IJCSC 0973-7391 Volume 7 • Number 1 Sept 2015 - March 2016 pp.138-149 available online at www.csjournalss.com

Feature Fusion of Colour Moment and Morphological Features for Enhanced Precision in Content Based Image Retrieval Systems

Swati Jain, Tanish Zaveri
Institute of Technology, Nirma University, Ahmedabad

Abstract
In this paper we propose an algorithm that learns to combine the distance measure using different feature sets in appropriate weights. The distance measure is the measure of all the images in the data set to the query image. This combined distance is them used for ranking the image for the retrieval purpose. One approach of combining the distance measure would be that the distance measures of the two feature set is added to find the combined distance, it is observed to retrieve many irrelevant images termed as false positive. To improve this scenario all the false positive retrieval are weeded out using a minimization algorithm. The outcome of this minimization processes is a weight matrix, which is used for finding the combined distance of the dataset images with the query image. This combined distance is observed to retrieve improved results when compared to the first case of simply adding the two distances. The algorithm is experimented with color moment and Circular Covariance Histogram (CCH). The dataset used is UC Merced LULC data set. The experimentation results show considerable gain in the number of relevant images retrieved in the top positions.
Keywords: CBIR, CCH, Color Moment Feature Fusion

1. Introduction

Space technology is improving day by day. This development has resulted in improved capabilities of image capturing. With the intention of achieving information with spectral, temporal and spatial completeness these years have seen the huge leap in all the three dimensions. Spectral-wise, with hyper spectral sensors resulting in every piece of information being represented in wide range of spectral bands. The authors have tried to characterize the spectral content to guide the search and answer the queries [1]. Speed of image acquisition has improved working as a multiplicative factor in the number of images and covering the temporal dimensions. The resolution have also improved from 30m to few centimetres per pixel. These enhancements has helped mankind tremendously in doing various predictions and exploring our earth in a better way, but at the same time have piled up a mountain of images to be managed and more difficult task of using them efficiently.

Remote sensing centres across the world are the storehouses of images and act as decimation centres of these images for those who want to use them. Most of these centres are managing the images using its meta data. Users specify the requirements which are mapped in meta data information which is usually less precise. Need of defining and managing the image data set based on its content has been very well understood and established in the field of remote sensing. Most of the Content Based Image retrieval (CBIR) systems have at least two modules feature extraction and feature matching. Researchers have explored various local features, and have tried to represent the images as holistic as possible [2].

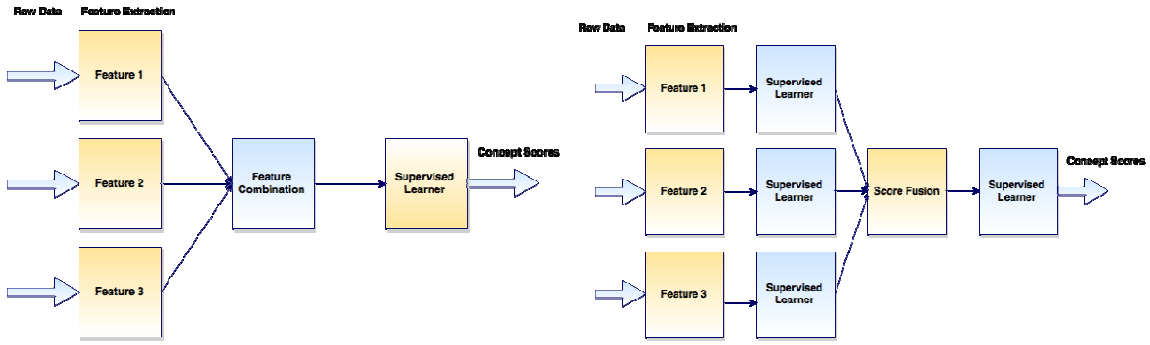
Over the years various features are defined and designed to capture distinguishing characteristics in an image and enabling content based image retrieval (CBIR) system. These features can be categorised into color, texture, and edge. Different color features are colour coherence vector (CCV) [3], colour moment [4], color histogram etc. Edge features concentrate

on edges of the local regions. They are more useful for the application that demands the task of object detection. Examples of edge features are edge direction histogram and edge coherence vector[5]. Gabor filters [6] and co-occurrence matrix [7] are most referred and cited texture descriptors. Both these texture descriptors are time tested in gray scale images. One more relatively new category of image descriptors are Morphological Texture. Erchan [8, 9] in his publications introduced morphological texture descriptors and later used them to describe satellite images. Morphological covariance as operator is used, in order to find textures. Circular Covariance Histogram (CCH) and Rotation Invariant Point Triplets (RIT) are morphological texture descriptors[9]. The process to calculate the morphological feature set is extremely compute intensive and hence that makes the feature extraction activity extended in terms of time.[10] shows how the parallel implementation can lead to considerable gain in computation time and hence resulting faster feature extraction. In the class of texture based features there are few more feature sets often used in defining images including satellite images,like Local Binary Patterns (LBP)[11] and Local Tetra Pattern (LTrP)[12]. Both the feature sets captures the relationship amongst the neighbouring pixels.

Recently introduced, texture descriptors are circular covariance histograms (CCH) and rotation invariant point triplets (RIT), on content based remote sensing image retrieval[13]. Author Erchan Aptoula has exhaustively performed a survey and is optimistic with the use of mathematical morphology in this domain[14]. The paper proposes the use of morphological operators as it is inherently good in exploiting the relation between the pixels which exactly is the texture and hence this makes it apt for using texture description. Feature set captures the information limited to a domain for instance the color information, texture information etc. To obtain the complete definition of the image the features can be combined also termed as fusion.
Par majorly two approaches in feature fusion prevails, namely Early Fusion and Late Fusion. As most of the applications referred here are for classification, the fusion categories are also defined in similar terms. In early fusion the distances using all the feature set is calculated and combined, this combined distance is then used for classification or learning the concept. Where as in the case of late fusion, the distance obtained by different features are used to independently learn the concept and obtain the class scores and then these scores are integrated to learn the concepts again. Since Early fusion combines the distance score of different feature so also known as feature fusion and, late fusion is termed as score fusion as it combines the class scores [15]. The early and late fusion schemes are illustrated in Fig 1a and 8b respectively.

Both fusion schemes have their own limitations and advantages. Early fusion goes through only one phase for learning so it is fast, while late fusion turns costly in terms of effort required for learning, as it requires two phases of learning. In case of early fusion since features are combined without any pre-processing it becomes difficult to achieve one common representation space for all the features[16].

Satellite Images are normally huge in size, taken frequently in various spectrum, so its rate of increase is also high. Authors have proposed several feature sets each capturing a particular dimension of the information in the image. An efficient way to combine these feature will facilitate a holistic comparison of the image content. The experimentation of fusion using proposed algorithm is done on the UC Merced Land Use Land Cover (LULC) dataset[17] and retrieval results are compared when features are plainly combined or when only one of the features are considered.



(a) Framework: Early fusion

Figure 1: Early and Late fusion

2. Feature Set

2.1. Circular covariance Histogram (CCH)

Texture can be defined as a relationship between the neighbourhood pixels that is generated by a common function. In author has discussed about morphological covariance [14]. As per his definition morphological covariance k is defined as the volume of the image eroded by pair of points at a distance. Various characteristics of co-variance with varying distance represented the texture properties like width, size and thinness of the pattern. Very soon the limitations were also established such as the features being variant to rotation and illumination. Mainly they were due to the structuring element (SE), morphological operator used and the evaluation method.

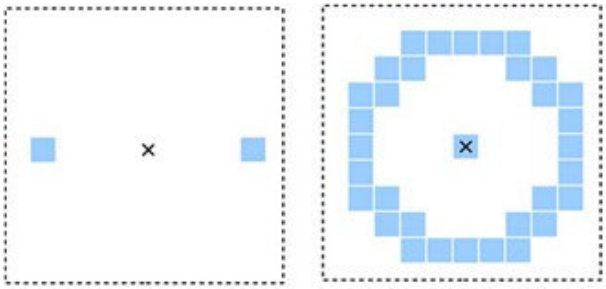


Figure 2: (a) Point Pair SE (b) Circular Shaped SE [18]

To overcome the above limitation a symmetric SE was proposed and hence shape of SE was identified as circular which makes it invariant to rotation and illumination. The various combination of dilation and erosion of a given set of pixel value is suggested as morphological operator. Later CCH which is circular covariance histogram was introduced.

The CCH is computed as below. The gray scale image I is processed with circular SE of S_j were S_j represents the various sizes of radii $j \in [1,n]$ and the range of morphological operator that could be used are erosion, dilation, opening and closing specified by M

$$= \quad (1)$$

This operation results in series of intermediate results, these image are then used to obtain an image of size f that maximize the difference with the original image and to be normalized by the cardinality of the SE.

$$\forall p, L(p) = \arg max_{1 \le j \le n} \{ |I(p) - [M_{S_j}(I)](p)| / |S_j| \} \quad (2)$$

where P denotes probability. In a way, CCH can be thought of as the histogram of the maximal outputs of locallycomputed

It is proposed that if CCH computation is done in parallel, a good performance gain can be obtained. After analysing the data dependency it was found that any stage does not use intermediate results. Thus it was motivating enough to use the parallel architecture to render GPU to this single instruction and multiple data scenario. Apart from parallelism in one image the same can be done on multiple images concurrently hence giving parallelism across images.

After morphological operator a series of intermediate images $[M_{S\,j}(I)]$ is obtained. To obtain the labelled image in parallel, each pixel label thread calculates the label by maximizing difference to original image, normalized by cardinality of B_i for corresponding pixel across image obtained for different radii.

2.2. Color Moments

One of the most common color feature is color Histogram. Histograms are useful because they are relatively insensitive to position and orientation changes and they are sufficiently accurate, however, they do not capture spatial relationship of color regions and thus, they have limited discriminating power. Many publications focus on color indexing techniques based on global color distributions. Color moments is one more approach which is found to be more robust with respect to the quantization parameter of the histogram. It is implemented by storing the first three moments of each color channel of an image in the index. For HSV image we store only 9 floating point numbers per image. The HSV color space has three components: hue, saturation and value. In HSV, hue represents color. In this model, hue is an angle from 0 degrees to 360 degrees. Saturation indicates the amount of white added to pure colour. It ranges from 0 to 100 percent. Sometimes the value is calculated from 0 to 1. When the value is 0, the color is white and when the value is 1, the color is a primary color. A faded color is due to a lower saturation level, which means the colour contains whiter. Value is the brightness of the color and varies with color saturation. It ranges from 0 to 100 percentage. When the value is 0, the color space will be totally black. With the increase in the value, the color space brightens up and shows various colors. The first three colour moment in an image are calculated as below. The first moment represents the average color, the second moment represents the standard deviation and the third color moment represents the

$$E_c = \frac{1}{N} \sum_{j=1}^{N} (I_c)$$

$$\sigma_c = \frac{1}{N} \left(\sum_{j=1}^{N} (I_c - E_c)^2 \right)^{1/2}$$

$$S_c = \frac{1}{N} \left(\sum_{j=1}^{N} (I_c - E_c)^3 \right)^{1/3}$$

Hence every image is represented as a (number of color channels X number of moments) that makes it a nine tuple.(E_h,σ_h,S_h, E_s,σ_s,S_s, E_v,σ_v,S_v). Color moment are indicative of the color

composition of the images. Hence using just one feature will retrieve false positive that is images with completely different content which just happen to have a similar color composition as the query image. Hence it is required to have the other description also in the image.

3. Proposed Work

We propose a weight learning algorithm that learns to combine the Morphological and color moment features appropriately to improve the retrieval results. Feature fusion is employed to minimize the cases where distance of irrelevant images is less than the distance of relevant images.

Formulation of the problem of finding the appropriate fusion weights, is discussed as below. The final distance vector which is combination of all the feature set is obtained by the equation 4. Where W is the vector that is to be learnt, and then final score is obtained as weighted (w) sum of different distances obtained between color and morphological features. In [19] score fusion is demonstrated, that maps all the distances in one scale. In the given algorithm late fusion is used and hence score values, obtained by classifier are combined. In case of score higher values signifies higher degree of similarity to a class and lower value signifies lesser belonging to the class. However, in our proposed work we have used distance vector instead of scores, and higher values signifies smaller degree of similarity between two images. Let all the images which are relevant be R_j and all the images non relevant as R_i. The idea is to minimize all the cases where the irrelevant image has less distance measure in comparison to relevant images hence trying to minimize them.

$$min_w \frac{1}{2}\|w\|^2 + \sum((w.(R_j - R_i)), s.t. w_n \geq 0 \quad (4)$$

The key idea is to add the distances between query image and all other images, using different features. All these calculated distances are in different scale. The objective is to map all the feature vectors into a common score space. Given a distance measure obtained by a particular feature set and w is the learned weight.

$$D_j = \sum_{j=1}^{n} d_j^j . w_i^j \quad (5)$$

In equation 5, D_i is the sum of the product of distance of i^{th} image w.r.t the query image and their respective weights. Distance d_i^j (distance for the i^{th} image using j^{th} feature), n is the number of feature sets considered. For a given feature set it is the vector of N (N is the number of images in the data set). This d_i^j for different values of j is not in the same scale and range, but in all the feature sets smaller distance value signifies higher degree of similarity to the query Image. As the score values are incomparable, fusion cannot be done directly. So to calculate the final distance
$D_j = \sum_{j=1}^{n} d_j^j . w_i^j$ is not appropriate, hence equation (1) is used. To calculate the weight matrix, w_i^j, we consider S_i as the set of images which are relevant and S_j as the set of non-relevant images. We minimize the cases where irrelevant images have distances less in comparison to relevant images with the query image.

Distance of all the images is calculated with respect to the query image, using one feature set at a time, and all the distances are added to find the final distance value. Pair i, j is marked, where i belongs to set relevant images and j belongs to the set of irrelevant images, such that the

distance calculated for irrelevant images is less than the distance calculated for relevant images. The set I contains all such marked pairs. Then we calculate a pair wise comparative matrix z_i^j which is the difference of the Euclidean distance of images in set I. Since the database used is classified images hence all images belonging to the same class as that of query image is considered as relevant and all other images are considered irrelevant. It is expected that distance for relevant images is smaller than the distance obtained for irrelevant, and all the cases that violates this necessary condition are marked in matrix I. Algorithm 1 is the variant of algorithm proposed in [19] where score fusion is replaced by distance measure.

1: *Initialize Array:* $w^t=1$
2: *Repeat n times, step 3 to 8*
3: $I^t \leftarrow \{(i,j)|w^t(x_i - x_j) > 0\}$
4: $Z_{i,j} \leftarrow x_i - x_j, \forall (i,j) \in I^t$
5: $\overline{w} \leftarrow w| \sum_{i,j \in I^n}(I + 2Z_{ij}^T Z_{ij})\mathrm{w} = 0$
6: $\overline{w} \leftarrow$ all positive values of w
7: $\alpha_t \leftarrow argmin_{0 \geq \alpha \geq 1} f(w^t + \alpha(\overline{w} + w^t))$ (finding alpha between 0-1 in an interval of 0.2)
8: $w^{t+1} \leftarrow w^t + \alpha(\overline{w} - w^t)$
9: Output: w^{t+1}

Algorithm 1: Calculation of Weight

Algorithm 1 describes the steps to calculate the weights for the distance scores. The algorithm is based on modified Newton method [19, 20, and 21].

In step 5 equation is solved linearly to obtain w. Size of I and hence Z is MxM where M is the number of image in the dataset. Step 6 removes all the negative values and hence results into negation of false negative values of the feasible set. The next two lines obtain the value of w^{t+1}. The loop is repeated n times n is found empirically such that the value of w becomes close to constant.

4. Experiment and Result

We have used the UC Merced LULC data set, which is the largest of its kind[22]. In particular, it consists of images categorized into 21 classes, with a pixel resolution of 30 cm. Each class contains 100 RGB color samples of size 256 × 256 pixels,. For CCH feature extraction, all data have been processed in grey level, with the conversion having been conducted through Grey = 0.299 × R + 0.587 × G + 0.114 × B. The color moment is calculated in HSV domain.

Table 1: Retrieval result for Different Class of Images

Image Class	No. of Relevant Images in Top 20			No. of Relevant Images in Top 40			No. of Relevant Images in Top 60		
	Without Weights	With Weights	% increase	Without Weights	With Weights	% increase	Without Weights	With Weights	% Increase
Agriculture	12.6	19.0	50.4	20.8	28.6	37.6	27.1	30.3	11.9
Aeroplane	5.1	19.9	288.7	7.8	28.5	267.1	10.1	30.3	198.9
Baseball	5.8	19.8	243.1	8.0	25.8	224.0	9.7	26.3	171.0

Beach	9.6	19.9	108.2	14.5	24.3	66.9	18.3	24.3	33.0
Building	6.2	19.7	215.9	10.2	27.5	170.3	13.5	29.1	115.5
Chaparral	18.0	18.1	0.5	34.7	21.3	-38.7	49.7	21.3	-57.2
Dense Residential	5.3	19.4	268.8	7.5	24.4	227.2	9.5	25.3	166.4
Forest	15.4	16.2	5.3	28.9	19.0	-34.3	39.1	19.0	-51.3
Free way	4.7	19.6	321.5	6.8	26.6	291.3	8.7	28.5	227.0
Golf Course	8.1	19.2	136.3	12.0	23.3	93.6	15.2	23.8	56.6
Harbour	14.1	19.6	39.8	22.8	24.5	7.6	28.6	25.5	-10.9
Intersection	4.8	19.9	316.9	7.0	27.9	297.6	9.5	30.1	217.8
Medium Residential	6.6	19.0	187.4	9.5	21.8	129.9	11.7	22.1	89.5
Mobile Homepark	8.7	19.7	127.2	11.7	25.8	121.0	14.1	26.5	87.9
Overpass	5.5	19.5	255.7	7.9	26.3	233.9	9.8	27.4	180.6
Parking	6.0	19.3	219.2	9.3	22.3	139.8	12.4	22.5	80.8
River	7.6	19.3	155.0	10.9	23.9	119.2	13.5	24.0	77.2
Runway	5.6	19.4	243.7	7.4	25.8	248.0	8.7	26.1	200.3
Sparse Residential	6.5	19.8	204.8	9.6	27.7	189.1	11.9	29.1	144.2
Storage Tank	4.2	19.3	361.2	6.0	24.7	310.0	7.7	26.0	239.9
Tennis Court	3.8	19.9	426.8	5.9	27.6	371.5	7.7	29.0	275.1
Average	7.8	19.3	198.9	12.3	25.1	165.4	16.0	26.0	116.9

- Step 1: CCH feature for image is a matrix of size (1 × 20) and color moment is of size (1 × 9). Euclidian distance of all the images in the database with the query image, using both the features are calculated, as d_1 and d_2.
- Step 2: Marking values in array$Z_{i,j}$, is the processes of marking all the pairs i, j when $i \in$ (*listofreleventimage*) and $j \in$ (*listofallirrelevantimages*) and distance measure of j is less than i with respect to the query Image.
- Step 3: Using the above mentioned algorithm the weights are updated such that the instances marked in $Z_{i,j}$are minimized.
- Step4: The updating process goes for N iteration.
- Step 5: Finally the distance value is calculated using 5, which is the weighted distance from the query image using both the feature set.

The table 1 shows the results obtained. There are total 21 classes in the database every class has 100 images. Average of retrieval of all the 100 images in a class is taken to summarize the performance. The results are compared with the retrieval result when the distance measures are just added and when the distance measures are added with appropriate

weights. These weights are obtained by the learning algorithm as shown in the Algorithm 1. Number of relevant image retrieved in top 20, top 40 and top 60 is compared for retrieval when distance values obtained with two features CCH and color moments are added and when the two distances are added with weights learnt from the proposed Algorithm. As is visible in the Table 1 When top 20 images are

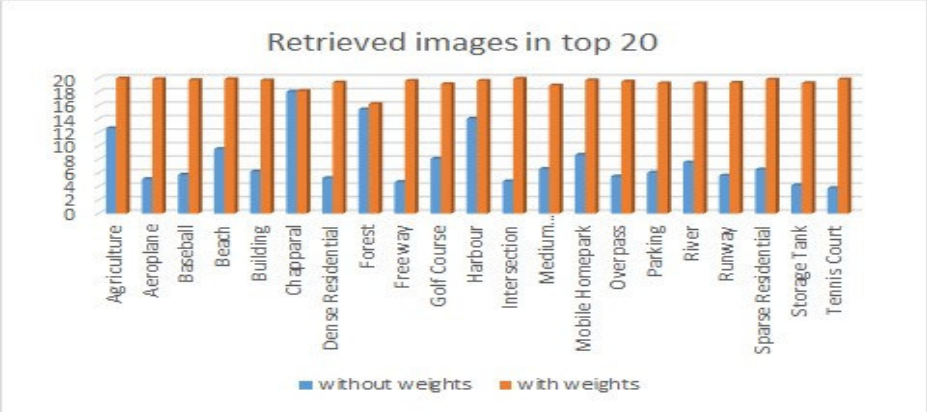


Figure 3: No of Relevant Images in Top 20

considered almost all the images retrieved using the proposed algorithm, are relevant. The highest percentage increase is 426% in case of Tennis court and lowest being 0.5% for the reason this had very small scope for improvement. When top 40 images are considered then also the average of 165% improvement is observed and when top 60 images are considered an average percentage increase of 116 % is observed. The retrieval results can be seen in the bar graph as in Figure 3 for top 20 retrieved images, in all the class of the image there is improvement in the number of relevant images retrieved in top 20 images. In Figure 4 for top 40 retrieved images except in two classes Chaparral and Forest still the fusion algorithm performs better in comparison to other approach. And in Figure 5 finally, for top 60 retrieved images three class namely Chaparral, Forest and Harbour the classic addition algorithm outperforms our fusion algorithm. But in all these three class the precision was already good n all these three class. The average amongst all the class when observer gives 198.9%, 165.4%, 116.9% improvement in the number of relevant images in top 20, 40 and 60 respectively . The graph clearly shows remarkable improved precision in the initial Positions in all the class and, for any information retrieval application the information at the top positions is very crucial and hence can be seen that in top 20 almost all the images retrieved are relevant.

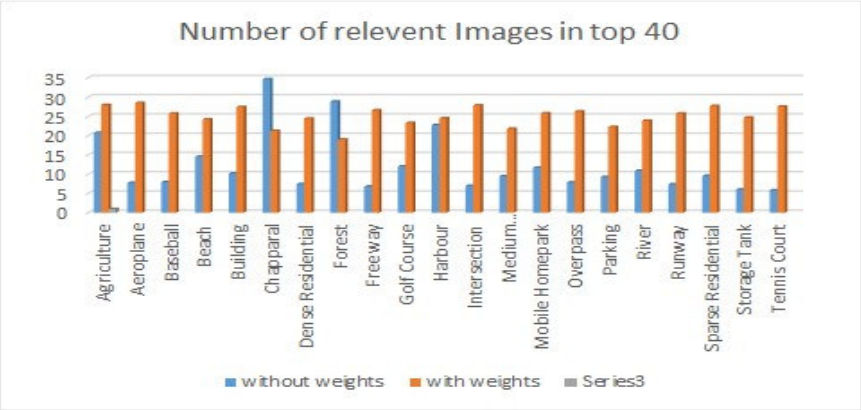


Figure 4: No of Relevant Images in Top 40

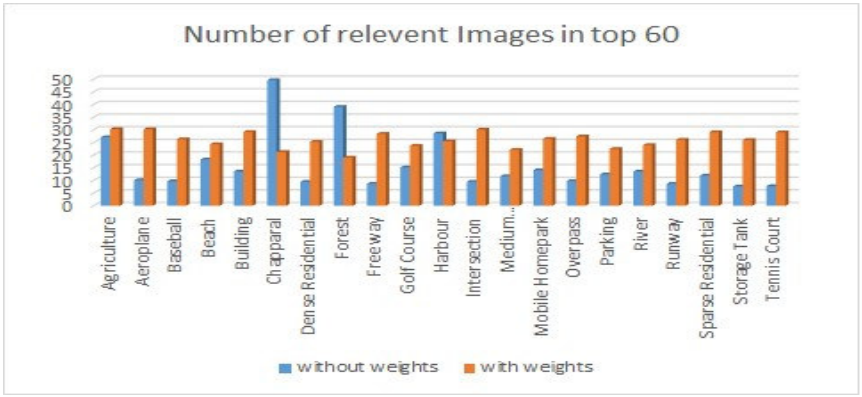


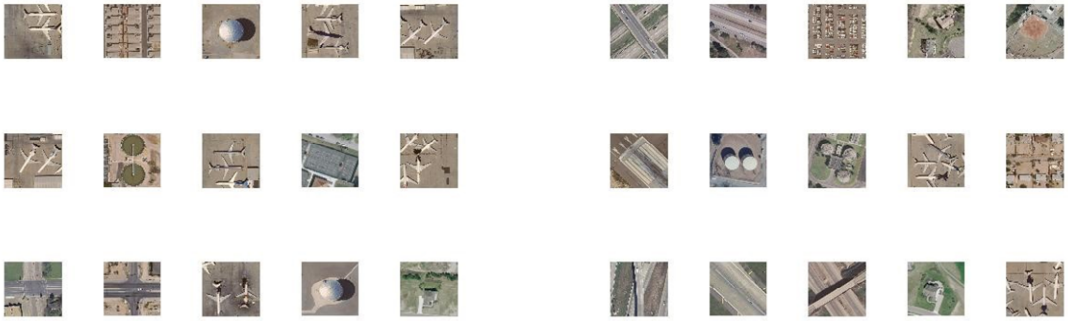
Figure 5: No of Relevant Images in Top 60

5. Conclusions

The color feature and texture feature belongs to two different class of the feature set. These features define two different aspect of the image. In this paper we have demonstrated a learning algorithm that learns the optimal

Figure 6: Query Image

weights for the fusion of the distance values obtained by two different class of the feature set. Principally combining the image description for color and texture should be effective and should retrieval good result. It is experimented to add them in some weighted forms and the results obtained are encouraging. The experiment is performed on UC Merced Land use Land cover data set. It is observed that when precision of the top 20 Images are compared and all the images retrieved are found to be relevant. The percentage increase in the number of relevant images in top 20, 40 and 60 are 198.9%, 165.4%, 116.9% respectively, which is substantial improvement.



(a) (b)

Figure 7: Retrieval Result without weights

(a) Retrieval Result with weights (b) Framework: Late Fusion

Figure 8: Retrieval Result with weights

References

[1] Veganzones, M. A., Maldonado, J. O. & Grana, M. : On content-based image retrieval systems for hyperspectral remote sensing images. *Computational Intelligence for Remote Sensing* 133, 125–144 (2008).

[2] Chen, L., Yang, W., Xu, K. & Xu., T. : Evaluation of local features for scene classification using vhr satellite images. *Remote Sens. Event, Munich, Germany* 385–388 (2011).

[3] Pass, G., Zabih, R. & Miller, J. Comparing images using color coherence vectors. In *Proceedings of the fourth ACM international conference on Multimedia*, 65–73 (ACM, 1997).

[4] Kodituwakku, S. &Selvarajah, S. Comparison of color features for image retrieval. *Indian Journal of Computer Science and Engineering* 1, 207–211 (2004).

[5] Gao, X., Xiao, B., Tao, D. & Li, X. Image categorization: Graph edit direction histogram. *Pattern Recognition* 41, 3179 – 3191 (2008).

[6] Zhang, D., Wong, A., Indrawan, M. & Lu, G. Content-based image retrieval using gabor texture features. In *IEEE Pacific-Rim Conference on Multimedia, University of Sydney, Australia* (2000).

[7] Haralick, R. M. Statistical and structural approaches to texture. *Proceedings of the IEEE* 67, 786–804 (1979).

[8] Aptoula, E. Extending morphological covariance. *Pattern Recognition* 45, 4524–4535 (2012).

[9] Aptoula, E. Remote sensing image retrieval with global morphological texture descriptors. *Geoscience and Remote Sensing, IEEE Transactions on* 52, 3023–3034 (2014).

[10] Kumar, S., Jain, S. & Zaveri, T. Parallel approach to expedite morphological feature extraction of remote sensing images for cbir system. In *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, 2471–2474 (2014).

[11] Banerji, S., Verma, A. & Liu, C. Novel color lbp descriptors for scene and image texture classification. In *15th International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, Nevada*, 537–543 (Citeseer, 2011).

[12] Murala, S., Maheshwari, R. &Balasubramanian, R. Local tetra patterns: a new feature descriptor for content-based image retrieval. *Image Processing, IEEE Transactions on* 21, 2874–2886 (2012).

[13] Huang, C. R., Chen, C. S. & Chung, P.-C. : Contrast context histogram - a discriminating local descriptor for image matching 4, 53–56 (2006).

[14] Aptoula, E. : Extending morphological covariance. *Pattern Recognition* 45, 4524–4535 (2012).

[15] Ye, G., Liu, D., Jhuo, I.-H. & Chang, S.-F. Robust late fusion with rank minimization. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 3021–3028 (IEEE, 2012).

[16] Snoek, C. G., Worring, M. &Smeulders, A. W. Early versus late fusion in semantic video analysis. In *Proceedings of the 13th annual ACM international conference on Multimedia*, 399–402 (ACM, 2005).

[17] Yang, Y. &Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL International Conference on*

Advances in Geographic Information Systems, GIS '10, 270–279 (ACM, New York, NY, USA, 2010).

[18] Aptoula, E. : Remote sensing image retrieval with global morphological texture descriptors. *IEEE Transactions on Geoscience And Remote Sensing* 99, 1–12 (2013).

[19] Ma, T., Oh, S., Perera, A. &Latecki, L. Learning non-linear calibration for score fusion with applications to image and video classification. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, 323–330 (2013).

[20] Keerthi, S. S. &DeCoste, D. A modified finite newton method for fast solution of large scale linear svms. In *Journal of Machine Learning Research*, 341–361 (2005).

[21] Lin, C.-J., Weng, R. C. &Keerthi, S. S. Trust region newton method for logistic regression. *The Journal of Machine Learning Research* 9, 627–650 (2008).

[22] Yang, Y. &Newsam, S. : Bag-of-visual-words and spatial extensions for land-use classification. *18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* 270–279 (2010).