Research Article

Active Discriminative Dictionary Learning for Weather Recognition

Caixia Zheng,[1,2] Fan Zhang,[1] Huirong Hou,[1] Chao Bi,[1,2] Ming Zhang,[1,3] and Baoxue Zhang[4]

[1]*School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China*
[2]*School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China*
[3]*Key Laboratory of Intelligent Information Processing of Jilin Universities, Northeast Normal University, Changchun 130117, China*
[4]*College of Statistics, Capital University of Economics and Business, Beijing 100070, China*

Correspondence should be addressed to Ming Zhang; zhangm545@nenu.edu.cn and Baoxue Zhang; bxzhang@nenu.edu.cn

Weather recognition based on outdoor images is a brand-new and challenging subject, which is widely required in many fields. This paper presents a novel framework for recognizing different weather conditions. Compared with other algorithms, the proposed method possesses the following advantages. Firstly, our method extracts both visual appearance features of the sky region and physical characteristics features of the nonsky region in images. Thus, the extracted features are more comprehensive than some of the existing methods in which only the features of sky region are considered. Secondly, unlike other methods which used the traditional classifiers (e.g., SVM and K-NN), we use discriminative dictionary learning as the classification model for weather, which could address the limitations of previous works. Moreover, the active learning procedure is introduced into dictionary learning to avoid requiring a large number of labeled samples to train the classification model for achieving good performance of weather recognition. Experiments and comparisons are performed on two datasets to verify the effectiveness of the proposed method.

1. Introduction

Traditional weather detection depends on expensive sensors and is restricted to the number of weather stations. If we can use existing surveillance cameras capturing images from the local environment to detect weather conditions, it would be possible to turn weather observation and recognition into a low-cost and powerful computer vision application. In addition, most of current computer vision systems are designed to execute in clear weather [1]. However, in many outdoor applications (e.g., driver assistance systems [2], video surveillance [3], and robot navigation [4]), there is no escape of "bad" weather. Hence, research of weather recognition based on images is in urgent demand, which can be used to recognize the weather conditions for many vision systems to adaptively turn their models or adjust parameters under the different weathers.

To date, despite its remarkable value, there are only a few of works that have been proposed to address the weather recognition problem. In [5, 6], some researchers proposed weather recognition models (sunny or rainy) from images captured by in-vehicle cameras. However, these models heavily relied on prior information of vehicles, which may weaken their performances. Song et al. [7] proposed a method to classify traffic images into sunny, fog, snowy, and rainy. Their method extracted several features such as the image inflection point, power spectral slope, and image noise and used K-nearest neighbor (K-NN) as classification model. This method applied only to classify the weather conditions of the traffic scene. Lu et al. [8] applied collaborative learning approach to label the outdoor image as either sunny or cloudy. In their method, an existence vector is firstly introduced to indicate the confidence in the corresponding weather feature being present in the given image. Then, the existence vector and weather features are combined for weather recognition. However, this method involved many complicated technologies (shadow detection, image matting, etc.); thus its performance largely depended on the accuracies of these technologies. Chen et al. [9] employed support vector machine (SVM) with the help of active learning to classify the

weather conditions of images into sunny, cloudy, or overcast. Nevertheless, they only extracted the features from sky part of the images and the useful information in nonsky part of images is neglected.

In general, the above methods have been successfully applied to the applications of weather recognition. However, they still suffer from the following limitations. Firstly, since these methods just extracted the features (e.g., SIFT, LBP, and HSV color) from whole images or only sky part of images, they neglected the different useful information in the sky part and nonsky part of images. In the sky part of images, the distribution of cloud and the color of sky are key factors for weather recognition, so the visual appearance feature such as the texture, color, and shape should be extracted. In the nonsky part of images, the features based on physical properties which can characterize the changes of images caused by the varying weather conditions and thus image contrast [5] and dark channel [10] should be considered. Secondly, these methods directly used K-NN or SVM based algorithms to classify the different weather conditions. Although K-NN is a simple classifier and it is easy to implement, it is not robust enough for the complicated real-world images in the practical application while SVM is a complex classifier and it is difficult to select the appropriate kernel function. Lastly, these methods required a large amount of labeled data for training, which is often expensive and seldom available.

To address the above problems, we propose a novel framework to recognize different weather conditions (sunny, cloudy, and overcast) from outdoor images acquired by a static camera looking at the same scene over a period of time. The proposed method extracts not only the features from the sky parts of images which are relevant to the visual manifestations of different weather conditions, but also the features based on physical characteristics in the nonsky parts. Thus, the extracted features are more comprehensive for distinguishing the images captured under various weather situations. Unlike other methods which used the traditional classifier (e.g., SVM, K-NN), the discriminative dictionary learning is used as the classification model. Moreover, in order to achieve good performance of weather recognition with a few labeled samples, active discriminative dictionary learning algorithm (ADDL) is proposed. In ADDL, the active learning procedure is introduced to the dictionary learning which selects the informative and representative samples for learning an impact and discriminated dictionary to classify weather conditions. As far as we know, the proposed framework is the first approach which combines the active learning technology into the dictionary learning for weather recognition.

The rest of this paper is organized as follows. Section 2 briefly reviews some related work. Section 3 presents the details of the feature extraction and the proposed ADDL algorithm. Extensive experiments and comparisons are conducted in Section 4, and Section 5 is the conclusion of the paper.

2. Related Work

2.1. Dictionary Learning.
Dictionary learning has emerged in recent years. It is one of the most popular tools for learning the intrinsic structure of images and has achieved state-of-the-art performances in many computer vision and pattern recognition tasks [11–13]. The unsupervised dictionary learning algorithms such as K-SVD [14] have achieved satisfactory results in image restoration, but they are not suitable for classification tasks because it only requires that the learned dictionary could faithfully represent the training samples. By exploring the label information of training dataset, some supervised dictionary learning approaches have been proposed to learn a discriminative dictionary for the classification task. Among these methods, one may directly use training data of all classes as the dictionary, and the test image can be classified by finding which class leads to the minimal reconstruction error. Such a naive supervised dictionary learning method is called sparse representation based classification (SRC) algorithm [15], which has shown good performance in face recognition. However, it was less effective in classification when the raw training images include the noise and trivial information and cannot sufficiently exploit the discriminative information in the training data. Fortunately, this problem can be addressed by properly learning a dictionary from the original training data. After the test samples are encoded over the learned dictionary, both the coding residual and the coding coefficients can be employed for identifying the different classes of samples [16]. Jiang et al. [17] proposed a method named label consistent K-SVD (LC-K-SVD), which encouraged samples from the same class to have similar sparse codes by applying a binary class label sparse code matrix. Fisher discrimination dictionary learning (FDDL) method [18] was proposed based on the Fisher criterion, which learned a structured dictionary to distinguish samples from different classes. In most existing supervised dictionary learning methods, l_0-norm or l_1-norm sparsity regularization was adopted. As a result, they often suffer heavy computational costs in both training and testing phases. In order to address this limitation, Gu et al. [19] developed a projective dictionary pair learning (DPL) algorithm, which learned an analysis dictionary together with a synthesis dictionary to attain the goal of signal representation and discrimination. DPL method can successfully avoid solving the l_0-norm or l_1-norm to accelerate the training and test process, so we adopt the DPL algorithm as the classification model in our work.

2.2. Active Learning.
Active learning, which aims to construct an efficient training dataset for improving the classification performance through iterative sampling, has been well studied in the computer vision fields. According to [20], the existing active learning algorithms can be generally divided into three categories: stream-based [21, 22], membership query synthesis [23, 24], and pool-based [25–27]. Among them, pool-based active learning is most widely used for real-world learning problems because it assumes that there is a small set of labeled data and a large pool of unlabeled data available. This assumption is consistent with the actual situation. In this paper, we adopt the pool-based active learning.

The crucial point of pool-based active learning is how to define a strategy to rank the unlabeled sample in the pool. There are two criteria, informativeness and representativeness, which are widely considered for evaluating unlabeled
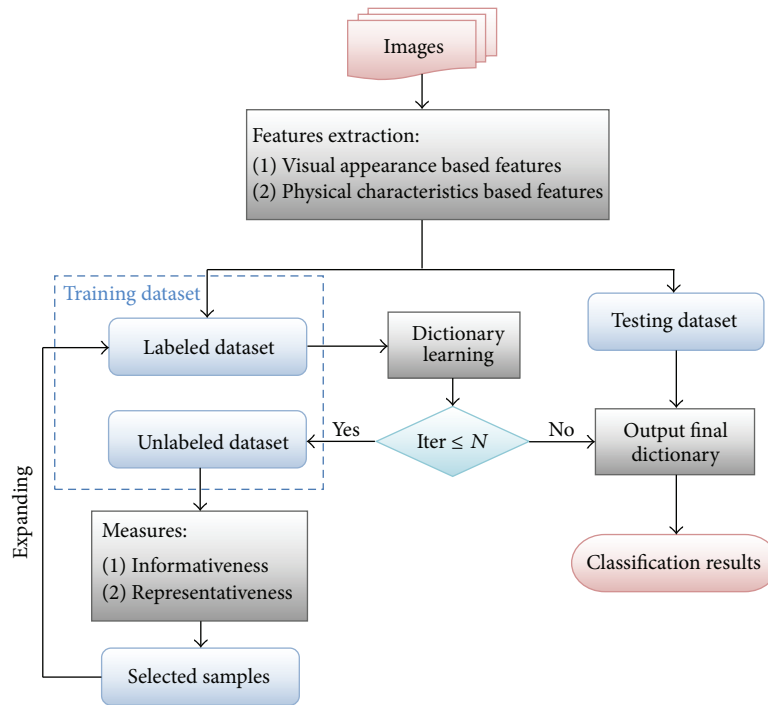
FIGURE 1: The flow chart of the proposed method.

samples. Informativeness measures the capacity of the samples in reducing the uncertainty of the classification model, while representativeness measures whether the samples well represent the overall input patterns of unlabeled data or not [20]. Most of active learning methods only take one of the two criteria into account when selecting unlabeled samples, which restricts the performance of the active learning [28]. Although several approaches [29–31] have considered both informativeness and representativeness of unlabeled data, to our knowledge, almost no researchers have introduced these two criteria of active learning into the dictionary learning algorithms.

3. Proposed Method

In this section, we present an efficient scheme, including the effective feature extraction and ADDL classification model, to classify the weather conditions of images into sunny, cloudy, and overcast. Figure 1 shows a flow chart of the overall procedure of our method. First, the visual appearance based features are extracted from the sky region and the physical characteristics based features are extracted from the nonsky region of images. Secondly, the labeled training dataset is used to learn an initial dictionary, and then the samples are iteratively selected from unlabeled dataset based on two measures: informativeness and representativeness. These selected samples are used to expand the labeled dataset for learning a more discriminative dictionary. Finally, the testing dataset is classified by the learned dictionary.

3.1. Features Extraction. Feature extraction is an essential preprocessing step in pattern recognition problems. In order to express the difference among the images of the same scene taken under various weather situations, we analyze the different visual manifestations of images caused by different weather conditions and extract the features which describe both visual appearance properties and physical characteristics of images.

From the viewpoint of visual appearance features of images, the sky is the most important part in an outdoor image for identifying the weather. In the case of sunny, the sky appears blue because the light is refracted as it passes through the atmosphere, scattering blue light, while, under the overcast condition, the sky is white or grey due to the thick cloud cover. The cloudy day is a situation between sunny and overcast. On a cloudy day, the clouds are floating in the sky, which exhibit a wide range of shapes and textures. Hence, we extract the SIFT [32], HSV color, LBP [33], and the gradient magnitude of the sky parts of images as the visual appearance feature of images. These extracted features are used to describe the texture, color, and shape of images for weather classification.

Unlike Chen et al. [9] who directly eliminated the nonsky regions of images, we extract two features in the nonsky parts of images based on the physical characteristics, which also can be used as powerful features to distinguish the different weather conditions. The interaction of light with the atmosphere has been studied as atmospheric optics. In the sunny (clear) day, the light rays reflected by scene objects reach to the observer without alteration or attenuation. However, under bad weather conditions, atmospheric effects cannot be neglected anymore [34–36]. The bad weather (e.g., overcast) causes decay in the image contrast, which is exponential in the depths of scene points [35]. So images of the same scene taken in different weather conditions should have different

contrasts. The contrast CON of an image can be computed by

$$\text{CON} = \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}}, \tag{1}$$

where $E_{\max}$ and $E_{\min}$ are the maximum and minimum intensities of the image, respectively.

Overcast weather may come with haze [8]. The dark channel prior presented in [10] is effective to detect the image haze. It is based on the observation that most haze-free patches in the nonsky regions of images should have a very low intensity at least one RGB color channel. The dark channel J^{dark} was defined as

$$J^{\text{dark}}(x) = \min_{c \in \{r,g,b\}} \left(\min_{x' \in \Omega(x)} \left(J^c\left(x'\right)\right)\right), \tag{2}$$

where J^c is one color channel of the image J and $\Omega(x)$ is a local patch centered at x.

In summary, both the visual appearance features (SIFT, HSV color, LBP, and the gradient magnitude) of the sky region and the physical characteristics based features (the contrast and the dark channel) of the nonsky region are extracted to distinguish the images under different weather situations. And then, the "bag-of-words" model [37] is used to code each feature for forming the feature vectors.

3.2. The Proposed ADDL Classification Model.

Inspired by the technological advancements of discriminative dictionary learning and active learning, we design an active discriminative dictionary learning (ADDL) algorithm to improve the discriminative power of the learned dictionary. In ADDL algorithm, DPL [19] is applied to learn a discriminative dictionary for recognizing different weather conditions, and the strategy of active samples selection is developed to iteratively select the unlabeled sample from a given pool to enlarge the training dataset for improving the DPL classification performance. The criterion of active samples selection includes the informativeness measure and the representativeness measure.

3.2.1. DPL Algorithm.

Suppose there is a set of m-dimensionality training samples from C classes, denoted by $X = [X_1, \ldots, X_c, \ldots, X_C]$ with the label set $Y = [Y_1, \ldots, Y_c, \ldots, Y_C]$, where $X_c \in R^{m \times n}$ denotes the sample set of cth class and Y_c denotes the corresponding label set. DPL [19] algorithm jointly learned an analysis dictionary $P = [P_1; \ldots; P_c; \ldots; P_C] \in R^{kC \times m}$ and a synthesis dictionary $D = [D_1, \ldots, D_c, \ldots, D_C] \in R^{m \times kC}$ to avoid resolving the costly l_0-norm or l_1-norm sparse coding process. P and D were used for linear encoding representation coefficients and class-specific discriminative reconstruction, respectively. The object function of DPL model is

$$\{P^*, D^*\} = \underset{P,D}{\arg\min} \quad \sum_{c=1}^{C} \left\|X_c - D_c P_c X_c\right\|_F^2 + \lambda \left\|P_c \overline{X}_c\right\|_F^2, \tag{3}$$
$$\text{s.t.} \quad \left\|d_i\right\|_2^2 \leq 1,$$

where $D_c \in R^{m \times k}$ and $P_c \in R^{k \times m}$ represent subdictionary pairs corresponding to class c, $\overline{X}_c$ represents the complementary matrix of X_c, $\lambda > 0$ is a scalar constant to control the discriminative property of P, and d_i denotes the ith element of dictionary D.

The objective function in (3) is generally nonconvex. But it can be relaxed to the following form by introducing a variable matrix A:

$$\{P^*, A^*, D^*\} = \underset{P,A,D}{\arg\min} \quad \sum_{c=1}^{C} \left\|X_c - D_c A_c\right\|_F^2$$
$$+ \tau \left\|P_c X_c - A_c\right\|_F^2 \tag{4}$$
$$+ \lambda \left\|P_c \overline{X}_c\right\|_F^2,$$
$$\text{s.t.} \quad \left\|d_i\right\|_2^2 \leq 1,$$

where τ is an algorithm parameter. According to [19], the objective function in (4) can be solved by an alternatively updated manner.

When D and P have been learned, given a test sample x_t, the class-specific reconstruction residual is used to assign the class label. So the classification model associated with the DPL is defined as

$$\text{label}(x_t) = \underset{c}{\arg\min} \quad \left\|x_t - D_c P_c x_t\right\|_2. \tag{5}$$

When dealing with the classification task, DPL requires sufficient labeled training samples to learn discriminative dictionary pair for obtaining good results. In fact, it is difficult and expensive to obtain the vast quantity of labeled data. If we can exploit the information provided by the massive inexpensive unlabeled samples and choose small amounts of the "profitable" samples (the "profitable" unlabeled samples are the ones that are most beneficial for the improvement of the DPL classification performance) from unlabeled dataset to be labeled manually, we would learn a more discriminative dictionary than the one learned only using a limited number of labeled training data. To achieve this, we introduce the active learning technique to DPL in the next section.

3.2.2. Introducing Active Learning to DPL.

When evaluating one sample is "profitable" or not, two measures are considered: informativeness and representativeness. The proposed ADDL iteratively evaluates both informativeness and representativeness of unlabeled samples in a given pool for seeking the ones that are most beneficial for the improvement of the DPL classification performance. Specifically, the informativeness measure is constructed based on the reconstruction error and the entropy on the probability distribution over the class-specific reconstruction error, and the representativeness is obtained from the distribution of the unlabeled dataset in this study.

Assume that we are given an initial training dataset $\mathbb{D}_l = \{x_1, x_2, \ldots, x_{nl}\}$ with its label set $\mathbb{Y}_l = \{y_1, y_2, \ldots, y_{nl}\}$ and an unlabeled dataset $\mathbb{D}_u = \{x_{nl+1}, x_{nl+2}, \ldots, x_N\}$, where $x_i \in R^{m \times 1}$ is an m dimensional feature vector and $y_i \in \{1, 2, 3\}$ is the corresponding class label (sunny, cloudy, or overcast).

The target of ADDL is to iteratively select N_s most "profitable" samples, denoted by $\mathbb{D}_s$, from $\mathbb{D}_u$ to query their labels $\mathbb{Y}_s$ and then add them to $\mathbb{D}_l$ for improving the performance of the dictionary learning classification model.

Informativeness Measure. Informativeness measure is an effective criterion to select informative samples for reducing the uncertainty of the classification model, which captures the relationship of the candidate samples with the current classification model. Probabilistic classification models select the one that has the largest entropy on the conditional distribution over its labels [38, 39]. Query-by-committee algorithms choose the samples which have the most disagreement among a committee [22, 26]. In SVM methods, the most informative sample is regarded as the one that is closest to the separating hyperplane [40, 41]. Since DPL is used as the classification model in our framework, we design an informativeness measure based on the reconstruction error and the entropy on the probability distribution over the class-specific reconstruction error of the sample.

For dictionary learning, the samples which are well-represented through the current learned dictionary are less likely to provide more information in further refining the dictionary. Instead, the samples have large reconstruction error and large uncertainty should be mainly cared about, because they have some additional information that is not captured by the current dictionary. As a consequence, the informativeness measure is defined as follows:

$$E\left(x_j\right) = \text{Error\_}R_j + \text{Entropy}_j, \qquad (6)$$

where $\text{Error\_}R_j$ and Entropy_j denote the reconstruction error of the sample $x_j \in \mathbb{D}_u$ with respect to the current learned dictionary and the entropy of probability distribution over class-specific reconstruction error of the sample $x_j \in \mathbb{D}_u$, respectively. $\text{Error\_}R_j$ is defined as

$$\text{Error\_}R_j = \min_c \quad \left\| x_j - D_c P_c x_j \right\|^2, \qquad (7)$$

where D_c and P_c represent subdictionary pairs corresponding to the class c, which is learned by DPL algorithm as shown in (4). The larger $\text{Error\_}R_j$ indicates that the current learned dictionary does not represent the sample x_j well.

Since the class-specific reconstruction error is used to identify the class label of sample x_j (as shown in (5)), the probability distribution of class-specific reconstruction error of x_j can be acquired. The class-specific reconstruction error probability of x_j in class c is defined as

$$p_c\left(x_j\right) = \frac{\left\| x_j - D_c P_c x_j \right\|^2}{\sum_{c=1}^{C} \left\| x_j - D_c P_c x_j \right\|^2}. \qquad (8)$$

The class probability distribution $p(x_j)$ for sample x_j is computed as $p(x_j) = [p_1, p_2, \ldots, p_C]$, which demonstrates how well the dictionary distinguishes the input sample. That is, if an input sample can be expressed well by the current dictionary, we will get a small value of $\left\| x_j - D_c P_c x_j \right\|^2$

to one of class-specific subdictionaries, and thus the class distribution should reach the valley at the most likely class. Entropy is a measure of uncertainty. Hence, in order to estimate the uncertainty of an input sample label, the entropy of probability distribution over class-specific reconstruction error is calculated as follows:

$$\text{Entropy}_j = -\sum_{c=1}^{C} p_c\left(x_j\right) \log p_c\left(x_j\right). \qquad (9)$$

The high Entropy_j value demonstrates that x_j is difficult to be classified by the current learned dictionary; thus it should be selected to be labeled and added to the labeled set for further training dictionary learning.

Representativeness Measure. Since the informativeness measure only considers how the candidate samples relate to the current classification model, it ignores the potential structure information of the whole input unlabeled dataset. Therefore, the representativeness measure is employed as an additional criterion to choose the useful unlabeled samples. Representativeness measure is to evaluate whether the samples well represent the overall input patterns of unlabeled data, which exploits the relation between the candidate sample and the rest of unlabeled samples.

The distribution of unlabeled data is very useful for training a good classifier. In previous active learning work, the marginal density and cosine distance are used as the representativeness measure to gain the information of data distribution [39, 42]. Li and Guo [43] defined a more straightforward representativeness measure called mutual information. Its intention is to select the samples located in the density region of the unlabeled data distribution, which is more representative regarding the remaining unlabeled data than the ones located in the sparse region. We introduce the framework of representativeness measure proposed by Li and Guo [43] into the dictionary learning.

For an unlabeled sample x_j, the mutual information with respect to other unlabeled samples is defined as follows:

$$M\left(x_j\right) = H\left(x_j\right) - H\left(x_j \mid X_{U_j}\right) = \frac{1}{2} \ln \left(\frac{\sigma_j^2}{\sigma_{j|U_j}^2} \right), \quad (10)$$

where $H(x_j)$ and $H(x_j \mid X_{U_j})$ denote the entropy and the conditional entropy of sample x_j, respectively. U_j represents the index set of unlabeled samples where j has been removed from U, $U_j = U - j$, and X_{U_j} represents the set of samples indexed by U_j. σ_j^2 and $\sigma_{j|U_j}^2$ can be calculated by the following formulas:

$$\sigma_j^2 = \mathcal{K}\left(x_j, x_j\right),$$

$$\sigma_{j|U_j}^2 = \sigma_j^2 - \sum_{jU_j} \sum_{U_j U_j}^{-1} \sum_{U_j j} . \qquad (11)$$

Inputs: Labeled set $\mathbb{D}_l$ and its label set $\mathbb{Y}_l$, Unlabeled set $\mathbb{D}_u$, the number of iteration I_t
     and the number of unlabeled samples N_s to be selected in each iteration.
(1) Initialization: Learn an initial dictionary pair D^* and P^* by DPL algorithm from the $\mathbb{D}_l$.
(2) **For** $i = 1$ to I_t, do
(3) Compute $E(x_j)$ and $M(x_j)$ by (6) and (10) for each sample x_j in the unlabeled dataset $\mathbb{D}_u$.
(4) Select N_s samples (denoted by $\mathbb{D}_s$) from the $\mathbb{D}_u$ by (13), and add them into $\mathbb{D}_l$ with their
     class labels which manually assigned by user. Then updates $\mathbb{D}_u = \mathbb{D}_u - \mathbb{D}_s$ and $\mathbb{D}_l = \mathbb{D}_l \cup \mathbb{D}_s$.
(5) Learn the refined dictionaries D^*_{new} and P^*_{new} over the expanded dataset $\mathbb{D}_l$.
(6) **End for**
Output: Final learned dictionary pair D^*_{new} and P^*_{new}.

ALGORITHM 1: Active discriminative dictionary learning (ADDL).

Assume that the index set $U_j = (1, 2, 3, \ldots, t)$ and $\sum_{U_j U_j}$ is a kernel matrix defined over all the unlabeled samples indexed by U_j; it is computed by the following form:

$$
\sum_{U_j U_j} = \begin{pmatrix} \mathscr{K}(x_1, x_1) & \mathscr{K}(x_1, x_2) & \cdots & \mathscr{K}(x_1, x_t) \\ \mathscr{K}(x_2, x_1) & \mathscr{K}(x_2, x_2) & \cdots & \mathscr{K}(x_2, x_t) \\ \vdots & \vdots & \vdots & \vdots \\ \mathscr{K}(x_t, x_1) & \mathscr{K}(x_t, x_2) & \cdots & \mathscr{K}(x_t, x_t) \end{pmatrix}. \tag{12}
$$

$\mathscr{K}(\cdot)$ is a symmetric positive definite kernel function. In our approach, we apply simple and effective linear kernel $\mathscr{K}(x_j, x_j) = \|x_i - x_j\|^2$ for our dictionary learning task.

The mutual information is used to implicitly exploit the information between the selected samples and the remaining ones. The samples which have large mutual information should be selected from the pool of unlabeled data for refining the learned dictionary in DPL.

Procedure of Active Samples Selection. Based on the above analysis, we aim to integrate the strengths of informativeness measure and representativeness measure to select the unlabeled samples from the pool of the unlabeled data. We choose the samples that have not only large reconstruction error and the entropy of probability distribution over class-specific reconstruction error with respect to the DPL classification model, but also the large representativeness regarding the rest of unlabeled samples. The sample set $\mathbb{D}_s = \{x_1^s, x_2^s, \ldots, x_{N_s}^s\}$ are iteratively selected from pool by the following formula:

$$
x^s = \arg\min_{x_j} \; \left(E(x_j) + M(x_j) \right), \tag{13}
$$
$$
j \in \{nl + 1, nl + 2, \ldots, N\}.
$$

The overall of our ADDL is given in Algorithm 1.

4. Experiments

In this section, the performance of the proposed framework is evaluated on two weather datasets. We first give the details about the datasets and experimental settings. Then, the experimental results are provided and analyzed.

4.1. Datasets and Experimental Setting

Datasets. The first dataset employed in our experiment is the dataset provided by Chen et al. [9] (denoted as DATASET 1). DATASET 1 contains 1000 images of size 3966×270, and each image has been manually labeled as sunny, cloudy, or overcast. There are 276 sunny images, 251 cloudy images, and 473 overcast images in DATASET 1. Figure 2 shows three images from DATASET 1 with the label sunny, cloudy, and overcast, respectively.

Because there are few available public datasets for weather recognition, we construct a new dataset (denoted as DATASET 2) to test the performance of the proposed method. The images in DATASET 2 are selected from the panorama images collected on the roof of BC building at EPFL (http://panorama.epfl.ch/ provides high resolution (13200×900) panorama images from 2005 till now, recording at every 10 minutes during daytime) and categorized into sunny, cloudy, or overcast based on the classification criterion presented in [9]. It includes 5000 images which were captured at approximately every 30 minutes during daytime in 2014, and the size of each image is 4821×400. Although both DATASET 1 and DATASET 2 are constructed based on the images provided by http://panorama.epfl.ch/, DATASET 2 is more challenging because it contains a large number of images captured in different seasons. In Figure 3, some examples of different labeled images in DATASET 2 are shown.

Experimental Setting. For the sky region of each image, four types of visual appearance features are extracted, including 200-dim SIFT, 600-dim HSV color feature (200-dim histogram of H channel, 200-dim histogram of S channel, and 200-dim histogram of V channel), 200-dim LBP, and 200-dim gradient magnitude feature. These features are only extracted from the sky part of each image, and the sky detector and feature extraction procedure provided by Chen et al. [9] are used in this paper. Besides the visual appearance features, two kinds of features based on physical characteristics of images captured under different weather are also extracted from the nonsky region, which consists of 200-dim bag-of-words representation of the contrast and 200-dim bag-of-word

(a)

(b)

(c)

FIGURE 2: Examples in DATASET 1. (a) Sunny; (b) cloudy; (c) overcast.

(a)

(b)

(c)

FIGURE 3: Examples in DATASET 2. (a) Sunny; (b) cloudy; (c) overcast.

representation of the dark channel. To be specific, we divide each nonsky region of the image into 32×32 blocks and extract the contrast and the dark channel features by (1) and (2) and then use bag-of-words model [37] to code each feature.

In our experiment, 50% images are randomly selected in each dataset for training and the remaining data is used for testing. The training data are randomly partitioned into labeled sample set $\mathbb{D}_l$ and unlabeled sample set $\mathbb{D}_u$. $\mathbb{D}_l$ is applied for learning an initial dictionary and $\mathbb{D}_u$ is utilized for actively selecting "profitable" samples to iteratively improve the classification performance. To make the experiment results more convincing, each following experimental process is repeated ten times, and then the mean and standard deviation of the classification accuracy are reported.

4.2. Experiment I: Verifying the Performance of Feature Extraction. The effectiveness of the feature extraction of our method is first evaluated. Many previous works merely extract the visual appearance features of the sky part for weather recognition [9]. In order to validate the power of our extracted features based on physical characteristics of images, the results of two weather recognition schemes are compared. One only uses the visual appearance features to classify the weather conditions, and the other combines the visual appearance features with features based on physical characteristics of images to identify the weather conditions. In order to weaken the influence of classifier on the results of weather recognition, K-NN classification (K is experientially set to 30), SVM with the radial basis function kernel, and the original DPL without active samples selection procedure are applied in this experiment. Figures 4 and 5 show the comparison results on DATASET 1 and DATASET 2, respectively.

In Figures 4 and 5, x-axis represents the different number of training samples and y-axis represents the average classification accuracy. The red dotted lines indicate just six visual appearance features of the sky area are used, and the blue solid lines indicate both six visual appearance features and

two features based on physical characteristics of the nonsky area are applied for recognition. From Figures 4 and 5, it is clearly observed that the combination of visual appearance features and physical features can achieve better performance for weather recognition task.

4.3. Experiment II: Recognizing Weather Conditions by the Proposed ADDL. In this section, the performance of the proposed ADDL algorithm is evaluated. First experiment is conducted to give the best parameters for ADDL. And then ADDL is compared against several popular classification methods.

4.3.1. Parameters Selection. There are three important parameters in the proposed approach, that is, k, λ, and τ. k is the number of atoms in each subdictionary D_c learned from samples in each class, λ is used to control the discriminative property of P, and τ is a scalar constant in DPL algorithm. The performances of our ADDL under various values of k, λ, and τ are studied on DATASET 1. Figure 6(a) lists the classification results when $k = \{15, 25, 35, 45, 55, 65, 75, 85, 95\}$. It can be seen that the highest average classification accuracy is obtained when $k = 25$. This demonstrates that ADDL is effective to learn a compact dictionary. According to the observation in Figure 6(a), we set k to be 25 in all experiments. The classification results obtained by using the different τ and λ are shown in Figures 6(b) and 6(c). From Figures 6(b) and 6(c), the optimal values of λ and τ are 0.05 and 25, respectively. This is because of the fact that a too big or too small λ value will lead the reconstruction coefficient in ADDL to be too sparse or too dense, which will deteriorate the classification performance. If τ is too large, the effect of the reconstruction error constraint (the first term in (4)) and the sparse constraint (the third term in (4)) is weakened, which will decrease the discrimination ability of the learned dictionary. On the contrary if τ is too small, the second term in (4) will be neglected in dictionary learning, which also reduces the performance of algorithm. Hence, we set $\lambda = 0.05$ and $\tau = 25$ for all experiments.
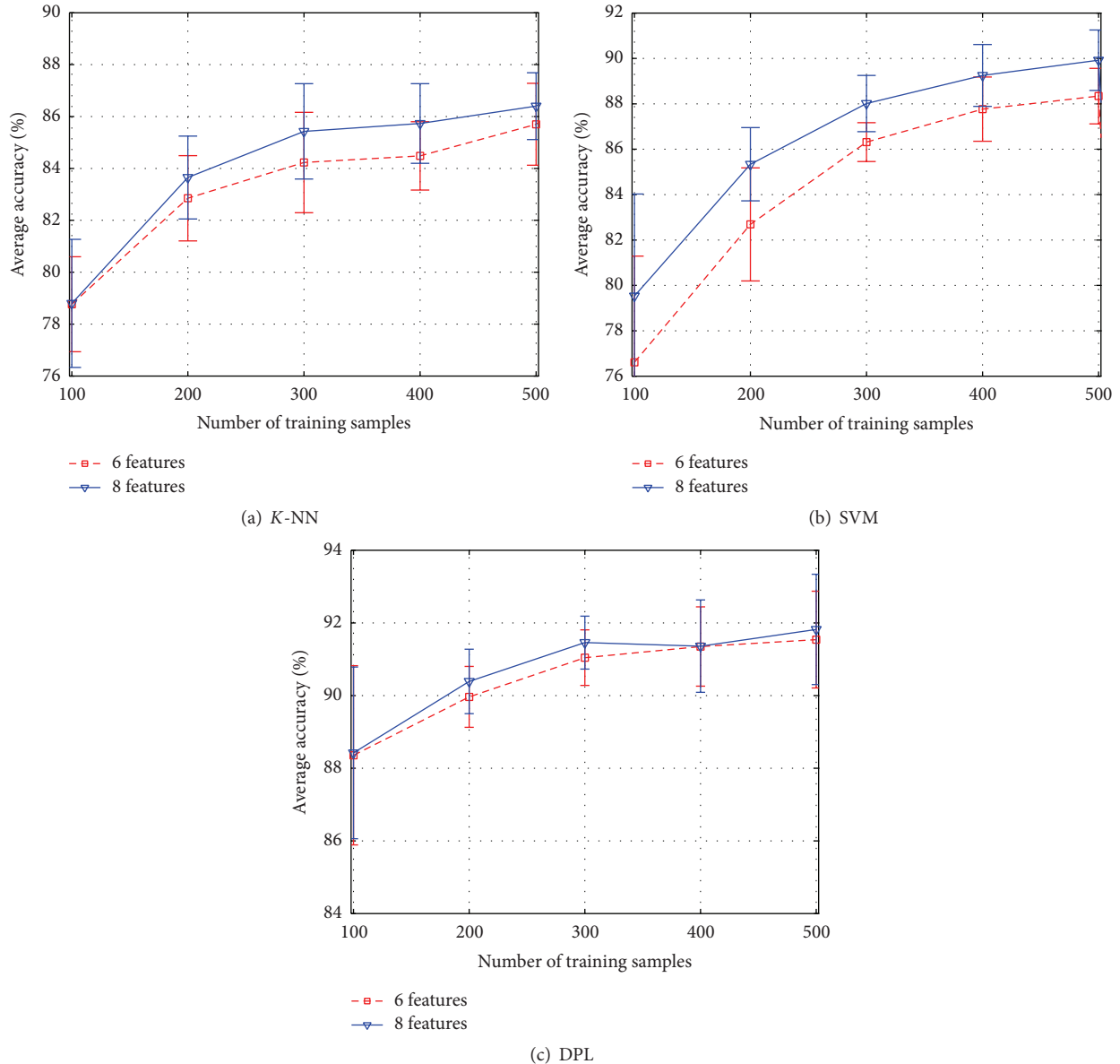
(a) *K*-NN

(b) SVM

(c) DPL

Figure 4: Weather recognition results over DATASET 1 by using different features.

Now we evaluate weather active samples selection can improve the recognition performance. 500 samples in DATASET 1 are randomly selected as training data and the remaining samples are used for testing. In training dataset, 50 samples are randomly selected as the labeled dataset $\mathbb{D}_l$ and the remaining 450 samples are selected as the unlabeled dataset $\mathbb{D}_u$. The proposed ADDL uses the labeled dataset $\mathbb{D}_l$ to learn an initial dictionary pair and then iteratively selects 50 samples from $\mathbb{D}_u$ to label for expanding the training dataset. Figure 7 shows the recognition accuracy versus the number of iterations.

In Figure 7, the 0th iteration indicates that we only use the initial 50 labeled samples to learn the dictionary, and the 9th iteration means using all 500 training samples to learn the dictionary for recognition. The recognition ability of ADDL

is improved by active samples selection procedure, and it achieves highest accuracy when the number of iterations is 3; total 200 samples are used for training. It is worth mentioning that ADDL obtains the best results when the number of iterations is set as 3. If iterations are larger than 3, the recognition rates will drop about 1%. This is because there are some noisy examples or "outliers" in the unlabeled dataset, and the more noisy examples or "outliers" will be selected to learn the dictionary along with the increase of iterations, which interferes with the dictionary learning and leads to the classification performance degradation. In the following the number of iterations is set to 3 for all experiments.

4.3.2. Comparisons of ADDL with Other Methods. Here, the proposed ADDL is compared with several methods. The first
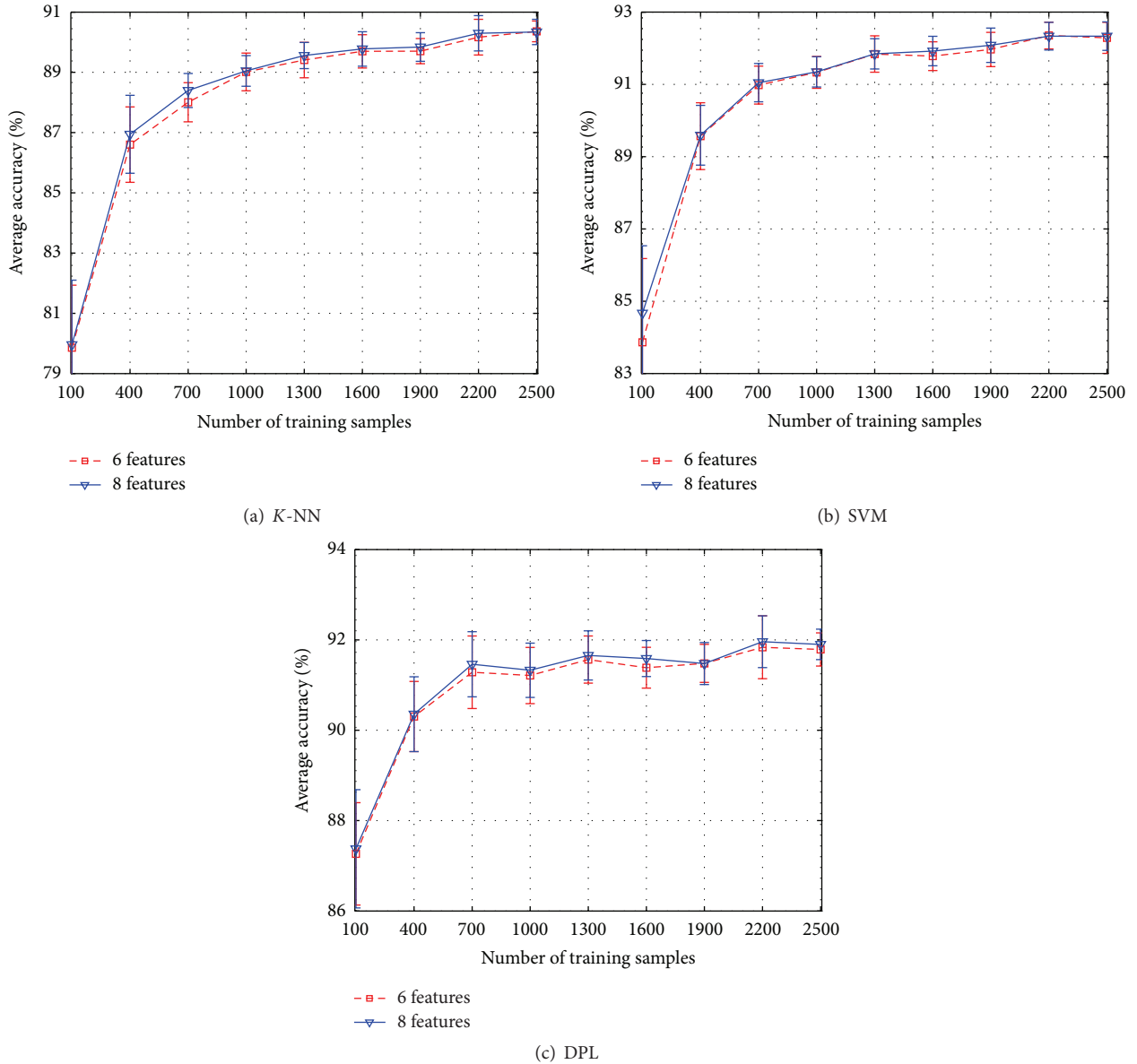
(a) K-NN

(b) SVM

(c) DPL

FIGURE 5: Weather recognition results over DATASET 2 by using different features.

two methods are K-NN algorithm used by Song et al. [7] and SVM with the radial basis function kernel (RBF-SVM) adopted by Roser and Moosmann [5]. The third method is SRC [15] which directly uses all training samples as the dictionary for classification. In order to confirm that the active samples selection in our ADDL method is effective, the proposed ADDL is compared with the original DPL method [19]. We also compare ADDL with the method proposed by Chen et al. [9]. As far as we know, the work in [9] is the only framework which addresses the same problem with our method, that is, recognizing different weather conditions (sunny, cloudy, and overcast) of images captured by a still camera. It actively selected useful samples to training SVM for recognizing different weather conditions.

For DATASET 1, 500 images are randomly selected as the training samples and the rest of images are used as the testing samples. In the training dataset, 50 images are randomly chosen as the initial labeled training dataset $\mathbb{D}_l$, and the remaining 450 images are regarded as the unlabeled dataset $\mathbb{D}_u$. ADDL and Chen's method [9] both include the active learning procedure; thus they iteratively choose 150 samples from $\mathbb{D}_u$ to be labeled based on their criterion of the samples selection and add these samples to $\mathbb{D}_l$ for further training the classification model. For K-NN, RBF-SVM, SRC, and DPL methods which are without the active learning procedure, 150 samples are randomly selected from $\mathbb{D}_u$ to be labeled for expanding the labeled training dataset $\mathbb{D}_l$. Table 1 lists the comparisons of our approach with several methods for
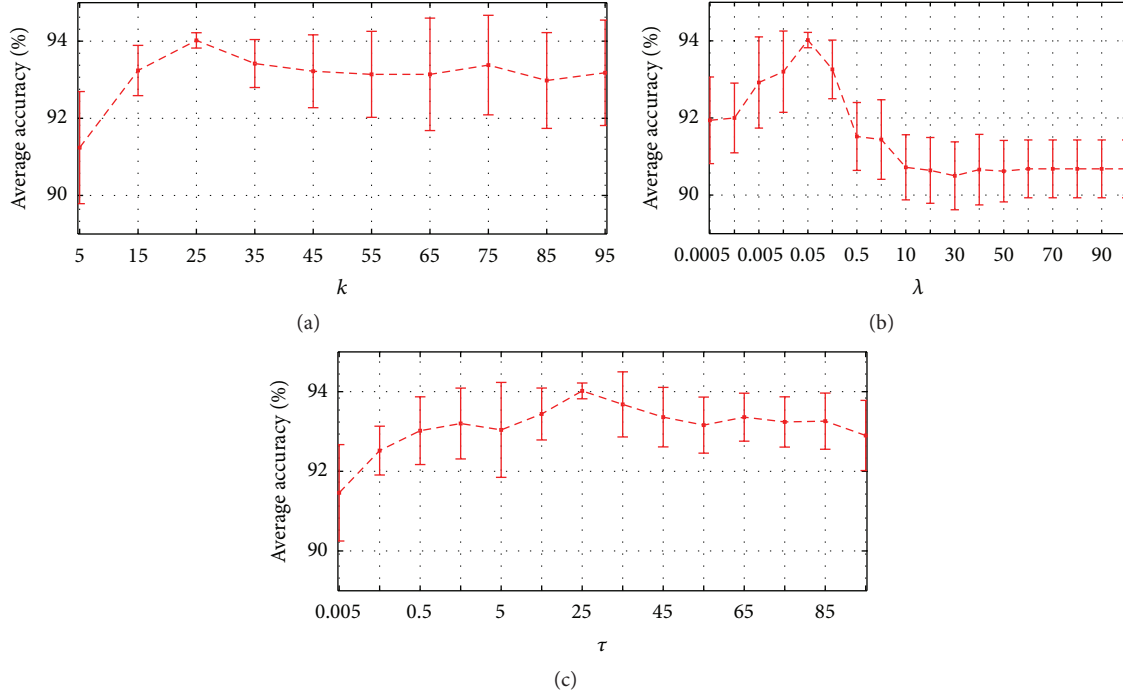
(a)

(b)

(c)

FIGURE 6: Selection of parameters. x-axis represents the different values of parameters and y-axis represents the average classification accuracy. (a) The average classification rate under different k. (b) The average classification rate under different λ. (c) The average classification rate under different τ.
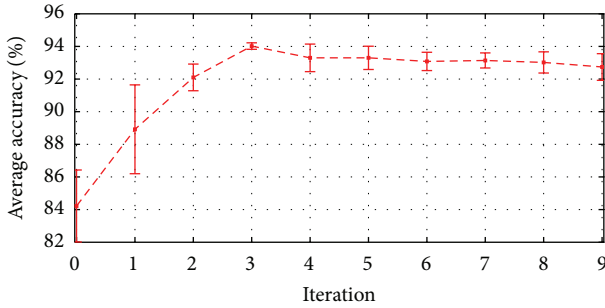

FIGURE 7: Recognition accuracy on DATASET 1 versus the number of iterations.

TABLE 1: Comparisons on DATASET 1 among different methods.

Methods	Classification rate (mean ± std)
K-NN	82.9% ± 1.2%
RBF-SVM	85.6% ± 1.6%
SRC	89.7% ± 0.9%
DPL	91.3% ± 1.6%
Chen's method [9]	92.98% ± 0.55%
ADDL	94.0% ± 0.2%

weather classification. As can be seen from Table 1, ADDL outperforms other methods. The mean classification rate of ADDL reaches about 94%.

In DATASET 2, 2500 images are randomly selected as the training samples and the rest of images are used as

TABLE 2: Comparison on DATASET 2 among different methods.

Methods	Classification rate (mean ± std)
K-NN	85.0% ± 1.6%
RBF-SVM	88.1% ± 1.4%
SRC	88.2% ± 1.1%
DPL	89.6% ± 0.8%
Chen's method [9]	88.8% ± 1.6%
ADDL	90.4% ± 1.0%

the testing samples. In the training dataset, 50 images are randomly chosen as the initial labeled training dataset $\mathbb{D}_l$; the remaining 2450 images are regarded as the unlabeled dataset $\mathbb{D}_l$. All parameters setting for DATASET 2 are the same as DATASET 1. Table 2 lists the recognition results of different methods, which indicates that the validity of the proposed ADDL is better than other methods.

From Tables 1 and 2, two points can be observed. First, we can find that the recognition performances of K-NN, RBF-SVM, SRC, and DPL are overall inferior to the proposed ADDL algorithm. This is probably because these four algorithms randomly select the unlabeled data from the given pool, which do not consider whether the selected samples are beneficial for improving the performance of the classification model or not. Second, although the proposed ADDL and Chen's method [9] both include the active learning paradigm, the proposed ADDL performs better than Chen's method [9]. This is due to the fact that Chen's method [9] only considers the informativeness and ignores representativeness

of samples when selecting the unlabeled samples from the given pool.

5. Conclusions

We have presented an effective framework for classifying three types of weather (sunny, cloudy, and overcast) based on the outdoor images. Through the analysis of the different visual manifestations of images caused by different weathers, the various features are separately extracted from the sky area and nonsky area of images, which describes visual appearance properties and physical characteristics of images under different weather conditions, respectively. ADDL approach was proposed to learn a more discriminative dictionary for improving the weather classification performance by selecting the informative and representative unlabeled samples from a given pool to expand the training dataset. Since there is not much image dataset on weather recognition, we have collected and labeled a new weather dataset for testing the proposed algorithm. The experimental results show that ADDL is a fairly effective and inspiring strategy for weather classification, which also can be used in many other computer vision tasks.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] F. Nashashibi, R. de Charette, and A. Lia, "Detection of unfocused raindrops on a windscreen using low level image processing," in *Proceedings of the 11th International Conference on Control, Automation, Robotics and Vision (ICARCV '10)*, pp. 1410–1415, Singapore, December 2010.

[2] H. Kurihata, T. Takahashi, I. Ide et al., "Rainy weather recognition from in-vehicle camera images for driver assistance," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 205–210, IEEE, Las Vegas, Nev, USA, June 2005.

[3] H. Woo, Y. M. Jung, J.-G. Kim, and J. K. Seo, "Environmentally robust motion detection for video surveillance," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2838–2848, 2010.

[4] H. Katsura, J. Miura, M. Hild, and Y. Shirai, "A view-based outdoor navigation using object recognition robust to changes of weather and seasons," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '03)*, vol. 3, pp. 2974–2979, IEEE, Las Vegas, Nev, USA, October 2003.

[5] M. Roser and F. Moosmann, "Classification of weather situations on single color images," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 798–803, IEEE, Eindhoven, The Netherlands, June 2008.

[6] X. Yan, Y. Luo, and X. Zheng, "Weather recognition based on images captured by vision system in vehicle," in *Advances in*

[7] H. Song, Y. Chen, and Y. Gao, "Weather condition recognition based on feature extraction and K-NN," in *Foundations and Practical Applications of Cognitive Systems and Information Processing: Proceedings of the First International Conference on Cognitive Systems and Information Processing, Beijing, China, Dec 2012 (CSIP2012)*, Advances in Intelligent Systems and Computing, pp. 199–210, Springer, Berlin, Germany, 2014.

[8] C. Lu, D. Lin, J. Jia, and C.-K. Tang, "Two-class weather classification," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 3718–3725, Columbus, Ohio, USA, June 2014.

[9] Z. Chen, F. Yang, A. Lindner, G. Barrenetxea, and M. Vetterli, "How is the weather: automatic inference from images," in *Proceedings of the 19th IEEE International Conference on Image Processing (ICIP '12)*, pp. 1853–1856, Orlando, Fla, USA, October 2012.

[10] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[11] A. Li and H. Shouno, "Dictionary-based image denoising by fused-lasso atom selection," *Mathematical Problems in Engineering*, vol. 2014, Article ID 368602, 10 pages, 2014.

[12] Z. Feng, M. Yang, L. Zhang, Y. Liu, and D. Zhang, "Joint discriminative dimensionality reduction and dictionary learning for face recognition," *Pattern Recognition*, vol. 46, no. 8, pp. 2134–2143, 2013.

[13] J. Dong, C. Sun, and W. Yang, "A supervised dictionary learning and discriminative weighting model for action recognition," *Neurocomputing*, vol. 158, pp. 246–256, 2015.

[14] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[15] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[16] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 791–804, 2012.

[17] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[18] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based fisher discrimination dictionary learning for image classification," *International Journal of Computer Vision*, vol. 109, no. 3, pp. 209–232, 2014.

[19] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS '14)*, pp. 793–801, December 2014.

[20] B. Settles, *Active Learning Literature Survey*, vol. 52, University of Wisconsin-Madison, Madison, Wis, USA, 2010.

[21] D. Cohn, L. Atlas, and R. Ladner, "Improving generalization with active learning," *Machine Learning*, vol. 15, no. 2, pp. 201–221, 1994.

[22] I. Dagan and S. P. Engelson, "Committee-based sampling for training probabilistic classifiers," in *Proceedings of the 12th*

Neural Networks—ISNN 2009, pp. 390–398, Springer, Berlin, Germany, 2009.

International Conference on Machine Learning, pp. 150–157, Morgan Kaufmann, Tahoe City, Calif, USA, 1995.

[23] D. Angluin, "Queries and concept learning," *Machine Learning*, vol. 2, no. 4, pp. 319–342, 1988.

[24] R. D. King, K. E. Whelan, F. M. Jones et al., "Functional genomic hypothesis generation and experimentation by a robot scientist," *Nature*, vol. 427, no. 6971, pp. 247–252, 2004.

[25] S. Chakraborty, V. Balasubramanian, and S. Panchanathan, "Dynamic batch mode active learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 2649–2656, Providence, RI, USA, June 2011.

[26] K. Nigam and A. McCallum, "Employing EM in pool-based active learning for text classification," in *Proceedings of the 15th International Conference on Machine Learning (ICML '98)*, pp. 350–358, Morgan Kaufmann, Madison, Wis, USA, 1998.

[27] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, "Active learning with Gaussian processes for object categorization," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, IEEE, Rio de Janeiro, Brazil, October 2007.

[28] S. J. Huang, R. Jin, and Z. H. Zhou, "Active learning by querying informative and representative examples," in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS '10)*, pp. 892–900, 2010.

[29] P. Donmez, J. G. Carbonell, and P. N. Bennett, "Dual strategy active learning," in *Machine Learning: ECML 2007*, J. N. Kok, J. Koronacki, R. L. de Mantaras, S. Matwin, D. Mladenič, and A. Skowron, Eds., vol. 4701 of *Lecture Notes in Computer Science*, pp. 116–127, Springer, Berlin, Germany, 2007.

[30] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, "Semi-supervised SVM batch mode active learning for image retrieval," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–7, Anchorage, Alaska, USA, June 2008.

[31] Z. Xu, K. Yu, V. Tresp, X. Xu, and J. Wang, "Representative sampling for text classification using support vector machines," in *Proceedings of the 25th European Conference on IR Research (ECIR '03)*, Springer, 2003.

[32] A. Bosch, A. Zisserman, and X. Muñoz, "Image classification using random forests and ferns," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, IEEE, Rio de Janeiro, Brazil, October 2007.

[33] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[34] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, pp. 598–605, Hilton Head Island, SC, USA, June 2000.

[35] S. G. Narasimhan and S. K. Nayar, "Removing weather effects from monochrome images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II-186–II-193, IEEE, Kauai, Hawaii, USA, December 2001.

[36] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.

[37] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proceedings of the International Workshop on Statistical Learning in Computer Vision (ECCV '04)*, pp. 1–22, Prague, Czech Republic, 2004.

[38] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in *SIGIR '94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organised by Dublin City University*, pp. 3–12, Springer, Berlin, Germany, 1994.

[39] B. Settles and M. Craven, "An analysis of active learning strategies for sequence labeling tasks," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '08)*, pp. 1070–1079, Association for Computational Linguistics, 2008.

[40] G. Schohn and D. Cohn, "Less is more: active learning with support vector machines," in *Proceedings of the 17th International Conference on Machine Learning (ICML '00)*, pp. 839–846, Morgan Kaufmann, 2000.

[41] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *The Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2002.

[42] M. Szummer and T. S. Jaakkola, "Information regularization with partially labeled data," in *Advances in Neural Information Processing Systems*, pp. 1025–1032, MIT Press, 2002.

[43] X. Li and Y. Guo, "Adaptive active learning for image classification," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 859–866, IEEE, Portland, Ore, USA, June 2013.