Improving High Resolution Satellite Images Retrieval Using Color Component Features

Houria Sebai and Assia Kourgli(✉)

USTHB, Faculté D'Electronique Et D'Informatique, LTIR,
B.P.32, 16111 El-alia, Bab-ezzouar, Algeria
assiakourgli@gmail.com

Abstract. This paper highlights multi-scale color component features that improve high resolution satellite images retrieval. Color component correlation across image lines and columns is used to define a revised color space. It is designed to take simultaneously both color and neighborhood information. From this new space, color descriptors namely RIULBP (Rotation Invariant Uniform Local Binary Pattern), LV (Local Variance) and a modified version of LV (smoothed LV) are derived through Dual Tree complex wavelet transform (DT-CWT) or scale-invariant features transform (SIFT) representations. The features obtained offer an efficient way to represent both color and texture/structure information. We report an evaluation of the proposed descriptors according to different similarity distances in our CBIR (Content-based image retrieval) schemes. We, also, perform comparison with recent approaches. Experimental results show that color LV descriptor combined to SIFT representation outperforms the other approaches.

Keywords: CBIR · DT-CWT · SIFT · LBP · Opponent color · SLV

1 Introduction

The interest in fast and accurate information retrieval systems over collections of remote sensing images is increasing as the volume of the available data grows exponentially creating new challenges in different fields including processing, archiving and retrieval. Given its importance, this problem received a lot of attention in the literature. Many approaches have been proposed to retrieve low and mid-satellite images using their content such as region level semantic features mining [1], Knowledge-driven information mining (KIM) [2], texture model [3], entropy-balanced bitmap (EBB) tree [4]. High resolution satellite retrieval schemes use different features according to color (spectral) features [5], texture features [6] [7] [9], structure features [5] [9]. However, the performance of most techniques is limited by the semantic gap between low level features and high level concepts. To reduce this gap, we explore the idea of color images encoding using relevant features according to texture and structure information to describe the images contents. Usually a color component descriptor is obtained through

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part II, LNCS 9280, pp. 264–275, 2015.
DOI: 10.1007/978-3-319-23234-8_25

the concatenation of the descriptors derived from component images in an opponent color space [10] [9]. Our approach is different, we use color component correlation across image lines and columns to define multiscale descriptors using a proposed measure named SLV (Smoothed Local Variance). These descriptors permit to encode simultaneously local texture/shape information according to three different color plans. To assess the efficiency of these descriptors, a comparison is made according to other color spaces usually used when dealing with satellite images retrieval. To take into account high resolution satellite images nature, two kind of multiscale approaches are considered. The former is based on DT-CWT representation that possesses a good directional selectivity for feature representation while the latter employs SIFT features that are well known for their efficiency for images retrieval. According to different similarity measures, the retrieval performances of the feature vectors obtained are compared to those reached by other studies showing that the descriptors LV and SLV computed on the color space derived permit to get more efficient retrieval schemes. Consequently, the paper is organized as follows : Section 2 introduces the different color component descriptors tested. Our CBIR scheme based on color component features and used through DT-CWT is presented also in this section. Section 3 introduces our bag of visual words model based on the combination of SIFT and our opponent color features. Section 4 presents the experimental results and discussions. Conclusions and future orientations constitute section 5.

2 Multiscale Opponent Color Descriptors

New descriptors obtained by combining color and texture (or shape) according to DT-CWT (or SIFT) representation are introduced. Both texture/structure and shape features such as LBP [11], HOG [12], Grey Level Co-occurrence Matrices, etc. are usually computed on each color channel separately. The features resulting are then concatenated to constitute one single vector feature. We extended the idea of Banerji *et al.* [13] to derive color component features that consider simultaneously color and texture information.

2.1 Opponent Cross-Correlation

Several opponent color spaces have been proposed and used for color representation based on the perceptual concepts. For high resolution satellite images retrieval, an appropriate space must be based on an intuitive combination of image characteristics (color and context). We referred to a recent work of Banerji et al [13] who suggested to define three new color plans by applying perpendicular encoding using the RGB color space. The first color image is the RGB_1, while the second color image RGB_2 encodes the rows across the red, green, and blue. The last image RGB_3 is obtained through columns encoding across the red, green and blue channels (more details can be found in [13]). Unlike classical color spaces that produce the same number of output images, Banerji's encoding scheme multiplies by three its outputs. This fact is prohibitive as it will induce

feature vectors that would be three times larger and thus a retrieval process taking three times more. To overcome this problem, a data reduction has been considered. It is based on an adaptive grayscale transformation that is used on each color plan to produce three new images. It is a kind of adaptive luminance transformation that lets the weights in the averaging process be dependent on the actual image to be convert,.i.e., be adaptive.

$$AL_i = R_i * \frac{\sqrt{\sum {R_i}^2}}{\sum {R_i}^2 + \sum {G_i}^2 + \sum {B_i}^2} + G_i * \frac{\sqrt{\sum {G_i}^2}}{\sqrt{\sum {R_i}^2 + \sum {G_i}^2 + \sum {B_i}^2}} + B_i * \frac{\sqrt{\sum {B_i}^2}}{\sqrt{\sum {R_i}^2 + \sum {G_i}^2 + \sum {B_i}^2}} \tag{1}$$

where i=1,2,3.

The new space obtained through the combination of cross-correlation and adaptive luminance transformation is named cross adaptive luminance (CAL). The performances of colors descriptors derived from CAL space are compared to those [13] [10] achieved using other popular color spaces used for images retrieval.

2.2 Color Features

Feature extraction is the basis of content-based image retrieval. Extraction of color feature is carried out by computing a visual feature on color images. Because of our image nature, some invariant rotation and translation features are required to characterize spatial color distribution and thus integrate structure information. To this aim, a modified version of local variance is suggested and compared to LBP variants. However, the color feature vector dimensions of typical color feature descriptors are quite large resulting in high computational cost in distance calculation for similarity retrieval. As a solution, we suggest to compute some uncorrelated statistical moments namely mean, kustosis, L1 and the quadratic mean instead of histograms to get a compact representation. The aforementioned color features are described below.

LBP The original LBP operator is defined in a rectangular $3 * 3$ neighborhood. It operates with eight neighboring pixels using the center as a threshold. The final LBP code is then produced by multiplying the threshold values by weights given by powers of two and adding the results. The authors [11] gave an improved uniform gray-scale and rotation invariance operator defined as:

$$LBP^{riu2}_{P,R} = \begin{cases} \sum_{i=0}^{P-1} s(g_i - g_c) & U \leq 2 \\ P + 1 & otherwise \end{cases} \tag{2}$$

the $LBP^{riu2}_{P,R}$ operator outputs are accumulated into a histogram of P + 2 bins. By changing P and R, the LBP code of diverse radius and neighborhood is computed to obtain texture features. LBP images obtained are decomposed using DT-CWT and statistical moments are computed using the sub-images resulting.

LV and SLV Measures of local variance have been widely used in image processing for texture and spatial image structure measures. As, this parameter is invariant to illumination changes, we suggest to compute the average value of local variances (LV) estimated around each pixel according to our color space. The LV computed in this manner will constitute a rotation invariant feature that permits to identify localized color distributions of each color plan. Nevertheless, it is sensitive to outlier as a single pixel with a different color will induce a significant increase of LV value. To surmount this shortcoming, we propose to compute the LV on a smoothed neighborhood. Smoothing is obtained through the use of edge preserving smoothing filters (EPSF) [15] [14]. Instead of processing filtering on each neighborhood, the whole image is filtered. Then, the LV is computed to produce a smoothed local variance (SLV), by this way, local image outliers are discarded.

2.3 Feature Extraction

In our CBIR system dedicated to HRS images, a muli-scale representation of color features descriptors is used to better reflect the objects of different sizes and shapes present in HRS images. Multiresolution DT-CWT technique [16] is adopted since it allows analysis that is localized in both space and frequency. It calculates the complex transform of a signal using two separate DWT decompositions (two trees). The prime motivation for introducing the dual-tree complex wavelet was shift invariance. DT-CWT [16] is able to discriminate between positive and negative orientation of the diagonal sub-bands. The horizontal and vertical sub-bands are divided giving six distinct sub-bands in each scale at orientations $\pm 15°$, $\pm 45°$, $\pm 75°$. DWT have three sub-bands in $0°$, $45°$, $90°$ only but DT-CWT having six sub-bands in $\pm 15°$, $\pm 45°$ and $\pm 75°$, for this reason DT-CWT improves the directional selectivity.

Statistical moments on DT-CWT decomposed subimages obtained either from LBP or LV are computed. For both features, different moments are tested to retain only the less correlated: For the LL block we kept the mean, kurtosis and Root Mean Square (RMS). While for the details, the standard L1 also called mean deviation MDni is calculated.

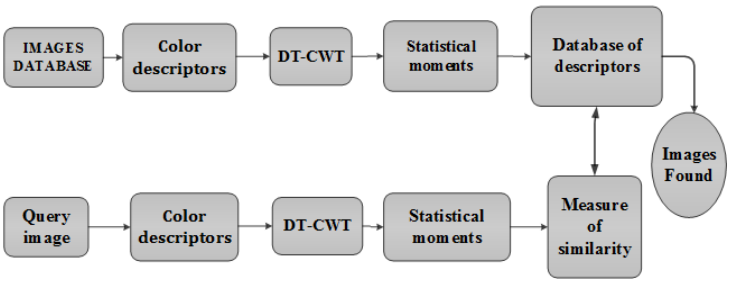

Fig. 1. CBIR schema

3 Bag of Visual Words Model

Bag of visual words is a vector of occurrence counts of a vocabulary of local image features. Representing an image using bag of visual words includes usually the following steps: Extract features, learn visual vocabulary, quantize features using visual vocabulary, represent images by frequencies of visual words. Feature representation methods deal with how to represent images as a collection of local properties calculated from a set of small sub-images named patches. These vectors are called features descriptors. One of the most famous descriptors is the Scale Invariant Feature Transform (SIFT) [17]. The final step for the BOW is to generate the code words (Word dictionary). The idea is to cluster the features descriptors of all patches based on given cluster number using K-means clustering. Codewords are then defined as the center of the learned clusters. The number of the clusters is the codebook size. Thus, each image can be represented by frequencies of visual words.

3.1 Features Extraction

SIFT. Scale invariant feature transform (SIFT) developed by David Lowe [17] to detect and describe local features in images is invariant to uniform scaling, orientation and partially invariant to affine distortion and illumination changes. To calculate SIFT feature vectors internal representation of the original image is created to ensure scale invariance. This is done by generating scale Space. Then, locations and scales of key points that are initially selected from local extrema in Difference of Gaussian (DoG) are determined. Each point needs to be compared with 8 neighboring pixels in the same scale and 18 neighboring pixels around the corresponding position of adjacent scales. The last step is to determine the orientation of key point using the direction of gradient of its neighboring pixels to assure the rotation invariance. Thus, each key point has three parameters: location, scale and orientation.

The SIFT descriptor is extracted from image patches around the interest point. To assure the rotation invariance, a key point rotation is made. Lowe suggests to describe each key point using 4x4 seed points to increase the stability of matching. The feature descriptor consists of histograms of gradient directions computed over a 4x4 spatial grid. The gradient directions are quantized into 8 bins so the final feature vector has dimension of 128 (4x4x8).

SIFT-Opponent Color Feature. To increase the performance of the retrieval system, SIFT and opponent color features LV (Local variance) are combined in our scheme. To obtain the SIFT-LV or SIFT-LBP descriptors, 128-dimensional SIFT descriptor is calculated for each key point while for local variance(LV) or LBP, region of 16x16 around the key point is considered. The region is subdivided into 9 overlapping blocks and 9 bins histograms of each block are calculated. Thus, 81-dimensional LV descriptor is obtained for each key point. The next step is to build a code books for SIFT descriptors and another one for LV

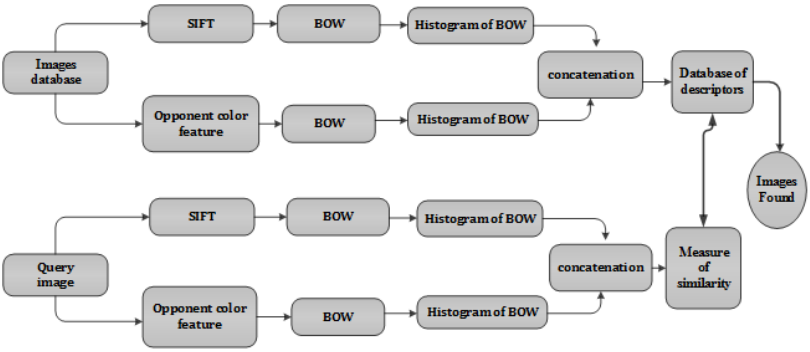

Fig. 2. CBIR schema using bag of visual words

descriptors. To consider both color and context information, LV is applied to CAL (cross adaptive luminance) images.

To evaluate the CBIR system proposed (Fig.2), other descriptors are tested. Following the CBIR scheme based on DT-CWT, different distance measures are also tested and compared.

4 Results and Discussions

The dataset [6] consists of images belonging to 21 categories: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts. Each category is composed of 100 samples (see Fig.3).
In a query by example scheme, we are interested in retrieving several similar images and this requires comparing two descriptors to obtain a measure of similarity (or dissimilarity) between the two image patterns. Some common similarity measures (Table 1) are tested and their performances are compared.

Table 1. Common Similarity measures

Manhattan distance(L1)	Euclidean distance(L2)	Khi square	Canberra distance
$\sum_{i=1}^{n} \lvert x_i - y_i \rvert$	$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$	$\sqrt{\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{(x_i + y_i)}}$	$\sum_{i=1}^{n} \frac{\lvert x_i - y_i \rvert}{\lvert x_i + y_i \rvert}$

CBIR performance is measured by precision and recall.
Precision Pas well as average precision and $\overline{P}$ are given as:

$$P = \frac{\text{Number of relevant images retrieval}}{\text{Total number of images retrieval}}, \overline{P} = \frac{1}{N_q} \sum_{k=1}^{N_q} P(k) \tag{3}$$

Where Nq represents the number of queries.

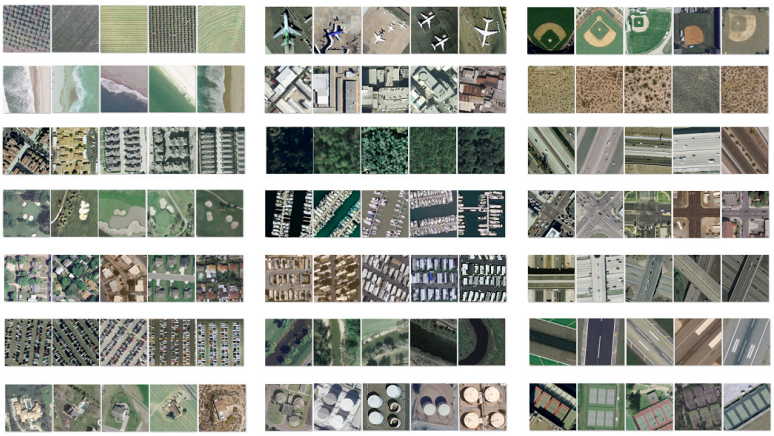

Fig. 3. Image patches of the 21 land-use/land-cover classes

Similarly recall R and average recall $\overline{R}$ are given as:

$$R = \frac{\text{Number of relevant images retrieval}}{\text{Total number of relevant images}}, \overline{R} = \frac{1}{N_q} \sum_{k=1}^{N_q} R(k) \tag{4}$$

To evaluate the performance of the proposed CBIR scheme, many tests were conducted. They aim, among others, to determine the best color space and the most appropriate measure of similarity for satellite images indexing using color descriptors as well as the best approach. We began by comparing the color descriptors derived from the DT-CWT sub-images considering the different color spaces previously presented. According to the paper of Shao et al [9], eight LULC classes are used in our experiments. These are agricultural, airplane, beach, buildings, chaparral, dense residential, forest, and harbor. Table 2 illustrates the overall recognition rate obtained for different precision values using the similarity measure of Canberra. The choice of the latter is motivated by the fact that Canberra distance is biased for measures around the origin and very sensitive for values close to zero. Thus, it is well suited for data scattered around an origin as it is the case for details images obtained from DT-CWT decomposition. The overall average precision obtained on whole images shows that the reduction of space that we offer through the calculation of the adaptive luminance allows us to have the benefit of the same performance (65.63 % instead of 65.73 %) using three times less images.

To know whether the distance of Canberra is the most suitable for our descriptors, we considered the use of other similarity measures while considering different color spaces. Average precision values are summarized in Table 3. In all cases, the measurement of Canberra outperforms other measures.

In order to increase the performance of the CBIR system, we combined our colors features derived from CAL space with colors descriptors resulting from RIULBP(uniform invariant LBP), LV (Local variance), SLV (Smoothed Local

Table 2. Precision values using DT-CWT applied to different spaces

	P% (N=10)	P% (N=20)	P% (N=30)	P% (N=40)	P% (N=50)	P% (N=60)	P% (N=70)	P% (N=80)	P% (N=90)	P% (N=100)	Average
RGB	85.88	77.35	71.73	67.31	63.73	60.48	57.60	54.90	52.30	49.81	64.11
opp [10]	84.97	73.85	65.53	59.59	54.49	50.39	47.13	44.27	42.03	40.15	56.24
opp[13]	87	78.93	73.02	68.63	65.19	62.21	59.53	56.9	54.24	51.69	65.73
CAL (proposed)	86.08	78.1	72.34	68.37	65.23	62.41	59.73	57.19	54.66	52.19	65.63

Table 3. Average precision using different distance measures

	Euclidean	Manhattan	Khi-square	Canberra
RGB	55.78	62.76	63.95	**64.11**
opp[13]	53.99	61.13	64.31	**65.73**
opp [10]	49.19	54.32	55.99	56.24
CAL	52.60	62.10	63.04	**65.63**

Variance). To obtain the feature SLV, smoothing is applied using anisotropic filter. Analysis of Table 4 highlights the fact that the parameters tested namely RIULBP and LV bring nothing (a gain of just 1%), while the proposed feature SLV can reach an average precision of almost 70%.

Table 4. Comparisons of precision values using RIULBP, LV and SLV with DT-CWT representation

	P% (N=10)	P% (N=20)	P% (N=30)	P% (N=40)	P% (N=50)	P% (N=60)	P% (N=70)	P% (N=80)	P% (N=90)	P% (N=100)	Average
CAL+RIULBP(CAL)	87.06	79.95	73.04	68.66	65.42	62.22	59.58	57.44	54.60	52.11	66.01
CAL + LV(CAL)	86.22	78.06	72.72	68.98	66.06	63.60	61.13	58.59	56.05	53.48	66.49
CAL+ SLV	87.38	80.3	75.59	71.99	69.45	67.14	65.06	62.90	60.48	57.72	**69.80**

Table 5. Comparisons of precision values using SIFT representation

	P% (N=10)	P% (N=20)	P% (N=30)	P% (N=40)	P% (N=50)	P% (N=60)	P% (N=70)	P% (N=80)	P% (N=90)	P% (N=100)	Average
SIFT	80.21	76.6	74.001	72.2	70.06	68.06	66.13	64.18	61.77	59.16	69.23
SIFT+ CAL	84.23	79.21	76.22	73.34	71.01	69.18	67.38	65.40	63.13	60.34	70.94
SIFT + ULBP(CAL)	83.01	78.75	75.8	73.57	71.88	70.15	68.65	67.14	65.19	62.43	71.65
SIFT+ SLV	85.1	80.93	78.27	76.37	74.52	72.64	70.71	68.67	66.26	63.36	73.68
SIFT + LV(CAL)	85.41	81.5	78.96	77.04	75.08	73.34	71.52	69.63	67.34	64.55	**74.44**

In order to evaluate our two CBIR schemes, average precision charts as well as precision-recall curves for the different features tested in [9] are compared with our descriptor SLV and reported in Fig.4 and 5. The charts indicate that our descriptors (CAL or CAL+SLV) yields to better results on all classes compared to the best feature CGOT proposed in ref [9]. Moreover, except for dense residential class, the combination CAL+ SLV performs better than CAL alone specially for highly textured classes i.e. agricultural and beach classes. Figure 5 depicts all the precision-recall curves corresponding to our descriptors (in solid lines) and those reported (in dashed lines) in ref [9]. It can be observed that our descriptors are more efficient. Indeed, as the correct returned images increase, precision decrease rapidly with the other techniques.

To further validate our CBIR scheme, we made a comparison with the results of Yang et al [6] using Local features based on SIFT representation (Figure 6), Gabor filter, color histogram and simple statistics on the whole database(2100

Table 6. Average precision using different measure using SIFT representation

	Canberra	Euclidean	Manhattan	Khi-square
SIFT	66.45	66.16	68.41	**69.23**
SIFT-CAL	66.45	67.61	66.44	**70.94**
SIFT-CAL(LV)	66.92	69.66	72.02	**74.44**

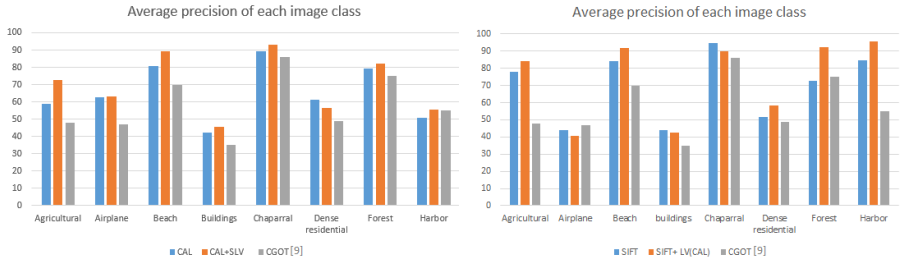

Fig. 4. Comparison of average precision for each image class for both DT-CWT and SIFT representations

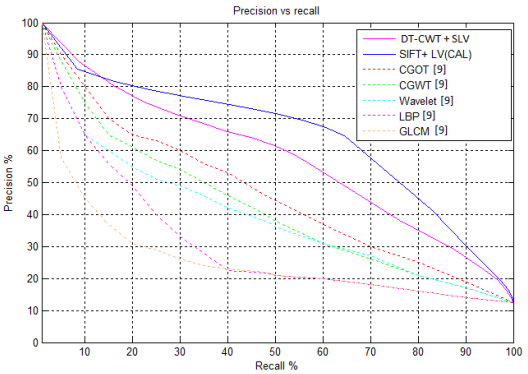

Fig. 5. Precision versus recall curves for the 8 classes

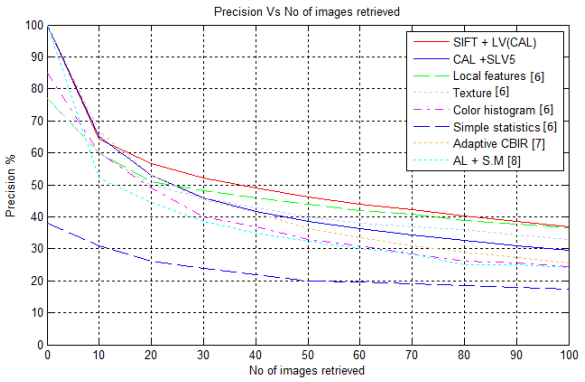

Fig. 6. Precision vs No of images retrieved for the 21 classes

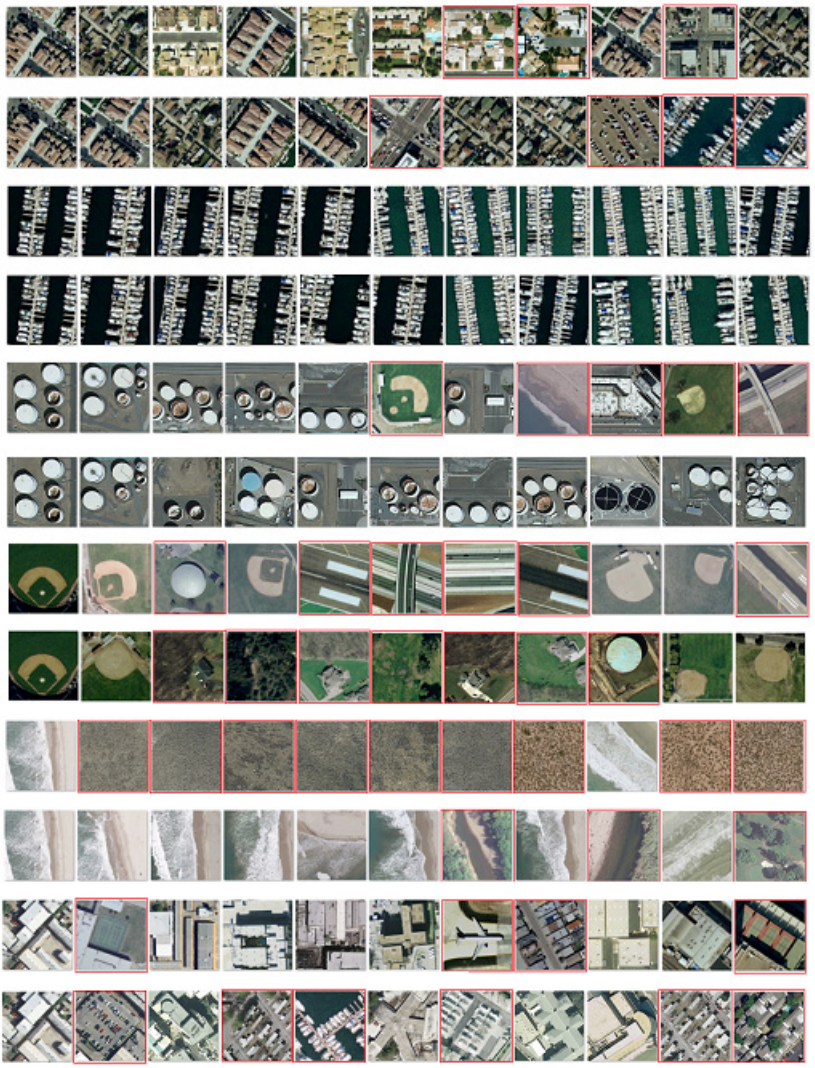

Fig. 7. Comparison of sample retrievals for different classes using local features [6] followed by ours results using SLV through DT-CWT: dense residential (72.22% [6], 63,63%), Harbor (100% [6], 100%), storage tanks (54.54% [6], 100%), baseball diamond (45.45% [6], 36.36%), beach (18.18%[6], 72.72%), building (63.63% [6], 45.45%)

.

images). We also made a comparison with two other studies. The first one based on 3D-LBP and HOG is adaptive [7] and the second one [8] compares different feature vectors such as statistical moments, Zernike moments, HOG, color histograms, LBP histograms extracted from wavelet decomposed images.

Considering all database images, average precisions have been computed obtaining 44.51% with local features based also on SIFT descriptors [6], 38.92%

using the Adaptive CBIR [7] and 47.01 % using our approach integrating color features. We should note that even if the CBIR scheme based on SIFT performs globally better but for some highly textured structured classes such as agricultural, beach and chaparral as well as for categories containing objects whose contours are curved such as airplane, baseball diamond and storage tanks we get more correct retrieval images using DT-CWT representation (see Fig. 4), specially in the first set of images found.

Fig. 7 gives some retrieval results obtained using local feature according to those presented and tested in [6] followed by our retrieval results. We can observe that storage tanks and beach classes are better recognized using the CBIR scheme based on DT-CWT representation. So, this second comparison, here, is in accordance with the previous comparisons, and both of them have validated the effectiveness and good performance of the proposed color descriptors to improve HRS images retrieval.

5 Conclusion

Rapid growth of remote sensed information generates a new research challenges in processing, transferring, archiving, and retrieving of these huge amounts of data. We introduced color component descriptors that are obtained through the creation of three new encoded color images. Even if the invariant to rotation color descriptor introduced is simple, experiments show that LV and SLV descriptors permit to enhance the retrieval performance of high resolution satellite images. We also quantitatively analyzed the effects of distance measures employed in the retrieval process, as well as the space color used to derive color descriptors. Comparison with other studies showed that the proposed features allied to DT-CWT and SIFT representations perform better than LBP, Grey Level co-occurrence matrices, wavelet and Gabor for image retrieval. Even if the color component descriptors perform better in all cases, for some categories, it is not sufficient. For those categories such as buildings, sparse residential and tennis-court, it would be interesting to define more appropriate descriptors taking into consideration their particularities (different shape and scale, surrounded by different classes, etc.). This should improve our CBIR system performances.

References

1. Liu, T., Zhang, L., Li, P., Lin, H.I.: Remotely sensed image retrieval based on region-level semantic mining. EURASIP Journal on Image and Video Processing, 1687–5281 (2012)
2. Daschiel, H., Pelizzar, A., Quartulli, M.: Information mining in remote sensing image archives: system concepts. IEEE Transactions on Geoscience and Remote Sensing **41**, 2923–2936 (2003)
3. Aksoy, S., Yalniz, I.Z., Nick, J., Tasdemirl, K.: Automatic Detection and Segmentation of Orchards Using Very High-Resolution Imagery. IEEE Transactions on Geoscience and Remote Sensing **50**, 3117–3131 (2012)

4. Scott, G., Klaric, M., Davis, C., Shyu, C.R.: Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases. IEEE Transactions on Geoscience and Remote Sensing **49**, 1603–1616 (2011)
5. Bao, Q., Guo, P.: Comparative studies on similarity measures for remote sensing image retrieval. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 1112–1116 (2004)
6. Yang, Y., Newman, S.: Geographic image retrieval using local invariant features. IEEE Transactions on Geoscience and Remote Sensing, 818–832 (2012)
7. Sebai, H., Kourgli, A.: An adaptive CBIR system for remote sensed data. In: 12th International Workshop on Content-Based Multimedia Indexing (CBMI), Klagenfurt (2014)
8. Sebai, H., Kourgli, A.: A comparative study of feature vectors derived from wavelets applied to high resolution satellite images retrieval. In: 4th International Conference on Image Processing Theory, Tools and Applications (IPTA), Paris, France (2014)
9. Shao, Z., Zhou, W., Zhang, L., Hou, J.: Improved color texture descriptors for remote sensing image retrieval. Journal of Applied Remote Sensing (JRS) **8**(1) (2014)
10. Van de Sande, K., Gevers, T., Snoek, C.M.: Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(9), 1582–1596 (2010)
11. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Trans. on Pattern analysis and Machine Intelligence **24**, 971–987 (2002)
12. Dala, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893 (2005)
13. Banerji, S., Sinha, A., Liu, C.: New image descriptors based on color, texture, shape, and wavelets for object and scene image classification. Neurocomputing **117**, 173–185 (2013)
14. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **12**(7), 629–639 (1990)
15. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: IEEE Sixth International Conference on computer Vision, pp. 839–846 (1998)
16. Kingsbury, N.G.: The dual-tree complex wavelet transform with improved orthogonality and symmetry properties. In: IEEE International Conference on Image Processing, pp. 375–378 (2000)
17. Lowe, D.G.: Object recognition from local scale-invariant features. In: ICCV, pp. 1150–1157 (1999)