-International Conference on Recent Trends in Information Technology, ICRTIT 2011
MIT, Anna University, Chennai. June 3-5, 2011

Compact and Efficient CBCD Scheme Based on Integrated Color Features

Roopalakshmi.R[#1], Ram Mohana Reddy.G[*2]

[#]*National Institute of Technology (NITK), Surathkal, Mangalore, INDIA.*
[1]r_roopalakshmi@hotmail.com

[*]*National Institute of Technology (NITK), Surathkal, Mangalore, INDIA.*
[2]profgrmreddy@gmail.com

***Abstract*— Content-Based Copy Detection (CBCD) is an emergent research topic and has been extensively studied recently. Color is one of the most important and easily recognizable features of visual content. To address efficiency and effectiveness issues of CBCD systems, in this article a compact and computationally inexpensive CBCD scheme using MPEG-7 Dominant Color Descriptor is proposed. This paper proposes a novel dominant color extraction technique, which solves the intrinsic problems of existing color clustering techniques. Experimental results show that the proposed scheme improves detection accuracy up to 38%, and also supports significant reduction in total computational cost up to the extent of 91%, when compared with that of existing schemes against various video transformations.**

***Keywords*— Video copy detection, Dominant color descriptor, Spatial coherency.**

I. INTRODUCTION

Since video copy detection is essential in various copyright-based applications, the exponential growth of online video services imposes severe challenges for the owners of the popular video web servers. In general, a video copy is defined as, a transformed video sequence, that is derived from another video. There are two general approaches for detecting copies of a digital media: digital watermarking and content based video copy detection (CBCD).The primary task of CBCD is, to detect whether a given video sequence is a copy or not, by utilizing the content based features of the media. The CBCD techniques are preferred when compared to watermarking techniques, because of the following key features [1]: i) The video signature generation will neither destroy nor damage video content, ii) More robust than fragile watermarking techniques, iii) Signature extraction can also be done after the distribution of digital media and iv) Capable of detecting copies, even if the original document is not watermarked.

The primary function of any CBCD system is, when a user uploads a query video, to find out the original video from which the query is taken, after matching query and original video features. CBCD techniques can be roughly classified into Global descriptors and Local descriptors techniques. Global descriptors like Ordinal measure [2], Color histograms [3] are compact and easy to extract, but they are less robust against region based attacks. Local descriptors like SIFT [4], SURF [5], PCA-SIFT [6] etc., use local interest points for feature extraction. The local descriptors are tolerant to many types of transformations, but they are typically not efficient, because of their high computational cost.

Since color is one of the dominant and distinguishing visual features of an image, in this paper, we employed Dominant Color Descriptors (DCD) of MPEG-7 standard [7], [8].GLA(Generalized Lloyd algorithm) is the one of the most extensively used algorithm to extract the dominant colors from an image [9]. But GLA suffers because of its intrinsic problems like high computational cost and requirement of initial seeds.So, the main focus of this paper is to propose a compact feature descriptor, in order to extract dominant colors of a given video sequence with low computational cost. The main contributions of this paper are as follows:

a) A compact technique -Frequency Image [10] is used to extract dominant colors that use color LUTs. It is easy to extract and computationally inexpensive compared to the existing color clustering techniques.
b) Integrated spatial coherent feature extraction- In this, along with dominant colors and their percentage of distribution, spatial coherency is also used in CBCD task. Inclusion of this spatial coherency factor, which uniquely characterizes the color distribution in the spatial domain, significantly improves the system performance.

The rest of the paper is organized as follows: Section II introduces framework of proposed scheme along with signature extraction and matching techniques; Section III shows the experimental results of the proposed scheme, followed by conclusion in section IV.

II. PROPOSED SCHEME

Fig.1 describes the schematic framework of proposed method, in which dominant colors are extracted for each key frame of the reference video. *In general, two images may have similar dominant colors, but the spatial distribution of pixels of same color in the two images, may not be same always*. So, in addition to dominant colors, the spatial coherency/ distribution of pixels, is also extracted and used in the present DCD scheme. The extracted video fingerprints are stored in fingerprint database. Whenever user uploads a query video,

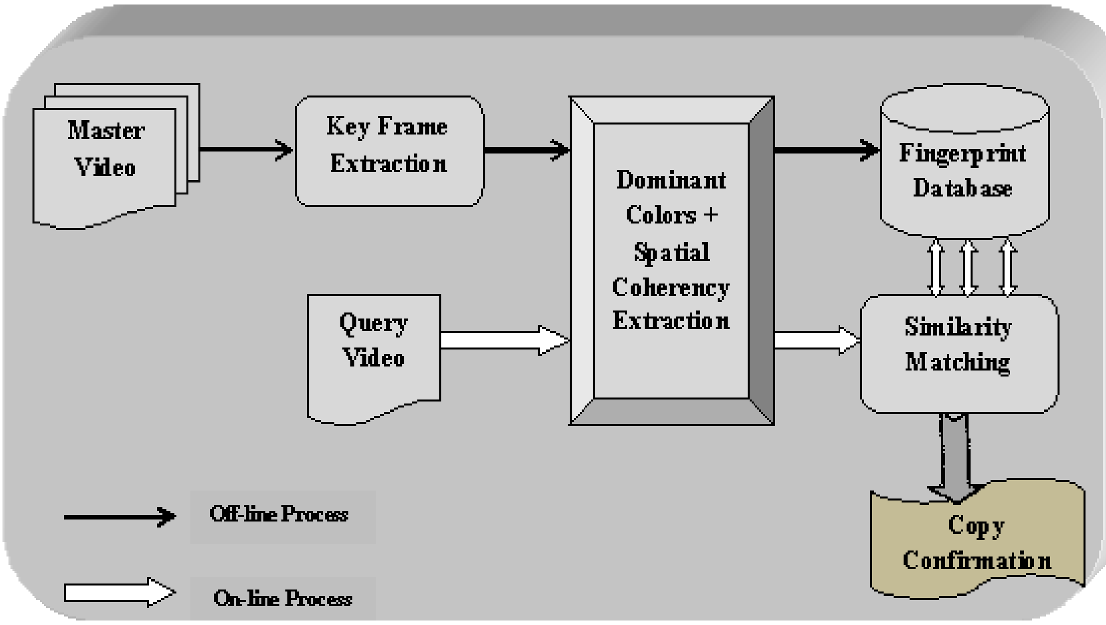

Fig. 1 Schematic Framework of Proposed Method

the respective feature descriptors are extracted, for each query video frame. Finally, the video fingerprints of query and reference videos are compared, in order to report copy confirmation.

A. Signature Extraction

The dominant color descriptor of MPEG-7 standard is defined as,

$$F = \{ \{ c_i , p_i , v_i \} ,s \} , i =1,2,3, .. N, \quad (1)$$

where (N) is the total number of dominant colors for an image, (c_i) is a 3-D dominant color vector,(p_i) is the percentage for each dominant color, and the sum of p_i is equal to 1.Two optional fields, spatial coherency (s) and color variance (v_i) provide more precise characterization of color distribution in spatial and color space domains. Spatial coherency describes the spatial distribution of pixels associated with each representative color, i.e. pixels of same color are how much co-located. Color variance explains the variation of color values of the pixels in the surroundings of a corresponding representative color. In this method, frame sampling technique is used to extract key frames of reference video, with a sampling rate of 10 frames/sec. In order to extract dominant color descriptors, RGB color space with a color look-up table of size 25 is used. Based on Euclidean distance measure, mapping of all pixels to the respective RGB component of color look-up table is performed and this information is stored as frequency image. Using this frequency image, the dominant color descriptors are extracted.

B. Signature Matching

Images can be effectively represented using few dominant colors, which varies from 3 to 8 colors. In the proposed scheme, feature descriptors are indexed based upon their RGB values. Fingerprint matching of proposed scheme involves searching the database for similar color distributions same as the input query that includes searching for each of the dominant colors separately.

If F_1 and F_2 are two dominant color descriptors given by,

$$F_1 = \{\{ c_i , p_i \},s_1 \} , \quad i =1,2,3,..,N_1, \quad (2)$$

$$F_2 = \{\{ b_j, q_j \}, s_2 \}, \quad j =1, 2,3,.., N_2, \quad (3)$$

then the distance D^2(F1,F2),between F_1 and F_2 is given by[11],

$$D^2(F1,F2) = \sum_{i=1}^{N_1} p_i^2 + \sum_{j=1}^{N_2} q_j^2 - \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} 2a_{i,j} p_i q_j \quad (4)$$

where p_i and q_j correspond to query and reference video descriptors and $a_{i,j}$ is the similarity coefficient between colors c_i and b_j. The similarity coefficient $a_{i,j}$ is given by,

if $d_{i,j} <= T_d$,
then $a_{i,j} = 1 - d_{i,j} / d_{max}$, (5)
else
$a_{i,j} = 0.$ (6)

TABLE I. COPY DETECTION RESULTS FOR DIFFERENT IMAGE TRANSFORMATIONS

Transforms	Baseline DCD Method (%)		Cho's Method (%)		Proposed DCD Method (%)	
	Precision	Recall	Precision	Recall	Precision	Recall
1) Blurring	98.1	83.2	90.1	78.8	100	95.6
2) Brightness	96.4	81.4	92.3	77.1	100	98.7
3) Noise Addition	90.5	73.3	83.5	65.3	100	92.5
4) Zooming Out	93.7	69.4	84.1	59.2	99.2	84.4
5) Image Ratio	88.6	61.9	66.6	60.8	98.1	79.6
6) Zooming In	91.3	57.8	59.7	52.4	100	63.7
7) Image Resize	92.4	58.1	73.6	51.9	100	69.3
8) Rotation	94.4	64.4	69.5	61.5	98.6	78.4

where distance $d_{i,j}$ is given by,

$$d_{i,j} = \| c_i - b_j \| \quad (6)$$

and $d_{max} = \alpha * T_d$,where α is set as 1.2 in our experiments. The threshold T_d is the maximum distance used to judge whether two color features are similar or not. Finally, the similarity between F_1 and F_2 can be described by the distance function D_{DC} [11],

$$D_{DC} = w_a |s_1 - s_2| D^2 (F_1, F_2) + w_b D^2 (F_1, F_2) \quad (7)$$

where s_1 and s_2 are spatial coherencies of query & reference videos, and w_a ,w_b are fixed weights, set as 0.3 and 0.7 respectively.

III. EXPERIMENTAL RESULTS

We used a video database of 101 video sequences, collected from Open Video Project [12], in order to evaluate the performance of our approach. The video database contains approximately 305297 frames. The video content includes news, documents, Education & life, movies, natural scenes, landscapes etc. The format of the original video data used is MPEG-1 with 352*240 pixels and 30 fps.

The detection result is treated as correct, if there is any overlap with the region, from which the query was extracted. We designed two sets of experiments to evaluate the detection accuracy and detection efficiency of proposed approach. From the video database, we randomly selected 15 videos, ranging from 5 to 8 seconds. Different kinds of transformations given by, 1) Blurring, 2)Brightness change, 3) Noise addition ,4) Zooming-out, 5)Image ratio ,6)Zooming-in ,7)Image resize and 8) Rotation are applied to those 15 videos to generate 120 video copies. Then, those 15 videos are used as query videos to search the database. To evaluate the efficiency, the computational cost of the single video copy detection is discussed.

A. Detection Accuracy

To measure the detection accuracy of proposed scheme, we used standard Precision and Recall metrics, which are given by,

$$\text{Precision} = TP / (TP + FP) \quad (8)$$

$$\text{Recall} = TP / (TP + FN) \quad (9)$$

True Positives (TP) are positive examples correctly labelled as positives. False Positives (FP) refer to negative examples incorrectly labelled as positives. False Negatives (FN) refer to positive examples incorrectly labelled as negatives.

Table I shows the comparison of precision & recall results of proposed DCD method, baseline DCD method and Cho's method of algorithm [13]. In the baseline DCD method, video fingerprint includes only the dominant colors & their percentage distribution values. The Table I results show that proposed DCD method provides better precision, recall values and thereby improves detection accuracy up to 24% compared to baseline DCD method. The results also prove that proposed DCD method produces better detection results and significantly increases detection accuracy upto 38.1% compared to Cho's method.

B. Detection Efficiency

In order to evaluate the efficiency of our approach, we have compared the computational cost of our approach with that of baseline DCD method and Kim's method [14]. Fig. 2 gives the computational cost details of proposed and reference methods. The experiments are conducted on a standard PC with 3.2 GHz CPU and 2 GB RAM. The results from Fig.2

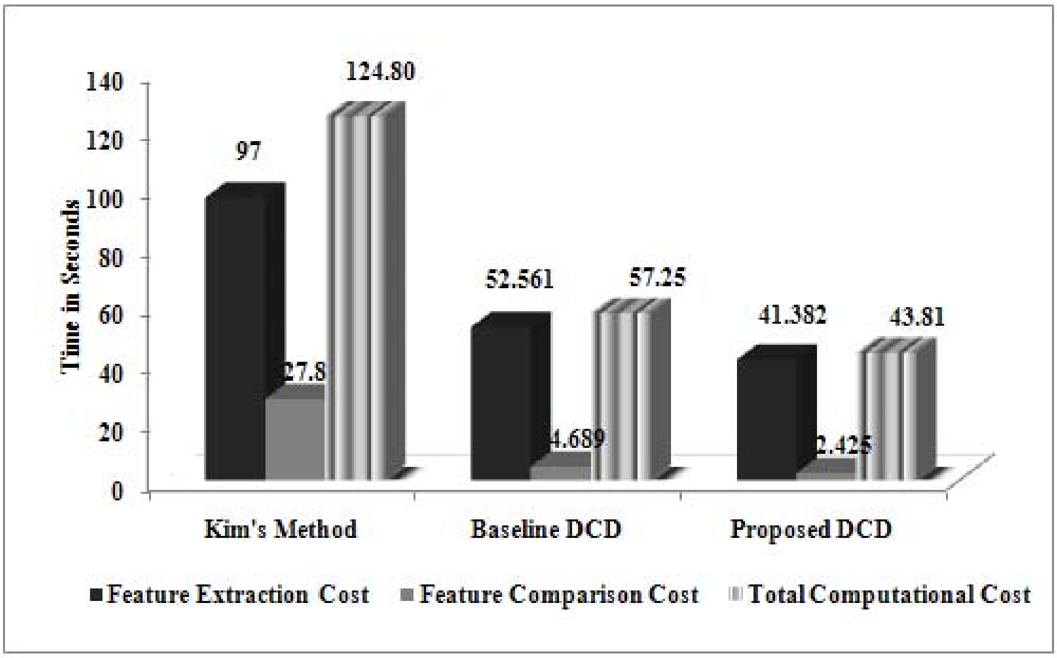

Fig. 2 Comparison of Computational Cost

show that, proposed DCD method is more efficient, when compared with baseline DCD method, by reducing total computational cost up to 26%. Fig.2 results also prove that proposed DCD approach is 3 times faster that of Kim's approach, which results in the reduction of total computational cost up to 91%.

IV. CONCLUSION

In this paper, we presented a simple and efficient video signature method using Dominant Color Descriptor of MPEG -7 standard. Experimental results show that our approach provides good performance in terms of detection accuracy rates and also reduces the computational cost, when compared with the existing approaches.

Further, our future work will be targeted at the incorporation of audio fingerprints to the existing approach, to improve overall system performance.

ACKNOWLEDGEMENT

We would like to thank the anonymous reviewers for their valuable comments and suggestions, which enhanced the quality of this research article.

REFERENCES

[1] R.Roopalakshmi and G.Ram Mohana Reddy, "Recent Trends in Content –Based Video Copy Detection" , i*n proc. of IEEE ICCIC ,Dec 2010*, [Online] Available: http://www.ieee.org/

[2] D.Bhat and S.Nayar, "Ordinal measures for image correspondence", *in proc. of IEEE Transactions on Pattern Analysis and Machine Intelligence*, 415– 423,1998.

[3] H. T. Shen, X. Zhou, Z. Huang, J. Shao, and X. Zhou.,"UQLIPS: A real-time near-duplicate video clip detection system", *in proc. of VLDB*, 2007.

[4] David G.Lowe ," Distinctive image features from scale-invariant key points," *in proc. of International Journal of Computer Vision*, pp. 91-110, 2004.

[5] Herbert Bay, Tinne Tuytelaars, Luc Van Gool, "SURF: Speeded Up Robust Features," *in proc. of Computer Vision and Image Understanding*, 346–359, 2008.

[6] Y.Ke and R. Sukthankar, "PCA–SIFT: A more distinctive representation for local image descriptors," i*n Proc. of Computer Vision and Pattern Recognition* (CVPR), pp. 506–513, 2004.

[7] B.S.Manjunath, P.Salembier, and T. Sikora, Introduction *to MPEG-7 - Multimedia Content Description Interface*, John Wiley and Sons, 2002.

[8] ISO/IEC/JTC1/SC29/WG11 MPEG 2001/N4358, *Text of ISO/IEC 15938-3/FDIS Information technology – Multimedia content description interface - Part 3 Visual*, Sydney, Australia, 2001.

[9] S. P. Lloyd. , "Least Squares Quantization in PCM," i*n proc. of IEEE Transactions on Information Theory*, Vol. 28, pp. 129-137, 1982.

[10] Toshiyuki Kashiwagi and Shunichiro Oe. 2007."Introduction of Frequency Image and applications**,"***SICE Annual Conference 2007*, Kagawa University, Japan.

[11] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and H. Shin, "An efficient color representation for image retrieval", *in proc. of IEEE Transactions on Image Processing*, Vol. 10, pp. 140–147,2001.

[12] http://www.open-video.org.

[13] Hye-Jeong Cho, Yeo-Song Lee, Chae-Bong Sohn, Kwang-Sue Chung, and Seoung, "A novel video copy detection method based on statistical analysis," *in proc. of International conference on Multimedia & Expo* ,2009.

[14] Joosub Kim and Jeho Nam, "Content-based Video Copy Detection using Spatio-Temporal Compact Feature," i*n proc. of International Conference on Advanced Communication Technology ICACT -2009*.

[15] J. Law-To, V. Gouet-Brunet, O. Buisson, and N. Boujemaa, "Video copy detection on the internet: The challenges of copyright and multiplicity," in proc. of *IEEE International Conference on Multimedia and Expo (ICME 2007)*, pp. 2082–2085, 2007.