Nonparametric Discrimination Using Proximities for Vehicle Detection

Abstract

Real time traffic information is crucial in traffic management. A Radio Frequency System-on-Chip (RF SoC) using the theory of Frequency Modulated Continuous Waves (FMCW) is practically employed in the experiment for this study. Data recorded via such detector is considered functional data. This study adopts nonparametric discrimination using three types of semi-metric, namely, Principal Component Analysis- (PCA-), Partial Least Square- (PLS-), and Derivative-type, for road-side microwave radar detection of vehicles. Two purposes are involved. The first is the recognition of vehicle types and the second is the recognition of the lane in which vehicles travel. This study shows that the PLS-type semi-metric performed the best among all for the first purpose, and the PCA- and Derivative-type of semi-metrics both yielded satisfactory results for the second purpose.

Keywords: Frequency Modulated Continuous Wave (FMCW); Functional data analysis; k-Nearest Neighbor (*kNN*); Partial Least Square (PLS); Radar recognition.

1. Introduction

Detecting objects such as vehicles via various devices, say microwave Radars, has been of vital importance in the field of the Intelligent Transportation System (ITS) for decades. Traditionally, target information gathered via radars is in image form. Techniques such as Inverse Synthetic Aperture Radar (ISAR) which generates two-dimensional high resolution images for target recognition have been widely utilized in maritime surveillance for the classification of ships and other objects, in addition to numerous other fields. Herman [7, 8] attempted target recognition directly based on raw data regardless of images. Traditionally, algorithms such as principal component analysis (PCA), Probabilistic neural network (PNN), and generic algorithm (GA) have been used directly to classify objects in wide range of research fields, see among others, Ramanan et al. [13], Sun et al [15], and Perez-Jimenez and Perez-Cortes[12]. Data from longitudinal studies or real time information captured from processes was termed functional data by Ramsay and Silverman [12]. Analysis of functional data is called functional data analysis (FDA), and provides us with a different perspective to traditional statistical data analysis and has been widely applied in numerous fields. This study classifies vehicles information obtained by microwave radar using FDA combining the appropriate measures of similarity.

In traffic management, a Radio Frequency System-on-Chip (RF SoC) using the theory of Frequency Modulated Continuous Waves (FMCW) can be used to collect traffic information and for future management or sustainable planning. This study thus adopts such a device for vehicle detection, and data collected via microwave radar detector is considered functional data. Jou et al. [11] also discussed the vehicle detection problem. In that experiment, all test vehicles traveled at the same speed, and the vehicle types were also limited. Moreover, both the parametric approach, namely, the multivariate analysis of variance (MANOVA) and the semi-parametric approach, namely, a linear mixed effects model, did not yield satisfactory results under such a design experiment. Therefore, this study aims to find a robust nonparametric methodology that can correctly classify objects into appropriate classes. The basic idea for subsequent data analysis is mainly derived from the work of Herman [7,8], with the additional adoption of nonparametric discrimination. In this study, two goals are set to achieve. The first is to recognize vehicle type, and the second is to recognize the lane in which vehicles is traveling.

The remainder of this paper is organized as follows: Section 2 contains the nonparametric discrimination of functional data using three types of proximity, namely PCA-, PLS- and Derivatives-type semi-metrics. The PLS regression method is also mentioned simultaneously. Section 3 presents an empirical example of vehicle recognition using microwave radar detector. Finally, conclusions and discussions are presented in the Section 4.

2. Model specification and methodology

Functional data sets appear in numerous scientific fields. Although each data point may be treated as a large finite-dimensional vector it is preferable to think of them as functional data. In this section, some notations of functional data and a brief introduction of nonparametric discrimination using three types of proximity are introduced.

2.1. Preliminaries

We first introduce some common notations and the terminology that generally used in mathematics.

Definition 1. A random variable *X* is called functional random variable (f.r.v.) if it takes values in an infinite dimentional space. An observation *x* of *X* is called a functional data. The *X* denotes a random curve, such that $X = \{X(t); t \in \mathrm{T}\}$.

Definition 2. || . || is a semi-norm on some space F as long as:

1. $\forall x \in \mathrm{F},\ \|x\| \geq 0$
2. $\forall (a,x) \in \Re \times \mathrm{F},\ \|ax\| = |a|\|x\|$
3. $\forall (x,y) \in \mathrm{F} \times \mathrm{F},\ \|x+y\| \leq \|x\| + \|y\|$

Definiton 3. *d* is a *semi-metric* on some space **F** as long as:

1. $\forall x \in \mathrm{F}, d(x,x) = 0$.
2. $\forall (x,y) \in \mathrm{F} \times \mathrm{F}, d(x,y) \geq 0$.
3. $\forall (x,y,z) \in \mathrm{F} \times \mathrm{F} \times \mathrm{F}, d(x,y) \leq d(x,z) + d(z,y)$.

Notice that a semi-norm || . || does not have the property that the fact || *x* || = 0 implies *x* = 0. Similarly, a semi-metric does not have the property that the fact *d*(*x*, *y*) = 0 implies *x* = *y*.

2.2. Three types of semi-metrics

2.2.1. *Semi-metrics based on functional PCA*

Functional Principal Components Analysis (FPCA) is a tool for computing proximities between curves in a reduced dimensional space. As long as $E[\int X^2(t)dt]$ is finite, the FPCA of the f.r.v. ***X*** has the following expansion [2]:

$$X=\sum_{k=1}^{\infty}(\int X(t)v_k(t)dt)v_k\ , \tag{1}$$

where $v_i,\ \forall i$ denotes the orthogonal eigenvectors of the covariance operator $\Gamma_X(s,t)=E(X(s)X(t))$ associated with the eigenvalues $\lambda_1\geq\lambda_2\geq\ldots$. And the truncated version of the above expansion of *X* is

$$\widehat{X}^{(q)}=\sum_{k=1}^{q}(\int X(t)v_k(t)dt)v_k\ , \tag{2}$$

The main goal is to find truncated version such that $E(\int(X(t)-p_qX(t))^2dt$ is minimized over all projections p_q of *X* into q-dimensional spaces. According to the classical L^2-norm, we can define a parametric class of semi-norms and semi-metrics as follows:

$$\|x\|_q^{PCA}=\sqrt{\int(\tilde{x}^{(q)}(t))^2dt}=\sqrt{\sum_{k=1}^{q}(\int x(t)v_k(t)dt)^2}\ , \tag{3}$$

and

$$d_q^{PCA}(X_i,x)=\sqrt{\sum_{k=1}^{q}(\int[X_i(t)-x(t)]v_k(t)dt)^2}\ . \tag{4}$$

In general, Γ_X is unknown and so is the v_k's. However, the covariance operator can be well approximated by its empirical version

$$\Gamma_X^n(s,t)=\frac{1}{n}\sum_{i=1}^{n}X_i(s)X_i(t)\,. \tag{5}$$

The eigenvectors of Γ_X^n are consistent estimators of eigenvectors of Γ_X. In practice, we never observe $\{x_i=\{x_i(t);\ t\in\mathrm{T}\},\ i=1,\ldots,n\}$ but a discrete version $\{x_i=\{x_i(t_1),\ldots,x_i(t_j);\ t\in\mathrm{T}\},\ i=1,\ldots,n\}$. So we can approximate the integral as follows:

$$\int[X_i(t)-x(t)]v_k(t)dt\approx\sum_{j=1}^{J}w_j(X_i(t_j)-x_i(t_j))v_k(t_j)\,, \tag{6}$$

where $w_1,\ldots,w_j$ are the weights which define the approximate integration. A standard choice is $w_j = t_j - t_{j-1}$. Similarly, the semi-metric $d_q^{PCA}(x_i,x_{i'})$ an be approximated by its empirical version:

$$d_q^{PCA}(xi,xi')=\sqrt{\sum_{k=1}^{q}(\sum_{j=1}^{J}w_j(x_i(t_j)-x_{i'}(t_j))v_k(t_j))^2}\ , \tag{7}$$

where $v_1, v_2,\ldots,v_q$ are the **W**-orthonormal eigenvectors of the covariance matrix $\Gamma^n\mathbf{W}={}^1\!/_n\sum_{i=1}^{n}x_ix_{i'}\mathbf{W}$ associated with the eigenvelues $\lambda_{1,n}\geq\lambda_{2,n}\geq\ldots\geq\lambda_{q,n}$, and $\mathbf{W}=diag(w_1,\ldots,w_j)$.

2.2.2. *The Partial Least Squares* (*PLS*)

The Partial Least Squares (PLS), first introduced by Herman Wold [5], is a widespread method for modeling relationship between a set of dependent variables and a large set of predictors. PLS generalizes and combines features from principal component analysis and multiple regression. PLS was first presented as an algorithm analogous to the power method and was subsequently suitably interpreted in a statistical framework [4,6,9,12].

Let **X** be the zero-mean (*n x N*) matrix and **Y** be the zero-mean (*n x M*) matrix, where *n* denotes the number of data sample. PLS decomposes **X** and **Y** into the form

$$\begin{aligned} \mathbf{X} &= \mathbf{TP'} + \mathbf{E} \\ \mathbf{Y} &= \mathbf{UQ'} + \mathbf{F} \end{aligned}, \tag{8}$$

where the **T**, **U** are (*n x p*) matrix of the *p* extracted components, the (*N x p*) matrix **P** and the (*M x p*) matrix **Q** represent loading matrices and the (*n x N*) matrix **E** and the (*n x M*) matrix **F** are the residual matrices. PLS is based on the NonLinear Iterative Partial Least Squares (NIPALS) algorithm and finds weight vectors **w**, **c** such that

$$[\mathrm{cov}(\mathrm{t},\mathrm{u})]^2 = [\mathrm{cov}(\mathbf{X}\mathrm{w},\mathbf{Y}\mathrm{c})]^2 = \max_{|r|=|s|=1}[\mathrm{cov}(\mathbf{X}\mathrm{r},\mathbf{Y}\mathrm{s})]^2, \tag{9}$$

where the term cov(**t**, **u**) = **t'u /** n denotes the sample covariance between the components **t** and **u.** The NIPALS algorithm starts with the random initial value of the component **u** and repeats sequentially till the convergence is achieved.

step 1. w = **X'u** / (**u'u**) (estimate **X** weights)

step 2. ||**w**|| → 1

step 3. **t** = **Xw** (estimate **X** component)

step 4. **c** = **Y't** / (**t't**) (estimate **Y** weights)

step 5. ||**c**|| → 1

step 6. **u = Yc** (estimate **Y** component)

Notice that if $M = 1$ we have the fact that **u** = **y**, and **Y**, denoted by **y**, is a one-dimensional vector. In this case the NIPALS procedure converges in a single iteration. Furthermore, it can be shown that the weight vector **w** also corresponds to the first eigenvector of the following series of equations:

$$\mathbf{w} \propto \mathbf{X'u} \propto \mathbf{X'Yc} \propto \mathbf{X'YY't} \propto \mathbf{X'YY'Xw}$$

This shows that the weight vector **w** is the right singular vector of the matrix **X'Y**. Similarly, the weight vectors **c** is the left singular vector of **X'Y**. The eigenvectors **t** and **u** are given as **t** = **Xw** and **u** = **Yc**, where the weight vector **c** is define in step 4 and 5 of NIPALS.

2.2.2.1. Forms of PLS

PLS is an iterative process. After the extraction of eigenvectors **t** and **u**, the matrices **X** and **Y** are derived from subtracting their rank-one approximations based on **t** and **u**. Different extractions yield several variants of PLS's. By equation (1) the eigenvectors **p** and **q** are computed as coefficients of regressing **X** on **t** and **Y** on **u**, respectively. Then, the eigenvectors can be solved by **p** = **X't /** (**t't**) and **q** = **Y'u /** (**u'u**).

1. *PLS Mode A*

The PLS Mode A is based on rank-one deflation of individual matrix using the corresponding loading and eigenvectors. In this case, the **X** and **Y** matrices are extracted by **X** = **X** - **tp'** and **Y = Y – uq'**. This method was originally proposed by Wold [13] to model the relationships between the different sets of data.

2. *PLS1 and PLS2*

PLS1 (either dependent variable or independent variable consists of a single variable) and PLS2 (both dependent and independent variables are multidimensional) are used as PLS regression method. The main feature of the approach is that the relation between of **X** and **Y** is asymmetric. The main assumptions of the form of PLS are as follows:

I. The eigenvectors $\{\mathbf{t}_i, i = 1 \ldots p\}$ are good predictors of **Y** where p denotes the number of extracted eigenvectors.

II. A linear inner relation between the eigenvectors **t** and **u** exists, that is, **U** = **TD** + **H**.

where **D** is the (p x p) diagonal matrix and **H** is the residual matrix. The asymmetric assumption of the relationship between the independent and the dependent variables is then transformed into a deflation scheme. The eigenvectors $\{\mathbf{t}_i, i = 1 \ldots p\}$ are good predictors of **Y**. Then the eigenvectors are used to deflate **Y**, that is, a component of the regression of **Y** on **t** is removed from **Y** at each iteration of PLS. **X** = **X** - **tp'** and **Y** = **Y** – **tt'Y / (t't**), where the weight vector **c** is defined in step 4 of NIPALS. This way of deflation ensures that the extracted eigenvectors $\{\mathbf{t}_i, i = 1 \ldots p\}$ are mutually orthogonal.

3. *SIMPLS*

Jong [10] introduced another form of PLS, denoted by SIMPLS. The SIMPLS approach directly finds the weight vectors $\{\hat{w}_i, i = 1 \ldots p\}$ which are applied to the original matrix **X**. The criterion of the mutually orthogonal eigenvectors $\{\hat{t}_i, i = 1 \ldots p\}$ still remains.

2.2.2.2. Semi-metrics based on functional PLS

Let $v_1^q, \ldots, v_p^q$ be the vectors of $\Re^J$ performed by multivariate partial least squares regression (MPLSR) where q denotes the number of the factors and p the number of scalar responses. The semi-metric based on the MPLSR is defined as:

$$d_q^{PLS}(x_i, x_{i'}) = \sqrt{\sum_{j=1}^{p} \left(\sum_{j=1}^{J} w_j (x_i(t_j) - x_{i'}(t_j)) v_k^q(t_j)\right)^2}, \quad (10)$$

where $w_1, \ldots, w_J$ are weights which define the approximate integration. A standard choice is $w_j = t_j - t_{j-1}$. When we consider only one scale response ($p = 1$), the proximity between two discrete curves is due to only one direction, which seems inadequate with regard to the complexity of functional data. However, as soon as we consider multivariate response, such a family of semi-metrics yields very good results, which is the case in the curves discrimination context.

2.2.3. Semi-Metrics Based on Derivatives

The semi-metric based on derivatives of two observed curves x_i and $x_{i'}$ can be defined as:

$$d_q^{deriv}(xi, xi') = \sqrt{\int (x_i^{(q)}(t) - x_{i'}^{(q)}(t))^2 \, dt} \,, \qquad (11)$$

where $x^{(q)}$ denotes the q-th derivative of x. Note that $d_0^{deriv}(x,0)$ is the classical L^2-norm of x. The computation of successive derivatives is very sensitive numerically. In order to overcome the numerical stability problem, we can use a B-spline [1] approximation for the curves. Once we have obtained an analytical B-spline expansion for each curve, the successive derivatives are directly computed by differentiating several times their analytic form. Let $\{B_1, \ldots, B_B\}$ be a B-spline basis, then the discrete approximate form of the curve $x_i = (x_i(t_1), \ldots, x_i(t_j))$ is as follows:

$$\hat{\beta}_i = (\hat{\beta}_{i1}, \ldots, \hat{\beta}_{i\mathrm{B}}) = \arg\inf_{(\alpha_1, \ldots, \alpha_\mathrm{B}) \in \Re^\mathrm{B}} \sum_{j=1}^{\mathrm{J}} (x_i(t_j) - \sum_{\mathrm{b}=1}^{\mathrm{B}} \alpha_\mathrm{b} \mathrm{B}_\mathrm{b}(t_j))^2 \,. \qquad (12)$$

This produces a good approximation of the solution of the minimization problem

$$\arg\inf_{(\alpha_1, \ldots, \alpha_\mathrm{B}) \in \Re^\mathrm{B}} \int (x_i(t) - \sum_{\mathrm{b}=1}^{\mathrm{B}} \alpha_\mathrm{b} \mathrm{B}_\mathrm{b}(t_j))^2 \, dt \,. \qquad (13)$$

Therefore, the approximate form of the curve $x_i = (x_i(t_1), \ldots, x_i(t_J))'$ is $\hat{x}_i(.) = \sum_{\mathrm{b}=1}^{\mathrm{B}} \hat{\beta}_{i\mathrm{b}} \mathrm{B}_\mathrm{b}(.)$. Because the analytic expression of the B_b's is well-known, the successive derivatives can be exactly computed and we can differentiate easily the approximated curves:

$$\hat{x}_i^{(q)}(.) = \sum_{\mathrm{b}=1}^{\mathrm{B}} \hat{\beta}_{i\mathrm{b}} \mathrm{B}_\mathrm{b}^{(q)}(.) \,. \qquad (14)$$

Then semi-metric based on derivatives of two observed curves x_i and $x_{i'}$, can be computed by

$$d_q^{deriv}(x_i, x_{i'}) = \sqrt{\int (\hat{x}_i^{(q)}(t) - \hat{x}_{i'}^{(q)}(t))^2 \, dt} \,. \qquad (15)$$

2.3. Nonparametric classification of functional data

Classification or discrimination of functional data is employed when we observe a f.r.v. $\mathcal{X}$ and a categorical response Y which gives the group of each functional observation. The main purpose is to classify new observations into appropriate groups. In this study, we hope to find a robust method for assigning each functional observation to some homogeneous group. We first review the nonparametric discrimination method [3].

2.3.1. A brief review

Let $(X_i, \mathrm{Y}_i)_{i = 1,\ldots,n}$ be a sample of n independent pairs, identically distributed as (X, Y) and valued in

$\mathbf{F} \times \mathbf{G}$, where $\mathbf{G} = \{1,\ldots,G\}$ and ($\mathbf{F}$, d) is a semi-metric vector space (X is a f.r.v. and d a semi-metric). The notation (x_i, y_i) denotes the observation of (X_i, Y_i).

1. **General classification rule (Bayes rule).** Given a functional observation x, the purpose is to estimate the posterior probability $p_g(x) = p(Y = y \mid X = x),\ g \in \mathbf{G}$. Once the G probabilities are estimated by $(\hat{p}_1(x),..., \hat{p}_g(x))$, the classification rule consists of assigning an incoming functional observation x to the group with highest estimated posterior probability $\hat{y}(x) = \arg\max_{g\in\mathbf{G}} \hat{p}_g(x)$. This classification rule is called Bayes rule. To use a suitable kernel estimator make precise discrimination of functional data.
2. **Kernel estimator of posterior probabilities.** Before defining the kernel-type estimator of the posterior probabilities, we remark that $p_g(x) = E(\mathrm{I}_{[Y=g]} \mid X = x)$, with $\mathrm{I}_{[Y=g]}$ equal to 1 if $Y = g$ and 0 otherwise. Therefore we can use kernel-type estimator introduced for the prediction via conditional expectation:

$$\hat{p}_g(x) = \hat{p}_{g,h}(x) = \frac{\sum_{i=1}^{n} \mathrm{I}_{[Y_i=g]} k(h^{-1}d(x, X_i))}{\sum_{i=1}^{n} k(h^{-1}d(x, X_i))}, \qquad (16)$$

where k is the kernel and h is the bandwidth (a strictly positive smoothing parameter). The kernel posterior probability estimate can be rewritten as

$$\hat{p}_{g,h}(x) = \sum_{\{i:Y_i=g\}} w_{i,h}(x), \qquad (17)$$

with

$$w_{i,h}(x) = \frac{k(h^{-1}d(x, X_i))}{\sum_{i=1}^{n} k(h^{-1}d(x, X_i))}, \qquad (18)$$

In order to compute the quantity $\hat{p}_{g,h}(x)$, we use only the X_i's belonging to both the group g and the ball centered at x and of radius h,

$$\hat{p}_{g,h}(x) = \sum_{i\in I} w_{i,h}(x), \qquad (19)$$

where

$$I = \{i : Y_i = g\} \cap \{i : d(x, X_i) < h\}, \qquad (20)$$

The closure X_i is to x the larger the quantity $k(h^{-1}d(x, X_i))$. Hence, the closer X_i is to x the larger is the weight $w_{i,h}(x)$. So, among the X_i's lying to the g-th group, the closer X_i is to x and the larger is its effect on the g-th estimated posterior probability. As long as k is nonnegative, the kernel estimator has the following interesting properties:

1. $0 \le \hat{p}_{g,h}(x) \le 1$,

2. $\sum_{g\in\mathbf{G}} \hat{p}_{g,h}(x) = 1$,

which ensure that the estimated probabilities are forming a discrete distribution.

3. **Choosing the bandwidth.** According to the shape of the kernel estimator, we have to choose the smoothing parameter *h*. As usual, *h* is constructed from minimizing a loss function *Loss* as $h_{Loss} = \arg\min_h Loss(h)$ where the function *Loss* can be built from $\hat{p}_{g,h}(x_i)$'s and y_i's. The misclassification rate is a nature choice among different types of *Loss* functions. Therefore, the functional classification can be performed as follows:

   Ⅰ. Training step

   for $h \in \boldsymbol{H}$

   for i =1, 2,…, n

   for g = 1, 2,…, G

$$\hat{p}_{g,h}(x_i) \leftarrow \frac{\sum_{\{i':y_{i'}=g\}} k(h^{-1}d(x_i, x_{i'}))}{\sum_{i'=1}^{n} k(h^{-1}d(x_i, x_{i'}))}$$

   endfor

   endfor

   endfor

   $h_{Loss} = \arg\min_{h \in \boldsymbol{H}} Loss(h)$.

   Ⅱ. Predicting step

   Let x be a new functional observation and $\hat{y}(x)$ be its estimated group:

$$\hat{y}(x) \leftarrow \arg\min g\{\hat{p}_{g,h_{Loss}}(x)\}$$

   where $\boldsymbol{H} \subset \Re$ is a set of suitable values for h and k is a known kernel.

2.3.2. k-Nearest Neighbors (kNN) estimator

The choices of the bandwidth *h* and the semi-metric *d* have great influence on the behavior of the kernel estimator. It is inefficient to choose bandwidth *h* among the positive real number subset from a computational perspective. So, let us consider a general way which is the *kNN* version of kernel estimator. We can thus replace a choice of real parameter among an infinite number of values with an integer parameter *k* (among a finite subset). The main idea of the *kNN* estimator is to replace the parameter *h* with h_k which is the bandwidth allowing us to take into account *k* terms in the weighted average. The p_g at *x* is estimated by

$$\hat{p}_{g,k}(x) = \frac{\sum_{\{i:y_i=g\}}^{n} k(h_k^{-1}d(x, x_i))}{\sum_{i=1}^{n} k(h_k^{-1}d(x, x_i))}, \tag{21}$$

where h_k is the bandwidth such that the number of $\{i: d(x, x_i) < h_k\}$ is k. The minimization problem on h over a subset of $\Re$ is replaced with a minimization on k over a finite subset {1, 2,…, K}:

$$k_{Loss} \leftarrow \arg\min_{k \in \{1,2,\ldots,K\}} Loss(k) \quad \text{and} \quad h_{Loss} \leftarrow h_{k_{Loss}},$$

where the loss function *Loss* is built based on $\hat{p}_{g,h}(x_i)$'s and y_i's.

Choosing the tuning parameter *k*, we must introduce a loss function *Loss* which allows us to build a local version of the *kNN* estimator. The main goal is to compute the quantity:

$$p_g^{LCV}(x) = \frac{\sum_{\{i:y_i=g\}}^{n} k\left(\frac{d(x,x_i)}{h_{LCV}(x_{i_0})}\right)}{\sum_{i=1}^{n} k\left(\frac{d(x,x_i)}{h_{LCV}(x_{i_0})}\right)}, \quad (22)$$

where $i_0 = \arg\min_{i=1,2,\ldots,n} d(x,x_i)$ and $h_{LCV}(x_{i_0})$ is the bandwidth corresponding to the optimal number of neighbors at x_{i_0} obtained by the following cross-validation procedure:

$$k_{LCV}(x_{i_0}) = \arg\min_k Loss_{LCV}(k,i_0), \quad (23)$$

where

$$Loss_{LCV}(k,i_0) = \sum_{g=1}^{G} (\mathrm{I}_{[y_{i_0}=g]} - p_{g,k}^{(-i_0)}(x_{i_0}))^2, \quad (24)$$

and

$$p_{g,k}^{(-i_0)}(x_{i_0}) = \frac{\sum_{\{i:y_i=g,i\neq i_0\}}^{n} k\left(\frac{d(x_{i_0},x_i)}{h_k(x_{i_0})}\right)}{\sum_{i=1,i\neq i_0}^{n} k\left(\frac{d(x_{i_0},x_i)}{h_k(x_{i_0})}\right)}. \quad (25)$$

As long as we set up the appropriate semi-metric and kernel function *k*(.), the prediction procedure is finished. The misclassification rate for the training sample $(x_i, y_i)_{i=1,2,\ldots,n}$ will be used to assess the performance of the predicted results. The procedure is described as follows:

for $i = 1, 2, \ldots, n$

$$y_i^{LCV} \leftarrow \arg\max_{g\in\{1,2,\ldots,G\}} p_g^{LCV}(x_i)$$

endfor

$$\text{Misclas} \leftarrow \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{[y_i \neq y_i^{LCV}]}.$$

3. Practical experiment and data analysis

In order to classify the received signals into the appropriate groups, a practical experiment was performed using a newly developed detector mounted beside a four-lane road in Jhubei city, HsinChu, Taiwan. Compared to traditional installment of radar detectors, the microwave radar detector was set, at the height of 4.2 meters, perpendicular to the road, thus avoiding so-called Doppler Effect. Figure 1 illustrates of installment of the radar detector in the experiment. Owing to the bandwidth limitation, the radar detection range included 512 points, each located approximately 0.78 meter from the next. Adjustments to the number of points within the radar detection range determine the resolution of images displayed on the monitor.

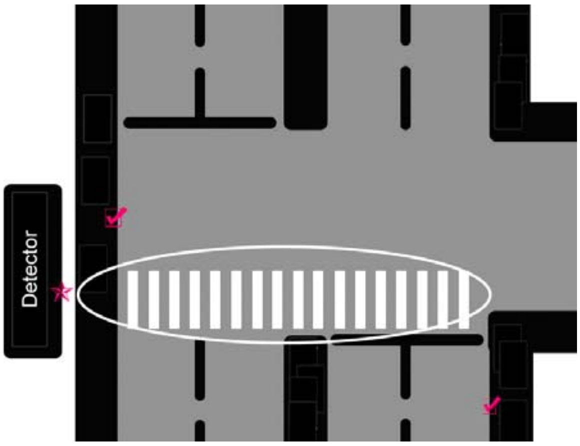


Fig 1. Illustration of installment of Radar detector and the range of Radar detection.

Due to the symmetry of the radar signal, travelling whether from the transmitter to the target or from the target to the receiver, the experiment only requires half of the 512 points within the radar detection range. Subsequent analysis thus focuses on the data set for this group of 256 points. Furthermore, in accordance with the radar equation, the intensity of the radar signal is in inverse square proportion to the distance between the transmitter and the target or the distance from the target to the receiver resulting in the fact that the intensity of the received signal is in inverse proportion to the distance with the power of 4. The weakest signals thus are those received from the fourth lane, while the strongest are those on the first lane. Combining this information on the distance and the size of a specific car contained in the received signal, this study aims at classifying vehicles into groups defined by the size and the lane on which vehicles are traveling using the measure of proximity described in the previous section.

3.1. Data description and data analysis

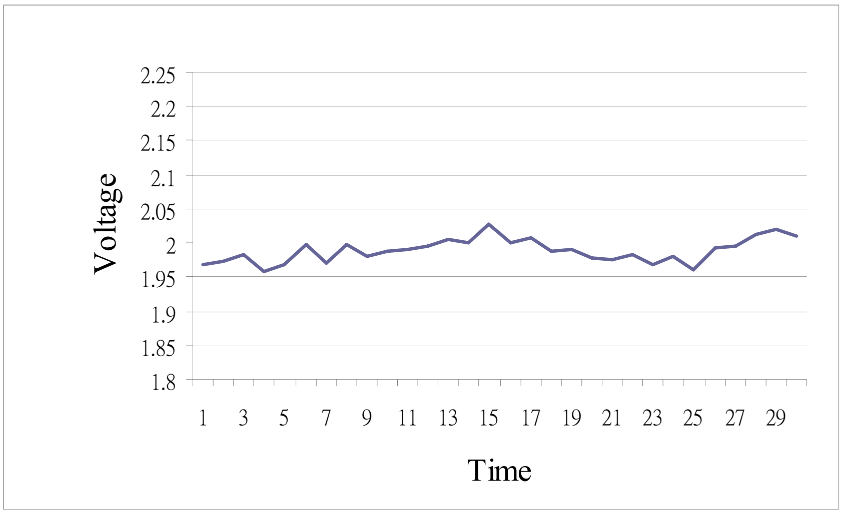
Vehicle data received via the microwave radar detector is presented in the form of intensities in the unit of voltage, and is considered functional data in this study. Notably, all the data are negative numbers. In this study, the absolute values of the data were used. Vehicles were divided into small and large vehicles, and a total 162 vehicles were sampled. Trucks and trailers are considered large vehicles while sedans are considered small vehicles. Due to the geographical condition of the place where the experiment was performed, large vehicles traveled only in the second and the third lanes while small vehicles traveled in all four lanes. Therefore, the data collected from Lanes 2 and 3 will be used for the recognition of vehicle types. Lane 2 contained 28 and 7 small and large vehicle, respectively. Meanwhile, Lane 3 contained 35 and 10 small and large vehicles, respectively. To achieve the second goal of this study, the data of small vehicles recorded in all lanes were used for the analysis. Lanes 1 to 4 were observed to contain 34, 28, 35, and 26 vehicles, respectively. All these data excluded equivocal data resulting from technical problems.

The number of files of each vehicle depends on the time spent within the radar detection range, which results from various factors including vehicle body length, speed and some technical problems experienced during recording. For instance, in absence of technical problems, small vehicles with short body length travelling at low speed would yield more files than large ones with long body length travelling at high speed. To reduce factor biases and retain complete graph, the individual vehicle data are adjusted to a fixed number, namely, 30, of files, as shown in Fig. 2. The reason to adjust these data to a fixed number of files is that with 30 files, an approximately complete wave is clearly formed, and therefore, such a handling procedure is convinced to be acceptable for subsequent analysis. This study assumes that there are k_i files and m vehicles, where $i = 1, 2, \ldots, m$ and employs the following procedure:

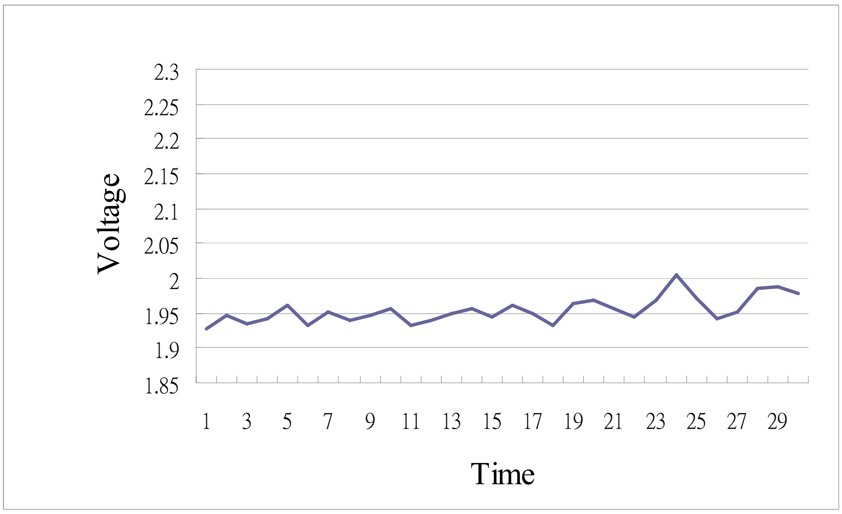
Step 1: Identify i maximum intensities of signals from k_i files.

Step 2: Repeat step 1 to obtain i_{max}, the maximum intensity of signal.

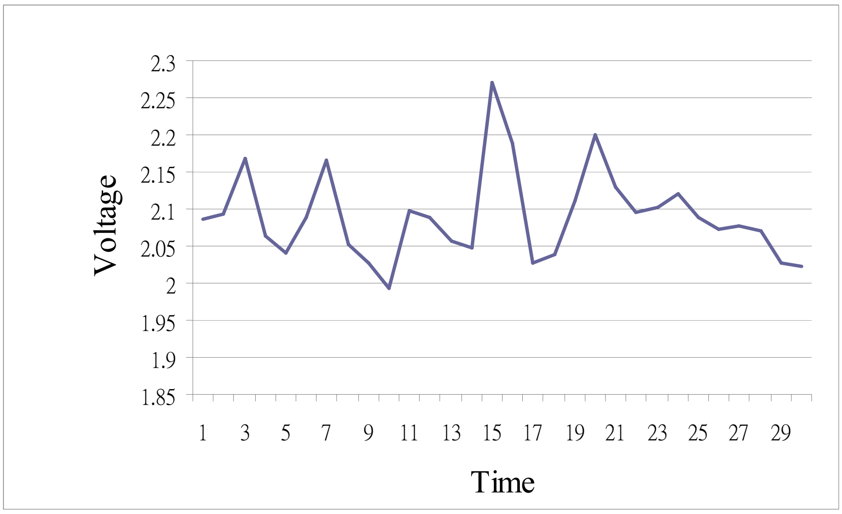
Step 3: Take 14 from the right, and 15 from the left of $i_{\max}$ to produce a total of 30 time-indexed data points.



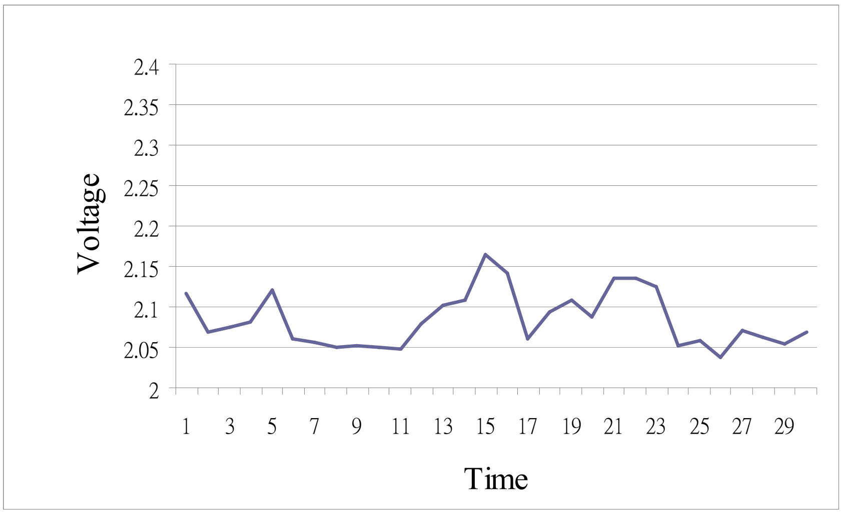
(a)



(b)



(c)



(d)

Fig 2. Illustration of the adjusted data. Small vehicles in Lanes 2 and 3 are shown in (a) and (b), respectively. Large vehicles in Lanes 2 and 3 are shown in (c) and (d), respectively

Based on the idea proposed by Herman [7,8], as mentioned in the previous section, this study directly analyze the adjusted data. The nonparametric discrimination using three types of proximity is used for the analysis. The forms of the three proximity measures are as follows:

1. PLS-type semi-metrics with 2, 3, 4, 5, 6, 7, 8 and 9 factors.
2. PCA-type semi-metrics with 2, 3, 4, 5, 6, 7 and 8 factors.
3. Derivative-type semi-metrics zero derivatives (classical L^2-norm).

Furthermore, to assess the capability of the nonparametric approach, two samples, one for training and the other for authenticating, were therefore randomly assigned, and their misclassification rate was calculated. Notice that approximately two third of the whole data were assigned as training data, and the others were for authentication. The analysis procedure was repeated 5000 times yielding 5000 misclassification rates, and the average misclassification rate was considered the criterion for the assessment.

For vehicle type recognition, four classes were built since there are two lanes and two vehicle types in this study. The results revealed that the average misclassification rates for all types of semi-metric are within the range of 17.5% ~ 27.3%. PLS-type was shown to be the best approach among all. Nevertheless, both the PCA- and Derivative-type semi-metrics still yielded satisfactory results. Regarding the recognition of lanes in which vehicles travel, the average misclassification rates were about 33%. Higher misclassification rate may result from more complicated analysis compared to the first purpose. The main reason for such a result could be the fact that the target information implicitly contains different sources of objects, such as background signals, noises from the environment, the interference among vehicles, and even the weather condition. Table 1 presents the average misclassification rates for each of the three semi-metrics.

Table 1.

Average misclassification rates for three types of semi-metric for purpose 1 and 2.

	PCA-type	PLS-type	Derivative-type
Vehicle type	27.20%	17.56%	27.30%
Lanes	32.54%	33.47%	32.70%

4. Conclusions

This study adopts three types of semi-metric for nonparametric discrimination for vehicle recognition using road-side microwave radar detector. For the first purpose, namely, the recognition of the vehicle type, all types of semi-metrics yield satisfactory results for such practical experiments and are recommended to be employed practically. Recognition of vehicle types is considered a relatively simple problem since the data for large vehicles is obviously different from that of small ones. Regarding the second purpose, namely, the recognition of lanes in which vehicles travel, lower classification rate was anticipated. The main reason for this result may be that the data contains other contents, such as background signals and noises from other objects. Consequently, more techniques in dealing with digital signal processing which is not within the discussion of this study are recommended to be used prior to any statistical analysis.

Acknowledgement

The authors would like to thank the National Science Council of Taiwan for partially financially supporting this research under Contract No. NSC 96-2221-E-009-120.

Reference

[1] Boor, C.A., 1978. A practical guide to splines. Springer, New York.

[2] Dauxois, J., Pousse, A., Romain, Y., 1982. Asymptotic theory for the principal component analysis of a random vector function: some application to statistical inference, Journal of Multivariate analysis, 12 136-154

[3] Ferraty, F., Vieu, P., 2006. Nonparametric functional data analysis. Springer, New York.

[4] Frank, I.E., Friedman, J.H., 1993. A statistical view of chemometrics regression tools. Technometrics, 35 109-148.

[5] Geladi, P., Kowlaski, B., 1986. Partial least square regression: A tutorial Analytica Chemica Acta, 35 1-17.

[6] Helland, I.S., 1990. PLS regression and statistical models. Scandivian Journal of Statistics, 17 97-114.

[7] Herman, S., A Particle Filtering Approach to Joint Passive Radar Tracking and Target Classification. Doctoral Dissertation. (Dept. of Electrical and Computer Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL, 2002).

[8] Herman, S., Moulin, P., 2002. A particle filtering approach to joint radar tracking and automatic target recognition. Proc. IEEE Aerospace Conference, (Big Sky, Montana), March 10-15.

[9] Hoskuldsson, P., 1988. PLS regression methods. Journal of Chemometrics, 2 211-288.

[10] Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. chemometrics and Intelligent Laboratory Systems, 18 251-263.

[11] Jou, Y.J., Huang, C.L., Yang, C.T., 2007 Semi-parametric Linear Mixed Effects Model for Vehicles Identification, ICCMSE Conference.

[12] Perez-Jimenez, A. J., Perez-Cortes, J.C., 2006. Genetic algorithms for linear feature extraction, Pattern Recognition Letters, Vol. 27, Issue 13, pp. 1508-1514.

[13] Ramanan, S., et al., 1995. pRAM nets for detection of small targets in sequences of infra-red images, Neural Networks, Vol. 8, No.7/8, pp.1227-1237.

[14] Ramsay, J., Silverman, B., 1997. Functional Data Analysis. Springer, New York.

[15] Sun, G. et al., 2006. Neurocomputing, Vol.69, Issue 4-6, pp. 387-420.

[16] Wold, H., 1966. Estimation of principal components and related models by iterative least squares. In: P.R. Krishnaiaah (Ed.), Multivariate Analysis, Academic Press, New York, 391-420.