for the associated shape entities.

*종신회원, 시스템공학연구소
**종신회원, 서울대학교 기계설계학과

Once the tolerance information is provided with an assembly model, the assemblability of the parts in the assembly can be verified. In fact, it would be desirable if the assemblability of parts could be verified in the computer when the geometric tolerances on the parts are stored together with the nominal geometric model of the parts and their assembled state. If so, tolerances can be assigned systematically in the design stage while considering the assemblability in advance. This would enable the design for assembly by allowing assemblability verification between toleranced parts in the design stage. This would be one example of the concurrent engineering activities.

To realize the concepts described above, we propose a new method to determine the assemblability when the tolerance information and the nominal geometric model of the components of an assembly are given. This method determines the assemblability by subdividing the ranges of relative position between parts until there exists the subdivided regions which do not cause interference. The continuity of the assembling path and assemblability can be inferred by analyzing these regions. These regions enclose the trajectory followed by the part during assembly. Thus separate regions imply the discontinuous trajectory and thus impossible assembling trajectory. The allowable range of relative motion is derived by assuming that the nominal parts are in an assembled state and the tolerance are small enough compared to the size of the parts. This method can be applied to simultaneous assemblability checking among several parts, which may be the toleranced parts, the actual part with shape error caused by manufacturing process, or the nominal parts.

2. Related Works

Requicha[1] introduced a new theory for geometric tolerances based on offsetting the boundary of a nominal solid model. Such offsets are termed tolerance zones, and they are used to constrain the allowable variations in size, form, orientation and position. Etesami[2] modeled the tolerance using a very similar approach to that of Requicha and he called his offset feature boundary solid.

In the vectorial tolerancing[3], the nominal part and the actual part are represented by using the vectors characterizing the orientation and the location of them. Thus the deviation vector between these vectors represents the tolerance.

Turner[4] developed a mathematical theory of tolerances in which tolerance specifications are interpreted as constraints that define a feasible region of model variations in a Cartesian space. Since the model variations may be caused by applying variations to the boundary of the part, the location, orientation and form of the boundary represents the model variations and they are constrained by tolerances.

Regarding the assemblability verification, the following publications can be found. First, data structure to represent an assembly was proposed[5]. This data structure assumes two mating conditions between parts; *against* condition between two nominal planar features and *fits* condition between two nominal cylindrical features. The positioning problem of nominal parts assembled with these mating conditions has been also studied[6]. This work has been extended to other mating conditions between features other than planar and cylindrical faces[7, 8].

Based on the theory of geometric tolerance proposed by Requicha[1], Fleming proposed the framework to represent the relative position of parts in an assembly by a network of tolerance zones and datums for the analysis of toleranced parts and their assemblies[9].

Srinivasan and Jayaraman derived the algebraic conditions about several assembly problems based on VBR (virtual boundary requirement)[10, 11]. They have been able to show that functional requirements in mechanical design can be stated in terms of VBR.

Turner treated a positioning problem of assembled parts as a constrained optimization problem[12]. Non-interference constraints are initially generated based on a vertex-face contacts. Inui and Kimura solved the positioning problem of non nominal parts in an assembly using simultaneous positioning[13].

3. Tolerance Representation

3.1 Interval Arithmetic

The interval arithmetic is a new branch of applied mathematics. A general treatment of the interval arithmetic can be found[14,15]. The theory on interval numbers is introduced because the mathematical representation of the tolerance in this work is based on the transformation matrix composed of interval numbers. It is also possible to do statistical tolerance analysis by introducing a statistical method to interval arithmetic.

An interval is a set of real numbers defined as

$$[a, b] \equiv \{x \mid a \leq x \leq b\} \tag{1}$$

The numbers a and b are called the bounds of the interval; a is called the lower bound and b is called the upper bound. of course, a should be smaller than b. The real number c is considered to be an interval $c=[c, c]$.

The interval arithmetic operators are defined as

$$[a, b] \circ [c, d] = \{ x \circ y \mid x \in [a, b] \text{and} y \in [c, d]\} \tag{2}$$

where $\circ$ represents addition, subtraction, multiplication or division such as $\circ \in \{+, -, \times, /\}$. Using the end points of the two intervals, the equation above can be rewritten as follows

$$[a, b]+[c, d]=[a+c, b+d] \tag{3}$$

$$[a, b]-[c, d]=[a-d, b-c] \tag{4}$$

$$[a, b]\times[c, d]=[\min(ac, ad, bc, bd), \max(ac, ad, bc, bd)] \tag{5}$$

$$[a, b]/[c, d]=[\min(a/c, a/d, b/c, b/d), \max(a/c, a/d, b/c, b/d)] \tag{6}$$

where $0 \notin [c, d]$ is required in the division relation.

3.2 Differential Matrix

An homogeneous transformation matrix is used to represent the spatial relationships between geometric elements, nominal or toleranced. Its column vectors represent the axis directions of a coordinate frame transformed from the reference coordinate frame and the transformed location of the frame origin. Any coordinate frame is free to move kinematically with degrees of freedom and can be positioned or constrained exactly relative to another frame by a transformation matrix. The transformation matrix can be derived by the concatenation of the translation and rotation.

If a new transformation matrix $(I+\Delta)$ is applied on a coordinate frame defined by the transformation matrix T, then the new transformed coordinate frame would be represented by $T(I+\Delta)=T+T\Delta$[16]. Here, the new transformation matrix gives the small variation to a shape element specified by applying translation and rotation in a very small amount, respectively. Therefore, a two dimensional differential matrix (Δ) can be defined as

$$\Delta(d_X, d_Y, \delta)=Q(d_X, d_Y)R(\delta)-I \tag{7}$$

where I is identity transformation matrix, Q is a translation matrix, R is a rotation matrix, d_X, d_Y are components to contrain the degree of freedom for X, Y translation respectively and δ is for rotation. On the assumption that a variation is small enough, the trigonometric functions can be linearized as follows

$$\lim_{\delta \to 0} \sin\delta = \delta, \lim_{\delta \to 0} \cos\delta = 1 \tag{8}$$

Substituting Relations (8) into Equation (7), the differential matrix can be derived as

$$\Delta(d_X, d_Y, \delta)=\begin{bmatrix} 0 & -\delta & d_x \\ \delta & 0 & d_Y \\ 0 & 0 & 0 \end{bmatrix} \tag{9}$$

Fig. 1 shows the relation of differential matrices

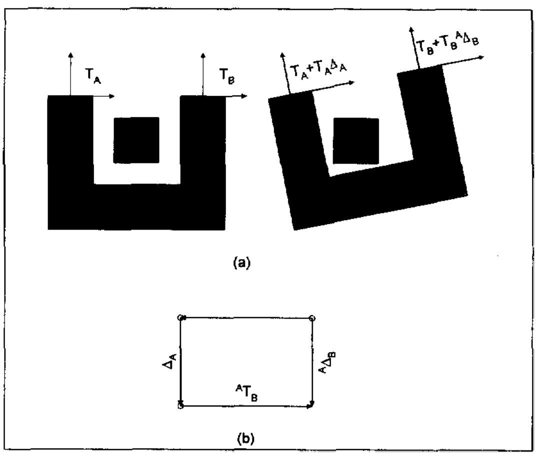


Fig. 1. Propagation of a differential matrix: (a) shows the propagation of the effect of a differential matrix and (b) shows the graph interrelating the transformation matrices and the differential matrices shown in (a).

between coordinate frames. In Fig. 1 (a), A and B frames are located at the ends of a concave polygon respectively. When A frame is moving with a differential matrix (Δ_A), B frame is also moving because A and B frames are attached on the same part. The new transformation matrix of frame B becomes $T_B + T_B\,{}^A\Delta_B$ and ${}^A\Delta_B$ is the matrix propagated to frame B by Δ_A applied to frame A. This relation can be rewritten by tracing the lower path in the graph of Fig. 1 (b) as below

$$T_B + T_B\,{}^A\Delta_B = (T_A + T_A\Delta_A)\,{}^A T_B \tag{10}$$

Since $T_B = T_A\,{}^A T_B$, the relation $T_B\,{}^A\Delta_B = T_A\Delta_A\,{}^A T_B$ is derived. From Equation (10), the differential matrix ${}^A\Delta_B$ can be derived as

$$^A\Delta_B = T_B^{-1} T_A \Delta_A\,{}^A T_B \tag{11}$$

$$= {}^B T_A \Delta_A\,{}^A T_B \tag{12}$$

3.3 Variational Entity

In this research, an *entity* is defined to be any shape element such as point, face and solid. An entity with the allowable ranges of variations is called a *variational entity*. Meanwhile a *nominal entity* is an entity without variations. A differential matrix, whose components are interval numbers, can describe a variational entity with respect to the corresponding nominal entity. The relation of the coordinate frames between the nominal entity and th variational entity can be represented by the transformation matrix and the differential matrix as follows

$$T_F = T_f + T_f\Delta \tag{13}$$

where T_F is a transformation matrix assumed to the coordinate frame of the variational entity, T_f is a transformation matrix of the nominal entity, and Δ is a differential matrix composed of intervals to give variations to the entity. Therefore, the geometric entities on a tolerance part can also be represented by the differential matrices whose elements are interval numbers. These interval numbers should correspond to the given tolerances including form, orientatuion and location tolerance because d_X, d_Y of differential matrix can constrain form and location, and δ can orientation.

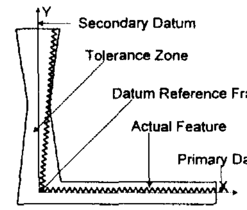


Fig. 2. Construction of a datum reference frame: This figure shows fitting t datum reference frame in 2 dimension according to their order of precedence.

The nominal entity can be regarded as a special case of the variational class where the differential matrix has no variations. In this research, a toleranced part can also be regarded as a variational part because the interval numbers of differential matrices given to the variational part can represent allowable ranges of tolerances.

3.4 Tolerance Propagation

The relationship of the measuring planes to the datum planes is illustrated in Fig. 2. Fig. 2 shows how to fit datum reference frame for a 2 dimensional example. In this case, the primary datum is fitted to the actual feature, then the secondary datum is fitted next according to the order of precedece of datums. This means that the primary datum constrains degrees of freedom, translation along Y direction and rotation, and the secondary datum does the remainder of degrees of freedom, translation along X direction. If the secondary datum was fitted earlier than the primary datum, then the secondary datum would constrain translation along X axis and rotation, and the primary datum would do only translation along Y axis. This contradicts the order of precedence. Therefore, in the case that the primary datum and the secondary datum can constrain the same degree of freedom, the primary datum should constrain this degree of freedom.

These influences can be explained by the propagation of differential matrices to specify the degrees of freedom of the datums as follows

$$\Delta_R(d_X, d_Y, \delta) = \Delta_R(d_Y, \delta) + {}^B\Delta_R(d_X) \tag{14}$$

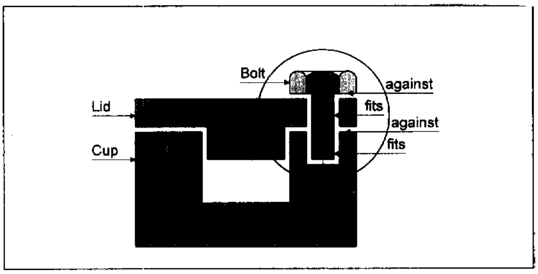


Fig. 3. Example of assembly: This figure shows an assembly *container* where *against* is a mate condition for planes and *fits* is a mate condition for cylimders.

where Δ is a differential matrix, subsctipt R indicates the datum reference frame and superscript A, B indicate the primary, the secondary datum planes respectively according to the order of precedence.

The rule of the prevoius illustration can be ecplained as follows

In the case that two or more datums can control the same degree of freedom, this degree is constrained by the datum which have the highest order of precedence among them.

4. Assemblability

Let us see Fig. 3 and 4 to explain how the assemblability is verified in this work. Fig. 3 shows an assembly of nominal parts, *container* composed of *lid, cup* and *bolt*. Fig. 4 shows the assembly of variational parts which have some variations due to shape errors or tolerances. Our assemblability analysis starts from the initial region shown in the left-hand side of Fig. 4. The origin of the coordinate frame attached to the bolt is expected to move within this region. Since this initial region is arbitrary provided by the user it includes the portions causing collision of the bolt with other parts in the assembly. If these portions causing the collision are removed somehow and the region is obtained in a form of 'T' as in Fig. 4 (a), the state can be said to be assemblable. If the resulting region is separated as in Fig. 4 (b), it implies that the assembly path of the bolt is not continuous and the bolt cannot be assembled to its final assembled state. In Fig. 4 (c), the

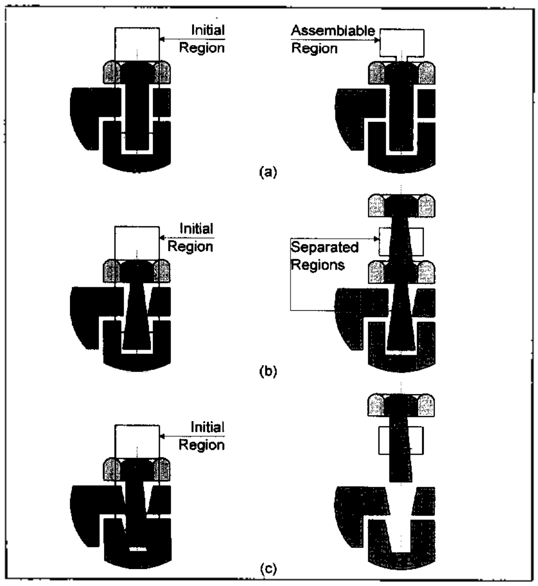


Fig. 4. Assemblability checking: If the parts of an assembly have variations, assemblability is classified by three cases as shown in (a) assemblable, (b) non-continuous assembling path and (c) no assemblabling position.

resulting region does not include the origin of the coordinate frame of the bolt at its final assembled state and it can be concluded that the bolt collides with other parts at the assembled state. Therefore, the assemblability can be verified by analyzing these regions derived from the initial region by eliminating regions causing interference.

4.1 Contact State

As explained earlier, the assemblability verification in this work resort to the identification of the position of parts causing interference. Thus we need to determine whether a part at a given position interferes other parts or not. We call this the contact state of the part.

To define contact states, two dimensional variational polygons are illustrated in Fig. 5. The gray region is the allowable range of variation within which the boundary of a polygon can move. The circle at the upper left corner can move freely without causing interference with the concave polygon and thus this state is defined as *Free*. However, the rectangle has an interference with the concave polygon even when it has the minimum boundary. Therefore, this state is de-

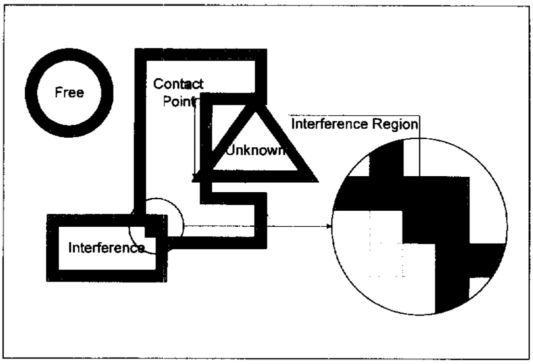


Fig. 5. The definition of contact states: where *Free* denotes no contact point, *Unknown* denotes a contact within the allowable ranges of variations, *Interference* denotes a interference between the inside regions of mate features.

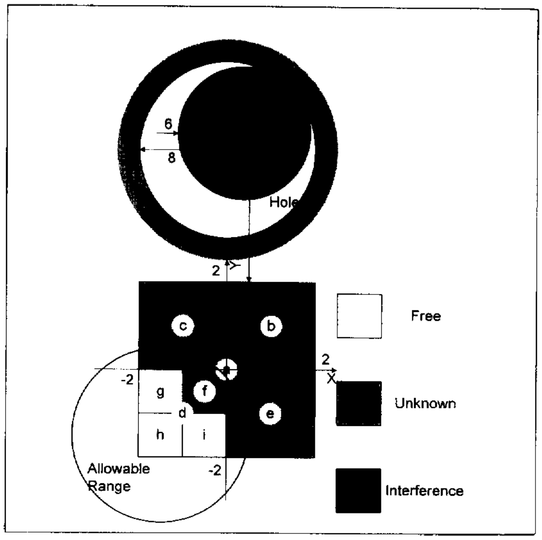


Fig. 6. An example for interval analysis: This figure shows the process to find the region in no interference between the peg and the hole. The bottom figure is the detailed picture of the center position of the peg.

fined as *Interference*. On the other hand, the triangle has no interference region with the concave polygon at its minimum boundary but has interference at its maximum boundary. If these polygons move within the given variations, the contact state can be either Free or Interference. Let us define this as *Unknown*.

These maximum and minimum boundaries can be regarded as MMC (Maximum Material Condition) and LMC (Least Material Codition) which are the fundamental and most important principles of geometric dimensioning and tolerancing.

4.2 Minimum Distance

A variational component may contact with the mating part in an assembly. A contact state, where and how many contacts happen, can be decided by the minimum distance between the two variational parts. The distance is defined as positive when its direction is in the outer normal of parts. Since each variational part has the maximum and the minmum boundary, the distance between the parts has the minimum and the maximum, and can be represented by an interval number, i.e. $[dista_i, distb_i]$. The minimum distance between the parts is also derived to be an interval number as below.

$$[mindista,\ minndistb] = [\min\ (dista_i),\ \min\ (distb_i)],\quad i = 1, 2, \ldots n \qquad (15)$$

where $[dista_i, distb_i]$ is i-th interval of distance of interest, n is number of these intervals between two parts, *mindista* is lower bound of minium distace and *mindistb* is upper bound.

If the minimum distance is positive, i.e. its lower bound is also positive, the two parts have no interference region (it means Free state) and the parts are as far apart as this minimum distance at least. If the minimum distance is negative, i.e. its upper bound is also negative, the component have some interference regions (it means Interference state). If the range of the minimum distance includes the zero, the contacts state is Unknown.

4.3 Interval Analysis

An interval analysis gets the solutions by subdividing intervals through the following three steps. These steps are explained with a two dimensional example in which the assemblability of a peg and a hole is verified based on the contact state. As shown in Fig. 6, the goal of the interval analysis is to get the circular allowable region in which the peg can move without interfering the hole.

First Step. Initialize

The initial value decides the limit region within

```
analysis (region)
1: if (Free) add region to solution
2: else if (Interference) discard region
3: else if (Small) add region to intermediate solution
4: else if (Unknown) {
5: subdivide region into subregion_i
6: analysis (subregion_i), i=1, 2...n
7: }
```

Fig. 7. The recursive algorithm of interval analysis for assemblability checking.

which the solutions should exist because the solutions ate derived by subdividing the initial region recursively. The largest rectangular region denoted by a in Fig. 6 is an initial region specified by the designer. Region a can be represented by a transformation matrix having two variables ($d_X = [-2, 2]$, $d_Y = [-2, 2]$) which represent the translations along X and Y axis. Region a is subdivided into four regions such as b, c, d and e by subdividing each of the two variables into two respectively. Region d is again subdivided into f, g, h and i, and regions c and e are also subdivided into four respectively. It means that initial region with two variables is subdivided into quadtree, three variables into octree, m variables into 2^m tree and so on.

Second Step. Subdivide Recursively

Fig. 7 shows the recursive algorithm to explain the interval analysis and Fig. 8 sows how the boundary of the peg varies for each subdivided region for the example in Fig. 6.

The first line of the algorithm in Fig. 7 explains how to treat Free region. For example, regions g, h and i in Fig. 6 belong to this category because the peg centered within these regions can move without interfering the hole boundary. Therefore, the set of these regions is the allowable range of peg locations and thus the solution we are looking for.

The second line explains that Interference region is not subdivided any more because its subdivided regions will also belong to Interference category. As shown in Fig. 8, region b is classified as Interference category.

The third line is a terminal condition of the algorithm. If the width of region is small enough, this region is not subdivided any more. Therefore, this region is classified as intermediate solution because its contact state is not classified as either Free or Interference. Actually, this means Unknown. In the

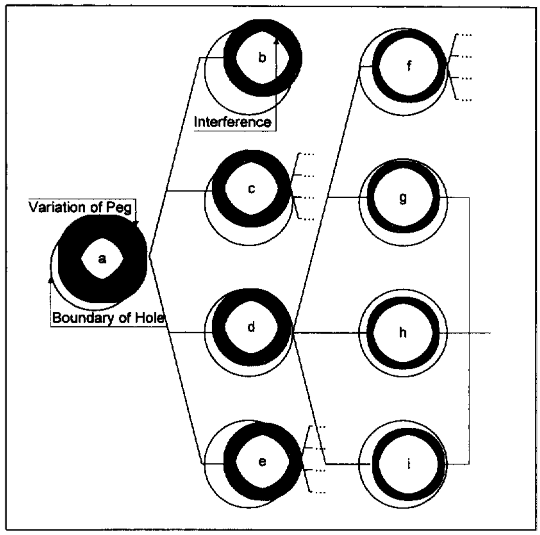


Fig. 8. The subdivision process of interval analysis: The position of a peg subdivided as quad trees as shown in Fig. 6 because the initial region has two variables along X and Y axis.

ideal case of treating nominal parts only, the intermediate solutions will not exist of the subdivision process continues to infinitely small regions. However, with the variation such as tolerance zone, these Unknown regions will always exist.

The fourth line classifies the region as Unknown if it is not classified as anyone of the above. Regions c, d, e and f of Fig. 8 belong to this category. After subdividing Unknown region (the fifth line), repeat the processes explained above (the sixth line) until the terminal condition is satisfied.

Third Step. Aggregate

The solutions obtained in the previous step are a set of regions. The solutions can be aggregated to result one solution or to reduce the number of solutions. The aggregated solution is an approximate solution which contains all the solutions and some wrong solutions. In Fig. 6, the solutions are g, h and i, and the aggregated solution of them cam be d. Region d contains not only the solutions g, h and i but also wrong solution f. This step should be omitted to get more exact solutions.

4.4 Assembly of Several Parts

The relative position between two parts can be constrained by the region which have six variables to constrain six degrees of freedom. One region of six variables is subdivided into 2^6 regions through the subdivision process where the region of each variable subdivided into two regions.

The assembly of several parts can be represented by their relative positions with reference to one fixed part. For n parts, n-1 relative positions should be specified and each relative position has six variables. So, n-1 regions with $6\times(n\text{-}1)$ variables are subdivided into $(2^6)^{n-1}$ regions. It means that computation time to progress subdivision increases exponentially in proportion to the number of parts. This problem will be solved as the performance of computer improves.

5. Conclusions

The proposed tolerancing method using the differential matrix, simply called *differential tolerancing*, is compatible with the current international standard. The effect caused by the variations of the datum can also be handled by summing the differential matrices while considering datum's order of precedence. Differential tolerancing can be applied to the problems such as three dimensional tolerance analysis by using the interval numbers for the elements of a differential matrix.

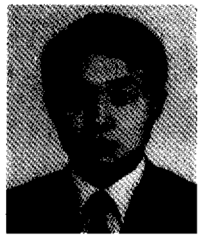
Using the mathematical representation of the tolerance, a method is presented for verifying the assemblability between toleranced parts by deriving the allowable ranges of relative motion. The allowable range of relative motion is derived by assuming that the nominal parts are in an assembled state and the tolerance are small enough compared to the size of the parts. The continuity of the assembling path and assemblability can be inferred by analyzing these ranges. This method can be applied to simultaneous assemblability checking among several parts, they may be the toleranced parts, the actual part with shape error caused by manufacturing process, or the nominal parts. If the assemblability of the parts can be verified in the design stage, the designer will be able to design the parts of an assembly considering their assemnblability in advance.

The following problems are left for further study. First, our method verifies the assemblability between two parts to be assembled by a 'clearcnce fit'. This implies that the parts are identified to be not assemblable even when they are originally designed to be assembled by a 'press fit'. This may be solved by allowing the negative minimum distance. Second, the computation time in this method increases exponentially as the number of parts and the maximum depth of subdivision increase. To overcome this problem, the current algorithm needs to be improved for efficiency. Finally, the equation to represent geometric tolerances was derived by linearizing the rotational terms assuming manufacturing variations of parts are small enough compared to the nominal geometry. For a big variation such as an infinite translation or a 360 degree rotation, the complete transformation matrix should be used instead of the differential matrix. To eliminate the assumption of smal variation, the equation should be derived in a more complex way and the algorithm may need more information.

References

1. Requicha, A.A. G., "Toward a theory of geometric tolerance," *International Journal of Robotic Research*, Vol 2, No. 4, pp. 45-60, 1983.
2. Etesami, F., "Tolerance verification through manufactured part modeling," *Jouranl of Manufacturing System*, Vol. 7, No. 3, pp. 223-232, 1988.
3. Wirtz, A., *Vectorial tolerancing for quality control and functional analysis in design*, CIRP International Working Seminar on Computer Aided Tolerancing, Penn State University, May 1991.
4. Turner, J.U., "A feasibility space approach for automated tolerancing," *Journal of Engineering for Industry*, Vol. 115, pp. 341-346, August 1993.
5. Lee, K. and Gossard, D.C., "A hierarchical data structure for representing assemblies: Part1," *Computer Aided Design*, Vol 17, No. 1, pp. 15-19, 1985.
6. Lee, K. and Andrews, G., "Inference of the positions of components in an assembly: Part2," *Computer Aided Design*, Vol, 17, No. 1, pp. 20-24, 1985.
7. Rocheleau, D.N. and Lee, K., "System for interactive assembly modelling," *Computer Aided Design*, Vol. 19, No. 2, pp. 65-72, 1987.
8. Kim, S.H. and Lee, K., "An assembly modelling system for dyamic and kinematic analysis," *Computer Aided Design*, Vol. 21, No. 1, pp. 2-12, 1989.

9. Fleming, A.D., *Analysis of Incertainties annd Geometric Tolerances in Assemblies of parts*, PhD thesis, University of Edinburgh, 1987.
10. Srinivasan, V. and Jayaraman, R., "Geometric tolerancing: II. conditional tolerances," *IBM Journal of Research and Development*, Vol. 33, No. 2, pp. 105-124, 1989.
11. Jayaraman, R. and Srinivasan, V., "Geometric tolerancing: I. virtual boundary requirements," *IBM Journal of Research and Development*, Vol. 33, No. 2, pp. 90-104, 1989.
12. Turner, J.U., "Relative positioning of parts in assemblies using matematical programming," *Computer Aided Design*, Vol. 22, No. 7, pp. 394-400, 1990.
13. Inui, M. and Kimura, F., "Algebraic reasoning of position uncertainties of parts in an assembly," *Solid Modeling Foundations and CAD/CAM Applications, pp. 419-428.*
14. Moore, R.E., *Interval Analysis*, Prentice-Hall, 1966.
15. Moore, R.E., *Methods and Applications of Interval Analysis*, Society for Industrial and Applied Mathematics, Philadelphia, 1979.
16. Paul, R.P., *Robot Manipulators: Mathematics, Programming, and Control*, The MIT Press, 1981.



박 상 호

1988년 서울대학교 기계설계학과 학사
1990년 서울대학교 기계설계학과 석사
1995년 서울대학교 기계설계학과 박사
1995년~현재 시스템 공학 연구소 선임 연구원
관심분야 : Solid modeling, Assembly modeling, Tolerancing, Computer Graphics

이 건 우

1978년 서울대학교 기계공학과 학사
1981년 미국 MIT 공학석사
1984년 미국 MIT 공학박사
1984년~1986년 Univ. of Illinois at Urbanachampaign 조교수
1986년~현재 서울대학교 기계설계학과 교수
관심분야 : Computer aided geometric design, NC tool path generation and verification, Solid modeling, Rapid prototyping