Optik 125 (2014) 7239–7243

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Order based feature description for high-resolution aerial image classification

Wei Huang[a,*], Lingda Wu[b], Yingmei Wei[c], Hanchen Song[c]

[a] China Satellite Maritime Tracking and Control Department, Jiangyin, 214431, China
[b] The Key Laboratory, Academy of Equipment, Beijing, 101416, China
[c] School of Information System & Management, National University of Defense Technology, Changsha, 410073, China

ARTICLE INFO

Article history:
Received 13 December 2013
Accepted 20 July 2014

Keywords:
Bag of visual words (BoVW)
Local invariant features
Feature description
Aerial image classification

ABSTRACT

We propose to use two intensity order-based descriptors for classification of high-resolution aerial images. By analyzing the characteristics of aerial images, it points out that the imagery does not have an absolute reference frame, and can be seen as a fusion of natural image and texture image. For these characteristics, it uses two intensity order based approaches to extract low-level features of aerial images. The two descriptors are inherently rotation invariant and encode complementary information about the image. It then uses a bag of visual words model to build a holistic representation of images. Experiments results on publicly available aerial scene imagery dataset show that linear combination of order based features encodes complementary information and is significantly better than SIFT. In addition, we make a consistent comparative analysis of different classification frameworks and several feature coding methods based on aerial image dataset.

1. Introduction

The high-fidelity aerial image provided by the advanced space-borne sensors can provide abundant spatial and textural information for land-use/cover classification. However, efficiently analysis and recognition these detailed images is an challenge task due to vast quantities of information. The conventional spectral-classification methods is to model aerial scenes using a bottom-up manner [1,2]. It firstly extracts spectral, textural, and geometrical attributes as features and then uses these features to classify pixels or local segments into their thematic classes. Over the last decade, the local invariant features have proven effective in computer vision community. Nowadays, many methods analyze the aerial images by extracting local invariant features [3,4].

Compared to natural images, aerial scene images generally do not have an absolute reference frame, which means that these images are not sensitive to rotation changes. The conventional descriptors, such as SIFT [5], GLOH [6], are not totally rotation invariant since they need to compute a dominant orientation for each image patch, but unfortunately the computation of dominant orientation is not reliable [7]. In this paper, we use intensity order based descriptors which are inherently rotation invariant to extract low-level features. In addition, aerial images can be seen as a fusion of natural image and texture image. As a result, our feature extraction methods include descriptors for natural images and texture images. They encode complementary information about the image. We begin by extracting low-level features from the aerial images, then we construct a holistic representation of the images by bag-of-visual-words model. The two features are combined linearly by multiple kernel learning.

The major contributions of this paper can be summarized as follows. First, we use two order based features to describe the aerial images according to their characteristics. Our approach is shown to result in higher classification rates on the aerial image dataset than SIFT. Second, we perform a thorough evaluation of the effects of different classification framework and encoding methods on aerial images by fixing the pipeline and its tuning.

The remainder of the paper is organized as follows. In Section 2 we review related work. In Section 3 we analyze the characteristics of aerial images and describe the two order-based features. Details of our experiments and results are presented in Section 4. Section 5 concludes the paper.

2. Related work

The conventional aerial image classification methods use a bottom-up manner for high-resolution aerial image classification.

* Corresponding author.
E-mail address: weeihuang@hotmail.com (W. Huang).

http://dx.doi.org/10.1016/j.ijleo.2014.07.127
0030-4026/

However, the effect of these algorithms relies heavily on the original image space segmentation [1].

More recently, BoVW (bag of visual words) based approaches have been used for aerial scene classification [8]. The basic bag of visual words model [9] typically quantizes low-level features using a visual dictionary typically constructed through k-means clustering. The set of visual words is then used to represent an image regardless of their spatial arrangement which is represented as a histogram. The visual word histograms computed over the image form the holistic representation of the image.

SPMK (spatial pyramid matching kernel) [10] is an efficient extension of the orderless BoVW representation. The visual words histograms of an image computed at different scales and spatial bins defined by the spatial pyramid representation are concatenated to construct the final vector. It takes into account the spatial information of visual words in the image space. Yang proposed SPCK (spatial pyramid co-occurrence kernel) [11] which characterizes both the absolute and relative spatial arrangement of visual words in an image. They combined the SPCK model with the BoVW approach to obtain a spatial extension in the latter. A higher classification accuracy was reported for their extended spatial co-occurrence kernel over the BoVW and SPMK approaches.

Recently, a large number of novel encoding methods for BoVW model have been proposed to improve on the standard hard quantization and histogram based encoding. ScSPM [12] proposes sparse coding (SC) instead of hard quantization to learn and code features. The method achieved state-of-the-art performance on several benchmarks. However, SC requires to solve L1-norm optimization problem, which is computationally expensive. LLC (locality-constrained linear encoding) [13] utilizes the locality constraints to project each descriptor into its local-coordinate system, and the projected coordinates are integrated by max pooling to generate the final representation. KCB (kernel codebook encoding) [14] applies techniques from kernel density estimation to allow a degree of ambiguity in vector quantization. Compared to SC, these methods are easy to compute. In this paper, we will evaluate these encoding methods on aerial image dataset.

The paper [15] uses local structural texture similarity descriptor as low-level features for aerial image classification. Similar to the classification of natural images, a lot of aerial image classification literatures use dense SIFT as the low-level features [4,16]. However, the local features, such as SIFT [5], GLOH [6], are not totally rotation invariant. Recently, MROGH has been proposed by Fan [7]. The descriptor is inherently rotation invariant thanks to a rotation invariant way of gradient computation and order based pooling. In addition, the SRP descriptor [17] which is proposed for texture image classification is theoretically rotation invariant and computationally efficiency.

3. Methods

In this section, we firstly analyze the characteristics of high-resolution aerial images compared with natural images. Then, we briefly review two intensity order based descriptors which we will use as low-level features for aerial images.

3.1. The characteristics of aerial images

Compared to natural images, aerial image has its own characteristics. Here we use UcMerced dataset [16] as an example to analyze the characteristics of high-resolution aerial image. UcMerced dataset has manually extracted aerial orthoimagery

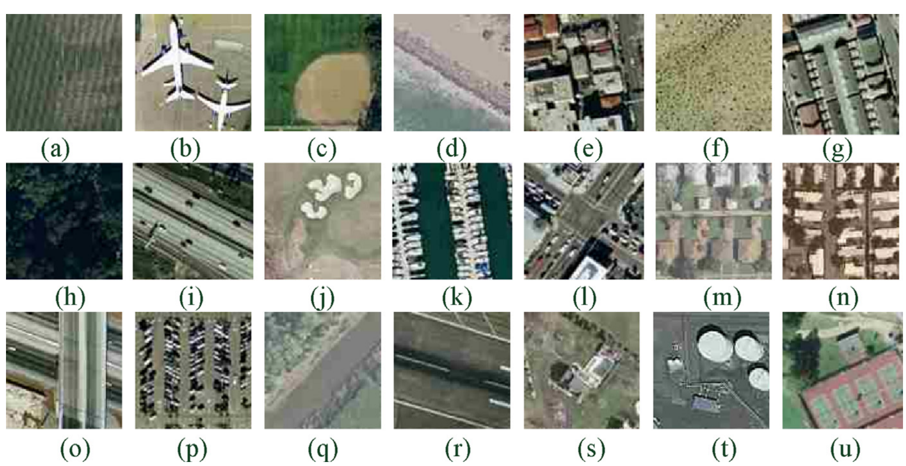

Fig. 1. UcMerced dataset examples. (a) agriculture, (b) airplane, (c) baseball diamond, (d) beach, (e) buildings, (f) chapa rral, (g) dense residential, (h) forest, (i) freeway, (j) golf course, (k) harbor, (l) intersection, (m) medium-residential, (n) mobile homepark, (o) overpass, (p) parking lot, (q) river, (r) runway, (s) sparse residential, (t) storage tanks, (u) tennis court.

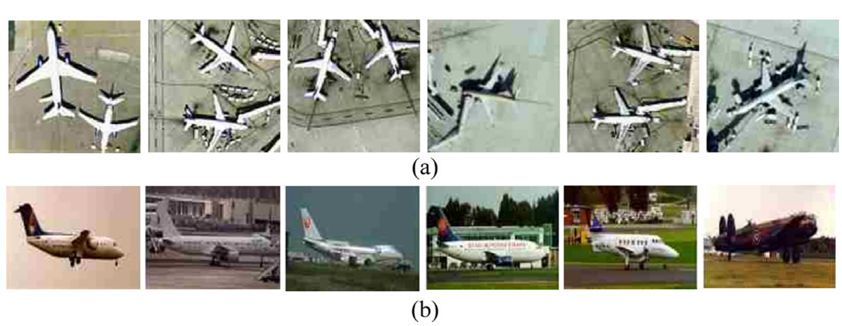

Fig. 2. Example images of airplane class. (a) from the UcMerced (b) from Caltech101.