Scene-free multi-class weather classification on single images
Zheng Zhang, Huadong Ma*, Huiyuan Fu, Cheng Zhang
Beijing Key Lab of Intelligent Telecomm. Software and Multimedia, Beijing University of Posts and Telecomm., Beijing 100876, China

Abstract
Multi-class weather classification is a fundamental and significant technique which has many potential applications, such as video surveillance and intelligent transportation. However, it is a challenging task due to the diversity of weather and lack of discriminate feature. Most existing weather classification methods only consider two-class weather conditions such as sunny-rainy or sunny-cloudy weather. Moreover, they predominantly focus on a fixed scene such as popular tourism and traffic scenario. In this paper, we propose a novel method for scene-free multi-class weather classification from single images based on multiple category-specific dictionary learning and multiple kernel learning. To improve the discrimination of image representation and enhance the performance of multiple weather classification, our approach extracts multiple weather features and learns dictionaries based on these features. To select a good subset of features, we utilize multiple kernel learning algorithm to learn an optimal linear combination of feature kernels. In addition, to evaluate the proposed approach, we collect an outdoor image set that contains 20 K images, called MWI (Multi-class Weather Image) set. Experimental results show the effectiveness of the proposed method.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction
Most of the existing methods in the field of computer vision are based on the assumption that the weather in outdoor images or videos is clear. However, different weather conditions such as rain, snow or haze will decrease the quality of images or videos, as shown in Fig. 1. Such effects may significantly degrade the performances of outdoor vision systems which relies on image/video feature extraction or visual attention modeling. Hence, the applications of weather classification are numerous, such as the detection and observation of weather conditions, image/video analysis, the reliability improvement of video surveillance systems. In this paper, we target at the problem of classifying multiple weather, such as sunny, rainy, snowy, and haze from single images.

Despite its remarkable value, multi-class weather classification has not been thoroughly studied. Some previous researches [1–3] focused on weather recognition from vehicle camera images for driver assistance. Most of these methods are only able to recognize rainy weather. Furthermore, the applications are limited due to the relatively fixed target scenes. Recently, the authors of [4,5] focused on two-class weather recognition, include sunny and cloudy. The authors of [4] estimated the weather conditions of popular tourism from images of the same scene. The authors of [5] proposed a collaborative learning framework via analyzing multiple weather cues for two-class weather recognition from single images. The authors of [6] proposed a method to label images of the same scene with three weather conditions including sunny, cloudy, and overcast. The authors of [7] proposed an approach for multi-class weather classification, which could be used for the traffic scene only. However, approaches for the fixed scene weather classification are not able to be applied in the practiced systems due to the following two reasons. First, it needs to learn different classifiers for different scenes. Second, it is hard to collect the training image set in any scene. The darkness in the night is another factor that results in the decrease of image quality. The authors of [8] proposed a Color Estimation Model for night removal from a single input image. They use a guided statistical Dark-to-Day prior to direct optimal performance.

Different from the above works, we propose a new framework for classifying multi-class weather from single images in any scene, which is based on dictionary learning and multiple kernel learning (MKL). Implementation of the kernel idea, however, entails substantial challenges. First, it is difficult to find suitable features to discriminate different weather conditions. Second, the features might be heterogeneous and the feature vectors are high-dimensional. Aiming at solving the above challenges, we first extract multiple features to represent different weather conditions. For example, the sky and shadow features can indicate

* Corresponding author.
E-mail addresses: zhangzheng@bupt.edu.cn (Z. Zhang), mhd@bupt.edu.cn (H. Ma), fhy@bupt.edu.cn (H. Fu), zhangcheng@bupt.edu.cn (C. Zhang).

http://dx.doi.org/10.1016/j.neucom.2016.05.015
0925-2312/© 2016 Elsevier B.V. All rights reserved.
sunny weather. The dark channel feature can indicate the haze weather. The HOG (Histogram of Oriented Gradients) based template matching feature can indicate rainy weather. The snowflake noise feature can indicate snowy feature. Some global features like contrast and saturation are used to distinguish multi-class weather. To improve the discrimination of image representation and enhance the performance of multiple weather classification, our approach extracts multiple weather features and learns multiple dictionaries based on these features. Then, we use multiple kernel learning algorithm to learn an optimal linear combination of feature kernels for selecting a good subset of features.

The contributions of this paper are as follows:

- We propose a scene-free multi-class weather classification framework by fusing multiple image features and learning multiple dictionaries. To our knowledge, this work is one of the first attempts towards single image multi-class weather classification.
- We propose two methods for detecting rain streak and snowflake from single images respectively. First, we propose a Histogram of Orientation Gradients (HOG) based template matching method to detect the rain streaks. Moreover, we regard snowflake as a kind of noise and define several rules to detect and describe snowflakes.
- We collect an outdoor image set contains 20 K images called MWI (Multi-class Weather Image) set, which provides an extensive testbed for the evaluation of existing methods and development of new approaches.

2. Related work

In this section, we give a brief review on dictionary learning and multiple kernel learning.

2.1. Dictionary learning

Dictionary learning is an effective feature learning technique. It has been successfully applied on a variety of problems in computer vision [9] and image analysis [10]. For image classification tasks, the learned dictionary can be used to represent images, which facilitates classification. The authors of [11] proposed a supervised learning method to learn a dictionary in the source image space and a corresponding transformation matrix. The learned global transformation matrix was used to map sparse features of source image patches to intensity values of the target patches. The authors of [12] proposed a sparse coding technique in a high dimensional feature space by using some implicit feature mapping. They applied it to image classification, face recognition, and kernel matrix approximation. However, these works focus on basic-level object classification that might fail in weather classification task, because the difference between different weather conditions is subtle and minute.

Some scholars put a lot of effort in the research on fine-grained classification via dictionary learning. The authors of [13] learned a hybrid dictionary with commonality and particularity that integrated an incoherence penalty term into the objective function for obtaining the class-specific sub-dictionary. The combination of the particularity and the commonality can faithfully represent the samples, and the particularities are more discriminative and more compact for classification. The authors of [14] learned both category-specific dictionaries and a shared dictionary to separate the different and common components of each image for fine-grained categorization. They proposed to impose cross-dictionary incoherent constraint and self-dictionary incoherent terms in the objective function for learning a discriminative dictionary.

2.2. Multiple kernel learning

In recent years, many good features have been designed to characterize various aspects of an object. However, simple classifiers cannot handle the high-dimensional feature space well. Thus, a proper feature selection and fusion is required for discarding irrelevant features and adapting the model to the specific problem. Multiple kernel learning [15] is viewed as an effective way to fuse features and design an optimal kernel [16,17]. The authors of [18] proposed a multiple kernel support vector machine...
(MK-SVM) scheme to discover knowledge from the gene expression data. This scheme improves the explanation capacity of SVM that consists of feature selection, predictive modeling and rule extraction. The authors of [19] extended the MKL framework to generalized MKL (GMKL) so that the combination form of kernels can be more flexible. GMKL allows learning general kernel parameterizations, including linear and non-linear kernel combinations, subject to general regularization. Recently, the authors of [20] presented a non-monotonic feature selection framework via multiple kernel learning that considers the number of selected features during searching for the optimal feature subset.

3. Our approach

In this section, we will propose a general framework for multiple weather classification from single images. The framework is composed by the extraction of multiple features including sky, shadow, rain streak, snowflake, dark channel, contrast and saturation; learning multiple dictionaries; selection and classification features via the learned sparse code and multiple kernel learning algorithm. Next, we will describe each component in detail.

3.1. Multiple feature extraction

For any pattern recognition problem, it is important to select proper features. We always implement the image classification by selecting interesting points as features or detecting the object emerging in the scene. However, weather classification from images is different from general image classification. It is impractical to use the traditional features for our problem because there can be the same object and interesting points under different weather conditions. Hence, it is not proper for applying the same kind of features as general image classification. We propose several low-level features by analyzing the property of images under different weather conditions. More specifically, we regard the features of sky, shadow, rain streak, snowflake and dark channel as local features so that each of them can indicate one specific weather, and regard the feature of contrast and saturation as global features that can indicate multiple weathers.

3.1.1. Sky

Sky might be the most obvious feature to indicate sunny weather in images. As shown in Fig. 2, sunny images have a clear sky or strong shadows, while the other weather images have gray sky or faint shadows. For the sky part, first, we detect the sky region in an image with the method suggested in [5]. Then, we extract the a and b channels in the Lab color space of the sky region to form a 200 dimensional feature vector.

3.1.2. Shadow

However, not all sunny images have a sky region. For the image without sky region, strong ground shadows can indicate the sunny weather. We apply the shadow detection tool in [21], and follow the implementation in [5] to form a 10 dimensional feature vector for each image.

3.1.3. Rain streak

Rain and snow detection from single images has been rarely studied in the literature, where no temporal information among successive images can be exploited, making the problem very challenging. Some works used HOG feature for rain streaks removing or detecting. Rain streaks are supposed to fall at the

Fig. 2. (a) Top: input images of different weather. Bottom: the detected sky regions. (b) Shadow detection result for a sunny image. (c) Shadow detection result for a haze image. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)
same orientation. Thus, we propose a HOG based template matching method to detect the rain streaks. In detail, we construct five pure rain HOG templates in different angles. For each image, we use the guided image filter [22] to decompose it into a low-frequency part and a high-frequency part which is suggested in [23], so that the rain streaks would be in the high-frequency part with nonrain textures. Then, we extract the HOG feature of the binary image from the high-frequency part. As shown in Fig. 3, there is no significant difference in the bottom HOG image which is computed from the original image. We use a sliding window whose size is the same with the templates to scan the whole image and compute the HOG similarity between templates and the patches. The similarity is calculated by Mahalanobis distance. Then, we choose the five best matched patches, and use their HOG features to form the 180 dimensional feature vector for the image.

3.1.4. Snowflake

As mentioned in the above section, it is difficult to detect snow, especially in single images. Snow is light and soft, and when there is wind, snowflakes will change their original direction and fly anywhere even to the upper sky. It is hard to describe the rule of the flying snowflakes, because the trajectory of snow is disordered.

We regard snowflakes as a kind of noise. Pixels are defined as snowflake noises when they have the following characteristics:

- The gray value of pixel $(x,y)$ is $L + \epsilon$, in which $L$ is the mean gray value of an image, $\epsilon$ is a threshold and greater than zero.
- The pixels inside the circle which is centered at pixel $(x, y)$ with $R$ as radius have the same gray value with $(x, y)$, while the pixels outside the circle do not have the same gray value with $(x, y)$.

For all the snowflake noise pixels in an image, we compute the histogram of intensity and hue of the patches they are located in to form the 200 dimensional feature vector.

3.1.5. Dark channel

Dark channel prior has been well studied in single image haze removal. The authors of [24] found that most local patches in haze-free outdoor images contain some pixels which have very low intensities in at least one color channel, i.e., the minimum intensity in such a patch should have a very low value. So, we utilize dark channel to indicate the haze weather. First, we divide an input image into patches. Then, we use the median value of dark channel intensities in these patches to form a 100 dimensional feature vector.

3.1.6. Global features

Contrast is a useful cue for weather classification as light is not the same in different weather conditions.

As the saturation is independent of illumination, it can represent different images under various illumination conditions. For an image $I$ we calculate the normalized saturation for each pixel by

$$s(x, y) = \frac{S(x, y) - \min(S)}{\max(S) - \min(S)}$$

where $\max(S)$ is the maximum saturation value and $\min(S)$ is the minimum saturation value of image $I$. For convenience of calculations in the following steps, we compute the histogram of the normalized saturation of an image to form the 10 dimensional feature vector.

3.2. Multi-feature dictionary learning

We first review the classical dictionary learning model. Suppose there are $N$ training data denoted as $X = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^{d \times N}$, $x_i \in \mathbb{R}^d$. From these data, we can learn a dictionary $D = [d_1, d_2, \ldots, d_\text{dK}] \in \mathbb{R}^{d \times K}$ by minimizing the objective function over $D$ and coefficient matrix $W = [w_1, w_2, \ldots, w_N] \in \mathbb{R}^{K \times N}$:

$$\arg \min_{D, W} \sum_{i=1}^{N} \|x_i - Dw_i\|_2^2 + \lambda \|w_i\|_1$$

s.t. $\|d_k\|_2 = 1, \quad \forall k = 1, \ldots, K$. (2)

Although the learned dictionary $D$ by Eq. (2) from $X$ is adapted for representation of the data, it is not suitable for classifying them. Outdoor images captured during daytime under different weather conditions cover a wide range of illuminations. The images are totally different in backgrounds but subtly different in intensity statistics [5]. They have some class-specific features which make them different from other weather conditions, and some common patterns which do not contribute to their discrimination. So, naive schemes are doomed to failure in this multi-class classification problem. If we use the same coding scheme, images of different classes would share lots of similar codes, and the proportion of discriminative codes would be very small. Moreover, we have to take advantage of multiple features to provide more information for multiple weather classification tasks.

Inspired by the work of [14], we learn multi-feature class-specific dictionaries and shared dictionaries for all the weather conditions. Suppose there are $Q$ class images in the training set, and each image has $Q$ different modalities of features. We denote the features with modality index $q(q = 1, \ldots, Q)$ that is associated with...
with the \(i\)-th \((i = 1, \ldots, N)\) class as \(X^q_i (X^q_i \in \mathbb{R}^{d_q \times m_i})\). \(d_q\) is the dimensionality of the \(q\)-th feature vectors. \(m_i\) is the number of features from the \(i\)-th class. The corresponding class-specific dictionary related to a single feature can be denoted as \(D^q_i (D^q_i \in \mathbb{R}^{k \times n_i})\). \(n_i\) is the number of codewords corresponding to this dictionary. The shared dictionary related to a single feature can be denoted as \(D^q (D^q \in \mathbb{R}^{k \times n})\) and the coefficients of \(X^q_i\) corresponding to \(D^q\) as \(W^q_i\). So, \(X^q_i\) can be denoted as \(X^q_i \approx D^q_0 W^q_0 + D^q_1 W^q_1 + \cdots + D^q_N W^q_N\). Then the \(q\)-th overall dictionary related to a single feature learning problem can be formulated as follows:

\[
\min_{D^q_0, D^q_i, W^q_i} \sum_{i=1}^{N} \left( \|X^q_i - [D^q_0^T D^q_i^T W^q_i]^T \|_2^2 + \lambda \|W^q_i\|_1 \right) + \sum_{i=0}^{N} m_i \left( \frac{\eta_i \|D^q_i^T D^q_i + \eta_{\lambda} \|W^q_i\|_1^2 \}^2}{n_i(n_i - n_i)} \right)
\]

s.t. \(\|D^q_i(:,j)\| = 1\), \(\forall i,j\),

in which \(W^q_i = [W^q_0, W^q_1, \ldots, W^q_N]\). \(D^q_i^T = [D^q_0^T, \ldots, D^q_{i-1}^T, D^q_{i+1}^T, \ldots, D^q_N^T]\), \(m_0 = \sum_{i=1}^{N} m_i\), \(I_{n_i}\) is an identity matrix whose size is \(n_i \times n_i\), \(\eta_i\) denotes the weight for self-incoherent term, and \(\eta_{\lambda}\) is the weight for cross-incoherent term, \(F\) denotes the Frobenius norm. We encode features via the global encoding strategy proposed by the authors of [14] based on the learned dictionaries for image.

![Fig. 4. Sample images in the MWI dataset.](image1)

<table>
<thead>
<tr>
<th>Type</th>
<th>Features</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local features</td>
<td>Sky</td>
<td>Channels a and b of Lab color space of the sky region</td>
</tr>
<tr>
<td></td>
<td>Shadow</td>
<td>The distance between a boundary and its k-nearest neighbors in sunny boundary set</td>
</tr>
<tr>
<td></td>
<td>Rain streak</td>
<td>The HOG similarity between rain streak patches templates</td>
</tr>
<tr>
<td></td>
<td>Snowflake</td>
<td>The histogram of intensity and hue of the snowflake patches</td>
</tr>
<tr>
<td></td>
<td>Dark channel</td>
<td>The minimum intensity values in patches</td>
</tr>
<tr>
<td>Global Features</td>
<td>contrast</td>
<td>The difference in color and brightness</td>
</tr>
<tr>
<td></td>
<td>saturation</td>
<td>One of three coordinates in the HSV color space</td>
</tr>
</tbody>
</table>

![Fig. 5. Performance of feature combination with new features progressively added to MKL.](image2)

| Table 1: The distribution statistics on MWI dataset. |
|----------|--------|--------|
| Label    | Sunny | Rainy | Snowy | Haze  |
| Number   | 10815 | 2342  | 2226  | 5004  |

<p>| Table 2: Features used in the experiments. |
|----------|--------|--------|</p>
<table>
<thead>
<tr>
<th>Type</th>
<th>Features</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local features</td>
<td>Sky</td>
<td>Channels a and b of Lab color space of the sky region</td>
</tr>
<tr>
<td></td>
<td>Shadow</td>
<td>The distance between a boundary and its k-nearest neighbors in sunny boundary set</td>
</tr>
<tr>
<td></td>
<td>Rain streak</td>
<td>The HOG similarity between rain streak patches templates</td>
</tr>
<tr>
<td></td>
<td>Snowflake</td>
<td>The histogram of intensity and hue of the snowflake patches</td>
</tr>
<tr>
<td></td>
<td>Dark channel</td>
<td>The minimum intensity values in patches</td>
</tr>
<tr>
<td>Global Features</td>
<td>contrast</td>
<td>The difference in color and brightness</td>
</tr>
<tr>
<td></td>
<td>saturation</td>
<td>One of three coordinates in the HSV color space</td>
</tr>
</tbody>
</table>

| Table 3: Weighting coefficients obtained by MKL. |
|----------|--------|--------|
| sk       | sh     | ra     | sn     | da     | co     | sa     |
| 0.310    | 0.262  | 0.285  | 0.264  | 0.302  | 0.255  | 0.282  |
| 0.6306   | 0.3565 | 0.5944 | 0.4528 | 0.6222 | 0.3480 | 0.5111 |

Z. Zhang et al. / Neurocomputing 207 (2016) 365–373
representation. By using the max pooling on the sparse codes, an image can be represented by \( Q \) feature vectors.

### 3.3. Feature fusion and classification

To select an optimal combination of features from the feature pool, we use the multiple kernel learning to learn the optimal weights for all features. If the weighting coefficient is larger than others, its corresponding feature contributes greater to the overall classification performance than other features. So, we choose the feature with the largest weighting coefficient as a start. Then, other features are respectively added in for testing. The combination of features which has the best performance, will be chosen as a new start. Repeat this process until the performance stops increasing. The selected feature combination provides the optimal choice.

Let \( T = \{x_i, y_i\}_{i=1}^N \) be the training image dataset, where \( x_i \) denotes the \( i \)-th sample, \( y_i \) denotes the corresponding class label, and \( N \) is the number of training images. Suppose each image has \( Q \) different modalities of features, and each type of feature has \( M \)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Sunny</th>
<th>Rainy</th>
<th>Snowy</th>
<th>Haze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.70</td>
<td>0.01</td>
<td>0.24</td>
<td>0.05</td>
</tr>
<tr>
<td>Recall</td>
<td>0.01</td>
<td>0.57</td>
<td>0.32</td>
<td>0.10</td>
</tr>
<tr>
<td>F-Measure</td>
<td>0.24</td>
<td>0.32</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.7139</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Classification accuracy confusion matrix of the proposed method on MWI dataset.

Fig. 7. Classification accuracy on each weather category via related methods.

Table 5

<table>
<thead>
<tr>
<th>Method</th>
<th>[2]</th>
<th>[3]</th>
<th>[6]</th>
<th>[27]</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.2267</td>
<td>0.1889</td>
<td>0.4158</td>
<td>0.5944</td>
<td>0.7139</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.2388</td>
<td>0.4282</td>
<td>0.3066</td>
<td>0.4280</td>
</tr>
<tr>
<td>MKL</td>
<td>0.5932</td>
<td>0.5943</td>
<td>0.5938</td>
<td>0.5944</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.7300</td>
<td>0.7137</td>
<td>0.7217</td>
<td>0.7139</td>
</tr>
</tbody>
</table>

Fig. 8. Classification accuracy on different number of training samples via related methods.

kernel functions. We denote the kernel function corresponding to the \( q \)-th feature as \( k_m^q(x_i, x_j) = k(f(x_i) f(x_j)) \). We aim to train a multi-kernels based classifier with a decision function \( f(x) \) to predict the weather class of an unlabeled image \( x \). In this work, we use some linearly combined base kernel functions to determine an optimal kernel function:

\[
k(x_i, x_j) = \sum_{q=1}^{Q} \sum_{m=1}^{M} \beta_m^q k_m^q(x_i, x_j),
\]

where \( \beta_m^q \) is one of the linear combination coefficients, and \( \beta_m^q \geq 0 \) \( \forall q \ \forall m \). Given the input feature \( x \) of the image, the decision function is defined as follows:

\[
f(x) = \sum_{q=1}^{Q} \sum_{m=1}^{M} \beta_m^q k_m^q(x) + b,
\]

where \( \alpha \) and \( b \) are the parameters of the standard SVM. Then, the objective function can be formulated as follows:

\[
\min_{\beta, \alpha, b} \frac{1}{2} \sum_{q=1}^{Q} \sum_{m=1}^{M} \beta_m^q \alpha^T K_m^q \alpha + C \sum_{i} \xi_i,
\]

s.t. \( y_i \cdot \sum_{q=1}^{Q} \sum_{m=1}^{M} \beta_m^q K_m^q(x_i) \alpha + y_i b \geq 1 - \xi_i \ \forall i \), \( \xi_i \geq 0 \ \forall i \), \( \beta_m^q \geq 0 \ \forall q \ \forall m \),

in which \( K_m^q(x_i) = [k_m^q(x_i, x_1), \ldots, k_m^q(x_i, x_p)] \), \( p \) is the number of weather classes, \( C \) is a positive constant. We adopt the gradient descent algorithm to solve the optimization problem in (6). The parameters \( \alpha, \beta \), and \( b \) can be calculated iteratively. We use one-against-all strategy to transform the multi-class classification into two-class classification. If there are \( P \) classes, then the objective function can be rewritten as:

\[
J = \sum_{p=1}^{P} J_p(\beta, \alpha_p, b_p),
\]

where \( J_p \) is a two-class classifier, the positive samples are the samples with the class label \( p \), the negative samples are the samples with other labels. We can obtain the class label by:

\[
y = \arg \max_{y_p} f_p(x)
\]

where \( f_p(x) \) has been described in Eq. (5).
4. Dataset and experiment

4.1. Dataset

We evaluate our approach on our dataset called MWI (Multi-class Weather Image) set. It contains 20 K images obtained from many web albums and films, such as Flicker, Picasa, MojiWeather, Poco, Fengniao. As shown in Fig. 4, most of the images have totally different backgrounds. The distribution statistics on MWI dataset is listed in Table 1. The images are collected by five volunteers, and they chose images with their own common sense. Then, they were asked to label all the images according to their knowledge. So, each image has five labels after the first round label. The majority voting mechanism is applied to predict the final label by choosing
the class that gets the highest number of votes. More information and examples of the dataset can be found in the released website\(^1\).

4.2. Experiment setting

We randomly select 4K images from the MWI dataset to evaluate our method. Then, we list the features used in the experiments in Table 2. In our experiment, we set K = 5 for the K-nearest neighbors in the shadow feature part. The following parameters are chosen by cross validation and fixed throughout all evaluations. For extracting rainy feature, we resize an image into 256 × 256, and set 8 × 8 as the cell size, the step is 4, and the size of the sliding window is 16 × 16. For extracting snowy feature, we set the threshold \( e = 0.3 \times L \), the radius \( R \) is \( \sqrt{2} \). For extracting dark channel feature, we resize the input images into 450 × 450, and set the size of a patch as 45 × 45. For learning dictionaries, we randomly sample about 12 K features from the training set as the initialization. For the weight of sparsity term \( \lambda \), we set it to be 0.3 because it has been experimentally shown that good performance in [25]. For the weight of incoherent terms, we set the ratio between the entries in self-incoherence term and cross-incoherence term to be 1:2, and set \( \eta_d = 0.1 \). For feature selection, we adopt the DOGMA toolbox [26] to learn the optimal weights for all features. To cope with multi-class classification task using SVM, we use the default setting in LIBSVM.

4.3. Experimental results

First, we use the MKL algorithm to learn the optimal weights for all features and select the optimal combination of features. Table 3 shows the weights obtained by MKL and the classification accuracy of each feature. We use \( sk, sh, ra, sn, da, co \), and \( sa \) to denote sky, shadow, rain streak, snowflake, dark channel, contrast, and saturation features, respectively. From the table, we can see that the feature with a higher weight has better classification accuracy. Sky feature holds the largest weight and the best accuracy. That is to say, sky feature makes the greatest contribution to the classification performance. Thus, we choose the sky feature as the start to test the performance and find out the optimal feature combination. We combine other features with the sky feature for testing respectively. Then, the combination with the best performance will be chosen as a new start. We repeat this process until the performance stops increasing. As shown in Fig. 5, the combination of sky, dark channel, rain streak, and saturation outperforms other combinations. In the following experiments, we choose this combination as the optimal feature selection for our task. This procedure can reduce complexity by avoiding unnecessary feature extractions and sparse coding.

Table 4 shows the comparison results of our approach and the baseline methods. To show the best performance of all methods, every method produced multiple results using a group of reasonable parameters. The first baseline is to implement SVM directly on the original feature vector without sparse coding and feature selection. The second baseline is implementing MKL on the original feature vector. Our method encodes the original features with the learned dictionaries and selects optimal feature combinations.

Fig. 6 shows the classification accuracy confusion matrix of MWI dataset. We list some samples from the successfully detected images shown in Fig. 9. We can see that 32% of the rainy images have been recognized as snowy images. Some failure cases are shown in Fig. 10. We check the extracted low-level features that

our rain streak feature extraction method is negatively impacted by the detected snowflake and saturation features.

We also compare our method with some related image classification methods as shown in Table 5. These results are substantially better than the performance in previous works, which demonstrate the effectiveness of our approach. We leverage the latest developments in machine learning, such as dictionary learning, and multiple kernel learning which are partly missing in the existing works.

To compare the classification performance with different number of training samples, we choose the method of [6,27], which have better performance in the previous experiment (Fig. 8). Fig. 7 illustrates the classification performance on each weather category by related methods. Our method outperforms other methods on each weather category.

5. Conclusion

We presented a framework for multi-class weather classification from single images in any scene. Our approach learned multi-feature dictionaries on each feature to improve the discrimination of image representation and enhance the performance of multiple weather classification. We utilized MKL algorithm to learn an optimal linear combination of feature kernels. For training and testing our approach, we collected the MWI (Multi-class Weather Image) set. We evaluated our approach on the dataset, and the results show the effectiveness of our method. Our work can be applied for the detection and observation of weather conditions, image/video analysis, the reliability improvement of video surveillance systems, and so on.

Acknowledgment

The research reported in this paper is supported by the National Natural Science Foundation of China under Grant Nos. 61332005 and 61402048; The Funds for Creative Research Groups of China under Grant No. 61421061; The Consponsored Project of Beijing Committee of Education; The Beijing Training Project for the Leading Talents in S&T (Jrc 201502).

References


\(^1\) The dataset is released in http://mwidataset.weebly.com.