Near Duplicate Image Matching using PCSLBP and SIFT based Varible Length Signatures

[1]Amruta Landge, [2]Pranoti Mane

[1,2]E&TC MESCOE, Pune, India

Email: [1]aklandge@gmail.com, [2]pranoti.mane@mescoepune.org

Abstract—Near duplicate image identification needs the matching of somewhat altered images to the original image. This will help in the detection of forged images. A great deal of effort has been dedicated to visual applications that need efficient image similarity metrics and signature. Digital images can be easily edited and manipulated owing to the great functionality of image processing software. This leads to the challenge of matching somewhat altered images to their originals, which is referred as near duplicate image identification. In this paper near duplicate images are matched using variable length signatures. These signatures are extracted from images using patch based approach. Similarity between two images is computed by comparing these signatures with help of earth mover's distance.

Keywords—near-duplicate images, near-duplicate image retrieval, SIFT, near-duplicate image detection.

I. INTRODUCTION

Near-duplicate images are created by taking independent pictures of the identical object under various conditions in resolutions, illuminations, and others. Moreover, they can be generated by modifying the original images with the help of some transformations such as image scaling and rotation. Consider basic flow of image matching technique. In image matching, image is represented to some feature spaces using suitable technique. Then some local or global features are extracted from image to represent them as vectors. These features are analyzed to compute the similarity between query image and reference database using different mathematical tools.

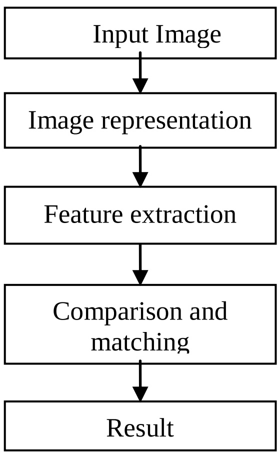

Fig.1. Basic block diagram of image matching technique

The widespread distribution of digital data and proliferation of digital images due to Internet has augmented problem related with copyright intrusion on digital images. Watermarking techniques have been introduced to protect copyrighted images. However watermarks are susceptible to geometric distortions and image processing and may not be very efficient. There are numerous applications of image matching algorithms that ranges from easy photogrammetry tasks like feature recognition to the development of 3D modeling software. Image matching and feature detection plays significant role in photogrammetry. Identifying near-duplicate images plays vital role in numerous applications such as postal automation, copyright protection and others. Their application persists to increase in a various fields day by day. In the recent decades, this has been a very active area of research and as shown by the huge amount of documentation and work published around this. As requirements change and become more demanding, researchers are encouraged to develop novel technologies to accomplish these requirements.

II. LITERATURE SURVEY

Image Representation and Feature Extraction Liu and Yang [1] represented image using color difference histogram. They encoded the edge and color orientations of the image in a uniform framework. Cha et al. [2] proposed distance among sets of measurement values as a measure of dissimilarity of two histograms. This method considers three versions of the univariate histogram and their computational time complexities. It has benefits over conventional distance measures regarding the overlap among two distributions. It considers similarity of the non-overlapping parts plus that of overlapping parts. In order to represent image Kim [3] used ordinal measures of the discrete cosine transform coefficients. Aksoy and Haralick [4] represented an image using co-occurrence variances and line-angle-ratio statistics. Those were ordered into a feature vector of 28 dimensions. In [5], Meng et al. represented an image using 279D feature vector. Enhanced Dynamic Partial Function used to measure similarity. It adaptively activated a several number of features in a pairwise manner in order to accommodate the characteristics of every image pair. In some works

[6], [7], the vectorial representations were first embedded into binary codes. Here, main issue was to make sure the images that were alike in the original vector space should be compact in the binary code space. Regardless of the simplicity, representing an image by a single vector generally fails to handle the variations between the near-duplicate images. Dimension of the features has to be determined a priori, in spite of the characteristics of image. Furthermore, vectors are not good at modeling the relationships amongst different parts of the image.

Chum et al. [8] represented an image based on its color histograms and then for fast retrieval used Locality Sensitive Hashing (LSH). Todorovic and Ahuja [9] employed tree of recursively embedded image regions to represent image. Tree was improved with new nodes generated by merging adjacent sibling nodes and producing directed acyclic graphs (DAGs). There are various advantages of using edge fragments or regions over interest points for image matching. The high dimensionality of regions makes them richer descriptors of target objects' geometric properties including shape and size, and gray-level contrast (photometric property). The higher dimensional character of regions makes their matching more stable to viewpoint changes and small illumination across given images. According to the image characteristics number of keypoints per image differs. Descriptor is used to characterize a support region surrounding the keypoint. The most commonly used descriptors are PCA-SIFT [10], Shape Context [11], Gradient Location and Orientation Histogram (GLOH) [12], etc. Xu et al. [17] introduced two-stage method for near-duplicate image detection and retrieval. First image divided into overlapped and non-overlapped rectangle blocks over multiple levels. In the first matching stage, the distances among two blocks from the two images were calculated with the help of SIFT features. In the next stage, multiple block alignment hypotheses considering scale variations and both piecewise spatial shifts were proposed. However, drawback of this method lies in the block divisions which fall short to cope with image rotations.

Zhao et al. [18] implemented PCA-SIFT and DOG for image representation. They used LIP-IS, for better nearest neighbor search. For matching one-to-one symmetric matching algorithm was proposed. It found to be highly reliable for near-duplicate keyframe identification. This is owing to its ability in excluding false LIP matches as compared with other matching strategies. The superiority of LIP-IS over LSH was verified from experiments. Nevertheless, both LIP-IS and LSH can not guarantee to locate the exact nearest neighbor. Chum et al. [19] used a visual vocabulary of vector quantized local feature descriptors. For retrieval employed enhanced min-Hash techniques. He implemented the bag-of-visual-words model for near-duplicate shot and image detection. To achieve efficient detection in a large data set to attain better detection in a big data set. However, bag-of-visual-words model has drawback that spatial layout of the visual words is totally disregarded. This will introduce ambiguity while matching. In addition, clustering process hampers the discriminative power of the local descriptors [20].

Sivic and Zisserman [21] represented image using the bag-of-visualwords model. A visual word vocabulary was learnt with the help of clustering local descriptors that are obtained from a training set. Then every descriptor from random image can be allocated to its nearest word in the vocabulary. Accordingly, an image can be characterized by a vector with every axis representing the occurrence statistics of the respective visual word in the image. There are some representations which are adapted to particular kinds of images, such as document images, face images and others. A renowned approach is to represent a face image using weighted sum of eigenfaces with the help of PCA [22].

A. Similarity computation

In [1] similarity between two images was calculated with the help of enhanced Canberra distance. In [3] image similarity computed using L1 norm. Aksoy and Haralick [4] calculated similarity between images with the help of distances between feature vectors. According to their experiments, probabilistic similarity measures performed better in terms of precision and recall compared to the cases where the geometric measures were employed. It was better to consider the feature distributions while deriving similarity measures. Image similarity computed using Hamming distance within the binary codes in [6] and [7]. In [28] Earth Mover's Distance was used for similarity computation which gives better results in case of variable-length signatures. In [31] Kang et al. formulated problem of image similarity computation in terms of sparse representation. Singha and Hemachandran [35] employed histogram intersection distance method for similarity measurement. For similarity computation Liu et al. [37], used Earth Mover's distance was employed.

III. METHODOLOGY

B. Algorithm 1: Patch based variable length signatures

We first generate patches from given image, based on which the image is represented by a variable-length signature. The length of signature is variable and depends on the number of patches in the image. Number of patches varies due to individual image characteristics.

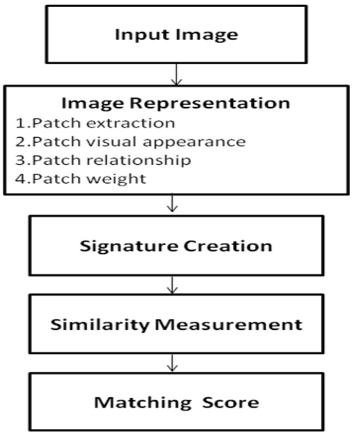

Fig.2. Algorithm for patch based signature generation

1. Image Representation

1.1 Patch extraction

Various methods can be used to represent an image based on characteristics like raw pixels, a keypoints and others. We exploit patches for image representation. A patch in the image is consists of pixels which are visually similar and spatially adjacent. The pixels which are similar in intensity and spatially adjacent are clustered, and then each resulting cluster is considered as a patch. Both spatial aspect and pixel intensity for patch extraction are taken into account.

A pixel p's K nearest intensity-similar neighbors in the image I are defined as a set

$$\wedge_p^k = \{q_i \in I \, || \, g_{qi} - g_p | \leq \eta, i = 1,2,3,....K\}$$

Where, g_{qi} and g_p are the intensities of the pixels qi and p respectively. Threshold η governs their intensity difference.

Given, $\wedge_p^k$ we obtain the distance d_p^k which is defined as $d_p^k = \max_{q_i \in \wedge_p^k} d(p,q_i)$

Where, d(•, •) measures the Euclidean distance between two pixels. For small d_p^k it is highly possible that the pixel p is densely surrounded by intensity-similar pixels. On the other hand, when d_p^k is large, it means that p is sparsely surrounded by intensity-similar pixels, under which context it is more probable to be an outlier.

In order to achieve scale invariance distance d_p^k is further normalized as $\tilde{d}_p^k = \frac{d_p^k}{D}$

Where, with D defined as the maximum d_p^k obtained from the image, i.e. $D = \max_{p \in I} d_p^k$

1.2 Patch visual appearance (HAi)

To illustrate patch visual appearance, good robustness to image illumination, orientation and scale variations is highly preferred. Patch visual appearance descriptor, Probabilistic Center-symmetric Local Binary Pattern (PCSLBP) which is an improvement of CSLBP is employed in this method.

CSLBP: It encodes the center-symmetric intensity differences in a pixel's local neighborhood. For given a pixel Pc its P neighbors that are evenly spaced around a circle of radius R centering on Pc are sampled, so the coordinates (xi, yi) of the i th (0 ≤ i ≤ P − 1) neighbor are given by

$$x_i = x_c + R * \cos(\frac{2\pi * i}{p})$$

$$y_i = y_c + R * \sin(\frac{2\pi * i}{p})$$

Where (xc, yc) are the coordinates of pc. Denoting the intensities of the P neighbors as g0, g1, . . . , gP−1, the pixel pc is then encoded as

$$CSLBP(p_c) = \sum_{i=0}^{\frac{P}{2}-1} s(g_i - g_{i+\frac{P}{2}}) * 2^i$$

With s(z) given as:

$$s(z) = \begin{cases} 1 & z > T \\ 0 & z \leq T \end{cases}$$

Where, T is a predefined threshold.

Histogram: Sampling P neighbors for a pixel, there will be a total number of $2^{\frac{P}{2}}$ different binary codes, each one represents a distinct micro-pattern. An image can be characterized by the distributions of the micro-patterns, a histogram of $2^{\frac{P}{2}}$ dimensions with each dimension counting the occurrences of the corresponding micro-pattern in the image.

Limitation of CSLBP: According to the definition of CSLBP two local neighborhoods will be considered as identical if they have the same CSLBP codes; otherwise, they will be regarded totally different. In this manner, it is not suitable particularly when the images are subject to variations, under which circumstance the local neighborhoods are likely to be distorted.

PCSLBP: this visual descriptor encodes the pixel intensity differences in a probabilistic manner. In spite of hard coding the intensity difference among two pixels as zero one it allocate a probability to characterize the difference**.** If the pixel intensity ranges from 0 to G, the pixel intensity difference will be in the interval of [−G, G]. Given a pixel intensity difference δ, we calculate its probability of being encoded as 1 and 0, which are denoted by Pr1(δ) and Pr0(δ) respectively.

They are defined according to

$$\Pr1(d) = \exp\left(\frac{(\delta - G)^2}{\sigma}\right)$$

$$\Pr0(d) = \exp\left(\frac{(\delta - (-G))^2}{\sigma}\right)$$

Where σ is positive real number.

From such a probabilistic perspective, the way to model micro-pattern distributions modified accordingly. In contrast to CSLBP, for which a local neighborhood contributes to only one histogram bin, each local neighborhood contributes to all the histogram bins in the proposed PCSLBP, with the magnitude of the contribution determined by the corresponding probability. Encoding pixel intensity differences in a probabilistic manner is preferred when compared with the hard coding employed in CSLBP. It is more informative about the intensity difference, with the larger (smaller) the difference, the higher the probability of being encoded as 1 (0).

1.3 Patch Relationship (HRi)

Beyond the patch visual appearance, we grab the relationships among the patches in the image. We compute the Euclidean distance between the gravity centers of two patches. The maximum distance among two patches in the image is quantized into L levels. Then a histogram HRi of L bins is associated with patch Oi, with the lth bin representing the number of patches in the image whose distances from Oi fall into the lth level.

1.4 Patch Weight (ωi)

We assign a weight ωi to each patch Oi in the image to indicate its contribution in identifying the image.

Weight = size of patch/size of image

It is defined to be proportional to the size of the patch, simply based on the assumption that the bigger a patch is, the more important it is for the image.

1.5 Signature creation

Overall, given an image composed of N patches, it is represented by a signature, S = {(HA1, HR1, ω1), (HA2, HR2, ω2), . . . , (HAN, HRN, ωN)}. So the length of the signature depends on the number of patches in the image Where, HAi describing the patch visual appearance, HRi patch relationship and ωi patch weight.

1.6 Similarity measurement and matching score

With each image represented by a signature, computing the similarity among two images transforms to the similarity computation among two signatures. Since the signature is of variable length, the commonly used similarity measures based on the fixed-dimension vectors are not applicable. Hence earth mover's distance is employed for matching.

Earth Mover's Distance: It is defined to integrate two parts related with the distance in terms of patch visual appearance, i.e. D ($H_{Ai}, \hat{H}_{Ai}$) and patch relationship, i.e. D ($H_{Ri}, \hat{H}_{Ri}$) and is formulated as

$$d_{ii} = \rho * \frac{D(H_{Ai}, \hat{H}_{Ai})}{C_A} + (1-\rho) * \frac{D(H_{Ai}, \hat{H}_{Ai})}{C_R}$$

Where, ρ is Weighting coefficient

C. Algorithm 2: Scale-invariant feature transform

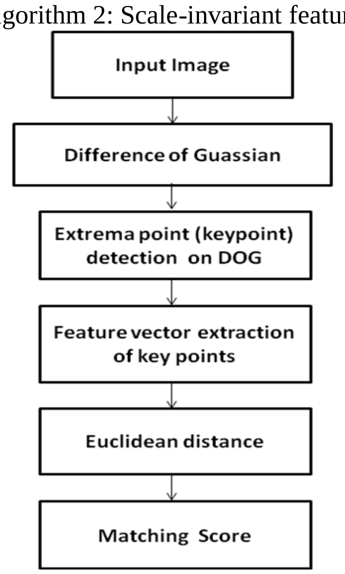

Fig.3. Algorithm for scale-invariant feature transform

Employing local keypoint detectors and descriptors has become a popular option for image representation. Researchers have observed lot of image local descriptors and detectors which are extensively explored in several applications. Invariant descriptors matching methods are more preferable under complex image deformations. The vital point based on descriptors approach is to get stable and unique feature descriptors. Scale-invariant feature transform performs reliable matching under rotation, illumination and scale, changes, so SIFT feature is greatly distinctive. The common practice is to first indentify some keypoints in the image using detectors such as Difference-of-Gaussian (DOG). The number of keypoints for each image changes in accordance with the image characteristics. Then a descriptor is used to characterize a support region around the keypoint. The most regularly employed descriptor is SIFT.

Algorithm:

1. Probable interest points that are invariant to scale and orientation are recognized using a difference-of-Gaussian function

2. Keypoints are selected based on measures of their stability

3. One or more orientations are assigned to each keypoint location based on local image gradient directions

4. The local image gradients are measured at the selected scale in the region about each keypoint

Nevertheless in the case of retrieval, for every descriptor in the input image, its nearest neighbor in the images from the database should be identified. Since there are usually hundreds or thousands of keypoints per image the computational overhead is prohibitive especially when the size of the database grows. In that scenario SIFT descriptor of keypoints must be selected that are from the patch of an image which has highest weight.

IV. RESULTS AND DISCUSSIONS

D. Datasets

In the image retrieval, there is generally a small query set and a large database, which is consists of relevant images with respect to the query plus irrelevant ones. The irrelevant images primarily serve as distractors. For every issued query, the relevant images in the database should be retrieved.

Two data sets are employed in the current work, namely Columbia Near Duplicate Image Database and California-ND.

Columbia Near Duplicate Image Database: It contains 150 near-duplicate pairs (300 images) and 300 non-near-duplicate images. California-ND: it is an annotated dataset for near-duplicate detection in personal photo collections. In order to learn some parameters involved in the current approach, data set is divided into two independent sets: test set, and training set each involving a query set and a database.

E. Performance evaluation

Precision-Recall (PR) curves: These are commonly used for performance evaluation in several retrieval systems. In information retrieval scenario precision is nothing but the he fraction of retrieved instances that are relevant, whereas recall is the fraction of related instances that are retrieved.

Mean Average Precision (MAP): It is single-figure measure, which takes into account precisions across all recall. MAP and PR curves can be used for performance evaluation and comparison on the test sets and MAP can be used for parameter selection on the training sets.

V. CONCLUSION

Patch based image matching method shows good robustness to image scale and orientation invariance. The keypoints obtained using SIFT are useful owing to their distinctiveness, which facilitate the correct match for a keypoint to be chosen from a large database of other keypoints. Scale-invariant feature transform is invariant to image rotation and scale and robust across addition of noise, and change in illumination.

REFERENCES

[1] G.-H. Liu and J.-Y. Yang, “Content-based image retrieval using color difference histogram,” Pattern Recognit., vol. 46, no. 1, pp. 188–198, 2013.

[2] S.-H. Cha and S. N. Srihari, “On measuring the distance between histograms,” Pattern Recognit., vol. 35, no. 6, pp. 1355–1370, 2002.

[3] C. Kim, “Content-based image copy detection,” Signal Process., Image Commun., vol. 18, no. 3, pp. 169–184, 2003.

[4] S. Aksoy and R. M. Haralick, “Probabilistic vs. geometric similarity measures for image retrieval,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2000, pp. 357–362.

[5] Y. Meng, E. Chang, and B. Li, “Enhancing DPF for near-replica image recognition,” in Proc. Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2003, pp. II-416–II-423.

[6] B. Wang, Z. Li, M. Li, and W.-Y. Ma, "Large-scale duplicate detection for web image search," in Proc. IEEE Int. Conf. Multimedia Expo, Jul. 2006, pp. 353–356.

[7] F. Zou et al., "Nonnegative sparse coding induced hashing for image copy detection," Neurocomputing, vol. 105, no. 1, pp. 81–89, 2013.

[8] O. Chum, J. Philbin, M. Isard, and A. Zisserman, "Scalable near identical image and shot detection," in Proc. 6th ACM Int. Conf. Image Video Retr., 2007, pp. 549–556.

[9] S.Todorovic and N. Ahuja, "Region-based hierarchical image matching," Int. J. Comput. Vis., vol. 78, no. 1,pp. 47–66, 2008.

[10] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun./Jul. 2004, pp. II-506–II-513.

[11] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 4, pp. 509–522, Apr. 2002.

[12] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[13] K. Mikolajczyk and C. Schmid,"A performance evaluation of local descriptors," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[14] K. Mikolajczyk and C. Schmid,"Scale and affine invariant interest point detectors," Int. J. Comput. Vis.,vol. 60, no. 1, pp. 63–86, 2004.

[15] D.-Q. Zhang and S.-F. Chang,"Detecting image near-duplicate by stochastic attributed relational graph matching with learning," in Proc.ACM Int. Conf. Multimedia, 2004, pp. 877–884.

[16] Y. Ke, R. Sukthankar, and L. Huston, "Efficient near-duplicate detection and sub-image retrieval," in Proc. ACM Int. Conf. Multimedia, 2004, pp. 869–876.

[17] D. Xu, T. J. Cham, S. Yan, L. Duan, and S.-F. Chang, "Near duplicate identification with spatially aligned pyramid matching," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 8, pp. 1068–1079, Aug. 2010.

[18] W.-L. Zhao, C.-W. Ngo, H.-K. Tan, and X. Wu, "Near-duplicate keyframe identification with interest point matching and pattern learning," IEEE Trans. Multimedia, vol. 9, no. 5, pp. 1037–1048, Aug. 2007.

[19] O. Chum, J. Philbin, and A. Zisserman, "Near duplicate image detection: Min-hash and tf-idf weighting," in Proc. Brit. Mach. Vis. Conf., 2008, pp. 493–502.

[20] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2008, pp. 1–8.

[21] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in Proc. Int. Conf. Comput. Vis., Oct. 2003, pp. 1470–1477.

[22] M. Turk and A. Pentland, "Eigenfaces for recognition," J. Cognit. Neurosci., vol. 3, no. 1, pp. 71–86, 1991.

[23] D. P. Lopresti, "A comparison of text-based methods for detecting duplication in scanned document databases," Inf. Retr., vol. 4, no. 2, pp. 153–173, 2001.

[24] L. Liu, Y. Lu, and C. Y. Suen, "Near-duplicate document image matching: A graphical perspective," Pattern Recognit., vol. 47, no. 4, pp. 1653–1663, 2014.

[25] D. Doermann, H. Li, and O. Kia, "The detection of duplicates in document image databases," Image Vis. Comput., vol. 16, nos. 12–13, pp. 907–920, 1998.

[26] S. Lu, L. Li, and C. L. Tan, "Document image retrieval through word shape coding," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 11, pp. 1913–1918, Nov. 2008.

[27] G. Meng, N. Zheng, Y. Zhang, and Y. Song, "Document images retrieval based on multiple features combination," in Proc. 9th Int. Conf. Document Anal. Recognit., Sep. 2007, pp. 143–147.

[28] Li Liu, Yue Lu,Ching Y. Suen, Variable-Length Signature for Near-Duplicate Image Matching,

IEEE Transactions On Image Processing, vol. 24, no. 4, April 2015.

[29] Arun Qamra, Yan Meng, Edward Y. Chang, "Enhanced Perceptual Distance Functions and Indexing for Image Replica Recognition", IEEE Transactions on Pattern Analysis & Machine Intelligence, vol.27, no. 3, pp. 379-391, March 2005, doi:10.1109/TPAMI.2005.54.

[30] Hsiao JH, Chen CS, Chien LF, Chen MS,"A New Approach To Image Copy Detection Based On Extended Feature Sets", IEEE Transactions On Image Processing, vol. 16, no. 8, August 2007,pp.2069-79.

[31] Li-Wei Kang, Chao-Yung Hsu , Hung-Wei Chen, Chun-Shien Lu, Chih-Yang Lin, Soo-Chang Pei, "Feature-Based Sparse Representation for Image Similarity Assessment",IEEE Transactions On Multimedia, vol. 13, no. 5, October 2011.

[32] Wei Dong, Zhe Wang, Moses Charikar, Kai Li, "High-Confidence Near-Duplicate Image Detection",:In ACM International Conference On Multimedia Retrieval(2012).

[33] Jingbo Zhang, Shu Chang,Man Luo, Li Yang, Jun Kong "A Multi-viewpoint Image Matching Method Based on Invariant Descriptors", International Conference on Information Engineering and Computer Science, 2009.

[34] Yue Wang, Zujun Hou, Chang, R.,Teck Wee Chua, "Keypoint-Based Near-Duplicate Images Detection Using Affine Invariant Feature And Color Matching", 19th IEEE International Conference on Image Processing (ICIP) 2012.

[35] Manimala Singha and K. Hemachandran, "Content Based Image Retrieval using Color and Texture", Signal & Image Processing: An International Journal (SIPIJ), Vol. 3, No. 1, February 2012.

[36] W Voravuthikunchai, B Crémilleux, F Jurie "Finding groups of duplicate images in very large dataset", Proceedings of the British Machine Vision Conference (BMVC 2012), 105.1--105.12.

[37] Li Liu1, Yue Lu1, Ching Y. Suen,"Novel Global and Local Features for Near-duplicate Document Image Matching",22nd International Conference on Pattern Recognition 2014.

[38] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," in Proc. Eur. Conf. Comput. Vis., 2004, pp. 469–481.

[39] D. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Distance measures for color image retrieval," in Proc. Int. Conf. Image Process., Oct. 1998, pp. 770–774.

[40] Shiv Vitaladevuni, Fred Choi, Rohit Prasad and Premkumar Natarajan, "Detecting Near-Duplicate Document Images using Interest Point Matching", 21st International Conference on Pattern Recognition (ICPR 2012)November 11-15, 2012. Tsukuba, Japan.

◈◈◈