Satellite Image Retrieval with Pattern Spectra Descriptors

Petra Bosilj
Univ. Bretagne Sud – IRISA
Vannes, 56000, France
petra.bosilj@irisa.fr

Erchan Aptoula
Okan University
Istanbul, 34959, Turkey
erchan.aptoula@okan.edu.tr

Sébastien Lefèvre
Univ. Bretagne Sud – IRISA
Vannes, 56000, France
sebastien.lefevre@irisa.fr

Ewa Kijak
Univ. Rennes 1 – IRISA
Rennes, 35000, France
ewa.kijak@irisa.fr

***Abstract*—The increasing volume of Earth Observation data calls for appropriate solutions in satellite image retrieval. We address this problem by considering morphological descriptors called pattern spectra. Such descriptors are histogram-like structures that contain the information on the distribution of predefined properties (attributes) of image components. They can be computed both at the local and global scale, and are computationally attractive. We demonstrate how they can be embedded in an image retrieval framework and report their promising performances when dealing with a standard satellite image dataset.**

I. Introduction

Proliferation and increasing performances (spatial precision, revisiting frequency) of Earth Observation satellites lead to massive amount of satellite image data. Mining such data is of primary importance and required to solve various problems. In this paper, we focus on image retrieval that allows retrieving geographic objects having a similar appearance when visually observed from EO satellites.

Image retrieval is typically achieved by means of computing descriptors, either globally for the whole image or on selected or predetermined parts of the image. Those descriptors (first aggregated in case when multiple descriptors per image are used resulting from using patches of the image) are further used in dedicated indexing/retrieval schemes [1], [2], [3], [4].

We propose to perform satellite image retrieval with morphological descriptors called pattern spectra. Morphological features have been applied to image retrieval, either in the general case [5] or in remote sensing [6]. We focus here on pattern spectra, used as global descriptors [7], [8] and more recently as local ones as well [9], [10]. They are histogram-like structures containing the information on the distribution of predefined properties (attributes) of image components. Their computation is very efficient when satellite images are represented through min-trees and max-trees, as already done with attribute profiles [11]. Indeed, while attribute profiles are pixelwise features, pattern spectra accumulate the distribution of these features into regions or larger extent.

The paper is organized as follows. Section II recalls the definition of pattern spectra, both as global and local descriptors, and also reviews the different attributes involved in our study. We describe the evaluation protocol used in our experiments in Sec. III, while results are discussed in Sec. IV before we conclude and give future research directions.

II. Pattern Spectra

Pattern spectra are histogram-like structures originating from mathematical morphology, commonly used for image analysis and classification [12], and contain the information on the distribution of sizes and shapes of image components. They can be efficiently computed using a technique known as (size or shape) granulometry [13], [14] on a max-tree and min-tree hierarchy [15], [16] (recent fast algorithms were also made available for 12-bit and arbitrary precision data [17]).

A. Global Pattern Spectra

The min and max-trees contain all the connected components of all the threshold sets of the image, and pruning those trees to remove all components below a certain size t realizes the operation known as area opening. A size granulometry [13] is a series of such openings with increasing size, $t_{i+1} > t_i$, which removes more components from the image in every step and can be seen as a set of sieves of increasing grades. Focusing instead on the amount of detail removed between consecutive area openings, we obtain a *size pattern spectrum* [12], a histogram over all the different size classes, where every size class is described by its Lebesgue measure. To measure the size of a component, any increasing attribute (giving increasing values when calculated on a nested sequence of regions) can be used. The extension of the technique to *shape granulometries* [14] allows the usage of non-increasing attributes, measuring the shape of the region, in construction of similar histograms across shape classes.

Combining shape and size pattern spectra into a unique 2D histogram, which shows the amount of image detail across dedicated size-shape bins, produces a *shape-size pattern spectra* [8]. When calculated for an image, they can be used as translation, scale and rotation invariant image descriptors and have been successfully applied to image classification [8] and retrieval [7]. As we want the descriptor to encode the difference between images with high and low content, we do not normalize the produced descriptors, but simply store the amount of image detail (Lebesgue measure of contributing components) as a ratio of total image size. The bin distribution over different shape and size classes is typically logarithmic over the chosen range of attribute values. When used in this manner, we will refer to the produced global image descriptors as *Global Pattern Spectra (GPS)*.

B. *Local Pattern Spectra*

Recently, a local extension of the pattern spectra has been proposed [10], [9], designed to characterize patches rather than the whole image, i. e. either specifically selected regions of interest or patches of predetermined shape and size. They were initially introduced on maximally stable extremal regions (MSER) [18] for reasons of computational efficiency, as MSERs can be extracted from max-tree and min-tree hierarchies in a straightforward manner. Initial experimental results show that these descriptors achieve competitive performances against SIFT [19] in the context of image retrieval [9], making the extension to *Local Pattern Spectra (LPS)* a promising option for satellite image retrieval.

In the straightforward extension from GPS to LPS, due to the logarithmic binning used, the descriptors lose their scale invariance property [10] if calculated on patches of different sizes. In the proposed work, we do not rely on any keypoint detection step such as MSER and instead compute the LPS on local rectangular patches defined on a regular grid over the image. This choice is motivated by the small size and limited content in the images considered in the experimental setup, that greatly alleviate the need for extracting regions of interests. If all the preselected local patches are of the same (or very similar) size, the scale invariance property holds. In an approach using multiple local patch sizes to achieve image description at multiple scales, the technique introduced in [9] is used to retain scale invariance.

C. *Attributes*

Both GPS and LPS measure the amount and distribution of image details for given classes of attributes. In this paper, the increasing attributes are considered to cope with size information, while shape-sensitive, non-increasing, attributes are used to capture shape information.

We rely on the standard approach of measuring the number of pixels of a region to describe its size. We tackle shape information through two different attributes. The first one is called Corrected Non-Compactness (CNC), and measured as the first moment invariant of Hu (i.e., a ratio between the moment of inertia and the squared area of a region). A correction factor is required when transitioning from the original formula in the continuous space to the discrete image space [20]. The second shape-related attribute is Shannon entropy (SE) (cf. [7]). Experimental results using combinations of these attributes embedded into global and local pattern spectra are presented henceforth.

III. Experimental setup

A. *Dataset and Evaluation Metrics*

We have conducted our experiments on two publicly available datasets: ImageNet Large Scale Visual Recognition Challenge 2010 (ILSVRC2010) Validation dataset [21] when a training set had to be used, while the validation was done on a satellite image retrieval dataset, namely the UC Merced Land Use Dataset [1] [22]. The Merced Dataset contains 2100 color RGB images organized into 21 classes (100 images per class), examples of which are shown in Fig. 1. All images are RGB color samples of size size equal to 256×256 pixels. We compute our descriptors firstly on the grayscale versions of the images, with the conversion $Gray = 0.299 \times R + 0.587 \times G + 0.114 \times B$. For our most successful global approach we have additionally tried tripling the size of the descriptor by concatenating the descriptors obtained when applying the same approaches separately on R, G and B channels.

The evaluation metrics chosen is ANMRR, as it is the most commonly metric used on this dataset and allows for straightforward comparison with other published results [6], [23], [22], [3]. ANMRR stands for *average normalized modified retrieval rank* and is commonly used to measure effectiveness of MPEG-7 retrieval [24]. Given a query q or all the queries of a same class, a number $\mathrm{K}(q)$ is defined, which denotes that only the first $K(q)$ returned images are considered as feasible in terms of retrieval evaluation and is often set as twice the size of the ground truth set $\mathrm{NG}(q)$. Assume that the k^{th} ground truth image is retrieved at $\mathrm{Rank}(k)$, a penalty function $\mathrm{Rank}^*(k)$ is defined for each retrieved item:

$$\mathrm{Rank}^*(k) = \begin{cases} \mathrm{Rank}(k), & \text{if } \mathrm{Rank}(k) \leq \mathrm{K}(q) \\ 1.25\,\mathrm{K}(q), & \text{if } \mathrm{Rank}(k) > \mathrm{K}(q) \end{cases} \tag{1}$$

From all the penalties $\mathrm{Rank}^*(k)$ for each query q, the average rank (AVR) for that q is defined:

$$\mathrm{AVR}(q) = \frac{1}{\mathrm{NG}(q)} \sum_{k=1}^{\mathrm{NG}(q)} \mathrm{Rank}^*(k) \tag{2}$$

After the intermediate step, ANMRR is directly defined as:

$$\mathrm{ANMRR} = \frac{1}{\mathrm{NQ}} \sum_{q=1}^{NQ} \frac{\mathrm{AVR}(q) - 0.5(1+\mathrm{NG}(q))}{1.25\,\mathrm{K}(q) - 0.5(1+\mathrm{NG}(q))} \tag{3}$$

where NQ is the number of queries. Thus ANMRR obtains values in range of 0 for best results, and 1 for worst results.

B. *Settings of Pattern Spectra Approaches*

Global Pattern Spectra. In the base approach, we calculate the GPS descriptors directly on the complete image samples. We chose to use 10 bins for the size (area) attribute, and 6 for the shape attribute. As we calculate a GPS from both a min-tree a max-tree, this produces global descriptors of size 120. We report the results using the CNC and SE as shape attributes, and observe a further improvement when combining them into a single descriptor of length 240. We also report an improvement over the base performance when applying the GPS to each of RGB channels separately.

Local Pattern Spectra. Further, we attempt a single-scale local approach. We densely sample the image, calculating the LPS on regular rectangular patches over a grid on the image. We attempt different patch dimensions and offsets between patch centers (resulting in different overlap between patches).

[1]available at: http://vision.ucmerced.edu/datasets/landuse.html

Fig. 1. Merced dataset

Our final choice of patch size is 80×80 with 16 pixels distance between patch center (cf. Fig. 2 for the results of the tuning experiments). The tuning experiments regarding patch and overlap size were done with the histograms of size 6×4, for the efficiency of the calculation. A larger histogram of size 8×6 was chosen for the final evaluation, giving an improvement over the smaller size and producing descriptors of length 96. We only report the results using CNC, as combination with other shape attributes as well as processing the RGB channels separately does not result in improvement over the base local approach and greatly increases the complexity of the approach.

Finally, we attempt a multi-scale approach based on a pyramid of patches. Here, we start with patch size 32×32 and the size of patch increases for each level of the pyramid ($2\times$ along each dimension), so the scale-invariance of LPS becomes relevant. We report the results of this approach both with scale-variant version of LPS [10] as well as with choosing a common reference scale to achieve scale invariance [9]. The distance between patch centers is again set to 16. The length of the descriptors is 96, the same as for the base local approach, but with three times more descriptors.

For both single-scale and pyramid approach using LPS, we use VLAD indexing to produce global image descriptors [25], using 8 cluster centers. We use a random subset of the ImageNet 2010 Validation set for building the visual vocabulary for VLAD, which is consequently formed both independent of the evaluation dataset as well as of its geographical context. The same approach to extracting descriptors as well as the same patch size (stopped at 256×256 for pyramid approaches) was used as for the evaluation dataset.

TABLE I
THE RETRIEVAL PERFORMANCES OF DIFFERENT LOCAL AND GLOBAL APPROACHES ON MERCED DATASET

approach	**ANMRR**
SIFT (on keypoints, [22])	0.601
dense SIFT ([3])	***0.4604*** (using VLAD)
global texture descriptors ([6])	0.575
local texture descriptors ([23])	0.585 (Bag of Words)
GPS - Area + CNC	0.579
GPS - Area + SE	0.670
GPS - both shape attributes (CNC + SE)	**0.557**
GPS - RGB decomposition (Area + CNC)	0.562
dense LPS (Area + CNC)	0.538
pyramid LPS (scale variant)	0.534
pyramid LPS (common scale 64×64)	**0.529**

IV. Results and discussion

With our base GPS approach, we outperform both previously proposed global and local morphological approaches based on texture [6], [23], as well as the seminal SIFT approach on this dataset [22]. Combining two different shape attributes is the preferred technique for improving this base results while still working with global descriptors. It results in a bigger performance improvement than decomposing the image into channels and is suited for possible use on images with more than three channels.

Further improvements are achieved by using a dense local approach, and that only by using 144 descriptors per image. However, the biggest improvement comes from using the scale-invariant LPS [9], reporting the best results with pattern spectra of 52.9% ANMRR. While it still remains to outperform the dense SIFT descriptors [3], which produce state of the art results on the dataset, we show here an improvement over previous morphology-based approaches as well as the seminal SIFT approach to retrieval on this dataset.

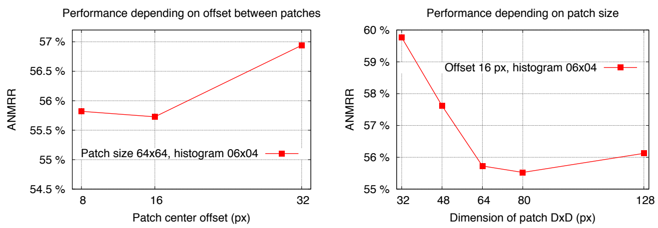

Fig. 2. Experiments for chosing the patch size and offset between patches. The tuning was done using smaller descriptors, with histograms of dimensions 6×4 and histograms size 48, while the chosen patch size (80×80) and offset between patches (16 pixels, resulting in 64 pixels of overlap for the chosen patch size) are later used with a larger histogram sizes.

V. Conclusion

Satellite image retrieval is required to exploit the vast amount of Earth Observation data available today. It requires both the definition of appropriate image descriptors and the design of an indexing/retrieval scheme. In this paper, we assess the relevance of morphological image descriptors, namely pattern spectra, in solving the satellite image retrieval issue. We rely on both global and local pattern spectra, the latter having been introduced in the literature only very recently. Experiments conducted on the standard UC Merced Land Use Dataset validate the usage of pattern spectra as the best performing morphological descriptors, both among global and local (patch-based) approaches.

In order to reach the performance of dense SIFT descriptors, the multi-scale pyramidal approach shows the most promise. In addition to reevaluating the construction of the scale pyramid and confirming the best preforming reference scale, introducing additional attributes based on shape (similarly to what was done in the global approach) could be considered, as well as using attributes based on region content or texture. Combining both LPS and SIFT descriptors in the approach using dense patches or multi-scale pyramids is also a viable option.

Acknowledgment

This work was supported in part by the Scientific and Technological Research Council of Turkey (TUBITAK) under Career Grant 112E210, and by the French National Research Agency (ANR) under Project Grant ANR-13-JS02-0005-01 (Asterix project).

The authors would also like to thank Giorgos Tolias for providing the framework and tools for executing VLAD indexing used in experimental evaluation of this paper.

References

[1] M. Wang, Q. M. Wan, L. B. Gu, and T. Y. Song, "Remote-sensing image retrieval by combining image visual and semantic features," *International journal of remote sensing*, vol. 34, no. 12, pp. 4200–4223, 2013.

[2] D. Espinoza-Molina and M. Datcu, "Earth-observation image retrieval based on content, semantics, and metadata," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 11, pp. 5145–5159, 2013.

[3] S. Ozkan, T. Ates, E. Tola, M. Soysal, and E. Esen, "Performance analysis of state-of-the-art representation methods for geographical image retrieval and categorization," *Geoscience and Remote Sensing Letters, IEEE*, vol. 11, no. 11, pp. 1996–2000, 2014.

[4] R. Bahmanyar, S. Cui, and M. Datcu, "A comparative study of bag-of-words and bag-of-topics models of eo image patches," *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, no. 6, pp. 1357–1361, 2015.

[5] E. Aptoula and S. Lefèvre, "Morphological description of color images for content-based retrieval," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2505–2517, November 2009.

[6] E. Aptoula, "Remote sensing image retrieval with global morphological texture descriptors," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 3023–3034, 2014.

[7] F. Tushabe and M. H. F. Wilkinson, "Content-based image retrieval using combined 2D attribute pattern spectra," in *Advances in Multilingual and Multimodal Information Retrieval*. Springer, 2008, pp. 554–561.

[8] E. R. Urbach, J. B. T. M. Roerdink, and M. H. F. Wilkinson, "Connected shape-size pattern spectra for rotation and scale-invariant classification of gray-scale images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 272–285, 2007.

[9] P. Bosilj, M. H. F. Wilkinson, E. Kijak, and S. Lefèvre, "Local 2D pattern spectra as connected region descriptors," in *Mathematical Morphology and Its Applications to Signal and Image Processing*, ser. Lecture Notes in Computer Science, vol. 9082. Springer, 2015, pp. 182–193.

[10] P. Bosilj, E. Kijak, M. H. F. Wilkinson, and S. Lefèvre, "Short Local Descriptors from 2D connected Pattern Spectra," in *IEEE International Conference on Image Processing (ICIP)*, 2015.

[11] M. Dalla Mura, J. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3747–3762, 2010.

[12] P. Maragos, "Pattern spectrum and multiscale shape representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 7, pp. 701–716, 1989.

[13] E. J. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Computer Vision and Image Understanding*, vol. 64, no. 3, pp. 377–389, 1996.

[14] E. R. Urbach and M. H. F. Wilkinson, "Shape-only granulometries and grey-scale shape filters," in *Proc. Int. Symp. Math. Morphology (ISMM)*, vol. 2002, 2002, pp. 305–314.

[15] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *Image Processing, IEEE Transactions on*, vol. 7, no. 4, pp. 555–570, 1998.

[16] R. Jones, "Component trees for image filtering and segmentation," in *IEEE Workshop on Nonlinear Signal and Image Processing, E. Coyle, Ed., Mackinac Island*, 1997.

[17] M. H. F. Wilkinson, "A fast component-tree algorithm for high dynamic-range images and second generation connectivity," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 1021–1024.

[18] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, no. 10, pp. 761–767, 2004.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[20] M. A. Westenberg, J. B. Roerdink, and Wilkinson, "Volumetric attribute filtering and interactive visualization using the max-tree representation," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2943–2952, 2007.

[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, pp. 1–42, April 2015.

[22] Y. Yang and S. Newsam, "Geographic image retrieval using local invariant features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 818–832, 2013.

[23] E. Aptoula, "Bag of morphological words for content-based geographical retrieval," in *International Workshop on Content-Based Multimedia Indexing (CBMI)*, 2014.

[24] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, no. 6, pp. 703–715, 2001.

[25] H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 3304–3311.