SIGNAL REPRESENTATION BASED ON A GAUSSIAN DECOMPOSITION

Ardeshir Goshtasby and Dan Schonfeld

Signal and Image Research Laboratory
Department of Electrical Engineering and Computer Science
The University of Illinois at Chicago
Chicago, Illinois 60680-4348

Abstract - In this paper, we study the representation of signals based on a finite Gaussian decomposition basis. An iterative optimization scheme is developed for the solution of the optimal Gaussian representation. The proposed scheme is based on the Marquardt algorithm for the solution of nonlinear least-squares estimation problems. The Gaussian decomposition is also utilized in the representation of image contours. Finally, an extension of the Gaussian decomposition to the representation of images is presented. Consequently, we demonstrate that the Gaussian decomposition results in a highly efficient method for the representation of signals.

I. Introduction

Many signal processing applications require an efficient signal representation. The efficiency of the representation is determined by the number of parameters needed to accurately approximate a signal. Representations resulting in fewer parameters often yield superior signal processing algorithms. Various signal representations may be formed by selection of different decomposition bases. A particularly interesting decomposition basis is formed by the Gaussian decomposition basis. The signal representation corresponding to the Gaussian decomposition basis shall be referred to as the Gaussian representation. Intuitively, the main advantage of the Gaussian decomposition basis is in resulting in optimal simultaneous spatial and frequency resolutions [1], [2]; thus, the Gaussian decomposition basis captures abrupt spatial and frequency variations in the signal. However, it is important to notice that, due to the nonorthogonality of the Gaussian decomposition basis, deriving the optimal Gaussian representation may be a very difficult task.

Proceedings of the 1991 Conference on Information Sciences and Systems, The Johns Hopkins University, March 20-22, Baltimore, Maryland.

In this paper, we study the representation of signals based on a finite Gaussian decomposition basis. In Section II, an iterative optimization scheme is developed for the solution of the optimal Gaussian representation. The proposed scheme is based on the Marquardt algorithm for the solution of nonlinear least-squares estimation problems. Initial parameter values for the Gaussian representation of a signal are estimated from the scale-space analysis of the signal. In Section III, the Gaussian decomposition is also utilized in the representation of image contours. In Section IV, an extension of the Gaussian decomposition to the representation of images is presented. Finally, in Section V, we conclude with a summary and discussion of our results. Several computer simulation experiments are used to demonstrate the efficiency of the Gaussian decomposition in the representation of signals.

II. Signal Representation

In the following section we propose a scheme for the representation of signals based on a finite Gaussian decomposition basis.

A. Gaussian Decomposition

Let us consider a signal $x(t)$. The finite Gaussian decomposition basis $b_i(t)$, for $i = 1, 2, \ldots, n$, is defined by

$$b_i(t) = e^{-(t-t_i)^2/2\sigma_i^2}, \qquad (1)$$

for $i = 1, 2, \ldots, n$. Let $\mathbf{p}$ denote the Gaussian representation parameters given by

$$\mathbf{p} = [(x_i \;\; t_i \;\; \sigma_i), i = 1, 2, \ldots, n]. \qquad (2)$$

Finally, the Gaussian approximation signal $\hat{x}_{\mathbf{p}}(t)$ is defined by

$$\hat{x}_{\mathbf{p}}(t) = \sum_{i=1}^{n} x_i b_i(t). \qquad (3)$$

Observe that the choice of the finite Gaussian decomposition basis will be critical in minimizing the approximation error. In the following section we develop a procedure to determine the optimal choice of the finite Gaussian decomposition basis.

B. Optimal Gaussian Representation

Let us denote a collection of N samples of the original signal $x(t)$ by $\mathbf{X} = [(x(t_k)), k = 1, 2, \ldots, N]$. The optimal Gaussian representation is obtained by minimizing the sum of the squared error between samples of the original signal and the Gaussian approximation signal with respect to the Gaussian representation parameters; i.e.,

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} E, \quad (4a)$$

where

$$E = \sum_{k=1}^{N} [x(t_k) - \hat{x}_{\mathbf{p}}(t_k)]^2. \quad (4b)$$

The solution of eq. (4) may be derived by solving the system of nonlinear equations obtained by setting $\partial E / \partial p_l = 0$, for $l = 1, 2, \ldots, 3n$, for the $3n$ unknown Gaussian representation parameters $\mathbf{p}$. Several approaches to this problem will be discussed next.

Various methods have been proposed for the solution of least-squares estimation of nonlinear parameters. The steepest-descent method [3] is a commonly used approach to the solution of nonlinear minimization problems. The method relies on an iterative procedure and updates the Gaussian representation parameters $\mathbf{p}$ by

$$\nabla \mathbf{E} = -[(\frac{\partial E}{\partial p_l}), \ l = 1, 2, \ldots, 3n]. \quad (5)$$

The steepest-descent method guarantees convergence, yet it is very slow [4].

An alternative approach is obtained by the modified Gauss-Newton method [5]. This method relies on an iterative procedure where a linearization of the Gaussian approximation signal is obtained from the Taylor series expansion; i.e.,

$$\hat{\mathbf{X}}_{\mathbf{p}+\mathbf{\Delta}} = \hat{\mathbf{X}}_{\mathbf{p}} + \mathbf{D}\mathbf{\Delta}, \quad (6a)$$

where

$$\hat{\mathbf{X}}_{\mathbf{p}} = [(\hat{x}_{\mathbf{p}}(t_k)), \ k = 1, 2, \ldots, N], \quad (6b)$$

and

$$\mathbf{D} = [(\frac{\partial \hat{x}_{\mathbf{p}}(t_k)}{\partial p_l}), \ k = 1, 2, \ldots, N, \ l = 1, 2, \ldots, 3n]. \quad (6c)$$

The Gaussian representation parameters $\mathbf{p}$ are updated by

$$\mathbf{\Delta} = [(\Delta_l), \ l = 1, 2, \ldots, 3n]. \quad (7)$$

Equation (4) is now a linear least-squares estimation problem. Setting $\partial E / \partial \Delta_l = 0$, for $l = 1, 2, \ldots, 3n$, we have

$$\mathbf{A}\mathbf{\Delta} = \mathbf{b}, \quad (8a)$$

where

$$\mathbf{A} = \mathbf{D}^T \mathbf{D}, \quad (8b)$$

and

$$\mathbf{b} = \mathbf{D}^T (\mathbf{X} - \hat{\mathbf{X}}_{\mathbf{p}}), \quad (8c)$$

where T is the matrix transpose operation. The modified Gauss-Newton method is a simple and fast procedure; however, it does not always converge [4].

Finally, the Marquardt algorithm relies on the steepest-descent and modified Gauss-Newton methods to obtain an iterative optimization procedure which results in a fast convergence [4]. In the Marquardt algorithm, the Gaussian representation parameters $\mathbf{p}$ are updated by

$$\mathbf{\Delta}_M = | \mathbf{\Delta} | \nabla \mathbf{E}. \quad (9)$$

Thus, the Gaussian representation parameters are updated in the direction determined by the steepest-descent method, and the size determined by the modified Gauss-Newton method.

The performance of the optimization procedures described is strongly dependent on the choice of the initial Gaussian representation parameters. Additionally, the number n of Gaussian decomposition basis functions employed in the representation will critically affect the approximation error. In the following, we use the scale-space image of the signal in order to estimate these parameters.

C. Scale-Space Image

The scale-space image of a signal is a very valuable tool in the analysis of the signal at varying resolutions [6]. The scale-space image exploits the fact that a Gaussian function has two inflection points causing its second derivative to produce a pair of zero-crossings. An example of a Gaussian function and its second derivative are depicted in Figs. 1(a) and 1(b), respectively. The arrows indicate the location of the zero-crossings. When a signal is filtered by a Gaussian filter a smoother signal is obtained. The scale-space image of a signal is formed by plotting the zero-crossings associated with the filtered signal versus the scale of the Gaussian filter [6]. We shall now investigate the relationship between the scale-space image and the Gaussian representation of a signal.

The second derivative of a signal that is formed by the summation of n Gaussian functions will have at most $2n$ zero-crossings. Figure 2 illustrates two examples where the second derivative of the summation of two Gaussian functions will have only a single pair of zero-crossings. Two spatially overlapping Gaussian functions are depicted in Fig. 2(a), and two Gaussian functions where one Gaussian function is strongly dominated by the other Gaussian function are depicted in Fig. 2(b). The second derivatives of the sum of the Gaussian functions shown in Figs. 2(a) and 2(b) are depicted in Figs. 2(c) and 2(d), respectively. The arrows indicate the location of the zero-crossings.

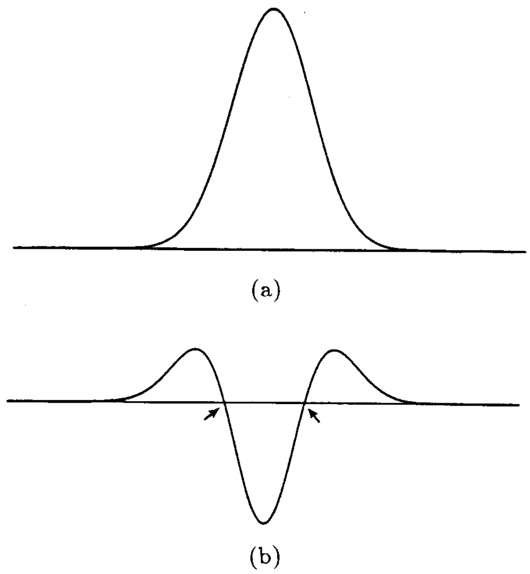

Fig. 1. (a) A Gaussian function; and, (b) its second derivative. The arrows indicate the location of the zero-crossings.

The scale-space image of a signal traces the zero-crossings of the signal at varying resolutions. As the scale of the Gaussian filter increases, the zero-crossings corresponding to a given Gaussian function will slowly move closer, merge at a particular scale, and vanish at higher scales. The scale of the Gaussian filter at which the zero-crossings corresponding to a given Gaussian function will vanish indicates the dominance of that particular Gaussian function in the original signal. The scale of the Gaussian filter may be increased up to a point where only two zero-crossings remain. These zero-crossings represent the most dominant Gaussian function in the original signal. Thus, the scale-space image contains an arch corresponding to each Gaussian function in the original signal. Figure 3 illustrates the relationship between the Gaussian functions in the original signal and its scale-space image. A collection of nine Gaussian functions are depicted in Fig. 3(a). Figure 3(b) depicts the signal formed by the summation of the Gaussian functions in Fig. 3(a). Finally, the scale-space image of the signal is depicted in Fig. 3(c). The arches in the scale-space image are ordered according to the dominance of the associated Gaussian functions.

Given an arbitrary signal we shall form the corresponding scale-space image. The sequence of pairs (t_{i_1}, t_{i_2}), for $i = 1, 2, \ldots, n$, represents the zero-crossings associated with n arches at the highest resolution (smallest scale) of the scale-space image of the signal. The parameters associated with these Gaussian

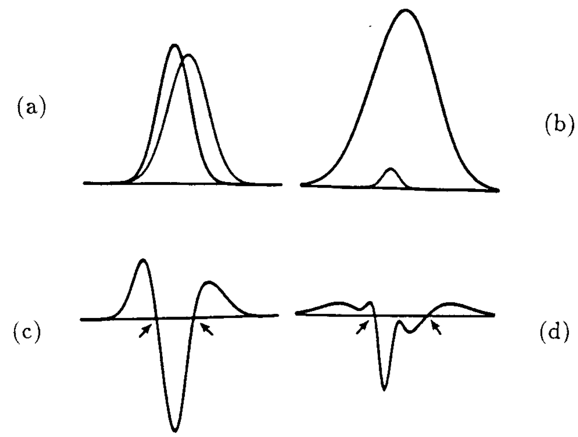

Fig. 2. (a) Two spatially overlapping Gaussian functions; and, (b) two Gaussian functions where one Gaussian function is strongly dominated by the other Gaussian function. (c)-(d) The second derivative of the sum of the Gaussian functions depicted in (a) and (b), respectively. The arrows indicate the location of the zero-crossings.

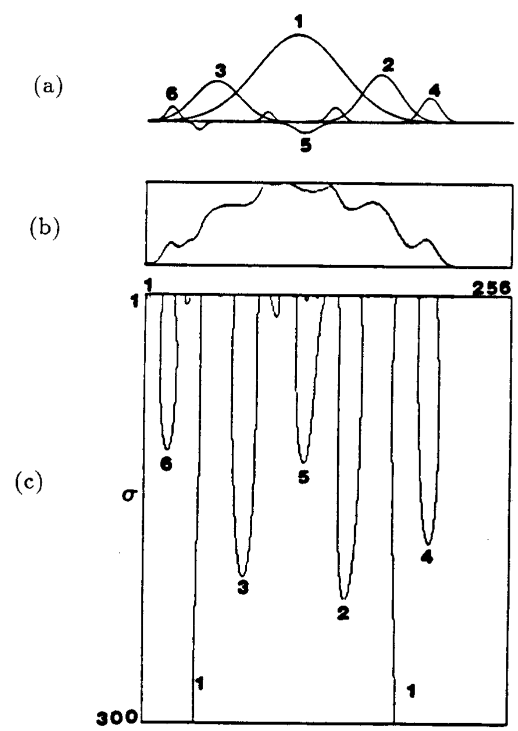

Fig. 3. (a) A collection of n Gaussian functions ($n = 9$); (b) the signal formed by the summation of the Gaussian functions in (a); and, (c) the scale-space image of the signal in (b).

functions may be estimated by

$$t_i = (t_{i_1} + t_{i_2})/2, \tag{10}$$

and

$$\sigma_i = |t_{i_1} - t_{i_2}|/2. \tag{11}$$

Thus, the Gaussian decomposition basis has been estimated directly from the scale-space image. The Gaussian decomposition coefficients x_i, for $i = 1, 2, \ldots, n$, associated with the estimated Gaussian decomposition basis may be obtained by the solution of a linear least-squares estimation problem. Hence, the scale-space image of the original signal allows us to estimate the initial Gaussian representation parameters. These initial Gaussian representation parameters may now be employed in the algorithms developed in the previous section for the determination of the optimal Gaussian representation. Also, notice that the number n of Gaussian decomposition basis functions required to obtain an efficient representation may be estimated from the relative dominance of the different arches in the scale-space image.

Figure 4 illustrates the Gaussian representation of a signal. The original signal is depicted in Fig. 4(a), whereas its corresponding scale-space image is depicted in Fig. 4(b). The Gaussian approximation of the original signal is depicted in Fig. 4(c) (n=9), whereas its corresponding scale-space image is depicted in Fig. 4(d). The initial Gaussian representation parameters used in the optimal Gaussian representation were obtained from the scale-space image depicted in Fig. 4(b).

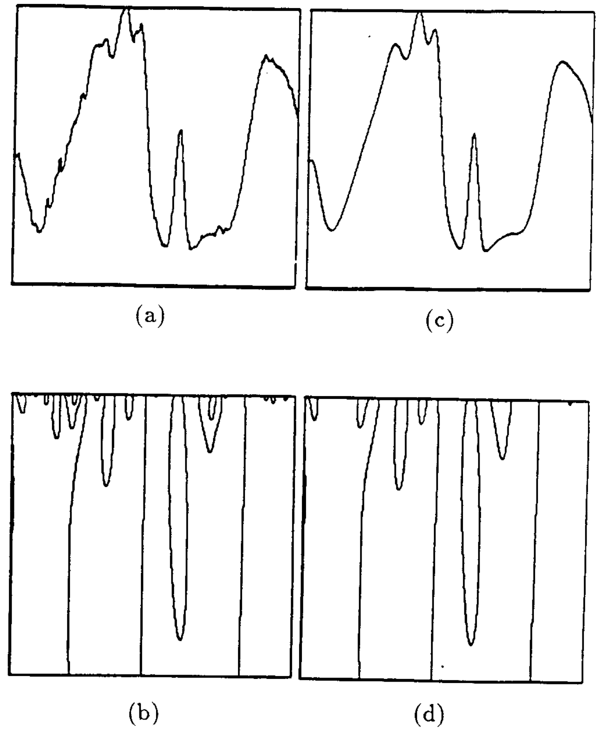

Fig. 4. (a) The original signal; and, (b) its scale-space image. (c) The Gaussian approximation of the original signal (n = 9); and, (d) its scale-space image.

III. Image Contour Representation

In the following section we use the Gaussian decomposition of signals to represent image contours. An image contour (curve) is given by

$$\mathbf{c}(t) = [x(t), y(t)], \tag{12}$$

where $x(t)$ and $y(t)$ are two one-dimensional signals indicating the position of the x and y coordinates, respectively, as the image contour is traced. Thus, parameter t is used to indicate the distance along the image contour. In practice, t may have to be estimated from the image contour [7], [8]. Figure 5 illustrates the representation of an image contour by two one-dimensional signals. The image contour is depicted in Fig. 5(a). The spatial coordinates, corresponding to $x(t)$ and $y(t)$, are depicted in Figs. 5(b) and 5(c), respectively. The start point for the parametrization along the image contour has been marked by "s".

Let us consider an image contour $\mathbf{c}(t)$. The finite Gaussian decomposition bases $b_{x_i}(t)$, for $i = 1, 2, \ldots, n_x$, and $b_{y_i}(t)$, for $i = 1, 2, \ldots, n_y$, are defined by

$$b_{x_i}(t) = e^{-(t-t_{x_i})^2/2\sigma_{x_i}^2}, \qquad i = 1, 2, \ldots, n_x, \tag{13a}$$

and

$$b_{y_i}(t) = e^{-(t-t_{y_i})^2/2\sigma_{y_i}^2}, \qquad i = 1, 2, \ldots, n_y. \tag{13b}$$

Let $\mathbf{p}$ denote the Gaussian representation parameters given by

$$\mathbf{p} = [\mathbf{p}_x \; \mathbf{p}_y], \tag{14a}$$

where

$$\mathbf{p}_x = [(x_i \; t_{x_i} \; \sigma_{x_i}), i = 1, 2, \ldots, n_x], \tag{14b}$$

and

$$\mathbf{p}_y = [(y_i \; t_{y_i} \; \sigma_{y_i}), i = 1, 2, \ldots, n_y]. \tag{14c}$$

Finally, the Gaussian approximation image contour $\hat{\mathbf{c}}_{\mathbf{p}}(t)$ is defined by

$$\hat{\mathbf{c}}_{\mathbf{p}}(t) = [\hat{x}_{\mathbf{p}_x}(t), \hat{y}_{\mathbf{p}_y}(t)], \tag{15a}$$

where

$$\hat{x}_{\mathbf{p}_x}(t) = \sum_{i=1}^{n_x} x_i e^{-(t-t_{x_i})^2/2\sigma_{x_i}^2}, \tag{15b}$$

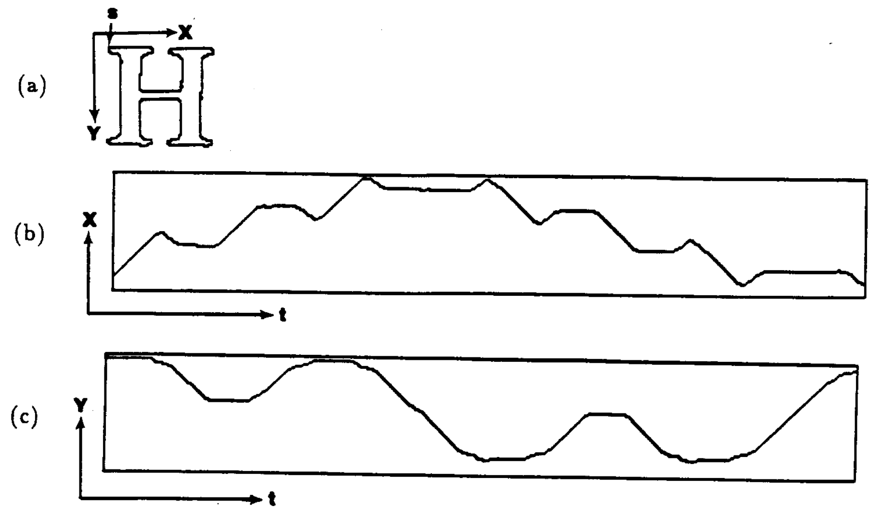

Fig. 5. (a) The image contour; and, (b)-(c) the spatial coordinates corresponding to $x(t)$ and $y(t)$, respectively.

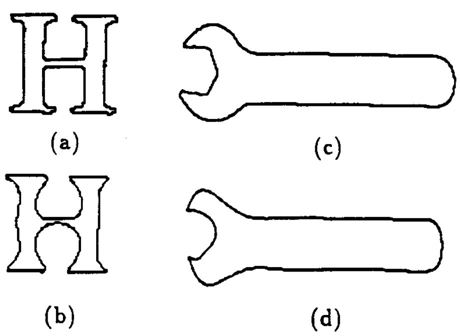

Fig. 6. (a) An image contour; and, (b) its Gaussian approximation ($n_x = 12$ and $n_y = 10$). (c) A second image contour; and, (d) its Gaussian approximation ($n_x = 4$ and $n_y = 8$).

and

$$\hat{y}_{\mathbf{p}_y}(t) = \sum_{i=1}^{n_y} y_i e^{-(t-t_{y_i})^2/2\sigma_{y_i}^2}. \tag{15c}$$

The Gaussian representation of image contours is given in terms of $\hat{\mathbf{p}} = [\hat{\mathbf{p}}_x \ \hat{\mathbf{p}}_y]$, where $\hat{\mathbf{p}}_x$ and $\hat{\mathbf{p}}_y$ correspond to the optimal Gaussian representation parameters of the spatial coordinates $x(t)$ and $y(t)$, respectively.

Figure 6 illustrates the Gaussian approximations of two image contours. The first image contour is depicted in Fig. 6(a), whereas, its Gaussian approximation ($n_x = 12$ and $n_y = 10$) is depicted in Fig. 6(b). A second image contour is depicted in Fig. 6(c), whereas, its Gaussian approximation ($n_x = 4$ and $n_y = 8$) is depicted in Fig. 6(d).

IV. Image Representation

In the following section we extend the Gaussian decomposition of signals to the representation of images. Let us consider an image $x(t_x, t_y)$. The finite Gaussian decomposition basis $b_i(t_x, t_y)$, for $i = 1, 2, \ldots, n$, is defined by

$$b_i(t_x, t_y) = e^{-[(t_x - t_{x_i})^2 + (t_y - t_{y_i})^2]/2\sigma_i^2}, \tag{16}$$

for $i = 1, 2, \ldots, n$. Let $\mathbf{p}$ denote the Gaussian representation parameters given by

$$\mathbf{p} = [(x_i \ t_{x_i} \ t_{y_i} \ \sigma_i), i = 1, 2, \ldots, n]. \tag{17}$$

Finally, the Gaussian approximation image $\hat{x}_{\mathbf{p}}(t_x, t_y)$ is defined by

$$\hat{x}_{\mathbf{p}}(t_x, t_y) = \sum_{i=1}^{n} x_i b_i(t_x, t_y). \tag{18}$$

The optimal Gaussian representation parameters $\hat{\mathbf{p}}$ are obtained by an extension of the iterative optimization procedure presented in Section II.

Figure 7 illustrates a Gaussian approximation of an image. The original image (128×128) is depicted in Fig. 7(a), whereas, its Gaussian approximation ($n = 55$) is depicted in Fig. 7(b). Allowing four-byte quantization for the Gaussian representation parameters, a bit rate of 0.43 bits/pixel and a SNR (peak-to-peak) of 25.5dB were obtained.

(a)

(b)

Fig. 7. (a) The original image; and, (b) its Gaussian approximation ($n = 55$) at 0.43 bits/pixel and SNR (peak-to-peak) of 25.5dB.

V. Summary

In this paper, we studied the representation of signals based on a finite Gaussian decomposition basis. An iterative optimization scheme has been developed for the solution of the optimal Gaussian representation. The proposed scheme is based on the Marquardt algorithm for the solution of nonlinear least-squares estimation problems. The Gaussian decomposition has also been utilized in the representation of image contours. Finally, an extension of the Gaussian decomposition to the representation of images has been presented. Several computer simulation experiments have been used to demonstrate that the Gaussian decomposition results in a highly efficient method for the representation of signals. The approach presented in this paper may be utilized in the representation of signals based on nonorthogonal decomposition bases.

Acknowledgements

The authors would like to thank Dr. Steven L. Eddins for his valuable comments.

References

[1] M. Porat and Y.Y. Zeevi, "The generalized Gabor scheme of image representation in biological and machine vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 10, no. 4, pp. 452-468, 1988.

[2] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by 2D visual cortical filters," *J. Opt. Soc. Amer. (A),* vol. 2, no. 7, pp. 1160-1169, 1985.

[3] H.B. Curry, "The method of steepest descent for nonlinear minimization problems," *Quarterly of Applied Mathematics,* pp. 258-261, 1944.

[4] D.W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of SIAM*, vol. 11, no. 2, pp. 431-441, 1963.

[5] H.O. Hartley, "The modified Gauss-Newton method for fitting of nonlinear regression functions by least-squares," *Technometrics,* vol. 3, no. 2, pp. 269-280, 1961.

[6] A.P. Witkin, Scale-space filtering: a new approach to multi-scale description," *Proc. Image Understanding,* pp. 79-95, 1984.

[7] M.P. Epstein, "On the influence of parametrization in parametric interpolation," *SIAM Journal of Numerical Analysis,* vol. 13, no. 2, pp. 261-268, 1976.

[8] E.T.Y. Lee, "Choosing nodes in parametric curve interpolation," *Computer-Aided Design,* vol. 21, no. 6, pp. 363-370, 1989.