Improving linearity of position-sensitive detector using support vector machines

Meiying Ye (叶美盈)

College of Mathematics and Physics, Zhejiang Normal University, Jinhua 321004

Received September 28, 2004

An intelligent method for improving position linearity of position-sensitive detector (PSD), based on support vector machines (SVMs), is developed. The SVM is established based on the structural risk minimization principle rather than minimizing the empirical error commonly implemented in neural networks. SVM can achieve higher generalization performance. Training SVM is equivalent to solving a linearly constrained quadratic programming problem, thus the solution of SVM is always unique and globally optimal. The improving position linearity procedure has been illustrated using a two-dimensional (2D) PSD. It is pointed out that the position linearity of the measuring system with a proper SVM correction is improved by two orders of magnitude in the measurement range.

OCIS codes: 040.5160, 130.6010, 130.4310.

The position-sensitive detector (PSD) is a precision semiconductor optical sensor, which produces output currents related to the "center of mass" of the light incident on the surface of the device. It is characterized by high position resolution, fast real-time operation, simple structure of processing circuit and so on. The PSD offers the potential for better resolution at a lower system cost.

The linearity of PSD output versus light spot position is of key importance[1]. Usually, the efforts are made to design the PSD as a linear device, which reduces the required mapping to a simple linear characteristic. Unfortunately, the accuracy of position sensing is affected by the nonlinear position error[2,3]. Due to non-idealities in the processing technology, the resistance of the current dividing layer is not uniform, and thus perfectly linear PSD cannot be manufactured. As a main problem, the nonlinear position error is unavoidable and even quite large in most of PSDs.

Therefore, users must correct the position error in many applications. The position error correction of PSD relies on the development of hardware or software linearizing methods to increment the system's accuracy. Analog circuits are frequently used for improving the linearity of PSD, but sometimes complex circuits are needed, and it is difficult to cope with component tolerances or temperature drift. The use of a lookup table is another solution, but the considerable amount of memory required will be an important problem in some microprocessor- or microcontroller-based systems. In recent years, utilizing its powerful ability in function approximation, the neural networks have been successfully used for position error correction of PSD[4−6]. Although the multilayer perceptron (MLP) or radial basis function (RBF) neural networks can be applied in extending the linear range of PSD, some inherent drawbacks, e.g., the multiple local minima problem, the danger of over fitting, the choice of number of hidden units, and an appropriate set of radial basis function centers, etc., would make it difficult to put the networks into practice in some situation.

In this letter, a new intelligent method for position error correction of PSD, based on support vector machines (SVMs)[7], is presented. The SVMs, which have become very popular as methods for learning from examples and have been recently introduced as a general alternative to neural networks[8], are established based on the structural risk minimization principle rather than minimizing the empirical error commonly implemented in neural networks. SVM achieves higher generalization performance than neural networks in solving these machine learning problems. Another key property is that unlike MLP training that requires nonlinear optimization with the danger of getting stuck into local minima, training SVM is equivalent to solving a linearly constrained quadratic programming problem. Consequently, the solution of SVM is always unique and globally optimal. A difference from the RBF neural networks is that no center parameter vectors of the Gaussians have to be specified and no number of hidden units has to be defined because of Mercer's condition. In this letter, the position error correction procedure has been illustrated using two-dimensional (2D) PSD as example. The results indicate that this approach is very effective.

We focus on the SVM regression and use it for improving position linearity of PSD. In the following, we briefly introduce the theory of SVM regression. For further details on SVMs we refer to Ref. [7].

The regression approximation addresses the problem of estimating a function based on a given set of data points $\{\mathbf{x}_i, d_i\}_{i=1}^N$ ($\mathbf{x}_i \in R^n$ is an input and $d_i \in R^1$ is a desired output), which is produced from an unknown function. SVMs approximate the function in the form of

$$f(\mathbf{x}, \alpha_i, \alpha_i^*) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b, \tag{1}$$

where $K(\mathbf{x}_i, \mathbf{x}) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x})$ is the kernel function. Common kernel function is the Gaussian kernel one,

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}\|^2 / (2\sigma^2)\right), \tag{2}$$

where σ is the bandwidth of the Gaussian kernel. It strongly affects the generalization performance of SVMs. Based on the Karush-Kuhn-Tucker (KKT) conditions of quadratic programming, only a number of coefficients

$(\alpha_i - \alpha_i^*)$ will assume nonzero values, and the data points associated with them could be referred to as support vectors. l is the number of support vectors. Reference [7] dealt with a SVM parameter C, which is referred to as the regularization constant. The parameter C determines the tradeoff between minimizing the training error and minimizing model complexity. When applying the SVMs to improve the position linearity of PSD, the optimal generalization still depends on the selection of the parameters (C and σ).

Figure 1 gives the architecture of the SVM used for improving position linearity of PSD. Here x' or y' (i.e. $f(\cdot)$ in Eq. (1)) is the output of the SVM as the value to be estimated, $\mathbf{x} = [x_d, y_d]$ are two inputs.

The 2D PSD, which is illustrated by Fig. 2, is able to detect a light spot moving over its surface in two dimensions. Photocurrent generated by light that falls on the active area of the PSD would be collected by the four perimeter electrodes. The amount of current between each perimeter terminal and the common return is related to the proximity of the incident light spot's center of mass to each collection surface.

If the sheet resistance of the semiconductor is uniform and the coordinate system is chosen as shown in Fig. 2, the position of the light spot on the surface of a 2D PSD at the point P can be found by

$$x = \frac{L}{2} \cdot \frac{X_2 - X_1}{X_1 + X_2}, \quad y = \frac{L}{2} \cdot \frac{Y_2 - Y_1}{Y_1 + Y_2}, \tag{3}$$

where L is equal to the length (width) of the PSD, X_1, X_2, Y_1, Y_2 represent the output signals (photocurrents) of each electrode, respectively, and x, y are the coordinate positions of the light spot.

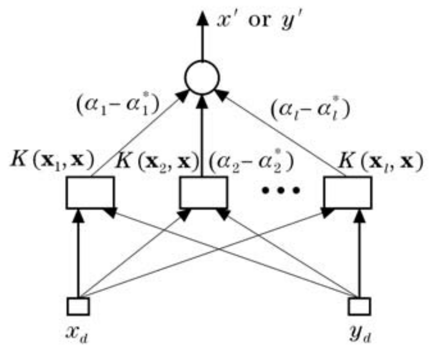

Fig. 1. Architecture of the SVM used for improving position linearity of PSD.

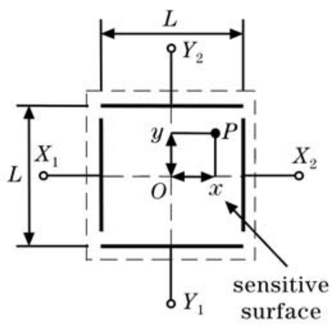

Fig. 2. A typical 2D PSD.

The accuracy of position sensing is affected by the nonlinearity of the sensor response characteristics. Due to non-idealities in the processing technology, the resistance of the current dividing layer is not uniform, and thus perfectly linear PSD cannot be fabricated.

For example, the response of a 2D PSD, S1869 (Hamamatsu), is nonlinear. There are different position detection errors in zone A (radius $r < 5$ mm) and B (radius 5 mm $\leq r \leq$ 10 mm) (see Fig. 3). According to the manufacturing specifications, the position detection errors are typically ± 300 μm and maximally ± 600 μm in zone A, typically ± 2400 μm and maximally ± 3000 μm in zone B. This PSD offers a larger photosensitive surface of 27×27 mm^2. In order to examine our measurement system, we use this PSD as an example. Figure 4 shows a plot of the normalized 2D PSD's readings, which denotes the positions of light spots on the PSD. Note that the values of all the normalized data will keep their range within $[-1, +1]$. A set of light spot positions is composed of 169 evenly spaced points on a grid: 13 rows of 13 points each. The spacing between adjacent columns (rows) was 2.25 mm. It should be immediately apparent that the output of this device is not a linear function of the light spot positions on the PSD surface.

Figure 5 shows the basic principle of improving the position linearity of PSD using a pair of SVMs (SVM1 and SVM2). The following nonlinear functions can be used to describe the variation of the 2D PSD's readings with the position of the light spot

$$x_d = g_1(x, y), \tag{4}$$

$$y_d = g_2(x, y), \tag{5}$$

where x_d and y_d are the 2D PSD's readings. Equations (4) and (5) are the PSD model. The light spot actual position (x, y) can be determined by the PSD inverse model:

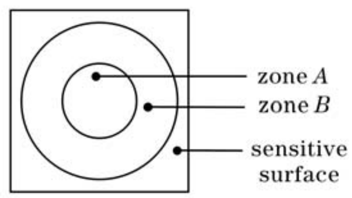

Fig. 3. Sensitive surface of the PSD S1869.

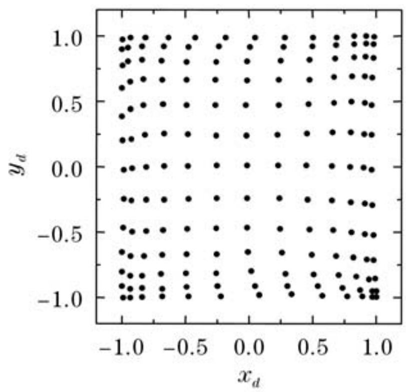

Fig. 4. The normalized 2D PSD's readings.

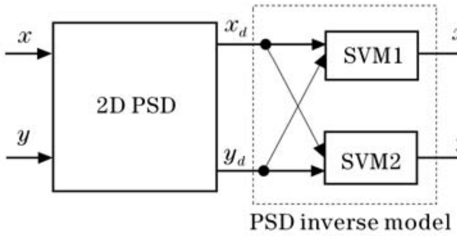

Fig. 5. Basic principle of improving the position linearity of PSD using SVMs.

$$x = f_1(x_d, y_d) = g_1^{-1}(x_d, y_d), \quad (6)$$

$$y = f_2(x_d, y_d) = g_2^{-1}(x_d, y_d). \quad (7)$$

Obviously, Eqs. (6) and (7) are complex nonlinear functions, and they cannot be directly solved. The SVMs can be trained by learning algorithms to properly represent the PSD inverse model. In the training phase, the nonlinear PSD's readings x_d, y_d can be used as inputs of the SVMs and the desired linearized response x or y as target. In the measuring phase, the PSD's readings are applied to the inputs of the trained SVMs, then the estimated values x', y' of the actual position appear at the outputs of the SVMs. Hence we can utilize the PSD inverse model based on SVMs to provide linear responses $x' \approx x$ and $y' \approx y$.

In this investigation, we use the k-fold cross validation[9] to automatically choose the appropriate parameters (C and σ). The k-fold cross validation is computed as follows. The training set is divided into k subsets (folds). The SVM is trained with $k-1$ subsets and then tested with the remaining subset to obtain the regression error. In following example, we use $k = 5$.

The result of this test is illustrated in Fig. 6. Comparing Fig. 4 with Fig. 6, it is obvious that the position error of SVM outputs is reduced greatly in comparison with that of the original PSD in a 27×27 mm^2 active area. The typical position detection errors resulting from the SVMs are ± 6 μm, and the position linearity of the measuring system with a proper SVM correction is improved by two orders of magnitude in the measurement range.

This experiment was repeated using the test PSD's readings, which was not used in the training phase. As a result, the position detection errors were within ± 15 μm, which is acceptable. The position linearity of the SVM outputs is raised greatly compared with the original PSD. This means that the SVMs have higher generalization ability.

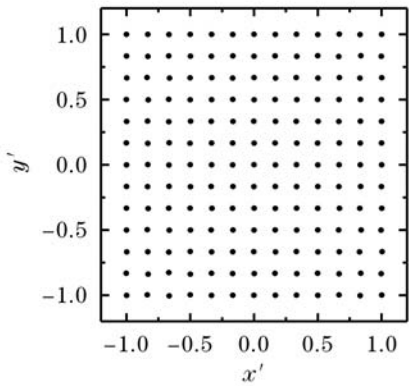

Fig. 6. The PSD inverse model (SVMs) outputs.

In order to better evaluate the performance of the proposed method, SVMs were compared with the neural networks. The position detection errors of neural networks were within ± 60 μm[4], so the SVM regression can obtain lower position detection errors than the neural networks. The result may be attributable to the fact that the SVMs implement the structural risk minimization principle, leading to better generalization than the neural networks.

In conclusion, as the SVMs have excellent approximating and generalizing capabilities, it can be used as an important alternative to neural networks in the position error correction of PSD.

This work was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. 602145.
M. Ye's e-mail address is ymy@zjnu.cn.

References

1. J. Henry and J. Livingstone, IEEE Sensors Journal **3,** 519 (2003).
2. B. Lin, M. Huang, X. Shun, and X. Cao, Proc. SPIE **4919,** 512 (2002).
3. P. Schaefer, R. D. Williams, G. K. Davis, and R. A. Ross, IEEE Trans. Instru. Meas. **47,** 914 (1998).
4. X. Wang and M. Ye, Proc. SPIE **4919,** 452 (2002).
5. X. Wang and M. Ye, Optical Technique (in Chinese) **28,** 174 (2002).
6. C. T. Mo, C. Z. Chen, L. L. Zhang, and F. J. Sun, Chin. J. Lasers (in Chinese) **31,** 427 (2004).
7. V. N. Vapnik, *The Nature of Statistical Learning Theory* (in Chinese) X. G. Zhang (trans.) (Tsinghua University Publishing Company, Beijing, 2000).
8. R. Dietrich, M. Opper, and H. Sompolinsky, Phys. Rev. Lett. **82,** 2975 (1999).
9. K. Duan, S. S Keerthi, and A. N. Poo, Neurocomputing **51,** 41 (2003).