mma 2012 ADVANCED PRODUCTION TECHNOLOGIES

11th INTERNATIONAL SCIENTIFIC CONFERENCE
NOVI SAD, SERBIA, SEPTEMBER 20-21, 2012

**Jakovljevic, Z.**

# POINT CLOUD REDUCTION USING SUPPORT VECTOR MACHINES

***Abstract:*** *This paper explores the possibilities of point cloud reduction using ε insensitive support vector regression (ε-SVR). ε -SVR is a technique that can carry out the regression using different kernel functions (sigmoid, radial basis function, B-spline, spline, etc.) and it is suitable for detection of flat regions and regions with high curvature in scanned data. Using ε -SVR the density of preserved points is adaptive – preserved points are denser at highly curved region and rare at flat regions. Adjusting the error cost in the regression, the number of preserved points can be fine tuned.*

***Key words:*** *Reverse engineering, point cloud reduction, support vector machines*

## 1. INTRODUCTION

The application of reverse engineering (RE) of the freeform shaped parts is rapidly dispersing over the years. Besides the reproduction of parts when original drawings are not available, RE is applied in the design of new products (e.g., in automotive industry where the sheet metal forming tools for car bodies are created based on wooden or clay models; in consumer products industry where aesthetic design is important; in generation of custom made accessories and prostheses for human) [1].

During the first step of RE, the surface of the physical object is digitalized and 3D point cloud is obtained. Contemporary measurement devices [2] and especially ones based on lasers have high measurement speed and resolution, giving large and dense point clouds at output. Although the sampling rate of measurement device can be adjusted according to the character of digitalized surface, the operator usually acquires as many points as possible because he is not sure about the needed density of points for adequate reconstruction of certain parts of scanned surface. Generally, a significantly larger amount of point cloud data is acquired than one that is sufficient and that can be efficiently handled during surface reconstruction. In order to operate with reconstructed surfaces at reasonable computational cost, the amount of point data should be reduced.

The easiest solution for data reduction is uniform downsampling. Nevertheless, in order to preserve the shape of original surface points, highly curved regions in point cloud should have high density, while for relatively flatter areas lover point density is acceptable. In order to address given issues a number of data point reduction techniques have been proposed.

The simplest way to create a surface model from point cloud is to generate a polygonal mesh over it. Consequently, the first methods for data point reduction were based on simplification of polygonal meshes [3] (the research has origins in image processing), while recent methods are based on direct cloud point data reduction. The most of the methods for direct cloud point data reduction are based on the estimation of the importance of each point in the cloud. For the importance evaluation different measures are used e.g., deviation of normal vectors in the vicinity of the point [4, 5], Hausdoff distance [6], and maximum deviation distance [7]. In order to improve decision making fuzzy logic has been employed [7].

Support vector machines are an emerging technique for data regression and classification. In this paper a possibility of ε insensitive support vector regression (ε-SVR) application in data point reduction is explored, and a method for point cloud reduction is proposed.

## 2. SUPPORT VECTOR REGRESSION

Given is the training data set $\{(\mathbf{x_1}, y_1), (\mathbf{x_2}, y_2) \ldots (\mathbf{x_l}, y_l)\}$ where $\mathbf{x_i}$ represent independent, and $y_i$ dependant variables. The goal of ε-SVR is to find a function f($\mathbf{x}$) that is *as flat as* possible and that *has maximum ε deviation* from $y_i$. The errors lower than ε are insignificant. In other words, all the $y_i$ should lie in the ε-tube around $f(\mathbf{x})$. In the case of linear dependence, the function $f(\mathbf{x})$ is in the form:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x_i} \rangle + b \qquad (1)$$

$f(\mathbf{x})$ is flat if $\mathbf{w}$ is small. In order to ensure the flatness:

$$\frac{1}{2}\|\mathbf{w}\|^2 \qquad (2)$$

should be minimized, subject to:

$$\left| y_i - \langle \mathbf{w}, \mathbf{x_i} \rangle - b \right| \le \varepsilon \qquad (3)$$

Nevertheless, in reality the scenarios in which all of the data lie within ε tube are extremely rare and optimization of the problem (2, 3) is infeasible. In order to create $f(\mathbf{x})$, anyway, the violation of the condition that all $y_i$ are within ε tube is allowed. To formalize this approach, slack variables $\xi_i$, $\xi_i^*$ are introduced and optimization problem (2, 3) is reformulated [8]:

$$\begin{aligned} &\textit{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right) \\ &\textit{subject to} \quad \begin{cases} y_i - \langle \mathbf{w}, \mathbf{x_i} \rangle - b \le \varepsilon + \xi_i \\ \langle \mathbf{w}, \mathbf{x_i} \rangle + b - y_i \le \varepsilon + \xi_i^* \end{cases} \end{aligned} \qquad (4)$$

Constant $C$ introduces the tradeoff between function flatness and number of points out of ε tube.

Optimization problem (4) can be represented in dual form [9]:

$$\begin{aligned} &maximize \quad \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{l}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\langle \mathbf{x_i},\mathbf{x}_j\rangle \\ -\varepsilon\sum_{i=1}^{l}\left(\alpha_i-\alpha_i^*\right)+\sum_{i=1}^{l}y_i\left(\alpha_i-\alpha_i^*\right) \end{cases} \quad (5) \\ &subject\ to \quad \sum_{i=1}^{l}\left(\alpha_i-\alpha_i^*\right)=0 \ and \ \alpha_i,\alpha_i^*\in[0,C] \end{aligned}$$

where $\alpha_i$, $\alpha_i^*$ represent Lagrange multipliers. Only for $|f(\mathbf{x_i})-y_i|\geq\varepsilon$ Lagrange multipliers are nonzero, while for vectors (points) inside ε tube $\alpha_i$, $\alpha_i^*$ vanish. The vectors with nonzero $\alpha_i$, $\alpha_i^*$ are called *support vectors*.

The solution of the problem (5) is given by:

$$\mathbf{w}=\sum_{ns}\left(\alpha_i-\alpha_i^*\right)\mathbf{x_i} \qquad (6)$$

where *ns* is the number of support vectors, leading to:

$$f(x)=\sum_{ns}\left(\alpha_i-\alpha_i^*\right)\langle \mathbf{x_i},\mathbf{x}\rangle+b \qquad (7)$$

The presented methodology can be applied for nonlinear regression by mapping data from the input space into a high-dimensional space where the regression is linear. It is worth noting that for optimization problem (5), it is enough to know only the inner product in the high-dimensional space i.e. it is not necessary to define the high-dimensional space in explicit form. Rather opposite, it can be defined using kernel $K(\mathbf{x},\mathbf{x}_i)$, which represents inner product in the space of higher dimension. Introducing $K(\mathbf{x},\mathbf{x}_i)$, problem (5) becomes:

$$\begin{aligned} &maximize \quad \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{l}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)K\left(\mathbf{x_i},\mathbf{x}_j\right) \\ -\varepsilon\sum_{i=1}^{l}\left(\alpha_i-\alpha_i^*\right)+\sum_{i=1}^{l}y_i\left(\alpha_i-\alpha_i^*\right) \end{cases} \quad (8) \\ &subject\ to \quad \sum_{i=1}^{l}\left(\alpha_i-\alpha_i^*\right)=0 \ and \ \alpha_i,\alpha_i^*\in[0,C] \end{aligned}$$

while the function $f(x)$ is defined by:

$$f(x)=\sum_{ns}\left(\alpha_i-\alpha_i^*\right)K\left(\mathbf{x_i},\mathbf{x}\right)+b \qquad (9)$$

and it is a hyperplane in the high-dimensional space.

The kernel can be any function that satisfies the conditions of Mercer's theorem [8]. For example, these are polynomial kernels, Gauss kernel, sigmoid kernel, some wavelets. New kernels can be defined by summing or multiplication of simpler kernels.

For the application at hand, two kernels are of the significance. The first is the B-spline of order $2n + 1$, defined by $2n + 1$ convolution of unit interval:

$$\begin{aligned} &K\left(\mathbf{x},\mathbf{x_i}\right)=B_{2n+1}\left(\left\|\mathbf{x}-\mathbf{x_i}\right\|\right) \\ &with\ B_k=\bigotimes_{i=1}^{k}\mathbf{1}_{\left[-\frac{1}{2},\frac{1}{2}\right]} \end{aligned} \qquad (10)$$

where $\mathbf{1}_X$ denotes indicator function on the set $X$ and ⊗ is the convolution.

The second is the spline kernel of order $k$ having $N$ knots located at $t_s$, which is defined by:

$$K\left(\mathbf{x},\mathbf{x_i}\right)=\sum_{r=0}^{k}\mathbf{x}^r\mathbf{x}_\mathbf{j}^r+\sum_{s=1}^{N}\left(\mathbf{x}-t_s\right)_+^k\left(\mathbf{x_j}-t_s\right)_+^k$$

## 3. APPLICATION OF ε-SVR IN DATA POINT REDUCTION

Point data cloud is usually, due to the nature of scanning process, structured into cross sectional curves. Otherwise, it can be restructured using projections on cross section planes. ε-SVR can be applied to each cross section and the function $f(\mathbf{x})$ can be obtained.

The fact that function $f(\mathbf{x})$ is as flat as possible in high dimensional space, i.e. it conforms as much as possible to selected kernel in initial space, can be used for determination of regions where the scanned line is not highly curved. In these regions the number of support vectors will be very small. In order to preserve the curvature, the points can be uniformly downsampled with predefined step.

On the other hand, in regions with high curvature, ε-SVR will not be able to fit all the points inside the ε tube and the number of support vectors will be higher. In these areas support vectors can be preserved points.

Due to the unknown curvature, the best kernels for ε-SVR of freeform surface scans are B-spline and spline. In this paper B-spline is opted to use. The other two parameters that should be set in order to carry out ε-SVR are error margin ε and error cost $C$.

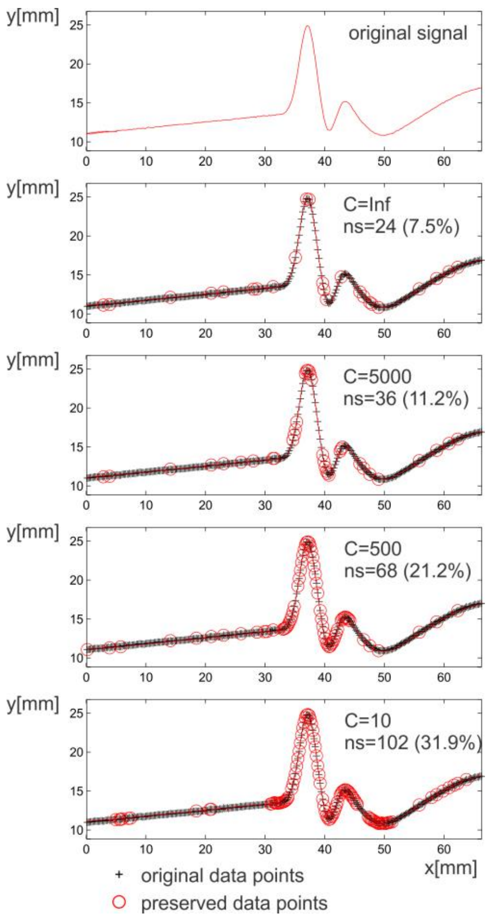


Fig. 1. Identified support vectors on synthesized signals for different values of parameter $C$ (ε=0.3)

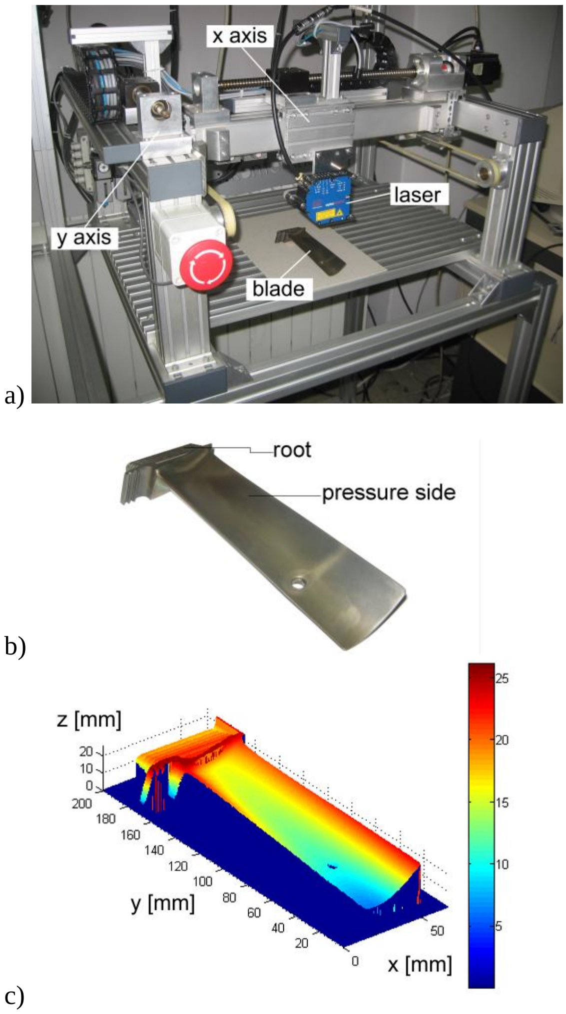


Fig. 2. a) Scanning device; b) A photo of scanned gas turbine blade; c) Obtained point cloud

The value of ε is related to the accuracy of scanning device and surface characteristics' tolerance and can be easily defined. The parameter *C* introduces the tradeoff between flatness and number of support vectors, i.e. preserved points, and it can be used for tuning the number of points that will be preserved in highly curved areas. The lower *C* will lead to higher number of preserved points, and vice versa.

Figure 1. shows an example of the ε-SVR carried out with different values of *C* on a curve synthesized in Matlab. In order to get closer to the reality the curve is noised with 20dB of white noise. It can be observed that with the decrease of *C* the curvature of regression line is lower and the number of preserved points (support vectors) in highly curved regions is higher.

## 4. TURBINE BLADE EXAMPLE

This Section considers a real world example of the gas turbine blade (Figure 2b). The pressure side of the blade represents smooth freeform surface, while its root has high curvature. The surface on the pressure side is scanned using set-up shown in Figure 2a. The scanning device – laser με OptoNCDT1700-100 is put on the 2d Cartesian manipulator. Laser measuring range is 100mm with 14 bit resolution. Measurement error due to the tilt angles is 0.5% at ±30°. The accuracy of manipulator is significantly lower: ±0.1mm. The surface is scanned along *y* axis in successive cross sections with the step of 0.2mm. Scanning speed was 100mm/s, and sampling rate 625Hz, which gives resolution of 0.16mm. The obtained point cloud has 417,500 points and it is presented in Figure 2c.

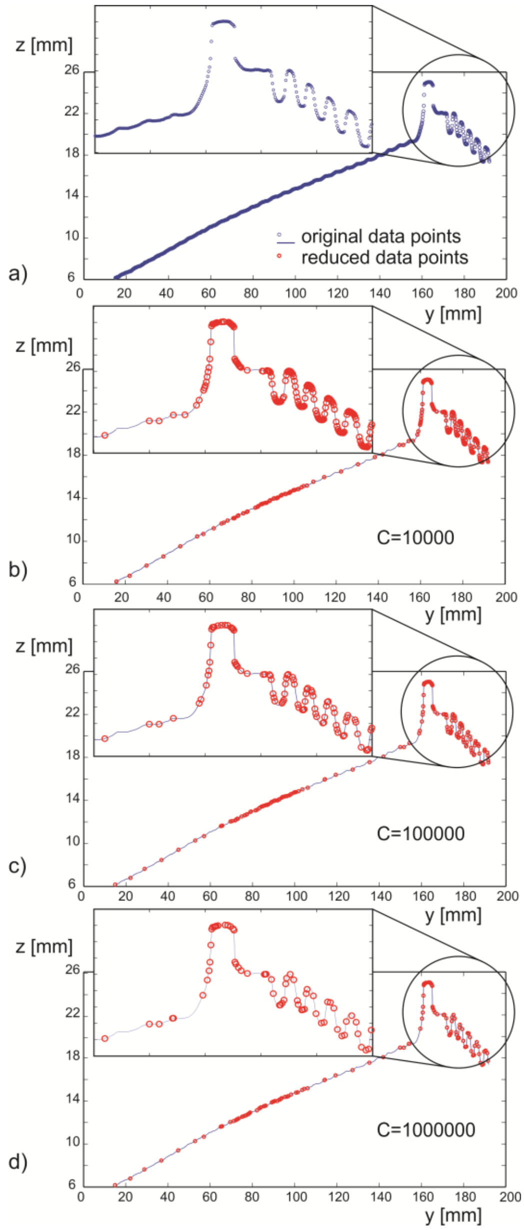


Fig. 3. a) Original points in cross section at *x*=28mm; b-d) Preserved points in the same cross section

Points are structured into cross sections along which the scanning is performed. Each of the obtained curves is subjected to ε-SVR as previously described. Parameter ε is set to 0.1mm, in accordance with scanning device accuracy. Parameter *C*, on the other hand is varied (Table 1).

ε-SVR gives support vectors that represent the points that should be preserved. Nevertheless in smooth areas support vectors are infrequent. Thus, in regions where the distance between two subsequent support vectors along abscissa was lower than 8mm the original signal was uniformly downsampled by 8mm (50 samples). In highly curved regions support vectors are dense and only they are preserved.

Point data scanned in one typical cross section at *x*=28mm are shown in Figure 3a. Applying ε-SVR together with uniform downsampling where needed, the number of points is adaptively reduced. The points on the pressure side surface are reduced by higher rate, while at highly curved area at blade root the number of preserved points is higher. Reduced point data for *C*=10,000, C=100,000 and *C*=1,000,000 are shown in Figure 3b, 3c and 3d, respectively.

The points in *x* direction are downsampled uniformly by 10 – the cross sections with the step of 2mm are taken. The number and percentage of preserved points for different values of the cost parameter *C* are shown in Table 1, while reduced point clouds together with polynomial meshes created in Catia are shown in Figure 4.

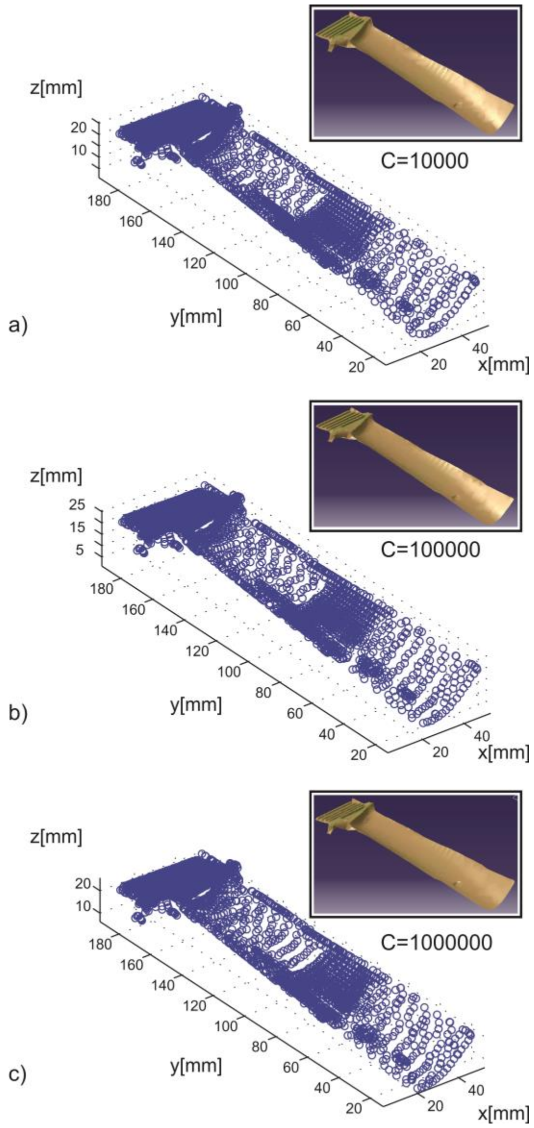


Fig. 4. Reduced point clouds and polygonal meshes

| C | # of preserved points | % of preserved points - *a*[*] | % of preserved points - *b*[**] |
|---|---|---|---|
| 10,000 | 5185 | 1.24 | 23.05 |
| 100,000 | 3351 | 0.80 | 14.9 |
| 1,000,000 | 3020 | 0.72 | 13.43 |

Table 1. Number and percentage of preserved points

[*] Initial number of points in cloud: *a* = 417500

[**] Number of points after downsampling along x and excluding points with z=0: *b*=22491

## 5. CONCLUSION

This paper proposed a method for adaptive data point reduction based on ε-SVR. It has been shown that SVR represents a tool that can be effectively used for higher reduction of data points at flat and lower reduction at highly curved areas. Cost parameter *C* is suitable for fine tuning of the number of preserved points. The B-spline kernel was selected as suitable for regression of freeform curves. Nevertheless, B-spline is prone to oscillation at smooth areas as can be observed in Fig. 3b-3d (*y*=70-100mm). In order to address this shortcoming the use of combination of spline and B-spline kernel could be explored.

The main shortcoming of the proposed method is the computation cost of SVR optimization problem, which is very high even with the application of sequential minimization algorithm.

## 6. REFERENCES


[1] Varady, T., Martin, R., R., Cox, J., *Reverse engineering of geometric models – an introduction*, Computer-Aided Design, 29, p.p. 255-268, 1997

[2] Savio, E., De Chiffre, L., Schmitt, R., *Metrology of freeform shaped parts*, Annals of the CIRP, 56/2, p.p. 810-834, 2007

[3] Cignoni, P.,Montani, C., Scopigno, R., *A comparison of mesh simplification algorithms*, Computers & Graphics, 22, p.p. 37-54, 1998

[4] Song, H., Feng, H., Y., *A progressive point cloud simplification algorithm with preserved sharp edge data*, International Journal of Advanced Manufacturing Technology, 45, p.p. 583–592, 2009

[5] Lee, K. H., Woo, H. Suk, T., *Point Data Reduction Using 3D Grids*, International Journal of Advanced Manufacturing Technology, 18, p.p. 201–210, 2001

[6] Ma, X., Cripps, R., J., *Shape preserving data reduction for 3D surface points*, Computer-Aided Design, 43, p.p. 902–909, 2011

[7] Budak, I., Sokovic, M., Barisic, B., *Accuracy improvement of point data reduction with sampling-based methods by Fuzzy logic-based decision-making*, Measurement, 44, p.p. 1188–1200, 2011

[8] Vapnik VN, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 2000

[9] Smola, A., J., Schoelkopf, B., *A tutorial on support vector regression*, Statistics and Computing, 14, p.p. 199-222, 2004



***Author:*** **Doc. Dr. Zivana Jakovljevic**, University of Belgrade, Faculty of Mechanical Engineering, Department for Production Engineering, Kraljice Marije 16, 11000 Beograd, Serbia, Phone.: +381 11 3302-264, Fax: +381 11 3370-364.
E-mail: zjakovljevic@mas.bg.ac.rs