Automatic Chinese Postal Address Block Location Using Proximity Descriptors and Cooperative Profit Random Forests

Xinghui Dong, Junyu Dong, Huiyu Zhou, Jianyuan Sun, and Dacheng Tao, *Fellow, IEEE*

***Abstract*—Locating the destination address block is key to automated sorting of mails. Due to the characteristics of Chinese envelopes used in mainland China, we here exploit proximity cues in order to describe the investigated regions on envelopes. We propose two proximity descriptors encoding spatial distributions of the connected components obtained from the binary envelope images. To locate the destination address block, these descriptors are used together with Cooperative Profit Random Forests (CPRF). Experimental results show that the proposed proximity descriptors are superior to two component descriptors which only exploit the shape characteristics of the individual components, and the CPRF classifier produces higher recall values than seven state-of-the-art classifiers. These promising results are due to the fact that the proposed descriptors encode the proximity characteristics of the binary envelope images, and the CPRF classifier uses an effective tree node split approach.**

***Index Terms*—Postal automation, postal address block location, proximity, random forests, cooperative game theory.**

I. Introduction

COMPUTER vision [4], [21], [25] and pattern recognition [9], [35], [41] algorithms have been widely applied to various industrial automation systems. Particularly, automated sorting of mails [2] plays an important role in mail delivery systems. The current automated sorting systems used in mainland China were mainly designed based on recognizing the postcode. However, recognition of postcodes [2] encounters problems as postcodes cover various sizes of regions in different countries and territories. For example, a postcode is allocated to a town rather than a building in mainland China. Hence, recognition of postcodes cannot lead to a specific address. In this situation, an optical character recognition (OCR) module [2] is required to recognize the destination address on letters or parcels. Ideally, a sorting system needs to precisely locate the destination address block in real time [22], [37] and send it to the OCR module [2]. An incorrectly located address block leads to immediate rejection of the mail.

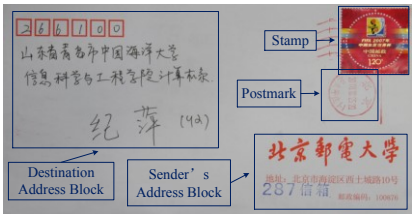


Fig. 1. Presence of parasitic objects with the destination address block on a Chinese envelope.

In this study, we aim to develop an automated method for locating the destination address block on Chinese postal envelopes used in mainland China. In an OCR-based automated mail sorting system [2], binary images are normally used as inputs. Therefore, we use a method deliberately designed for Chinese envelope image segmentation [16] before address block location is performed. The envelopes (see Fig. 1) used in mainland China are different from those used in other countries and territories. The significant difference is that the Chinese envelopes contain two different addresses, one is for the sender and the other is for the addressee. The presence of stamps and postmarks (see Fig. 1) also makes the locating task challenging. However, the destination address block normally lies far from that of the sender (more details can be found in the China National Standard for Postal Envelope Writing/Printing [8]). Evidence shows that the spatial layout of local image features is important to texture perception [15]. In addition, objects or shapes tend to form a single group when they stay close to each other, according to the Gestalt Law of Grouping [1].

Inspired by the characteristics of Chinese envelopes and the importance of proximity cues to perceptual grouping [1], we here propose two proximity descriptors computed using connected components (or components) within a local region. The computational complexity and sensitivity to noise can be reduced using components instead of pixels. One descriptor is a variant of Shape Context [3] which is originally designed based on contour points. The other descriptor is created by capturing local self-similarity characteristics [32]. However, we exploit

J. Dong is supported by the National Natural Science Foundation of China (NSFC) (No. 61271405 and 41576011). H. Zhou is supported by UK EPSRC under Grants EP/N508664/1, EP/R007187/1 and EP/N011074/1, and Royal Society-Newton Advanced Fellowship under Grant NA160342. (Corresponding author: Junyu Dong).

X. Dong is with the Center for Imaging Sciences, the University of Manchester, M13 9PT, UK (e-mail: dongxinghui@gmail.com).

J. Dong and J. Sun are with the Department of Computer Science, Ocean University of China, Qingdao (e-mail: dongjunyu@ouc.edu.cn).

H. Zhou is with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, and Department of Informatics, University of Leicester, UK (e-mail: h.zhou@qub.ac.uk).

D. Tao is with the UBTECH Sydney Artificial Intelligence Center and the School of Information Technologies, University of Sydney, NSW 2008, Australia (email: dacheng.tao@sydney.edu.au).

the self-similarity based on the shape of components rather than the appearance of local image patches. Both the descriptors encode the proximity characteristics over a local region. To our knowledge, component-wise proximity descriptors have not been used in locating postal address blocks.

Over the past decades, random forest (RF) classifiers [5], [12] have been applied to various tasks. The merits of using a RF classifier include that: (1) it is efficient; (2) it does not require tuning of parameters; and (3) it minimizes the chance for overfitting. However, the tree node split functions that existing RF classifiers use only utilize strong discriminant attributes while ignoring weak discriminant attributes. As known, the solution to the cooperative game theory [6] is able to produce a reasonable approach that allows analyzing the payment ability or the power of players without an assumption. As a result, this approach can explore the contributions of all the players. Therefore, we propose a new RF classifier, i.e., Cooperative Profit Random Forests (CPRF), using a split method based on the Shapley value [31] used in cooperative game [6]. This split method exploits both strong and weak discriminant attributes.

The contributions of this paper include that: (1) the application of proximity cues to locating postal address blocks on Chinese envelopes; (2) the introduction of two new component-wise proximity descriptors for encoding binary images; (3) the proposal of a Shapley value based CPRF classifier; and (4) the comparison between the CPRF and seven state-of-the-art classifiers. The rest of the paper is organized as follows. In Section II, we review the related work. In Section III, we outline our framework. The proposed proximity descriptors and CPRF classifier are detailed in Sections IV and V respectively. In Sections VI and VII, we describe the experimental setup and report the results, respectively. Finally, conclusions are provided in Section VIII.

II. Related Work

Automatic extraction of postal address blocks has been widely studied. Jain and Bhattacharjee [22] treated an envelope image as a combination of textured regions and converted postal address block location to texture classification. The method proposed by Yu *et al.* [42] concentrates on address block location for complex postal items with an arbitrary layout of printed entities. Xue *et al.* [37] used geometric constraints to segment strings and proposed an optimization-based address interpretation method. Govindaraju and Tulyakov [18] applied the features computed from the contours of the connected components labeled in images to address block location. The method of Kagehiro *et al.* [23] consists of two stages: address block candidate nomination and candidate evaluation.

An address block location method based on hierarchical graph coloring and the pyramidal organization of data was presented by Gaceb *et al.* [17]. A clustering method was then applied to these features to separate the address cluster from the other clusters. In the same year, Menotti and Borges [26] developed image segmentation and address block location methods based on feature selection. Schmidt *et al.* [30] proposed a website business address extraction method using both the patterns and gazetteers derived from freely available knowledge sources. Radha and Aparna [28] developed an automatic Indian postal address block detection method based on text block extraction. Recently, Cheng and Xu [7] located Chinese postal address blocks using a binary classifier.

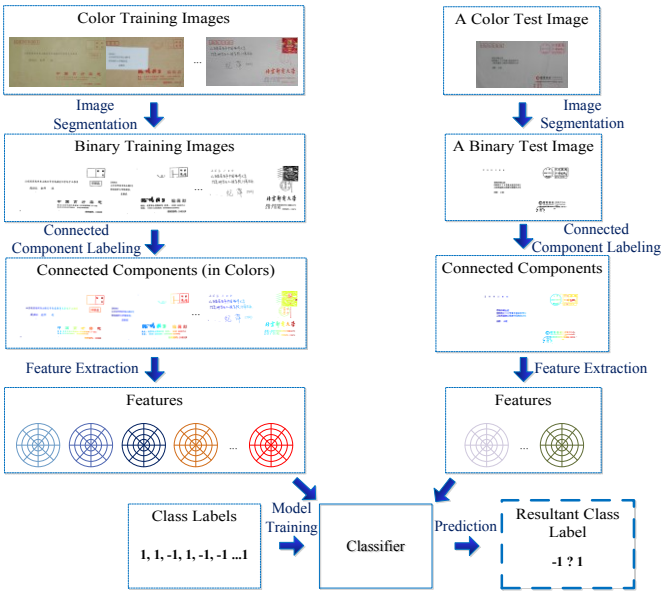


Fig. 2. Pipeline of the proposed postal address block location method.

In contrast, text (line) detection [43], [39] has been given more attention than address block location in the literature. Zhang *et al.* [43] conducted text line detection in natural scene images by using the symmetry characteristic of character lines. An extensive survey has been performed by Ye and Doermann [39], in which analysis and comparison of various challenges, approaches, as well as the performance of text detection and recognition studies were investigated. However, text (line) detection cannot be applied to locating Chinese postal destination address blocks as the address of the sender is written or printed on the envelopes used in mainland China.

To summarize, the aforementioned methods do not exploit proximity cues. It has been shown that, however, these cues are important to the perceptual grouping used by the human visual system [1]. We therefore introduce two proximity descriptors for locating postal address blocks. Compared with local image features, the proposed descriptors encode the proximity characteristics manifested in large spatial regions.

III. The Framework of the Proposed Method

The proposed Chinese postal address block location method is conducted in five stages. Fig. 2 shows the pipeline of the proposed method. We will describe the five stages in detail.

A. Image Segmentation

The application of image segmentation to original images will reduce the interference of the background information to address block location. It is also necessary to remove the boxes surrounding the postcodes on the Chinese envelopes before address block recognition is conducted. We hence use the method that Dong *et al.* [16] proposed to perform envelope image segmentation. The advantage of this method over the other approaches [27], [38] is that it does not require post-processing and is also efficient. The Wiener filter [36] was first applied to an image in order to reduce the interference of

noise. Only gray level images are used for efficiency purposes. Finally, a binary envelope image is derived.

B. Connected Component Labeling

Since we aim to locate the postal address blocks containing various Chinese characters, we therefore take the connected components obtained from the binary envelope images as basic elements. In this case, the extracted features based on the components will encode the context information and be less sensitive to the noise contained in the binary images than the features computed based on pixels. Besides, the computational speed of feature extraction, model training and address block prediction is enhanced. After connected component labeling [13] is complete, we obtain a set of component labels. Each of these labels is associated with a foreground pixel subset.

C. Feature Extraction

Due to the characteristics of Chinese envelopes and the importance of proximity cues to perceptual grouping [1], we propose two proximity descriptors whose details are presented in Sections IV-C and IV-D respectively. In addition, the two descriptors introduced in Section VI-A-1 are used for baselines. For each component, four sets of features are computed.

D. Model Training

We use "-1" and "1" as class labels where "1" represents a destination address component (positive sample) while "-1" denotes a non-address component (negative sample). Since the address blocks contained in training images have been labeled using a bounding box, the connected components that locate in this box are labeled as "1"; otherwise they are labeled as "-1". After feature extraction is complete, we create a training dataset for the training images, including feature vectors and class labels. We use this training dataset to train our classifier.

E. Destination Address Block Prediction

Once the model is trained, it can be used for locating address blocks by predicting the class label ("-1" or "1") of the connected components labeled in the test images. At this stage, the same type of features are used as that used in the model training stage. We traverse the whole test image in a component-by-component manner for prediction. Only the class label of one component is predicted every time. After the labels of all the components have been predicted, the location corresponding to a component whose class label is "1" is assigned the gray level of 0; otherwise it is set to the gray level of 255. The locations with the gray level of 0 are considered as the destination address block predicted in the test image.

IV. Component-wise Proximity Descriptors

Evidence [8] shows that destination address blocks usually position at the top-left region of Chinese envelopes (see Fig. 1). As discussed in [1], the human visual system uses proximity cues to separate a region from others. Here, we develop two proximity descriptors by encoding the proximity characteristics of the connected components extracted from a local region, which are described in Sections IV-C and IV-D respectively.

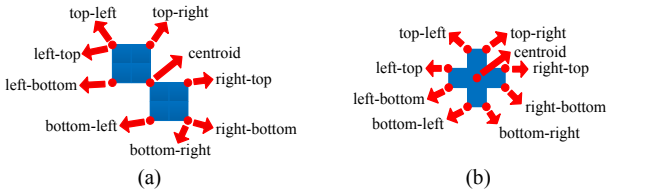


Fig. 3. The illustration of the reference points (red points) of two different connected components (blue regions).

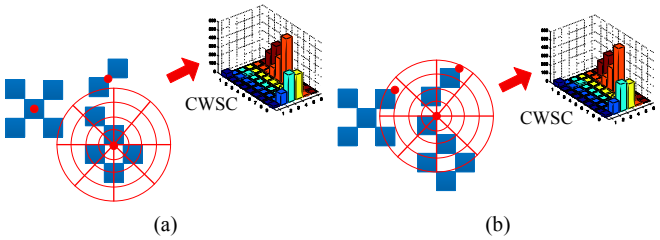


Fig. 4. The computation process of component-wise shape context features in terms of: (a) a "centroid" reference point (the center of the disks), and (b) a "top-right" reference point (the center of the disks).

A. Connected Component Labeling

We analyze binary envelope images based on connected components (or components) rather than image pixels. The advantages of using components over pixels include that (1) locating address blocks becomes more efficient because in an image, the number of the components is significantly less than that of the pixels. Compared to image patches, the components are more perceptually intuitive; and (2) the locating operation is immune to noise. To this end, we conduct connected component labeling [13] on the binary envelope images.

B. Describing Connected Components Using a Set of Reference Points

Connected components containing a set of pixels normally present different shapes. Different image properties can be used to describe connected components, e.g., size of the bounding box, perimeter and centroid. We use the coordinates of the centroid and eight extrema points to represent a component. That is to say, each component is described by nine points, i.e., the "centroid", "left-top", "left-bottom", "top-left", "top-right", "right-top", "right-bottom", "bottom-left" and "bottom-right" points. These points are referred to as "Reference Points". Fig. 3 shows the reference points of two components in detail.

C. Component-wise Shape Context

The original shape context descriptor [3] was designed to represent the shape of contours via encoding the co-occurrence of the distance and orientation of contour point pairs. Since the appearance of the destination address block on Chinese envelopes is different from that of other regions, we describe local regions rather than contours by exploiting the proximity characteristics of these regions. We here use an improved shape context algorithm by computing shape context co-occurrence histograms based on connected components.

To be specific, for each of the nine types of reference points, we obtain a reference point map. Given this reference point map, we compute an $A \times D$ shape context histogram at the location of each reference point. A shape context histogram is

computed based on a circular neighborhood with the radius of R pixels, whose center locates at the current reference point. Before the histogram is computed, we pad the reference map with a blank surrounding boundary area (whose thickness is equal to R) in order to guarantee that the shape context can be computed at the original boundary locations. Given a $W \times H$ image, the size of the padded image is $(W + 2R) \times (H + 2R)$. The distance and orientation between the current reference point and the other reference points within the circular neighborhood are calculated. All the distance and angle values are quantized into a 2D co-occurrence matrix, i.e., the shape context histogram, which consists of D distance bins and A angle bins. Similar to the method that Belongie *et al.* [3] proposed, we use the *log* space to quantize the distances in order to amplify the influence of the closer reference points. Fig. 4 shows two examples in which the component-wise shape context histograms are calculated in terms of a "centroid" point and a "top-right" point respectively. Eventually, the shape context histogram captures the spatial relationship between the central reference point and its neighboring reference points.

For each connected component, we compute nine shape context histograms in total (with regard to its nine reference points). These histograms are then concatenated into a single feature vector in order to generate a representation of the proximity characteristics of the local region surrounding the current component. This feature vector is referred to as "Component-wise Shape Context" or "CWSC". The dimensionality of the CWSC feature vector is $A \times D \times 9$.

D. Component-wise Local Self-Similarity

The component-wise shape context descriptor considers the nine reference points of a connected component individually and enables a single representation in terms of each reference point to be calculated. Hence, this descriptor does not take into consideration the local shape of the components. However, we have observed that the font size of the characters in a block is usually similar to each other. We thus propose a second proximity descriptor: component-wise local self-similarity.

According to [32], local self-similarity is computed based on the *Euclidian* distance between the local circular neighborhood of a pixel q and the neighborhoods of the other pixels in the surrounding region (with the radius of R pixels) of this pixel. After the distance computation is performed, a distance surface $d_q(x, y)$ is obtained. This surface is normalized and converted to a correlation surface $c_q(x, y)$. The correlation surface is mapped to the *log*-polar space whose center locates at the pixel q. This space is quantized into A orientation angle bins and D distance bins. The maximal correlation value in each bin is concatenated into an $A \times D$ dimensional feature vector and is further normalized to the range of [0, 1] using linear stretching. The normalized feature vector is referred to as "Local Self-Similarity Descriptor" of the pixel q. This descriptor encodes the spatial layout of the local self-similarity data.

In this study, we modify the original local self-similarity descriptor in order to encode the local self-similarity of connected components instead of local neighborhoods. Each component is represented by its nine reference points (see Fig. 3). All the reference points of each component are comprised of a reference point map. We pad the reference point map using a blank boundary area (whose thickness is equal to R pixels). The padding operation guarantees that the local self-similarity can be fully computed at each original boundary location.

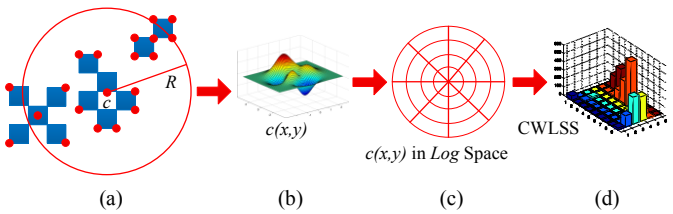


Fig. 5. The computation pipeline of the component-wise local self-similarity descriptor in terms of the connected component c. In (a), red points mean the reference points of different components.

In terms of the "centroid" reference point (see Fig. 3) of a given component c, a circular region with the radius of R pixels is defined (see Fig. 5 (a)), whose center locates at this point. Each component whose "centroid" reference point falls in this region is considered. The Sum of Square Differences (SSD) between the nine reference points of the central component (i.e., c) and those of each neighboring component is computed. It should be noted that the coordinates of the reference points are subtracted from the coordinates of the "centroid" reference point, to remove the influence of the position. In fact, only the eight boundary reference points are used for computing the SSD. All the SSD values computed from the overall neighboring components are comprised of a distance surface $d(x, y)$. This surface is further converted to a correlation surface $c(x, y)$ (see Fig. 5 (b)). The remaining computation procedures (see Figs. 5 (c) and (d)) are the same as those used to compute the original local self-similarity descriptor [32]. Finally, an $A \times D$ dimensional descriptor is derived to represent the component c. We refer to this descriptor as "Component-wise Local Self-Similarity" or "CWLSS".

V. The Cooperative Profit Random Forests Based on the Shapley Value

In this section, we introduce the proposed Cooperative Profit Random Forests (CPRF) classifier in detail.

Given a binary decision tree l ($l \in \{1, \ldots, L\}$), it classifies a sample $x_i \in R^p$ (i =1,…, n) via traversing the tree till encountering a leaf node. A binary split function is associated with the tree node i. The split function is expressed as:

$$h(x_i, \theta_i) \in \{0,1\}, \tag{1}$$

where θ_i is a parameter. If $h(x_i, \theta_i) = 0$, x_i is considered as a left child; otherwise it is treated as a right child. This process is repeated till a leaf node is reached. In terms of x_i, the output of $\hat{f}_l(x_i)$ is the prediction label $y_i \in \{-1,1\}^n$ of x_i and is stored at a leaf node. Regarding the split function $h(x, \theta)$, a simple implementation is fulfilled based on thresholding τ:

$$\theta = (j, \tau), \text{ and} \tag{2}$$

$$h(x, \theta) = [x(j) \leq \tau], j = 1, \ldots, p, \tag{3}$$

Algorithm 1: Shapley Value Based Cooperative Profit Random Forests

Given n labeled training samples: $D = \{(x_1, y_1), \dots (x_n, y_n)\} \in R^{n\times(p+1)}$ with each attribute variable $f_j = (x_{1,j}, \dots, x_{n,j}), j = 1, \dots, p$,
For each decision tree l, $l \in \{1, \dots, L\}$, do:
1. Take a bootstrap sample subset D_l of size n from D with a replacement;
2. Extending tree nodes:
While
i. At each node, randomly select m ($m \ll p$) attribute variables from the p available attribute variables;
ii. Compute the "best" binary splitting points $\varepsilon_{i,j}$ among all possible splits on the m attributes, where
$$\varepsilon_{i,j} = \frac{x_{i,j}+x_{i,j+1}}{2}, i = 1, \dots n;\ j \in (1, \dots, m);$$
iii. For the split point $\varepsilon_{i,j}$of each attribute $f_j, j \in (1, \dots m)$, do:
Compute:
$$r(N_{left}) = \sum_{j=1}^{p} \emptyset_{f_j}(\Gamma),\ f_j \in N_{left\,(f_j)}, \text{ and}$$
$$r(N_{left}) = \sum_{j=1}^{p} \emptyset_{f_j}(\Gamma),\ f_j \in N_{left\,(f_j)},$$
where $N_{left\,(f_j)} = \{(f_1, \dots, f_p) \in N_{f_j} : x_{i,j} \le \varepsilon_{i,j}\}$ and
$N_{right\,(f_j)} = \{(f_1, \dots, f_p) \in N_{f_j} : x_{i,j} > \varepsilon_{i,j}\}$.
Then,
$$Split(\hat{f}_j, \hat{\varepsilon}_{i,j}) \leftarrow arg\ max(r(N_{left}) + r(N_{right})),$$
where $N_{left\,(\hat{f}_j)} = \{(f_1, \dots, f_p) \in N_{\hat{f}_j} : x_{i,j} \le \hat{\varepsilon}_{i,j}\}$ and
$N_{right\,(\hat{f}_j)} = \{(f_1, \dots, f_p) \in N_{\hat{f}_j} : x_{i,j} > \hat{\varepsilon}_{i,j}\}$.
Until the minimal number of samples in a node is reached;
3. Making a classification prediction for a test sample x based on
$$\hat{f}(x) = \arg\ \max_y \sum_{l=1}^{L} I(\hat{f}_l(x) = y),$$
where $\hat{f}_l(x)$ is the classification prediction of the response variable at x using the l-th decision tree. The proposed random forests algorithm predicts the class at x as that receives the greatest vote from individual trees.

where $[\cdot]$ denotes the indicator function.

A decision forest contains a series of independent decision trees l ($l \in \{1, \dots, L\}$). In terms of a sample $x_i \in R^p$ (i=1,…,n), an ensemble model is used to integrate all the predictions $\hat{f}_l(x_i)$ ($l \in \{1, \dots, L\}$) obtained using these trees for individual predictions. Majority voting is a popular ensemble model for classification applications. Since the variance of the predictions of different decision trees is high, the overfitting issue is usually encountered [5]. In this situation, decision forests allow us to train a set of de-correlated decision trees and combine their predictions using an ensemble model in order to reduce the overfitting possibility. This type of decision forests is referred to as random forests [5].

The Gini index or information gain ratio [5] split functions are normally used by random forests algorithms. However, these functions only select strong discriminant attributes. We therefore propose a new random forests classifier by using a split method based on the Shapley value [31] known in the cooperative game theory [6] in order to exploit the merits of both weak and strong attributes.

A cooperative game $\Gamma = (N, \gamma)$ consists of a player set $N = \{1, 2, \dots, n\}$ and a characteristic function $r: 2^N \to R$. Given a subset $S \subseteq N$, $r(S)$ is interpreted as the profit achieved by the players in S. One of the key problems in the cooperative game theory is how to allocate the total income $r(N)$ to each player i ($i \in N$) in the grand player coalition N in a fair and reasonable way. The advantages of the Shapley value [31] were demonstrated for feature selection [33] because it not only measures the distribution of the incomes allocated to players, but also estimates the contributions of players [6]. Given the Shapley value is denoted as $\emptyset(\Gamma) \in R^n$, the payoff to the i-th player: $\emptyset_i(\Gamma)$ can be computed as [31]:

$$\emptyset_i(\Gamma) = \sum_{S\subset N} \Delta_i(S) \times \frac{|S|!(n-|S|-1)!}{n!}, \text{ and} \quad (4)$$

$$\Delta_i(S) = r(S \cup i) - r(S), \quad (5)$$

where n represents the total number of players and $\Delta_i(S)$ denotes the contribution of player i to the coalition $S \subseteq N$.

The Shapley value considers possible intrinsic and intricate correlative interactions between players. It can be incorporated into random forests to estimate the best split point and the corresponding attributes. Specifically, we traverse every possible split point corresponding to each candidate attribute. The best split point is selected as that which produces the greatest sum of the Shapley values with regard to the attributes at both left and right children nodes. From the perspective of a cooperative game, the formation of cooperation between the parent and children nodes yields the maximum income.

In order to guarantee that the proposed split method owns good discriminant ability for target classes, we use mutual information [29] to estimate the contribution of each attribute (player) in Equation (5). By convention, if more than a half of the attributes of the coalition S are interdependent with $i \notin S$, then i joining the coalition S produces 1 for the total income of S; otherwise it yields 0. Also, conditional mutual information [29] is used to measure the interdependence between a single attribute $i \notin S$ and the other attributes $j \in S$. Let $\varphi(i, j)$ denote an interdependence index which is expressed as:

$$\varphi(i,j) = \begin{cases} 1, & I(j; class|i) > I(j; class) \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

$\Delta_i(S)$ can be calculated as:

$$\Delta_i(S) = \begin{cases} 1, & I(j; class|i) \ge 0\ and\ \sum_{j\in S} \varphi(i,j) \ge \frac{|S|}{2} \\ 0, & \text{otherwise} \end{cases}, \quad (7)$$

where conditional mutual information is computed as

$$I(j; class|i) = p(j, class, i) \log \frac{p(j,class|i)}{p(j|i)p(class|i)}, \quad (8)$$

and mutual information is computed as

$$I(j; class) = p(j, class) \log \frac{p(j,class)}{p(j)p(class)}. \quad (9)$$

We can obtain the Shapely value of each attribute according to Equations (4) and (6). The proposed CPRF classifier is described in Algorithm 1 in great detail. Given a test sample x (i.e., an connected component labeled in a binary envelope image in our experiment), the class label predicted using the CPRF classifier is obtained as $\hat{f}(x)$. The procedure described in Section III-E is used to fulfill address block location.

VI. Experimental Setup

In this section, we introduce the experiments for locating Chinese postal address blocks. In addition to the proposed descriptors, we implement two component descriptors. These descriptors and seven different classifiers are used as baselines.

A. The Baseline Descriptors and Classifiers

1) The Baseline Descriptors

For comparison purposes, we implement two different descriptors for representing the single component. First, we use

the coordinates of the nine reference points (see Fig. 3) to describe a component. We normalize the x and y coordinates via dividing these by the width and height of the image respectively. This removes the influence of the image size. The normalized coordinate descriptor is named "Component-wise Positions" or "CWP". Second, we compute a set of shape measurements for each component, including "Area", "Convex Area", "Eccentricity", "Equiv Diameter", "Euler Number", "Extent", "Filled Area", "Major Axis Length", "Minor Axis Length", "Orientation", "Perimeter" and "Solidity". The descriptor that comprises these measurements is termed as "Component-wise Shape Measurements" or "CWSM".

2) The Baseline Classifiers

Support Vector Machines (SVM) We test SVM [11] using three kernel functions: linear (SVM-LIN), radial basis function (SVM-RBF) and histogram intersection (SVM-HI). Parameters are obtained using cross-validation on the validation dataset.

Extreme Learning Machines (ELM) We use 1000 hidden neurons and the sigmoid function for the ELM classifier as proposed by Huang *et al*. [20].

Decision Trees (DT) The optimal subset is selected for each split based on exact search [10]. We use the Gini impurity measure [5] as the criterion.

Naive Bayes (NB) The multinomial distribution is used for a naive Bayes classifier [14]. Prior probabilities are estimated from the frequencies of the training class labels.

Random Forests (RF) Given an F dimensional feature vector, a subset of $\lceil\sqrt{F}\rceil$ features are randomly selected as Breiman [5] proposed. We use the Gini impurity measure [5] to conduct feature and decision boundary selections for each branch node of the subset. In total, 200 decision trees are used.

When the proposed cooperative profit random forests (CPRF) classifier is used, feature vectors are quantized into [1, 8] for CWSC and CWLSS, as the computation of mutual information only accepts discrete inputs. Considering the dimensionality of CWSC feature vectors is high, a subset of $\lceil\sqrt{F}/3\rceil$ features is randomly selected for efficiency purposes. The other setup is the same as that used for the original random forests.

B. Dataset

To our knowledge, there is no publicly available Chinese envelope image dataset. One possible reason is due to the privacy issue. In our experiments, we captured 800 envelope images in total. These images contain handwritten and/or machine-printed text, various spatial layouts, different sizes and orientations of fonts, and the text with different watermarks or shading. The destination address block in each binary envelope image is manually labeled using a bounding box as the ground-truth data. The 800 images are randomly divided into three subsets: training, validation and test, which contain 300, 300 and 200 images respectively.

C. Performance Measures

We use *Precision* and *Recall* as performance measures. In the context of binary (positive or negative) classification, (1) *True Positives* (TP) mean the positive samples which are classified into the positive class; (2) *False Positives* (FP) stand for the negative samples which are labeled as the positive class; and (3) *False Negatives* (FN) denote the positive samples which are classified into the negative class. *Precision* ($\in [0,1]$) is defined as the fraction of the number of true positives and the total number of true positives and false positives (see Equation (10)), while *Recall* ($\in [0,1]$) is defined as the fraction of the number of true positives and the total number of true positives and false negatives (see Equation (11)). These measures are computed across all the test images in this study.

$$Precision = \frac{TP}{TP+FP}, \tag{10}$$

$$Recall = \frac{TP}{TP+FN}. \tag{11}$$

VII. EXPERIMENTAL RESULTS

In this section, we report the results obtained for destination address block location. First, our method is assessed in five experiments. In Experiment I, we test the proposed proximity descriptors using different region radii. In Experiments II and III, we examine the proposed proximity descriptors using different numbers of distance and angle bins quantized for computing these descriptors, respectively. We report the computational time cost and examine the noise resistance ability of our method in Experiments IV and V respectively. Then, we compare the results derived using the proposed method with those obtained using five popular baseline classifiers, those reported in existing studies [7], [37], those obtained using convolutional neural networks (CNN) [24], [34] features, and those derived using our method after text line detection is applied to the images. In addition, we generalize the proposed method to a new Chinese envelope image dataset.

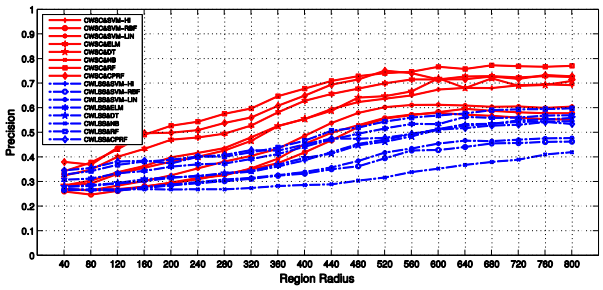
A. Address Block Location Using Proximity Descriptors

1) Using Different Region Radii (R)

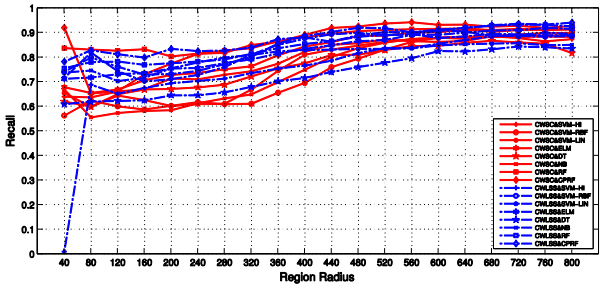
We examine the two proximity descriptors using different region radii. The angle and distance of both the descriptors are quantized into six and four bins respectively. The CPRF and seven baseline classifiers are tested along with each descriptor.

Fig. 6 (a) shows the precision values derived using different combinations of proximity descriptors and classifiers. It shows that: (1) the CWSC descriptor normally outperforms CWLSS when the same classifier is used; (2) the proposed CPRF classifier performs comparably to, or slightly worse than, the original random forests classifier [5]; (3) the two random forests and the ELM classifiers [20] perform better than their counterparts; (4) the performances of the histogram intersection, linear and radial basis function SVM classifiers distribute in a descending order; (5) the combination of the CWSC descriptor and the original RF classifier [5] performs the best while the CWLSS descriptor and naïve Bayes [14] performs the worst; and (6) the performance obtained using different combinations of descriptors and classifiers normally increases when large regions are used. However, the performance becomes relatively stable when the region radius is over 680 pixels.

Furthermore, the recall values produced by different combinations of proximity descriptors and classifiers are shown in Fig. 6 (b). We observe that: (1) the CWSC descriptor yields the similar performance to that it generates when the precision measure is applied, while the performance of CWLSS greatly varies; (2) the CWLSS descriptor performs better, when the radial basis function or linear SVM is used, than the case



(a)



(b)

Fig. 6. The performance measures: precision (a) and recall (b) obtained using different combinations of proximity descriptors ($A = 6$ and $D = 4$) and classifiers when various region radii are used.

TABLE I
THE BEST PERFORMANCE OBTAINED USING DIFFERENT COMBINATIONS OF DESCRIPTOR AND CLASSIFIER ACROSS DIFFERENT REGION RADII WHEN TWO MEASURES ARE CONSIDERED SEPARATELY

		SVM-HI	SVM-RBF	SVM-LIN	ELM
Precision	**CWP**	0.20	0.22	0.21	0.25
	CWSM	0.21	0.17	0.17	0.28
	CWSC	0.70	0.60	0.61	0.73
	CWLSS	0.55	0.46	0.48	0.57
Recall	**CWP**	0.06	0.77	0.83	0.36
	CWSM	0.77	0.48	0.48	0.71
	CWSC	0.89	0.90	0.89	0.91
	CWLSS	0.90	0.92	0.90	0.92
		DT	**NB**	**RF**	**CPRF**
Precision	**CWP**	0.24	0.20	0.00	0.24
	CWSM	0.43	0.23	0.63	0.16
	CWSC	0.72	0.58	**0.77**	0.75
	CWLSS	0.56	0.42	0.60	0.54
Recall	**CWP**	0.09	0.60	0.00	0.33
	CWSM	0.21	0.34	0.14	0.38
	CWSC	0.87	0.88	0.92	**0.94**
	CWLSS	0.84	0.86	0.93	**0.94**

Bold fonts indicate the best result when each performance measure is considered (this continues in the following tables).

when the histogram intersection SVM is used; (3) the proposed CPRF classifier outperforms its counterparts in most of the cases, when the same descriptor is used; (4) the ELM classifier [20] normally outperforms the decision trees [10] and three SVM classifiers [11]; (5) different combinations of descriptors and classifiers perform stably when the region radius is over 680 pixels; and (6) the recall values derived are normally higher than the precision values obtained in the same condition.

Finally, Table I reports the best precision and recall values obtained across varying region radii. Meanwhile, the performance of the two baseline descriptors is also shown. As can be seen, the two proposed descriptors always significantly outperform the baseline descriptors. The best recall value: 0.94 is obtained using the proposed CPRF classifier while the best precision value 0.77 is provided by the original RF classifier [5].

Precision: 1 and Recall: 1

(a)

Precision: 1 and Recall: 0.99

(b)

Fig. 7. Two groups of binary envelope images and the corresponding address block location resultant images obtained using CWSC and CPRF. In each group, the image displayed at the left side is the binary image while the image shown at the right side is the resultant image. The precision and recall values are shown below the images.

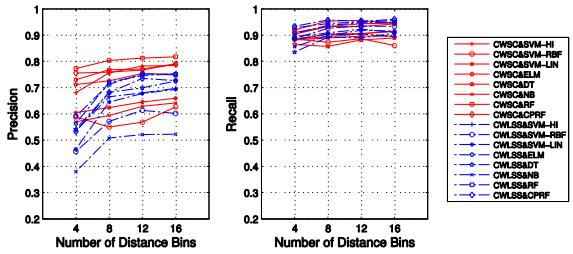


Fig. 8. The values of the precision and recall obtained using different combinations of proximity descriptors and classifiers when different numbers of distance bins (D) are used ($A = 6$).

Fig. 7 shows two groups of binary envelope images and destination address block images obtained using the CWSC descriptor with the CPRF classifier. It can be seen that this combination yields good location results.

2) Using Different Numbers of Distance Bins (D)

In this experiment, we investigate the effect of the number of distance bins on the proposed descriptors. Only the region radius of 680 pixels is considered as the performance of the proposed descriptors becomes stable when the used radius is larger than this value in the previous experiment. We test four different distance bin numbers, i.e., D ($D \in \{4,8,12,16\}$), while keeping the number of angle bins constant ($A = 6$). The precision and recall values obtained using the CWSC and CWLSS descriptors are plotted in Fig. 8. It shows that: (1) the recall values obtained become relatively stable, or even drop, when over 12 distance bins are quantized while this does not stand when the precision values are considered; (2) given that the same classifier is used, the CWSC descriptor normally outperforms CWLSS when precision is considered; (3) the recall values obtained are higher than the precision values for the same combination of descriptors and classifiers; (4) the two random forests and ELM classifiers [20] normally outperform the other approaches no matter which proximity descriptor is used; and (5) the proposed CPRF classifier produces better recall values than its counterparts.

3) Using Different Numbers of Angle Bins (A)

We also examine the effect of the number of angle bins (i.e., A) on the CWSC and CWLSS descriptors. Here, $R = 680$ is

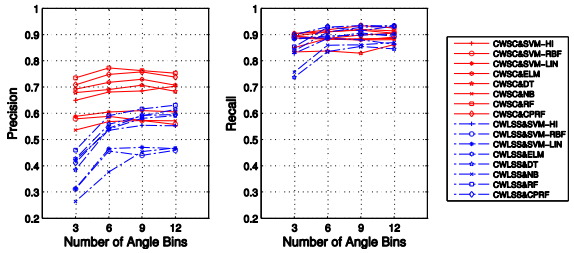


Fig. 9. The values of the precision and recall obtained using different combinations of proximity descriptors and classifiers when different numbers of angle bins (A) are used ($D = 4$).

TABLE II
THE TOTAL TRAINING TIME (SECONDS) AND AVERAGE TEST TIME PER IMAGE REQUIRED BY CWSC AND CWLSS ($R = 680$, $A = 6$ AND $D = 16$) WHEN DIFFERENT CLASSIFIERS ARE USED

		SVM-HI	SVM-RBF	SVM-LIN	ELM
CWSC	Training	590.02	2452.23	926.28	14.09
	Test	2.42	10.55	3.77	0.26
CWLSS	Training	256.17	481.77	209.83	20.89
	Test	0.43	1.38	0.55	0.29
		DT	**NB**	**RF**	**CPRF**
CWSC	Training	34.40	3.76	1117.50	5736.13
	Test	0.01	0.06	0.03	0.18
CWLSS	Training	2.29	0.14	125.64	4537.68
	Test	0.004	0.04	0.02	0.07

used. The number of angle bins is set to 3, 6, 9 and 12. On the other hand, the number of distance bins is set as $D = 4$. Fig. 9 illustrates the precision and recall values obtained using the CWSC and CWLSS descriptors. As can be seen, (1) the precision or recall values look similar, or even drop, when more than six angle bins are used; (2) given a classifier, CWSC normally performs better than CWLSS when precision is considered while this is not the case when recall is considered; (3) in terms of a combination of descriptors and classifiers, the recall values obtained are higher than the precision values; (4) the performance of the other classifiers is inferior to that of the two random forests and ELM classifiers [20] when each of the two proposed descriptors is used; and (5) the CPRF classifier usually yields higher recall values than the other classifiers.

4) Computational Time Cost

The experiments have been conducted on a laptop with a 64-bit, 2.50 GHz Intel(R) i7-4710MQ CPU and 16.0 GB memory. The total training time and the average test time per image required by CWSC and CWLSS ($R = 680$, $A = 6$ and $D = 16$) along with different classifiers are reported in Table II. As can be seen, the test time required by CPRF is reasonable even if the training of this classifier is time-consuming. However, the training can be off-line performed. Therefore, the CPRF classifier provides a proper efficiency and effectiveness for our application.

5) Noise Resistance

In order to examine the noise resistance ability of the proposed postal address block location methodology, we add four different levels (SNR or Signal-to-Noise Ratio) of Gaussian noise to the training and test images. The address block location experiment is performed on these images. For simplicity, only the CWSC descriptor with the "optimal" parameters (i. e., $R = 680$, $A = 6$ and $D = 16$) is tested together with both the RF and CPRF classifiers. Table III presents the results in detail. Compared to the precision: 0.79 and recall: 0.96 obtained using the original images, the proposed address block location methodology manifests strong noise resistance ability when no less than 25 dB Gaussian noise is added to the images. Although when the noise level: SNR reaches to 0, the recall obtained using the proposed methodology is almost not affected.

TABLE III
THE PRECISION AND RECALL VALUES OBTAINED USING RF AND CPRF TOGETHER WITH THE BEST CWSC FEATURES WHEN FOUR DIFFERENT LEVELS (SNR: DB) OF NOISE IS ADDED TO IMAGES

	SNR (dB)	0	25	50	100
RF	Precision	0.38	0.80	0.80	0.80
	Recall	0.94	0.94	0.95	0.94
CPRF	Precision	0.37	0.77	0.77	0.77
	Recall	0.96	0.95	0.96	0.95

TABLE IV
THE PRECISION AND RECALL VALUES OBTAINED USING THE CWSC AND CWLSS DESCRIPTORS ($R = 680$, $A = 6$ AND $D = 16$) TOGETHER WITH DIFFERENT WEKA'S [19] BASELINE CLASSIFIERS

	Classifier	C4.5	ZeroR	REPTree	LMT	RandomForest
CWSC	Precision	0.78	0.00	0.74	**0.81**	0.80
	Recall	0.90	0.00	0.93	0.92	**0.95**
CWLSS	Precision	0.72	0.00	0.72	0.73	**0.75**
	Recall	0.91	0.00	0.93	0.92	**0.96**

B. Comparison with Popular Baseline Classifiers

We also test the proposed CWSC and CWLSS descriptors along with the free classifier tool: Weka [19]. For simplicity, only the "optimal" parameters: $R = 680$, $A = 6$ and $D = 16$ are used for both descriptors. In total, five baseline classifiers provided by Weka [19] are tested, including C4.5, ZeroR, REPTree, LMT (Logistic Model Trees) and RandomForest. The results are shown in Table IV. Compared with the results displayed in Fig. 8, it can be seen that Weka's random forest classifier generates almost the same result as that we have obtained. However, the other Weka's classifiers usually produce inferior results to those derived using the random forest classifier.

C. Comparison with Existing Chinese Postal Address Block Location Studies

In this study, both the highest recall values obtained using the proposed CWSC and CWLSS descriptors are 0.96 (see Fig. 8). In contrast, the highest recall values (which are identical to the accuracy [7] and the extraction rate [37]) reported in [7] and [37] are 0.88 and 0.82 respectively. Although we cannot use these values as the baselines, our experimental results suggest that the proposed destination address block location methods, i.e., the combinations of the CWSC or CWLSS descriptors and cooperative profit random forests, produce better results than those derived in the existing studies [7], [37].

D. Comparison with Convolutional Neural Networks Features

We extract features from the penultimate fully-connected layer of two different pre-trained CNN models: Alex-Net [24] and GoogLeNet [34]. Also, we fine-tune these models using our labeled dataset. Correspondingly, two additional sets of features are extracted from the fine-tuned Alex-Net [24] and GoogLeNet [34] models. The CPRF classifier is used because its advantage over the other classifiers has been shown. The

TABLE V
THE PRECISION AND RECALL VALUES OBTAINED USING FOUR SETS OF CNN FEATURES AND THE BEST CWSC FEATURES ($R = 680$, $A = 6$ AND $D = 16$)

	Pre-trained Alex-Net	Fine-Tuned Alex-Net	Pre-trained GoogLeNet	Fine-Tuned GoogLeNet	Best CWSC
Precision	0.75	**0.79**	0.70	0.78	**0.79**
Recall	0.94	0.94	0.85	0.95	**0.96**

TABLE VI
THE PRECISION AND RECALL VALUES OBTAINED USING THE CWSC DESCRIPTOR WHEN $R = 680$, $A = 6$ AND $D = 16$ ARE USED

	Maximum Sliding Window Scale [43]						Best CWSC
	1	2	3	4	6	8	
Precision	0.50	0.67	0.75	0.76	0.75	0.75	**0.79**
Recall	0.92	0.93	0.93	0.95	0.95	0.95	**0.96**

training and prediction operations are the same as those performed for the proposed descriptors. The precision and recall values obtained using the four sets of CNN features are shown in Table V. It is observed that fine-tuning improves the performance of the CNN features; while the best performance of these features is close to the best results obtained using the proposed CWSC descriptor. One possible reason is that the insufficient number of the envelope images may limit the performance of Alex-Net [24] and GoogLeNet [34].

E. Using Text Line Detection as Pre-processing

Since text line detection techniques normally detect the text information only, we use the method that Zhang *et al*. [43] proposed as a pre-processing step. This method is used before the image segmentation operation is performed. We use all the default parameters excepting "Maximum sliding window scale", which is tuned. We only test the CWSC descriptor using its "optimal" parameters (i.e., $R = 680$, $A = 6$ and $D = 16$) along with the CPRF classifier. Table VI reports the precision and recall values obtained using this combination. It can be seen that the best performance is obtained when "Maximum sliding window scale" is set to 4. In this case, the precision and recall values obtained are 0.76 and 0.95 respectively. Correspondingly, these performance values obtained without using the text line detection pre-processing are 0.79 and 0.96 respectively. This result shows that the pre-processing of text line detection does not boost the performance of the proposed CWSC descriptor. This may be due to the fact that the text line detection model is trained over limited samples.

F. Generalization to a New Dataset

To augment the results that we have derived, we further collect 600 Chinese envelope images. We randomly divide these images into two equal-sized subsets: training and test. The CWSC and CWLSS descriptors with the "optimal" parameters: $R = 680$, $A = 6$ and $D = 16$ are then applied to the new dataset along with different classifiers. We report the results obtained using the new dataset in Table VII. It can be observed that the performance values are higher than those (see Fig. 8) obtained using the original images. By checking the new dataset, it is found that the proportion of machine-printed envelope images is higher than that of those images included in the original dataset. Since the destination address blocks shown on machine-printed envelopes are more compact and standard than those occur on hand-written envelopes, it is more effective

TABLE VII
THE PRECISION AND RECALL VALUES OBTAINED USING DIFFERENT COMBINATIONS OF DESCRIPTOR ($R = 680$, $A = 6$ AND $D = 16$) AND CLASSIFIER ON THE NEW CHINESE ENVELOPE IMAGE DATASET

		SVM-HI	SVM-RBF	SVM-LIN	ELM
CWSC	**Precision**	0.95	0.76	0.92	0.94
	Recall	0.98	0.95	0.98	0.98
CWLSS	**Precision**	0.89	0.93	0.96	0.97
	Recall	0.97	0.95	0.95	0.97
		DT	**NB**	**RF**	**CPRF**
CWSC	**Precision**	0.93	0.88	**0.97**	0.95
	Recall	0.97	0.97	0.98	**0.99**
CWLSS	**Precision**	0.89	0.88	**0.98**	0.97
	Recall	0.97	0.95	0.97	**0.98**

to train a classifier using the new dataset than the original one. This fact should account for the better results shown in Table VII. It can also be seen that the CPRF classifier still produces higher recall values than its counterparts. This finding is consistent with that we observed in the original experiments.

VIII. CONCLUSIONS

In this paper, we introduced two proximity descriptors to represent the specific regions on Chinese envelope images. We also proposed a random forests classifier by using a new split method based on the Shapley value [31] known in the cooperative game theory [6]. This classifier is referred to as "Cooperative Profit Random Forests" or "CPRF". The proposed descriptors were applied to locating destination address blocks on Chinese envelope images along with the CPRF classifier. The results showed that both the proposed descriptors were superior to two baseline descriptors. The joint use of each of the proposed descriptors and CPRF generated the best recall performance. This performance was even higher than those derived using four sets of CNN features at the same conditions. The promising performance should be attributed to the fact that the proposed descriptors are able to encode the spatial layout of the components contained in a local region by exploiting proximity cues rather than only capturing the shape characteristics of individual components. Besides, the proposed CPRF classifier produced higher recalls than those yielded by seven state-of-the-art classifers. This result should be due to the effective tree node split method that the CPRF classifier uses.

REFERENCES

[1] J.C. Banerjee, "Gestalt theory of perception," *Encyclopaedic Dictionary of Psychological Terms*, Publications Pvt. Ltd, 1994.
[2] S. Basu, N. Das,R. Sarkar, M. Kundu, M. Nasipuri, and D. K. Basu, "A novel framework for automatic sorting of postal documents with multi-script address blocks," *Pattern Recognit*, vol. 43, no. 10, pp. 3507-3521, 2010.
[3] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509-522, 2002.
[4] C. Bertoncini, K. Rudd, B. Nousain, and M. Hinders, "Wavelet Fingerprinting of Radio-Frequency Identification (RFID) Tags," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 12, 4843-4850, 2012.
[5] L. Breiman, "Random forests," *Mach Learn.*, vol. 45, no. 1, pp. 5-32, 2001.
[6] G. Chalkiadakis, E. Elkind, and M. Wooldridge, "Computational Aspects of Cooperative Game Theory", *ser. Synthesis Lectures on Artificial Intelligence and Machine Learning*. *Morgan and Claypool Publishers*, 2011.

[7] M.L. Cheng and J.H. Xu, "Address block localization for Chinese postal envelopes with clutter background," in: *Proc. International Conference on Fuzzy Systems and Knowledge Discovery*, 2014, pp. 638-643.
[8] China National Standard for Postal Envelope Writing/Printing (GB/T 22657.1-2008).
[9] C. Cho, B. Chung, and M. Park, Development of Real-Time Vision-Based Fabric Inspection System, *IEEE Transactions on Industrial Electronics*, vol. 52, no. 4, 1073-1079, 2005.
[10] D. Coppersmith, S. J. Hong, and J. R. M. Hosking, "Partitioning Nominal Attributes in Decision Trees," *Data Mining and Knowledge Discovery*, vol. 3, no. 2, pp. 197-217, 1999.
[11] C. Cortes and V. Vapnik, "Support-vector networks," *Mach Learn.*, vol. 20, no. 3, pp. 273-297, 1995.
[12] A. Criminisi, J. Shotton and E. Konukoglu, "Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," *Foundations and Trends in Computer Graphics and Vision*, vol. 7, no. 2-3, pp. 81-227, 2012.
[13] M.B. Dillencourt, H. Samet and M. Tamminen, "A general approach to connected-component labeling for arbitrary image representations," *Journal of ACM*, vol. 39, no. 2, pp. 253-280, 1992.
[14] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, no. 2, pp. 103-137, 1997.
[15] X. Dong and M.J. Chantler, "The importance of long-range interactions to texture similarity," *Computer Analysis of Images and Patterns*, LNCS, 8047 Springer Berlin/Heidelberg, pp. 425-432, 2013.
[16] X. Dong, J.Y. Dong and S.K. Wang, "Segmentation of Chinese postal envelope images for address block location," *Advances in Visual Computing*, LNCS, 5876 Springer Berlin/Heidelberg, pp. 558-567, 2009.
[17] D. Gaceb, V. Eglin, F. Lebourgeois and H. Emptoz, "A new pyramidal approach for the address block location based on hierarchical graph coloring," *Image Analysis and Recognition*, LNCS, 4633 Springer Berlin/Heidelberg, pp. 1276-1288, 2007.
[18] V. Govindaraju and S. Tulyakov, "Postal address block location by contour clustering," in: *Proc. Seventh International Conference on Document Analysis and Recognition*, 2003, pp. 429-432.
[19] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10-18, 2009.
[20] G. Huang, L. Chen, and C. Siew, "Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes," *IEEE Trans. Neural Networks*, vol. 17, no. 4, pp. 879-892, 2006.
[21] H. Iwata, and S. Sugano, "Human-Robot-Contact-State Identification Based on Tactile Recognition," *IEEE Transactions on Industrial Electronics*, vol. 52, no. 6, 1468-1477, 2005.
[22] A.K. Jain and S.K. Bhattacharjee, "Address block location on envelopes using Gabor filters," *Pattern Recognit.*, vol. 25, no. 12, pp. 1459-1477, 1992.
[23] T. Kagehiro, M. Koga, H. Sako, and H. Fujisawa, "Address-block extraction by Bayesian rule," in: *Proc. the 17th International Conference on Pattern Recognition*, 2004, pp. 582-585.
[24] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," in: *Proc. Advances in Neural Information Processing Systems*, 2012, pp. 1106-1114.
[25] X. Li, S. K. Tso, X. Guan, and Q. Huang, "Improving Automatic Detection of Defects in Castings by Applying Wavelet Technique," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 6, 1927-1934, 2006.
[26] D. Menoti and D.L. Borges, "Segmentation of envelopes and address block location by salient features and hypothesis testing," *Journal of Computer Science*, vol. 6, no. 1, pp. 66-79, 2007.
[27] W. Niblack, *An introduction to digital image processing*, N.J.. Prentice Hall, 1986.
[28] R. Radha and R.R. Aparna, "A novel approach for Postal address block and Pincode segmentation," *International Journal of Imaging and Robotics*, vol. 14, no. 3, pp. 86-96, 2014.
[29] D.L. Schilling, "Elements of Information Theory," *Hoboken, NJ, USA: Wiley*, 2003.
[30] S. Schmidt, S. Manschitz, C. Rensing and R. Steinmetz, "Extraction of Address Data from Unstructured Text using Free Knowledge Resources," in: *Proc. the 13th International Conference on Knowledge Management and Knowledge Technologies*, 2013.
[31] L.S. Shapley, "A value for n-person games," in: A.W.T.H.W. Kuhn (Ed.), *In Contributions to the Theory of Games*, vol. 2, *Princeton University Press*, 1953, pp. 307–317.
[32] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in: *Proc. the 25th IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
[33] X. Sun, Y. Liu, J. Li, J. Zhu, X. Liu, and H. Chen, "Using cooperative game theory to optimize the feature selection problem," *Neurocomputing*, vol. 97, no. 15, pp. 86-93, 2012.
[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," in: *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
[35] C. Tong and X. Shi, "Decentralized Monitoring of Dynamic Processes Based on Dynamic Feature Selection and Informative Fault Pattern Dissimilarity," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 6, 3804-3814, 2016.
[36] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, The MIT Press, 1964.
[37] J.L. Xue, X.Q. Ding, C.S. Liu, R. Zhang, and W.W. Qian, "Location and interpretation of destination addresses on handwritten Chinese envelopes," *Pattern Recognition Letter*, vol. 22, no. 6-7, pp. 639-656, 2001.
[38] S.D. Yanowitz and A.M. Bruckstein, "A new method for image segmentation," in: *Proc. International Conference on Pattern Recognition*, 1988, pp. 270-275.
[39] Q. Ye and D. Doermann, "Text detection and recognition in imagery: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 7, pp. 1480-1500, 2015.
[40] E.A. Yonekura and J. Facon, "Postal envelope segmentation by 2-D histogram clustering through watershed transform," in: *Proc. International Conference on Document Analysis and Recognition*, 2003, pp. 338-342.
[41] D. You, X. Gao, and S. Katayama, "WPD-PCA-Based Laser Welding Process Monitoring and Defects Diagnosis by Using FNN and SVM," *IEEE Trans. on Industrial Electronics*, vol. 62, no. 1, 628-636, 2015.
[42] B. Yu, A.K. Jain, and M. Mohiuddin, "Address block location on complex mail pieces," in: *Proc. International Conference on Document Analysis and Recognition*, vol. 2, 1997, pp. 897-901.
[43] Z. Zhang, W. Shen, C. Yao, and X. Bai, "Symmetry-based text line detection in natural scenes," in: *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2558-2567.

Xinghui Dong received his PhD degree from Heriot-Watt University, UK, in 2014.He is currently working as a Research Associate in the Centre for Imaging Sciences, The University of Manchester, UK. His research interests include automatic defect detection, image representation, texture analysis and visual perception.

Junyu Dong received his BSc and MSc from the Department of Applied Mathematics at Ocean University of China in 1993 and 1999 respectively. He received his PhD in Image Processing in 2003 from Heriot-Watt University. Dr. Dong joined Ocean University of China in 2004 and he is currently a professor and the Head of the Department of Computer Science and Technology. His research interests include machine learning, computer vision and underwater image processing.

Huiyu Zhou received his BEng from Huazhong University of Science and Technology of China, and an MSc degree from University of Dundee, UK, respectively. He was then awarded a PhD degree in computer vision from Heriot-Watt University, UK. He currently is a lecturer at Queens University Belfast, UK. He has taken part in the consortium of a number of research projects in medical image processing, computer vision, intelligent systems, and data mining. He has published widely in the field.

Jianyuan Sun is currently pursuing the Ph.D. degree in computer application technology with the Department of Computer Science and Technology. Her research interests include pattern recognition, image processing, and machine learning.

Dacheng Tao (F'15) is Professor of Computer Science and ARC Laureate Fellow in the School of Information Technologies and the Faculty of Engineering and Information Technologies, and the Inaugural Director of the UBTECH Sydney Artificial Intelligence Centre, at the University of Sydney. He mainly applies statistics and mathematics to Artificial Intelligence and Data Science. His research interests spread across computer vision, data science, image processing, machine learning, and video surveillance.