

Content based Video Retrieval Systems Performance based on Multiple Features and Multiple Frames using SVM

Mohd.Aasif Ansari, Hemlata Vasishtha
Engineering & Technology,
Shri Venkateshwara University,
Gajraula, India

Abstract—In this paper, Content Based Video Retrieval Systems performance is analysed and compared for three different types of feature vectors. These types of features are generated using three different algorithms; Block Truncation Coding (BTC) extended for colors, Kekre's Fast Codebook Generation (KFCG) algorithm and Gabor filters. The feature vectors are extracted from multiple frames instead of using only key frames or all frames from the videos. The performance of each type of feature is analysed by comparing the results obtained by two different techniques; Euclidean Distance and Support Vector Machine (SVM). Although a significant number of researchers have expressed dissatisfaction to use image as a query for video retrieval systems, the techniques and features used here provide enhanced and higher retrieval results while using images from the videos. Apart from higher efficiency, complexity has also been reduced as it is not required to find key frames for all the shots. The system is evaluated using a database of 1000 videos consisting of 20 different categories. Performance achieved using BTC features calculated from color components is compared with that achieved using Gabor features and with KFCG features. These performances are compared again with the performances obtained from systems using SVM and the systems without using SVM.

Keywords—CBVR; KFCG; Multiple Frames; SVM; BTC; Gabor filter

I. LITERATURE REVIEW AND RELATED WORK

Researchers have developed a number of techniques, methods and systems in the field of content based video retrieval systems. They are required to effectively search, index and retrieve videos from databases but the reliable and effective systems are still awaited for huge databases [6]. For this reason, text based searches are still in practice for the video retrieval systems [5]. A content based retrieval system was developed for commercial use [15]. Face detection method was used for image and video searches in this system. But this method also proved to be very poorly performing [8] by the automatic systems participated in the video retrieval track [16]. A hope emerged when low level features were utilized. Comparison of low level features extracted from key frames of the query and the videos from database provide better results for video retrieval systems [6]. Other useful and much more important information from videos can bring performance of the video retrieval systems to a great level of success. Researchers still face a challenge to utilize important

information such as sequence of shots, temporal and motion information [5]. To compensate this problem and to get better retrieval performance, a video retrieval system [2] utilized all frames of a shot instead of only the key frames so that more visual features are extracted. Another system [12] integrated color and motion features for better utilization of spatio-temporal information but a fact is still relevant that an efficient image retrieval technique results in an efficient video retrieval technique where image from the query video is used as a query [8]. The system proposed here utilizes visual features from multiple frames instead of a single frame, key frames of the shot or all of its frames. The proposed system provides the much required solutions to the problems mentioned above which are, lower efficiency when only a single image is used, high computational cost when key frames are used and unavailability of proper tools for clustering algorithm. This system provides reasonable efficiency along with low computation cost.

In section II, features extraction algorithms and classification are discussed; section III discusses about similarity measure; section IV shows the methodology to calculate result parameters in the proposed CBVR system, while the proposed CBVR system is elaborated in section V. Result analysis is presented in section VI; problems and challenges faced by the CBVR system are discussed in section VII and it is concluded by section VIII.

II. FEATURES EXTRACTION AND CLASSIFICATION

Color, texture and motion features are the most useful features for classification and retrieval of videos. Color histogram proves to be useful to represent color content while extraction of Gabor features is a popular way to represent texture features [4].

A. Extraction of BTC Features

Block truncation coding (BTC) is basically a compression technique for images [14]. BTC features are calculated for small blocks formed by dividing an image instead of calculating for each pixel [17], [18]. BTC is used to obtain features from color information of pixels belonging to the small blocks. BTC features from multiple frames are employed to obtain very high precision and recall values. These features can also be used for image classification and retrieval purpose. The BTC technique can be extended to RGB

images by considering each color component (red, green and blue) as a separate plane [14]. BTC features are obtained as shown in the equations (1-5).

- An inter band average image (IBAI) is formed as shown in (1)

$$I_{avg}(x, y) = I_R(x, y) + I_G(x, y) + I_B(x, y) \quad (1)$$

- Threshold values for the three color components are calculated as shown in (2) for one of the components (red).

$$R_{threshold} = \frac{\sum_{r=0}^x \sum_{c=0}^y I_R(r, c)}{x * y} \quad (2)$$

- Binary bitmaps are created for each of the three components as shown in (3) for the red component

$$R_{bm}(r, c) = \begin{cases} 1, & I_R(r, c) \geq R_{threshold} \\ 0, & otherwise \end{cases} \quad (3)$$

- m_1 and m_2 are the mean values found for the three components as shown in (4) and (5) for the red components.

$$m_{R1} = \frac{\sum_{r=0}^x \sum_{c=0}^y R_{bm}(r, c) I_R(r, c)}{\sum_{r=0}^x \sum_{c=0}^y R_{bm}(r, c)} \quad (4)$$

$$m_{R2} = \frac{\sum_{r=0}^x \sum_{c=0}^y (1 - R_{bm}(r, c)) I_R(r, c)}{x * y - \sum_{r=0}^x \sum_{c=0}^y R_{bm}(r, c)} \quad (5)$$

where, $m_1 = \{m_{R1}, m_{G1}, m_{B1}\}$ and $m_2 = \{m_{R2}, m_{G2}, m_{B2}\}$

m_1 and m_2 represent the entire block. Mean values of all the blocks considered together represent the entire image.

B. Extraction of Gabor Features

Gabor features provide good representation of edge and texture features for objects and texts and help to distinguish them effectively from the background [7]. Gabor filters are capable of extracting features from edges or regions of different objects inside an image directed towards desired orientations with different frequencies [22].

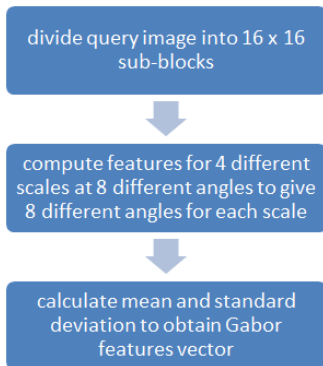


Fig. 1. Gabor Filter Algorithm

Method to extract Gabor features is shown in Fig. 1 while the mathematical expressions are given from equation 6 to equation 11 [20].

For a given image, discrete Gabor wavelet transform is given by a convolution using equation (6) for an image $I(r, c)$ where, $r = 0, 1, 2, \dots, R$ and $c = 0, 1, 2, \dots, C$.

$$W_{uv} = \sum_p \sum_q I(r - p, c - q) G_{uv}^*(p, q) \quad (6)$$

Where, G_{uv}^* is complex conjugate of G_{uv} . G_{uv} is generated by some morphological operations on mother wavelet. $p * q$ is the size of filter mask, u and v are scale and orientations.

Gabor filters are applied on the image with different orientations and different scales to find a set of magnitudes $E(u, v)$ containing the energy distribution in the image in different orientations and scales as shown in (7).

$$E(u, v) = \sum_r \sum_c |W_{uv}(r, c)| \quad (7)$$

To obtain texture features Standard deviation σ and mean are required and calculated as shown in equations (8) and (9) respectively

$$\text{Standard Deviation, } \sigma_{uv} = \sqrt{\frac{\sum_r \sum_c (|W_{uv}(r, c)| - \mu_{uv})^2}{R * C}} \quad (8)$$

$$\text{Mean, } \mu_{uv} = \frac{E(u, v)}{R * C} \quad (9)$$

Texture features vector F is formed by a set of feature components [25], [26] i.e., different values of σ_{uv} and μ_{uv} calculated by varying u and v as shown in equation (10).

$$f = [\sigma_{u_0v_0}, \sigma_{u_1v_1} \dots \sigma_{u_{UV}v_U}] \quad (10)$$

$$f_{Gabor} = \frac{f - \mu}{\sigma} \quad (11)$$

C. Extraction of KFCG Features

Compression is achieved in vector quantization by using some bits to represent a closest codeword for small blocks formed by dividing an entire image [27]. Linde-Buzo-Gray (LBG) is most commonly used algorithm to generate codebook [28]. In LBG algorithm, vectors found in the blocks are training vectors which are separated to form different clusters. They are divided again and again by process of iteration. Codebook vectors are centroid of these clusters [29]. A training vector is represented by codebook vector closest to it [30]. Codebook vectors are represented by a set of codewords which are used to encode and decode the images [31]. Kekre's Fast Codebook Generation (KFCG) Algorithm is basically used for image compression [32] [23]. It requires less time to generate the codebook through vector quantization method. The codebook generated is used in the proposed system as a feature vector for video retrieval purpose [20].

D. Classification of Features using Support Vector Machine

Support Vector Machine (SVM) improves performance of content based image retrieval (CBIR) significantly [11]. It is the inspiration to use SVM for CBVR too. SVM can utilize the features representing a video similarly it does for CBIR. Here, the feature vector can be the features extracted from frames, shots, scenes or events. Features from known categories of videos are labeled to train the svm. Similar features extracted from other videos are used by SVM for classification of videos. Use of SVM is a milestone in automatic classification of videos [19] with better efficiency.

III. SIMILARITY MEASURE

Features extracted from the images provide most convenient method for similarity measurement [1]. The query video is retrieved by finding similarity between its feature

vector [9], [10] and feature vector of the videos stored in database. Video similarity is measured at different resolution [13]. So the selection of features becomes relevant for calculating similarity. Similar videos can also be obtained by using SVM. The videos classified by SVM to form one category show greater similarity among them. The most similar video can be obtained by finding euclidean distance between the query video and the videos classified to form that category. Again, the feature vector is used to calculate the euclidean distance.

The equation for Euclidean distance between a query frame q and a database frame d is shown in (12)

$$\text{Euclidean Distance} = \sqrt{\sum_{n=1}^N (V_{dn} - V_{qn}) \cdot (V_{dn} - V_{qn})} \quad (12)$$

Where V_{dn} are the feature vectors of database frame d and V_{qn} are the feature vectors of query frame q each having size N [20].

IV. RESULT EVALUATION METHOD

The performance of video retrieval is evaluated with the same parameters as it is evaluated in image retrieval [11]. Recall and precision are the two parameters [2] as given in (13) and (14).

$$\text{Recall} = \frac{DC}{DB} \quad (13)$$

$$\text{Precision} = \frac{DC}{DT} \quad (14)$$

DC = number of similar clips detected correctly

DB = number of similar clips in the database

DT = total number of detected clips

V. PROPOSED CBVR SYSTEM

A CBVR system is proposed in this paper in which multiple frames are obtained for the query videos and the videos' database instead of using single frame or key frames or all frames [2]. BTC, Gabor and KFCG features are obtained as mentioned above in features extraction section. The similar and most relevant videos are obtained from the output directory containing videos of that category. Significantly higher results have been obtained using this system. A typical methodology is used in this system where a video is retrieved from its category. Here, database is processed offline. The videos are represented by feature vectors formed from any one or a combination of more than one from three types of features extracted from their multiple frames. Feature vectors are then labelled and stored in the features database. An SVM is trained for the categories registered in the system using labelled feature vectors stored in the database. Variables are obtained from the trained SVM. Feature vectors from the query videos are used for classification using SVM variables already saved. Videos obtained in the output folder are the videos of the desired category. For a query clip, videos stored in the given category can be ranked according to the distance measures and most similar videos are retrieved. Euclidean distance is used to measure similarity [20]. Retrieval system without using SVM is shown in Fig. 2 [20]. Most similar

videos are obtained based on minimum distance between feature vectors stored in the database and feature vectors of the query image. As mentioned above, multiple frames based classification and retrieval yields acceptable results without the complexity of finding key frames to represent a shot.

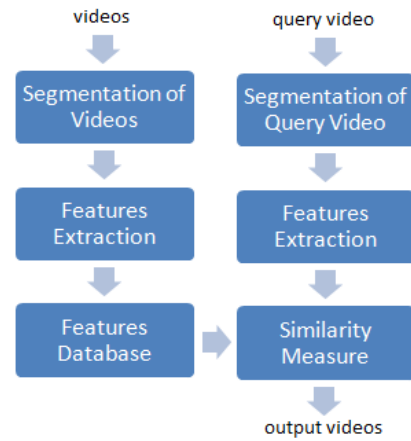


Fig. 2. Proposed CBVR system

A process flow of the CBVR system using SVM is shown in Fig. 3. Multiple frames are obtained during segmentation. Features are then extracted for each of those frames and stored in feature vectors database. Feature vectors are labelled for the pre-decided categories.

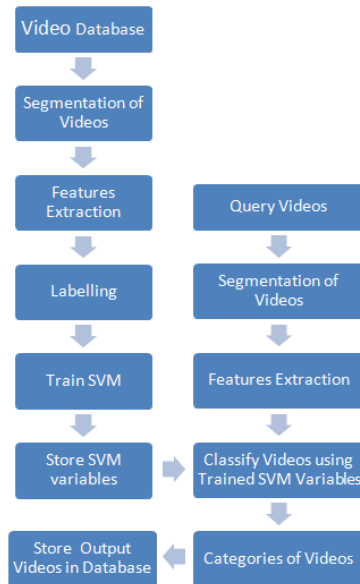


Fig. 3. Proposed CBVR system

SVM is trained and its variables are stored. This process is done offline. The query videos are separated into the categories based on stored SVM variables using feature vectors of the query videos. Videos obtained for different categories are stored with different categories in the output database. The query video can be retrieved by exact similarity matching from the classified videos using Euclidean distance method.

VI. RESULTS

A. Database

The technique using multiple frames with one or multiple features using SVM is applied to a video database having 1000 videos with 20 categories of 50 videos each as shown in Fig. 4. Videos similar to the query video are stored in output folder after classification using SVM classifier. The precision and recall values are computed by grouping the number of classified videos belonging to the category of query video and then finding minimum distance between them and the query video.

B. Analysis of Results

The charts shown from Fig. 5 to Fig. 10 for different features represent the retrieval results obtained for retrieving and classification of video clips from different categories. These categories are among the 20 categories of video clips from the video database of 1000 videos. The results obtained are highly appreciable for all the categories. The results are obtained using SVM based on Gabor features extracted from multiple frames of the video clips. Similarly, results are also obtained using block truncation coding method extended for color images [24] and KFCG algorithm. The charts compare the performance obtained by the system using SVM with the performance obtained from system based on same features without using SVM [21]. Comparison of systems performance is also done using three different features without using SVM and while using SVM.



Fig. 4. Sample video database of 1000 videos with 20 categories

1) Results for video clips using Gabor features

Fig. 5 shows results (precision values) obtained by CBVR system based on Gabor features extracted from multiple frames using SVM. There is a significant improvement in results using SVM as compared to results obtained without using SVM except for one case.

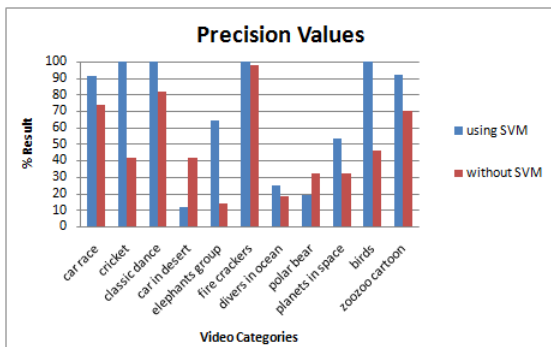


Fig. 5. Comparison of Precision values shown for given categories of videos using SVM and without using SVM using Gabor features

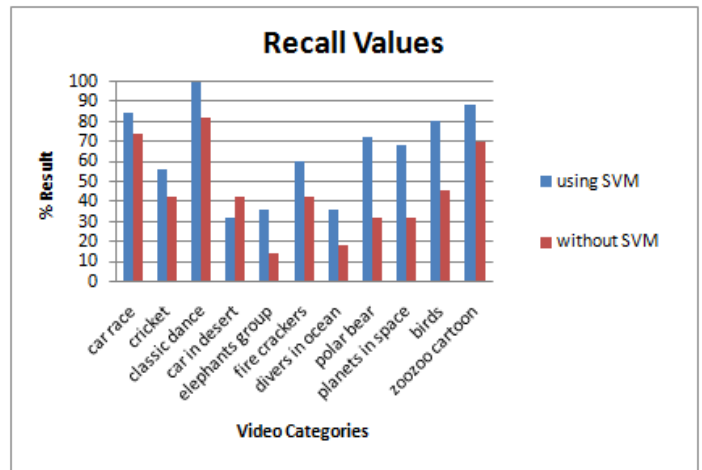


Fig. 6. Comparison of Recall values shown for given categories of videos using SVM and without using SVM using Gabor features

Fig. 6 shows results (recall values) obtained by CBVR system based on Gabor features extracted from multiple frames using SVM. There is significant improvement in results using SVM as compared to results obtained without using SVM except for one case.

2) Results for video clips using KFCG features

Fig. 7 shows results (precision values) obtained by CBVR system based on KFCG features extracted from multiple frames using SVM. There is significant improvement in results using SVM as compared to results obtained without using SVM except for one case.

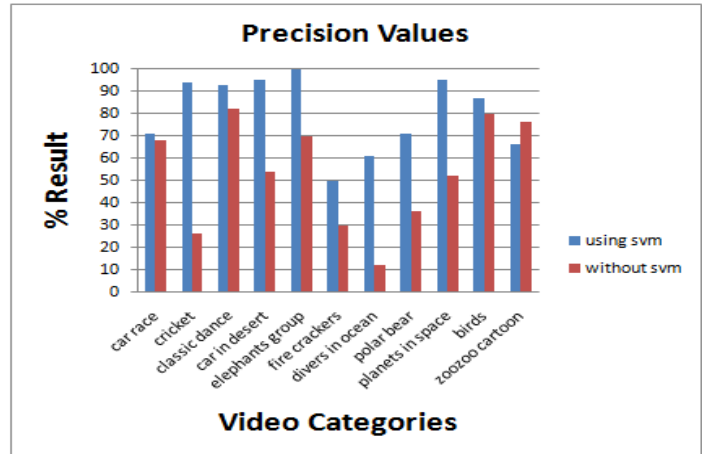


Fig. 7. Comparison of Precision values shown for given categories of videos using SVM and without using SVM using KFCG features

Fig. 8 shows results (recall values) obtained by CBVR system based on KFCG features extracted from multiple frames using SVM. There is significant improvement in results using SVM as compared to results obtained without using SVM except for one case.

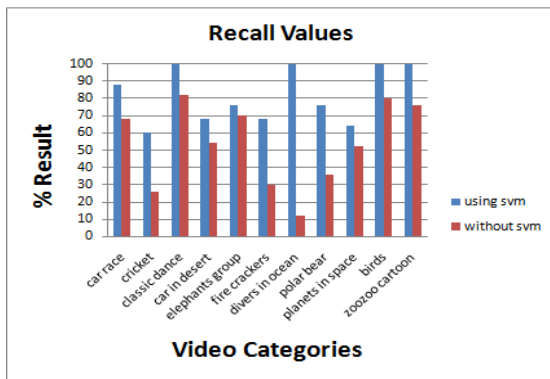


Fig. 8. Comparison of Recall values shown for given categories of videos using SVM and without using SVM using KFCG features

3) Results for video clips using BTC features

Fig. 9 shows results (precision values) obtained by CBVR system based on BTC features extracted from multiple frames using SVM. There is significant improvement in results using SVM as compared to results obtained without using SVM.

Fig. 10 shows results (recall values) obtained by CBVR system based on BTC features extracted from multiple frames using SVM. There is significant improvement in results using SVM as compared to results obtained without using SVM.

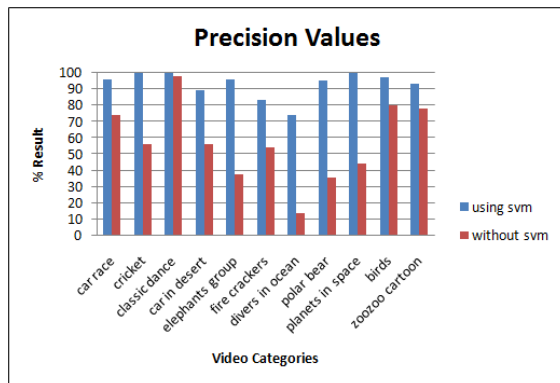


Fig. 9. Comparison of Precision values shown for given categories of videos using SVM and without using SVM using BTC features

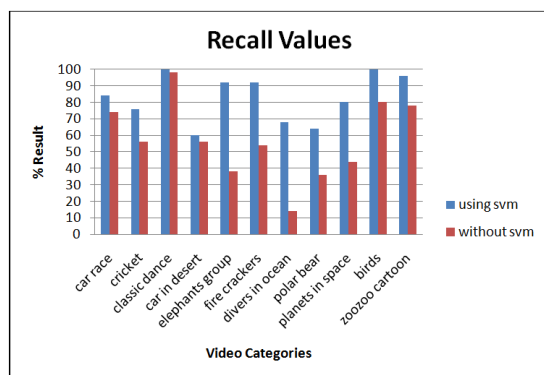


Fig. 10. Comparison of Recall values shown for given categories of videos using SVM and without using SVM using BTC features

VII. PROBLEMS AND CHALLENGES

Low level features representing the frames are used in implementation of CBVR systems using query by image or query by clips like the one shown in the proposed system. These low level features extracted from frames are used to measure similarity between different videos. Due to this, different types of videos containing distinct objects but with similar backgrounds may produce false retrievals. For example, videos showing players playing football may be retrieved along with videos showing players playing cricket due to similar background of the field or a video showing a person delivering speech may be retrieved with videos showing a different person delivering speech with a similar background of the stage. Low level features are utilised for content based image retrieval when query is done by example image. Performance and efficiency of such systems searching video is quite acceptable when non-identical features are present in them but the performance is very poor when low level features belonging to different videos are identical. This is due to the fact that they are unable to utilise the semantic features e.g., different videos having different electronic equipments but with distinct low level features.

VIII. CONCLUSION

The proposed system shows better classification and enhanced video retrieval results. The higher efficiency has become possible due to utilization of distinct features representing distribution of color information (BTC method), inclination of edges in multiple directions (Gabor algorithm), codewords representing blocks (KFCG algorithm). Though, the result is appreciable for all the three types of features but the Precision and Recall values are much higher for BTC and KFCG features as compared to Gabor features. The performance is boosted further due to use of features from multiple frames instead of using single key frame representing a shot. Additional improvement is achieved by the use of SVM which makes the system highly efficient.

IX. FUTURE SCOPE

Computational cost of the proposed system is better as there is no requirement to find the key frames for each shot. Though we have enhanced classification and retrieval of videos, an attention and focus is required to eliminate the drawback of producing false result when videos have similar backgrounds. Another scope of future research is the recognition and grouping of videos belonging to same category but having different low level features.

REFERENCES

- [1] Weiming Hu, Nianhua Xie, Li Li, Xianglin Zeng, Maybank S., "A Survey on Visual Content-Based Video Indexing and Retrieval", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41-6,797-819, 11/2011
- [2] Liang-Hua Chen, Kuo-Hao Chin, Hong-Yuan Liao, "An integrated approach to video retrieval", Proceedings of the nineteenth conference on Australasian database-Volume 75, 49-55, 2008
- [3] Dengsheng Zhang, Aylwin Wong, Maria Indrawan, Guojun Lu, "Content-based Image Retrieval Using Gabor Texture Features", IEEE Transactions PAMI, pages 13-15, vol. 12.
- [4] Yining Deng, B.S. Manjunath, "Content-based Search of Video Using Color, Texture, and Motion", IEEE, pg 534-537, 1997

- [5] Ja-Hwung Su, Yu-Ting Huang, Hsin-Ho Yeh, Vincent S. Tseng, "Expert Systems with Applications", 37, pg 5068-5085, 2010
- [6] Nicu Sebe, Michael S. Lew, Arnold W.M. Smeulders, "Video retrieval and summarization", Computer Vision and Image Understanding, vol. 92, no. 2-3, pg 141-146, 2003
- [7] C. V. Jawahar, Balakrishna Chennupati, Balamanoahar Paluri, Nataraj Jammalamadaka, "Video Retrieval Based on Textual Queries", Proceedings of the Thirteenth International Conference on Advanced Computing and Communications, Coimbatore, Citeseer, 2005
- [8] Alexander G. Hauptmann, Rong Jin, and Tobun D. Ng, "Video Retrieval using Speech and Image Information", Electronic Imaging Conference (EI'03), Storage Retrieval for Multimedia Databases, Santa Clara, CA, January 20-24, 2003.
- [9] Swain M.J. and Ballard, B.H. "Color Indexing," Int'l J. Computer Vision, vol. 7, no. 1, pp. 11-32, 1991.
- [10] Hafner, J. Sawhney, H.S. Equitz, W. Flickner, M. and Niblack, W. "Efficient Color Histogram Indexing for Quadratic Form Distance," IEEE Trans. Pattern Analysis and Machine Intelligence, 17(7), pp. 729-736, July, 1995.
- [11] Aljahdali, S., Ansari, A., Hundewale, N., "Classification of Image Database Using SVM With Gabor Magnitude", International Conference on Multimedia Computing and Systems (ICMCS), Tangier, Morocco, 2012, vol., no., pp.126,132, 10-12 May 2012
- [12] Mohd. Aasif Ansari, Hemlata Vasishtha, "CBVR and Classification of Video Database-Latest Trends, Methods, Effective Techniques, Problems and Challenges", International Journal of Computer Applications (ISSN : 0975 - 8887), Volume 125 - No. 6, September 2015
- [13] R. Lienhart, "A System For Effortless Content Annotation To Unfold The Semantics In Videos," in Proc. IEEE Workshop Content-Based Access Image Video Libraries, pp. 45-49, Jun. 2000.
- [14] Guoping Qiu, "Colour Image Indexing Using BTC", IEEE Transaction on Image Processing, Vol. 12, January 2003.
- [15] Visionics Corporate Web Site, FaceIt Developer Kit Software, <http://www.visionics.com>, 2002.
- [16] The TREC Video Retrieval Track Home Page, <http://www.nlpir.nist.gov/projects/trecvid/>
- [17] Edward J. Delp, Martha Saenz, and Paul Salama, "BLOCK TRUNCATION CODING (BTC)", Handbook of Image and Video Processing: BTC, Video and Image Processing Laboratory, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana.
- [18] Delp, E.J., "Image Compression Using Block Truncation Coding", IEEE Transactions on Communication, ISSN : 0090-6778, Volume:27, Issue:9, pg 1335-1342, Sep 1979,.
- [19] Markos Zampoglou, Theophilos Papadimitriou, IEEE Member, and Konstantinos I. Diamantaras, IEEE Member, "Support Vector Machines Content-Based Video Retrieval based solely on Motion Information", IEEE, ISSN : 1551-2541, Print ISBN: 978-1-4244-1566-3, 2007.
- [20] Mohd. Aasif Ansari, Hemlata Vasishtha, "Enhanced Video Retrieval and Classification of Video Database Using Multiple Frames Based on Texture Information", International Journal of Computer Science and Information Technologies (ISSN : 0975 - 9646), Volume 6 (2), pg 1740 - 1745, April 2015
- [21] Mohd. Aasif Ansari, Hemlata Vasishtha, "Enhanced Video Retrieval and Classification of Video Database Using Multiple Frames Based on MBTC features", International Journal of Advances in Electronics and Computer Science (ISSN : 2393 - 2835), Volume 2 (12), pg 32 - 39, Dec 2015
- [22] Arti Khaparde, B. L. Deekshatulu, M. Madhavilath, Zakira Farheen, Sandhya Kumari V, "Content Based Image Retrieval Using Independent Component Analysis", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.4, April 2000.
- [23] H. B. Kekre, Tanuja K. Sarode, "Fast Codebook Generation Algorithm for Color Images using Vector Quantization," International Journal of Computer Science and Information Technology, Vol. 1, No. 1, pp.: 7-12, Jan 2009.
- [24] Mohd. Aasif Ansari, Hemlata Vasishtha, "CBVR and Classification of Videos based on BTC Color Features using SVM", Proceedings of the International Conference on Emerging Research Trends in Applied Engineering and Technology, KGCE, Karjat, ISBN: 978-93-5258-409-3, 1-2 April, 2016.
- [25] Flickner M. et al, "Query by image and video content: the QBIC system", IEEE Computer 1995, Volume 28, Number 9, pp 23-32, 1995.
- [26] C. Faloutsos, R. Barber, M. Flicker, J. Hafner, W. Niblack, D. Petkovic and W. Equitz, "Efficient and effective querying by image content", J. Intell. Inf. Systems 3, 231-262, 1994.
- [27] Mukesh Mittal, Ruchika Lamba, "Image Compression Using Vector Quantization Algorithms: A Review", International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277 128X), Volume 3, Issue 6, June 2013.
- [28] Manoj Kumar, Poonam Saini, "Image Compression With Efficient Codebook Initialization Using LBG algorithm", International Journal of Advanced Computer Technology (IJACT), ISSN:2319-7900, vol. 2, no. 2
- [29] Jari Kangas, "Increasing the Error Tolerance in Transmission of Vector Quantized Images by Self-Organizing Map", Helsinki University of Technology, Finland.
- [30] Yu-chen, Hu, Bing-Hwang, Su Chih-Chiang, Tsou. "Fast VQ codebook search algorithm for grayscale image coding", Science Direct Image and vision computing 26(2008) 657-666.
- [31] Asmita A. Bardekar, Mr. P.A. Tijare, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (6), 2011, 2584-2589. ISSN:0975-9646
- [32] H. B. Kekre, Tanuja K. Sarode, "New Fast Improved Codebook Generation Algorithm for Color Images using Vector Quantization," International Journal of Engineering and Technology, vol.1, No.1, pp.: 67-77, September 2008.