A Subblock Partition Of Multi-Layer Pattern Based Image Classification Approach

LV Xuan^{1,2} MA Zezhong LIU Qing

¹Chongqing Land Resources Housing Surveying and Planning Institute

² School of Remote Sensing and Information Engineering, Wuhan University

Huangshandadao Road No.64, yubei, Chongqing, China

+86 13668093020

Chongqing Land Resources Housing Surveying and Planning Institute

Huangshandadao Road No.64, yubei, Chongqing, China

+86 13896829668

Chongqing industrial& commercial school

baisha, jiangjin, Chongqing, China

+86 13618381908

ABSTRACT: Since traditional partition approach may construct very different image representation because of the changed locations of objects in the same image, a subblock partition of multi-layer pattern method for image representation is proposed. The saliency windows straddled by superpixels are utilized to partition the image into multi-layer pattern subblocks. Then all the subblocks are combined to a three order tensor. Comparing to the results of image classification item of Pascal Voc 2007 Challenge, it indicates that the proposed representation method is robust to the varied object locations and achieves better performance than other approaches.

CCS Concepts

Computing methodologies→Computer vision • Computing methodologies→ Machine learning

1 INTRODUCTION

Accurate depiction of the pattern information in the image is the premise to realization of the accurate description and classification of the image. In a traditional method, the image is usually partitioned into rectangular blocks^{[1][2]} at different scales, and the apparent information and position distribution information in the image is expressed by description of the information of each subblock. The Bag of Feature, a well-known method for describing pattern information in an image, is derived from the Bag of Words (BOW) method in text categorization. The traditional BOW method^[3-6] describes images by means of statistics of the frequency of occurrence of all feature words in the image, and as a disordered expression based on a local feature, it ignores the spatial position information in the image. In literature^[1], by constructing a spatial multi-layer pyramid, the image is partitioned into different numbers of square subblocks at each layer to express the position information, which improves the accuracy of the classification to a certain extent, but the position information is expressed weakly. The transformation stability is poor and the high dimension vector description of the resulting image is not easily measurable. In literature^[7], a method of oblique line projection and circumferential projection is proposed to describe the

position information of image features. With more position information than the SPM method, it shows some robustness to the rotation transformation in the image, but this method focuses more on the "linear" and "rotating" distribution of the image, and fails to fully reflect the various transformations that actually exist in the image. Literature^[8-11] analyzes the symbiotic information of overall or local feature phrases (Phrases) in the image, and does not reflect the hierarchical information in the image. Moreover, the feature phrase information will grow exponentially with the increase in the number of words, so only part of the information is usually selected, resulting in the loss of important discrimination information in the image. However, sub-lattice partition similar to SPM method is used in the text, there is still a problem that the same object is split so that the occurrence of the same object in various positions in a scenario will form a completely different feature description, lacking in the description of the overall information of the target in the image.

In light of the above problems, the paper proposes a new image classification method. Firstly, multiple potential target windows are acquired through the salient area distribution^{[12][13]} calculation, the spanning characteristics of window and superpixel block^[14], and similarity of a bridging superpixel within and outside the window and within the window are analyzed to obtain a

*Corresponding author: shixuanlv305@126.com, 313782848@qq.com, 43036358 @qq.com

target window with the best target fitness and the probability of occurrence of the target in each window. Finally, according to the different layers set, the mode subblocks are partitioned according to the probability of the occurrence of a target in the window, and the tensor description is performed on each hierarchical mode subblock to form a final image description for classification.

2 SUBBLOCK PARTITION OF MULTI-LAYER PATTERN

(a) Layer 0 (b) Layer 1 (c) Layer 2
Fig.1 Spatial pyramid partition

(a) Layer 1 (b) Layer 2 (c) Layer 3 (d) Layer 4
Fig.2 Pattern subblock

In order to better express the information in an image, this paper proposes a new partitioning method for image subblock partition: image multi-layer pattern subblock partition. The pattern sub-block partition at different layers of the image means that the image is assumed to be partitioned into three layers to extract the pattern subblocks respectively. The original picture is regarded as a whole in the first layer, the image is roughly partitioned into two pattern blocks in the second layer, each representing a similar pattern in the image (e.g. grass, land), and the image is partitioned into three pattern blocks (e.g. land, grassland and food) in the third layer, as shown in Figure 2. Different from the SPM model in the classical method, the image is partitioned into sub-squares at different layers in SPM, and all pictures are partitioned in the same way. Despite of more spatial position information than a traditional dictionary model, such a hard partition suffers a problem that the same object is partitioned into different subblocks, resulting in an inaccurate description of the picture and failure of the final classification. The partition of the pattern subblock is also different from the image segmentation. When the texture, color and other features of the target are different saliently, the image segmentation result will partition the same object into multiple blocks, and the pattern subblock partition is intended to partition the same target in the image into one block (that is, other backgrounds are allowed in the block, but the main part of the block is a certain target), rather than to carry out the exact segmentation of the target. There may be a plurality of target objects in an image, and to capture and partition a plurality of target objects in an image into different subblocks, a method based on salient area detection and superpixel feature analysis is used in this paper to achieve the partitioning of pattern subblocks.

Fig.3 Saliency map

Fig.4 Analysis of window and superpixel

First, the salient area distribution of the image under multi-scale is detected through the Fourier conversion spectrum residual [12] method, as shown in Figure 3, and the T windows in different positions and sizes are sampled according to the salient area distribution; the Lab features and position features of the image are analyzed and clustered to achieve superpixel segmentation, using Kmeans, Ncut, mean shift, graph cut model, etc. Finally, by analyzing the position and distribution relationship between each window and its neighboring super-pixels, the probability of occurrence of each window is determined, and then the pattern subblock is extracted. The key problem is to position a plurality of target objects in the image and their corresponding external windows. To achieve accurate positioning, the relationship between windows and superpixels is analyzed as follows:

As shown in Figure 4, all potential target windows are generated by the distribution of salient areas, reflecting the potential target areas in the image, so each window w in the saliency map contains information on the extent of salient targets within the metric window. Superpixels partition an image into multiple small areas of similar color and texture, a target to meet the pixels in each superpixel block belongs to the same object, but one object may comprise a plurality of superpixels. Superpixels under the strong boundary target will not cross the boundary of the target, so superpixels have the effect of maintaining the target boundary. Due to different distributions of positions between the window generated by the saliency map and superpixels, in order to ensure the integrity of the target object in the window and to maximize the fit of the window to the target, a metric way as shown in formula (1) is designed to calculate the fit of the window to the target object in the window, the higher the fit of the window with the target object, the larger the value SW .

$$SW(w, \lambda_0, \lambda_1, \lambda_2) = \lambda_0 \left(1 - \sum_{s \in S_N} \frac{\min(|s_{out-w}|, |s_{in-w}|)}{|w|} \right) + \lambda_1 \sum_{s \in S_N} \text{sim}(s_{in-w}, S_{all_in-w}) - \lambda_2 \sum_{s \in S_N} \text{sim}(s_{out-w}, S_{all_in-w}), \quad (1)$$

In the above formula, w is for a window, s for a superpixel, S_N for a superpixel set adjacent to w , s_{in-w} for a portion of the superpixel located in w , s_{out-w} for a portion

of the superpixel located outside w , and $S_{all_in_w}$ for all superpixels in w . The first term in formula (1) represents the extent to which all superpixels adjacent to window w spanning the window boundary. The larger the value, the closer the window is to the target object. As shown in Figure 4, assume that s_1 , s_2 , s_3 are some three superpixels adjacent to the window w_l . When all the superpixels in S_N are completely in w_l (such as superpixel s_3), the partial area of s_{out_w} for each superpixel is 0,

$$\sum_{s \in S_N} \frac{\min(|s_{out_w}|, |s_{in_w}|)}{|w|} = 0 \quad (2)$$

in which case, the first item gets the maximum value of 1. When most of the superpixels are located in w_l (such as superpixel s_2) or a small portion of superpixels are located in w_l (such as superpixel s_1), the superpixel has a small spanning to the boundary of window w_l , and the value obtained by the first term is larger. When the portion of a superpixel located inside the window is equivalent to that outside the window (such as superpixel s_1 vs. w_2), the window has poor fit to the target, and the value obtained by the first item is smaller. The second term in the formula indicates the degree of similarity in distribution between the portion of the superpixel s within the window and all the superpixels within the window. The larger the value, the greater the overall inclusion by the window of the target. The third term in the formula indicates the degree of similarity in distribution between the portion of the superpixel s outside the window and all the superpixels outside the window. The larger the value, the poorer the overall inclusion by the window of the target. When similarity is calculated, s_{in_w} , s_{out_w} , $S_{all_in_w}$ is expressed as a Gaussian distribution described by mean and variance. $sim(\cdot, \cdot)$ is measured by χ^2 distance and is defined as follows:

$$sim(P, Q) = \frac{1}{2} \sum_i \frac{(P_i - Q_i)^2}{(P_i + Q_i)} \quad (3)$$

By comparison of SW values of each window, the first n windows $w_1, w_2, w_3, \dots, w_n$ with the highest probability of occurrence of targets are obtained, and each window represents a pattern subblock. Moreover, pattern subblocks at different layers of the image can be constructed. If the layer is 3, the image is partitioned into three blocks. The first two blocks are the two most salient target windows in the figure, and the third block is the remaining area $I - w_1 \cup w_2$ of the image except the first two windows. The coarse partitioning method based on the pattern subblock can not only express the salient information in the image, but also has the basic pattern information in it, which is beneficial to form an accurate description of the image under the dictionary model.

3 TENSOR DESCRIPTION OF COMPLEX IMAGE PATTERN INFORMATION

3.1 Multi-feature dictionary creation and feature quantization

The traditional feature dictionary creation method is partitioned into three steps: 1) feature detection; 2) feature description; 3) feature clustering. There are many feature detection methods, such as Harris-Laplace feature detection, Gaussian difference, edge area-based and salient area feature detection methods, and the feature points detected by different feature detection methods can describe the feature information of different aspects of the image. The method of feature description is the expression of the detected feature points, e.g. SIFT descriptor and HOG descriptor. The feature dictionary created by different feature detection methods and feature description methods can express different feature information (e.g. gradient information, color information, etc.) in the image, so that the image content is described in a more sufficient way. This paper constructs three types of feature dictionaries: SIFT feature dictionary, color distribution feature dictionary and texture feature dictionary.

1) SIFT feature dictionary. The first type of feature dictionary is constructed by Difference of Gaussian feature detection and SIFT feature descriptor. The SIFT features detected in all training set images are clustered to obtain a feature dictionary $D_1 = [d_1, \dots, d_K] \in R^{N \times K}$ of size K . If the feature set $X = [x_1, \dots, x_M] \in R^{N \times M}$ is a set of M N -dimensional SIFT feature points detected in an image, as the image features have obvious sparse characteristics, under the overcomplete dictionary description, the sparse representation can obtain lower reconstruction errors, and can make the feature expression be more specific to salient features [7]. Therefore, the following sparse coding is used in this paper to cast features onto sparse feature words:

$$\min_U \sum_{m=1}^M \|x_m - D_1 u_m\|_2^2 + \lambda \|u_m\|_1, \quad (4)$$

$$s.t. \|d_k\| \leq 1, \quad \forall k = 1, 2, \dots, K.$$

Let $U = [u_1, \dots, u_M] \in R^{K \times M}$, the elements in the K -dimensional vector represent the linear coefficient of the feature point x_m represented by the K words in the dictionary, and U is the sparse coding of the feature set X under the feature dictionary D_1 .

2) Feature dictionary of color distribution. A second type of feature dictionary is constructed by dividing the image into blocks and counting the subblock Lab color space features. First, according to the multi-layer pattern subblock partition method proposed in the first part, each image in the training map set is partitioned into sub-pattern subblocks of different layers. The Lab color space feature of each subblock is counted, and each color component is represented by mean and variance to obtain a 6-dimensional feature vector $[\mu_L, \sigma_L, \mu_a, \sigma_a, \mu_b, \sigma_b]^T$. If there are n training set images and each image is partitioned into m layers, there are a total of subblocks. All subblock feature vectors are clustered into K centers (the same number as

cluster centers in the first type of dictionary) to obtain a feature dictionary $D_2 \in R^{6 \times K}$. The coding method for the new feature is the same as the sparse coding method in 1).

3) Texture feature dictionary. The texture feature dictionary is constructed by K-SVD^[15] method. Firstly, the training set sample picture is partitioned into subblocks of 8×8 , K subblocks are randomly selected from all subblocks to obtain the texture feature dictionary $D_2 \in R^{64 \times K}$, and then the atoms are updated using the K-SVD method to get the final texture feature dictionary.

After the feature set X is represented by the feature dictionary D in a sparse coding manner, the sparse expression coefficient of each element x_i in the set is obtained. Therefore, the coefficient matrix U reflects the sparse representation relationship between all the features in X and the dictionary D . In order to better express the salient features in the image, the MaxPooling method used in literature^[7] is adopted, which demonstrates a good stability to the spatial variation of local features.

The image is first transformed into multi-layer pattern subblocks, and the feature set X in each subblock is calculated. Since the feature information distribution is counted in the partitioned pattern subblocks, the feature information of the subblocks after being described by the dictionary is concentrated on the feature words related to the corresponding patterns, so that the distinguishability of the feature description is greatly enhanced. In the Figure 2, after the hedgehog pattern subblock is described by the dictionary, most of the features are concentrated on the dictionary words related to the hedgehog feature. The positional movement of the hedgehog in the figure has less influence on the final image description. Therefore, the resulting dictionary feature description has more distinguishing sparse characteristics. Then, using the feature set X , after description by the dictionary, the sparse coefficient matrix $U = [u_1, \dots, u_M] \in R^{K \times M}$ is obtained, and the feature vector $F = [f_1, f_2, \dots, f_k]$ of X is constructed as follows:

$$f_j = \max \{|u_{j1}|, |u_{j2}|, \dots, |u_{jM}|\} \quad (5)$$

The feature vector F describes a relatively salient portion of the subblock feature set X that corresponds to the feature dictionary word. The feature selection method is not susceptible to local feature changes, and the expression of the feature distribution is more stable.

3.2 Tensor description of image pattern information

Compared to the vector pattern, the tensor can better reflect the pattern in the original state of the data, and the dimensionality and complexity will also be reduced. In the traditional partitioning method, the image is usually constructed as a multi-scale pyramid, and then the checkerboard partitioning is made in each layer. After each subblock is described by a feature dictionary, all the feature vectors are concatenated to represent one image, as shown in Figure 5. When the number of words in the dictionary is large, a feature vector with high dimensionality is often formed, and the measurement of similarity between high-dimensional feature vectors is

very difficult. In addition, when the target position changes alone in the image, a huge feature vector will also be formed. Therefore, based on the pattern subblock partition, a tensor description of the image is used in this paper.

First, the histogram dictionary description of the multi-layer pattern subblock is represented as a Tensor Representation^[16]. As shown in Figure 6, the row vector of two-dimension tensor in the figure represents the feature information of each pattern subblock. Since the feature information is counted in partitioned pattern subblocks, the feature information is concentrated on the feature words related to the corresponding pattern (such as grassland) after the subblock is described by the dictionary, so that the distinguishability of the feature description is greatly enhanced. For example, in the figure, with regard to the first behavior pattern subblock 1 of the two-dimension tensor (the main pattern is grassland), after the subblock is described by the dictionary, most of the features are concentrated on the dictionary words related to the grassland features, and thus, the resulting feature description has more distinguishing sparse characteristics (the number of features associated with grassland is much higher than the number of features associated with other patterns).

Secondly, in combination with a multi-dictionary model that reflects different feature information of images, an image is described as a third-order tensor, where the first order is the total number N of pattern subblocks at different layers, the second order is a K dimension feature vector described by a feature dictionary, and the third-order is the number P of feature dictionaries reflecting different feature information in the image. After the image is partitioned by the multi-layer pattern subblock, multiple pattern subblocks can be obtained. After each pattern subblock is described by the feature dictionary, it can form a second-order tensor, and the tensor row vector represents the feature information of each pattern subblock. Moreover, after each pattern subblock is described by other dictionaries (color feature dictionary, texture dictionary), the corresponding two-dimension tensor can also be constructed. Therefore, under the multi-dictionary model, an image can be described as a third-order tensor $\Psi \in R^{N \times K \times P}$, as shown in Figure 5. Assume that $\Omega = [V_1, \dots, V_P]$ is the set consisting of P feature dictionaries (where P is 3), where $V_p \in R^{D \times K}$ is the feature dictionary in the D dimensional feature space in which the p -th size is K , and all feature dictionaries in Ω have a size of K . Each pattern subblock can be represented by P feature dictionaries as P K -dimension feature vectors.

Finally, under the tensor feature description, the tensor canonical correlation analysis^[17] is used to extract the image features, and then the image features are accurately described. The image is described by the feature dictionary set Ω to obtain a tensor description $\Psi \in R^{N \times K \times P}$. Canonical correlation analysis in the tensor pattern can measure the similarity between two tensor data. The calculation pattern can be partitioned into a joint multi-mode sharing pattern and a single-mode sharing pattern.

For two third-order tensors $\Psi_1, \Psi_2 \in R^{N \times K \times P}$, joint multi-mode sharing means that any two of the three patterns are shared, and only a model transformation is performed on

one unshared pattern. That is, look for $\Theta = \{(u_p, v_p), (u_k, v_k), (u_n, v_n)\}$, make

$$\rho = \max_{\Theta} \langle X', Y' \rangle \quad (6)$$

Where,

$$(X')_{NKP} = (\Psi_1 \times_p u_p)_{nk} (\Psi_1 \times_k u_k)_{np} (\Psi_1 \times_n u_n)_{kp} \quad (7)$$

$$(Y')_{NKP} = (\Psi_2 \times_p v_p)_{nk} (\Psi_2 \times_k v_k)_{np} (\Psi_2 \times_n v_n)_{kp} \quad (8)$$

is the inner product of two tensors. For example, two third-order tensors $A, B \in R^{I \times J \times K}$, then

$$\langle A, B \rangle = \sum_{i,j,k} (A)_{ijk} (B)_{ijk} \quad (9)$$

The single-mode sharing pattern means that only one pattern is shared, and a model transformation is performed on the other two unshared patterns. That is, look for

$$\Phi = \left[\left\{ (u_k^1, v_k^1), (u_p^1, v_p^1) \right\}, \left\{ (u_n^2, v_n^2), (u_p^2, v_p^2) \right\}, \left\{ (u_n^3, v_n^3), (u_k^3, v_k^3) \right\} \right] \quad (10)$$

Let $\rho = \max_{\Phi} \langle X', Y' \rangle$, where,

$$(X')_{NKP} = (\Psi_1 \times_k u_k^1 \times_p u_p^1) (\Psi_1 \times_n u_n^2 \times_p u_p^2) (\Psi_1 \times_n u_n^3 \times_k u_k^3) \quad (11)$$

$$(Y')_{NKP} = (\Psi_2 \times_k v_k^1 \times_p v_p^1) (\Psi_2 \times_n v_n^2 \times_p v_p^2) (\Psi_2 \times_n v_n^3 \times_k v_k^3) \quad (12)$$

For example, when the first patten of tensor $\Psi \in R^{N \times K \times P}$ is shared, the following formula can be solved by overlapping SVD method:

$$\max_{U_k, V_k, U_p, V_p} \langle \Psi_1 \times_k U_k \times_p U_p, \Psi_2 \times_k V_k \times_p V_p \rangle \quad (13)$$

The canonical orthogonal set $\{(U_k, V_k), (U_p, V_p)\}$ (where, $\{(u_k, v_k \in \mathbb{R}^K), (u_p, v_p \in \mathbb{R}^P)\}$ is the middle column vector) is obtained to get d canonical correlation values $(\rho_1, \dots, \rho_d)$, $d < \min(K, P)$.

Hence, through the tensor correlation analysis^{[17][18]}, a canonical correlation feature can be obtained (the two sharing patterns produce $3 \times d$ features separately). Each feature represents the similarity of different semantic aspects of the data and can be used for creation of classifiers.

The advantage of tensor description lies in that it is an extension and supplement of the traditional vector pattern. Compared with the vector pattern, the tensor can better reflect the pattern in the original state of the data, and the dimensionality and complexity will also be reduced. Therefore, this description can avoid the generation of high dimensional feature descriptions. It can not only depict the relationship within the same order, such as the relationship between subblocks of each pattern, but also the correspondence between different orders, e.g. the relationship between the feature descriptions of the same subblock in different feature dictionaries. Therefore, the third-order tensor description of the image pattern information not only makes the representation of different pattern information in the image more concentrated, but also presents a sparse characteristic with strong resolution, showing the stereoscopicity of the depiction of the image feature information, leading to full utilization of a variety of feature information in the characterization image.

Fig.5 3-D tensor Image representation

4 EXPERIMENT AND ANALYSIS

4.1 Pattern subblock partition

The pattern subblock partition method proposed in this paper is intended to extract an important target area from the image rather than to segment the target accurately and is achieved in accordance with the spanning characteristics of a sub-window generated by the superpixel and saliency area, as shown in Figure 6. Figure 6 (a) shows the result of superpixel segmentation. When superpixel segmentation is performed, the pixels in the same superpixel block must belong to the same texture pattern of the same target. Figure 6 (b) is a multi-scale saliency map for generation of multiple potential target windows.

When the pattern subblock is partitioned, for example, the layer is 3, the image is partitioned into three areas: the two most prominent target windows in the image are extracted as the foreground window, and the other areas in the image are used as the third pattern subblock (when statistics of features is made, subblock intersections are not counted repeatedly, and only included in blocks with relatively salient targets). Figure 7 shows the result of multi-layer pattern subblock partition. As seen from the figure, layer 1 is the original image; layer 2 partitions the image into saliency and non-saliency areas; and layer 3 detects the two most salient target areas in the image, dividing the image into two target areas and one background area; layer 4 partitions the image into three target areas and one background area, and the pattern subblock partition is the basis of the subsequent tensor description.

Fig.6 Superpixel and multi-scale saliency map

(a) Layer1

(b) Layer2

Fig.7 Multi-Layer pattern subblock detection

4.2 Testing of database

In order to evaluate the effectiveness of the proposed image description method, Pascal VOC2007^[19] image database is selected in this paper for experiments. The PASCAL VOC 2007 image library contains 20 types of images, 9963 pictures. The pictures in this image library are characterized by both indoor and outdoor images, close-ups and outdoor scenery, and various shooting angles. In addition, the target size changes more significantly and there are multiple interference targets in addition to the target to be identified, as show in Figure 8, which indicates 20 types of images in the image library.

Fig.8 Twenty image categories of Pascal VOC 2007

4.3 Experimental results and analysis

In constructing the first type of feature dictionary, this paper adopts the Difference of Gaussian (DOG) to implement feature detection and uses 128 -dimension SIFT feature descriptors to represent feature points. The feature quantization method trains the feature dictionary in the manner of sparse coding constraints proposed in Literature^[7]. As for the evaluation index of classification results, the Average Precision (AP) method is used. This method is the index of standard in PASCAL game, which is obtained by calculating the area under the Precision /Recall curve, the higher the value, the better the effect. The Precision is the number of identified true positive samples partitioned by the number of all samples identified as positive samples. The Recall is the number of identified true positive samples partitioned by the number of all true positive samples.

Figure 9 and Figure 10 show the experimental comparison of the method (a subblock partition with multi-layer pattern, SPMP) with other related methods, where the category number corresponds to that in Figure 8. As shown from the experimental results, in types 4, 6, 7, 8, 10, 15, 19 and 20, this method achieves better classification results. The method in this paper is suitable for dealing with the case where there are multiple salient targets in the image, rather than for the case where the target occupied area is relatively small. Figure 11 is a comparison of the classification of subblock partitions in different layer patterns. SPMP-1 is a 1- layer partition, that is, the original map; SPMP-2 is a 2- layer pattern subblock partition, that is, one target and one background partition; SPMP-3 is two- target one-background partition. According to the experimental results, the three-layer partition effect is the best, followed by the two-layer partition, and one-layer partition is poor because there is no partition pattern information, which is equivalent to the original dictionary model method.

5 CONCLUSION

In terms of partitioning of multi-layer pattern subblocks, the pattern subblock partition method based on salient area distribution and super-pixel distribution is proposed. By analysis of the potential target window extracted from the multi-scale saliency map and the spanning characteristics of the super-pixel, the multiple windows close to the target are obtained, and the target areas at different layers in the image are extracted. On the basis of multi-dictionary model and pattern subblock partition, this paper depicts the image pattern subblocks with multiple dictionary models, and describes the image pattern information with third-order tensor space. In combination with the method based on canonical correlation analysis, a feature design classifier is extracted, and used for classification of complex natural scene images. As a new image description method, it can be applied to the content-based image classification, retrieval and other fields.

Fig.9 1-10 category classification results of Pascal VOC 2007

Fig.10 11-20 category classification results of Pascal VOC 2007

Fig.11 Classification results of different pattern partition layers

ACKNOWLEDGMENTS

This work was supported by Image Recognition Based High Resolution Remote Sensing Data Analysis Research, KJ-2017002.

REFERENCES

1. Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2006: 2169–2176, New York City, NY
2. Cao Y, Wang C G, Li Z W, et al. Spatial-bag-of-features. In: Computer Vision and Pattern Recognition (CVPR), 2010: 3352-3359
3. Grauman K, Darrell T. Pyramid match kernels: Discriminative classification with sets of image features. In: IEEE International Conference on Computer Vision (ICCV), 2005, 2: 1458-1465
4. Wallraven C, Caputo B, Graf A. Recognition with local features: the kernel recipe. In: IEEE International Conference on Computer Vision (ICCV), 2003, 1: 257–264
5. Willamowski J, Arregui D, Csurka G, et al. Categorizing nine visual classes using local appearance descriptors. In ICPR Workshop on Learning for Adaptable Visual Systems, 2004
6. Zhang J, Marszalek M, Lazebnik S, et al. Local features and kernels for classification of texture and object categories: An in-depth study. Technical Report RR-5737, INRIA Rhˆone-Alpes, 2005
7. Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2009: 1794-1801
8. Torresani L, Szummer M, Fitzgibbon A. Learning query-dependent prefilters for scalable image retrieval. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2009: 2615 - 2622
9. Yuan J S, Wu Y, Yang M. Discovery of collocation patterns: from visual words to visual phrases. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2007: 1-8
10. Zitnick C L, Sun J, Szeliski R, et al. Object instance recognition using triplets of feature symbols. Tech. Report, Microsoft Research, 2007
11. Wu Z, Ke Q F, Isard M, et al. Bundling features for large scale partial-duplicate web image search. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2009: 25-32
12. Hou X D, Zhang L Q. Saliency detection: A spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007: 1-8
13. Alexe B, Deselaers T, Ferrari V. What is an object? In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010: 73-80
14. Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation. International Journal of Computer Vision (IJCV), 2004, 59(2): 167-181
15. Aharon M, Elad M, Bruckstein A. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing, 2006, 54(11): 4311-4322
16. Kolda T G, Bader B W. Tensor Decompositions and Applications. Siam Review, 2009, 51(3): 455-500
17. Kim T K, Wong S F, Cipolla R. Tensor canonical correlation analysis for action classification. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2007: 1-8
18. Kim T K, Cipolla R. Canonical correlation analysis of video volume tensors for action categorization and detection. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2009, 31(8): 1415-1428
19. Everingham M, Gool L, Williams C, et al. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results