International Journal of Computer Applications Technology and Research
Volume 4– Issue 8, 608 - 613, 2015, ISSN: 2319–8656

Utilization of Support Vector Machine for Efficient CBVR and Classification of Video Database using Gabor Features from Multiple Frames

Mohd. Aasif Ansari
Engineering and Technology
Shri Venkateshwara University
Gajraula, India

Hemlata Vasishtha
Shri Venkateshwara University
Gajraula, India

Abstract: Content Based Video Retrieval (CBVR) systems are used for retrieval of desired videos from a large collection on the basis of features extracted from videos. The extracted features are used to index, classify and retrieve desired and relevant videos while filtering out undesired ones. Videos can be represented by their audio, texts, faces and objects in their frames. An individual video possesses unique motion features, color histograms, motion histograms, text features, audio features, features extracted from faces and objects existing in its frames. Videos containing useful information and occupying significant space in the databases are under-utilized unless exist CBVR systems capable of retrieving desired videos by sharply selecting relevant while filtering out undesired videos. Results have shown performance improvement when features suitable to particular types of videos are utilized wisely. Various combinations of these features can also be used to achieve desired performance. Many researchers have an opinion that result is poor when images are used as a query for video retrieval. Here, instead of using a single image or key frames, multiple frames of the video clip being searched are used. Also, instead of using Euclidean Distance to measure similarity Support Vector Machine (SVM) is used. This method used for CBVR system shown in this paper yields an enhanced and higher retrieval results. Also, multiple frames based classification and retrieval yields significantly higher results without the complexity of finding key frames to represent a shot. The system is implemented using MATLAB. Performance of the system is assessed using a database containing 1000 video clips of 20 different categories with each category having 50 clips. The performance is tested using features extracted using Gabor filters as these are most frequently used to represent texture features.

Keywords: CBVR; Multiple Frames; Gabor; SVM; MATLAB

1. INTRODUCTION

With lack of satisfaction from textual based video retrieval, the idea of content based video retrieval has been the attention for researchers since long time. In the beginning of content based video retrieval, they tried to retrieve videos using an image. However, video retrieval using query by image is not successful as it cannot represent a video. A video is a sequence of images and audio. A query video provides rich content information than that provided by a query image. Finding the relevant video by sequentially comparing the low level visual features of key frames of the query video with those of key frames of videos in database provide long pending solution to yield better result [6] of video retrieval. Finding similarity measure requires key frames matching and hence computing key frame features including color histogram, texture and edge features, etc., to calculate distance parameter. These huge computations cause long response time to the users and thus, the problem of high computation cost in computing visual features of videos is persistent. Apart from this, considerations for motion features, temporal, sequence and duration of shots in a video pose a challenge for the research area [5]. The structural and content attributes obtained through content analysis, segmentation, video parsing, abstraction processes and the attributes entered manually are referred to as metadata. Video is indexed on a table using the metadata using clustering process which categorizes video clips or shots. Clustering process categorizes video clips or shots using metadata to form an index table of videos into different visual categories. Researchers have developed various tools and schemes to index, enquire, browse, search and retrieve videos from large databases but effective and robust tools are still lacking to test with large databases [6]. Due to these limitations [5], [6] a majority of video searches and retrievals still relies on keyword or text attributions. Face detection is assessed for image and video analysis. It was experimented in a commercial system [15]. It was found that accuracy of face recognition in video collection of the type mentioned in the system [8] was too poor to prove to be useful. Overall a large number of queries do not yield satisfactory results as mentioned [8] about one third of the queries were unanswerable by any of the automatic systems participating in the video retrieval track [16]. No system or method was able to provide relevant results. An integrated video retrieval system is proposed [2] where a video shot is represented not by key frame only but by all frames to extract more visual features of a shot. Color and motion features are integrated to fully exploit the spatio-temporal information contained in a video [29]. To overcome these drawbacks, i.e. considering lower efficiency of CBVR systems using a single image and very high computational cost of CBVR systems using key frames and the problem of availability of effective tools for CBVR systems using clustering process and to strike a balance between the efficiency and computational cost, visual features from multiple frames of a video clip are used in the system proposed here instead of a single frame or key frames or all frames of a clip. Also, it is learnt from the evaluation of video information retrieval that good image retrieval leads to good performance of video retrieval system when query is an

image or an image from the query video [8]. Computational cost point of view, the system proposed in this paper is cost effective along with acceptable as well as significantly higher results.

In section 2 features and features extraction algorithms are discussed; section 3 discuses about similarity measure; section 4shows the methodology to calculate result parameters in the proposed CBVR system. Proposed CBVR system is elaborated in section 5 and the result charts are shown in section 6; problems and challenges posed to this CBVR system are discussed in section 7 and the conclusion is presented in section 8.

2. FEATURES AND FEATURES EXTRACTION

2.1 Extraction of Gabor Features

For effective video indexing, classification and retrieval visual features embedded in video data is exploited. Three primary features to be extracted are color, texture and motion for effective video indexing. These features are represented by color histogram, Gabor texture features and motion histogram respectively [4]. Edge histogram and texture features are one of the most reliable data for effective video retrieval application. Gabor filters can also be used to obtain textural properties of texts which are distinct and distinguish them from its background in the image [7]. Extraction of Gabor features involves finding local energy of the signal i.e., localized frequency parameters are obtained. Gabor filter consists of multiple wavelets obtaining energy in multiple orientations with multiple frequencies with each of them tuned to a particular direction and frequency. Thus, texture features are obtained. The texture features are used to find images or regions inside the images having similar textures. The filters of a Gabor filter bank are designed to detect different frequencies and orientations [30]. They can be used to extract features on key points detected by interest operators [17]. From each filtered image, Gabor features are calculated and used to retrieve images. The algorithm for extracting the Gabor feature vector is shown in fig. 1 and the related equations (1 - 4) are also shown below [18], [20].

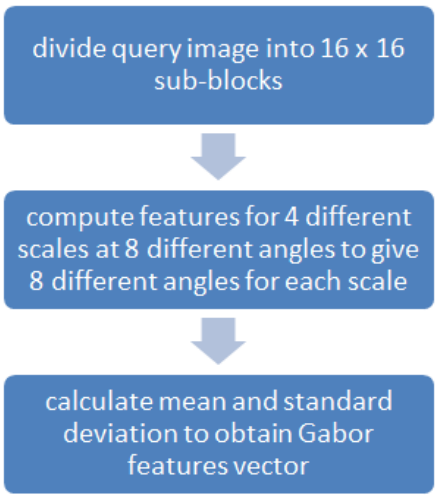

Figure 1. Gabor Filter Algorithm

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

$$W_{uv} = \sum_{p}\sum_{q} I(r-p, c-q)G_{uv}^{*}(p,q) \quad -(1)$$

where, G_{uv}^{*} is complex conjugate of G_{uv} . G_{uv} is generated by some morphological operations on mother wavelet. P X 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 U(u, v)containing the energy distribution in the image in different orientations and scales.

$$E(u,v) = \sum_{r}\sum_{c} |W_{uv}(r,c)| \quad - \quad (2)$$

Since we are interested to obtain texture features Standard deviation σ and mean is calculated using equations (3) and (4) respectively

$$\text{Standard Deviation, } \sigma_{uv} = \sqrt{\frac{\sum_r \sum_c (|W_{uv}(r,c)| - \mu_{uv})^2}{R\ X\ C}} \quad - \quad (3)$$

$$\text{Mean, } \mu_{uv} = \frac{E(u,v)}{R\ X\ C} - (4)$$

Texture features vector F is formed by a set of feature components [19], [14] i.e., different values of σ_{uv} and μ_{uv} calculated by varying u and v as shown in equation (5).

$$f = [\sigma_{u0v0}, \sigma_{u1}\sigma_{v1} \dots \sigma_{uUvV}] - (5)$$

$$f_{Gabor} = \frac{f - \mu}{\sigma} \quad - \quad (6)$$

2.2 Classification of features using Support Vector Machine

Use of Support Vector Machine (SVM) can be of great help for video classification. The frames from a video or a key frame representing a shot can be used to represent a video. It can also be represented by other components such as shots, scenes or events. Features are extracted from these video components. Corresponding features of videos from different categories are labeled to train SVM. Once the SVM is trained for these classes, it can be used to classify another group of videos having features extracted similarly. It is a big achievement towards automatic classification of videos [21]. Enhanced results can be obtained to classify a group of videos into their corresponding categories as it has been already obtained for features representing images. It has been observed that SVM can improve the results for CBIR problems [11]. SVMs are kernel based techniques used for classification. They can perform linear as well as non-linear classification as per the kernel design. The training process of a SVM is shown according to equations mentioned below.

Let's have a data (which may be feature vectors) V_K of m points spreaded over a d dimensional plane is used to train a SVM.

$$X = \left\{(V_K, C_K) \mid V_K \in R^d,\ C_K \in \{-1, +1\}\right\}_{K=1}^{m} \quad -\ (7)$$

$X$is termed as the training data. The data V_K is to be classified among two different categories as denoted by $C_K \in \{-1, +1\}$ and V_K is a d dimensional real vector.

We need to find a hyper plane separating the data V_K as shown in the fig. 2

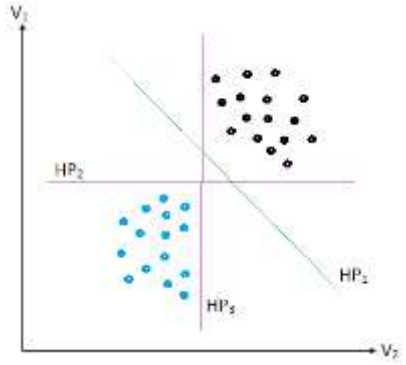

Figure 2. Hyper planes and two classes

Fig. 2 shows three hyper planes separating the two classes of variables. It can be observed that hyper planes HP_2 and HP_3 are separating the two classes but the margins are very less as they are very closed to some of the variables while the hyper plane HP_1 separates the two classes with a good margin. So HP_1 is selected while training the SVM. The hyper plane is shown by equation 8.

$$R \cdot V - q = 0 \quad - \quad (8)$$

Where, R is a normal vector to the hyperplane, $\cdot$ denotes dot product and a variable $\frac{q}{||R||}$ is used to find the offset of the hyperplane from the origin along the normal vector R. The given classification is linear classification. Linear classification is always not possible. In such cases, non-linear classification is required using non-linear equations for the kernel used for SVM training.

3. SIMILARITY MEASURE

Queries are classified by categories sorted out according to type of features used or type of example data. The query is found out by calculating similarity between feature vector [9], [10] stored in the database and the features of the query videos. The similarity is obtained by classification of videos using these features. Measuring similarity by using features is most convenient and direct method [1]. It is found by obtaining groups of videos classified by an SVM using their features. In query by example frames like the one used in the system shown in this paper similarity measure to find relevant and similar videos usually low level feature matching is used. Video similarity can be measured at different levels of resolution or granularity [13]. A video clip is retrieved by finding most similar video from the group of videos classified by the SVM. Furthermore, the most similar video can be obtained using the frames separated out from the enquired clip with those of the videos stored in the database. Video retrieval result depends greatly on video similarity measures. The videos are retrieved by finding similarity between the features extracted from multiple frames associated with query video and videos from the database.

4. 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 equations (9) and (10).

$$Recall = \frac{DC}{DB} \quad -(9)$$

$$Precision = \frac{DC}{DT} \quad -(10)$$

$DC = number\ of\ similar\ clips\ detected\ correctly$

$DB = \ number\ of\ similar\ clips\ in\ the\ database$

$DT = total\ number\ of\ detected\ clips$

Crossover points are calculated using the above mentioned two parameters to find the performance of the proposed system.

5. 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]. Features are extracted from these frames. 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 features extracted from their multiple frames. Features are then labelled and stored in the features database. An SVM is trained for the categories registered in the system using the labelled features stored in the database. Variables are obtained from the trained SVM. Features 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. The proposed system is shown in fig. 3. As mentioned above, multiple frames based classification and retrieval yields acceptable results without the complexity of finding key frames to represent a shot. A process flow of the proposed CBVR system is shown in fig. 3. Multiple frames are obtained during segmentation. Features are then extracted for each frame and stored in features database. Features are labelled for the pre-decided categories. 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 features of the query videos. Videos obtained for different categories are stored with different categories in the output database.

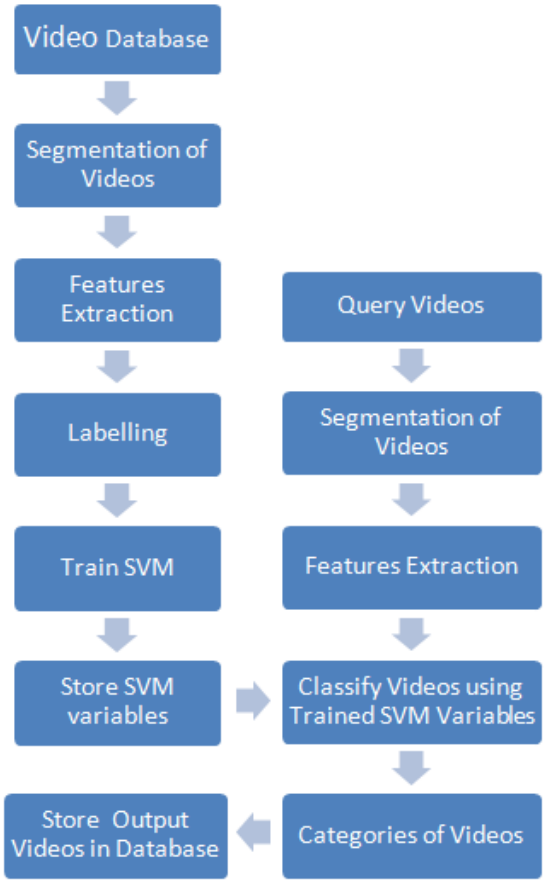

Figure 3. Proposed CBVR system

6. RESULTS

6.1 Database

The technique using multiple frames with Gabor 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.

Figure 4. Video database of 1000 videos with 20 categories

6.2 Results

The charts shown below in fig.5 and fig.6 represent the retrieval results obtained for retrieving and classification of video clips from ten different categories. These categories are among the 20 categories of video clips from the video database of 1000 videos. The results obtained are much appreciable for all the categories but these ten categories of them are demonstrated here. The results are obtained using SVM based on features extracted using Gabor wavelet transform from multiple frames of video clips.

6.2.1 Results(Precision Values) for the video clips

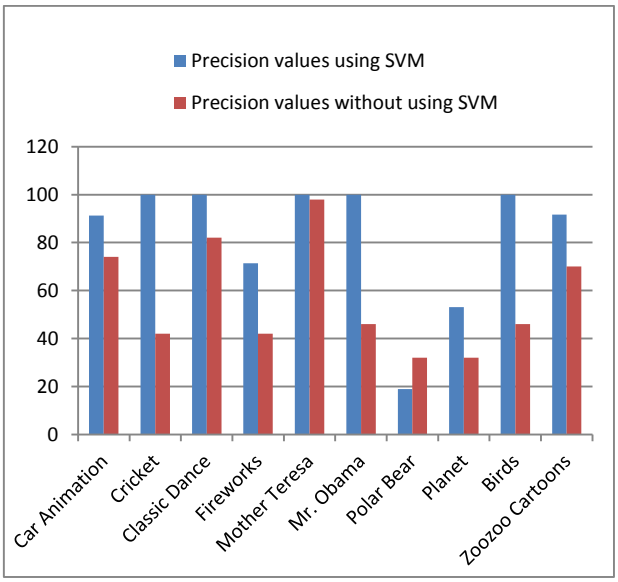

Figure 5. Comparison of Precision values shown for ten categories of videos using SVM and without using SVM using Gabor features

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

6.2.2 Results(Recall Values) for the video clips

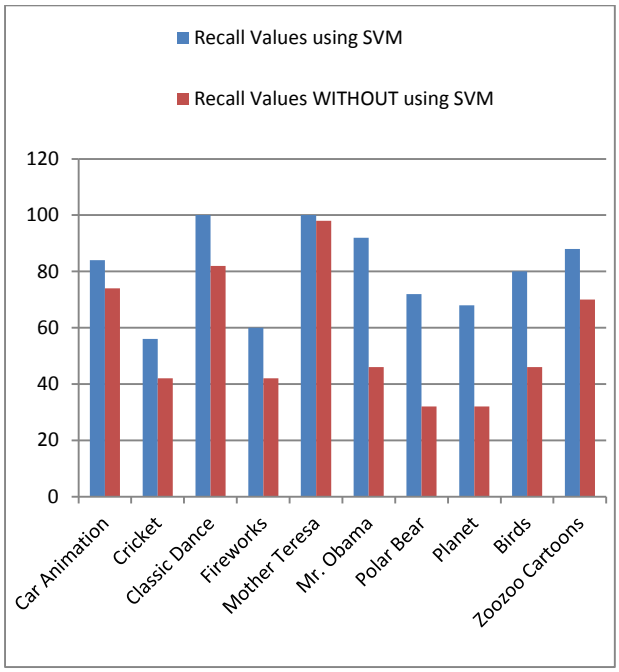

Figure 6. Comparison of Recall values shown for ten categories of videos using SVM and without using SVM using Gabor features

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.

7. PROBLEMS AND CHALLENGES

The video content is represented by spatial and temporal characteristics of videos. In spatial domain, features are obtained from frames to form feature vectors from different parts of the frames. In temporal domain, video is segmented into its elements like frames, shots, scenes and video clips and features like histograms, moments, textures and motion vectors represent the information content of these video segments [7]. Drawback of techniques employing key frames matching is that temporal information and the related information between the key frames in a shot is lost. Content based video retrieval systems using query by image or query by clips using images or frames is implemented with low level features present in these images. Because of this, different objects present against similar backgrounds in frames belonging to different videos can yield confusing or false retrievals. Also, the low level features of the frames belonging to different videos can also yield false retrievals due to their corresponding low level features matching though use of SVM enhances result to a significant level.

8. CONCLUSION

It can be concluded from discussion in the previous sections that encouraging results are obtained and comparatively higher efficiency is achieved by using features in support vector machines from multiple frames instead of single frame or key frames representing a shot. Also, computational cost is lower for the system proposed here than that when using key frames to represent shots of a video. Query by example image is popular for content based image retrieval. Low level features are used for retrieval. The retrieval performance and the usefulness of these systems is restricted to the queries having distinct low level visual features but they do not address to the problems of video retrievals using semantic information for the query. Also, an efficient solution is needed to address the problems for the queries having similar backgrounds and showing confusing results. Automatic retrieval systems should be the focus and it requires more attention from researchers for improved retrieval results. A trend to reduce computational cost is needed to project commercialized systems for video indexing, classification and retrieval to facilitate the availability of low cost, fast and efficient CBVR systems. Capability of these systems can be magnified by reaching huge video databases that exist and are accessible on the web. The accessible databases should empower the users with options to accurately select the desired videos only while filtering out the relevant but undesired as well as irrelevant videos so that valuable, moral, ethical and informative data becomes accessible efficiently, quickly and at low cost.

9. 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, Balamanohar Paluri, Nataraj Jammalamadaka, "Video Retrieval Based on Textual Queries", Proceedings of the Thirteenth InternationalConference 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 forMultimedia 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), 2012 , vol., no., pp.126,132, 10-12 May 2012

[12] Aasif Ansari, Muzammil H. Mohammed,"Content Based Video Retrieval Systems - Methods, Techniques, Trends and Challenges", International Journal of Computer Applications (ISBN : 973-93-80885-36-9),Volume 112 – No. 7, February 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] C. Faloutsos, R. Barber, M. Flicker, J. Hafner, W. Niblack, D. Petkovic and W. Equitz, "Efficient and effective querying by image content", J. lntelL Inf. Systems 3, 231-262, 1994.

[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] Arti Khaparde, B. L. Deekshatulu, M. Madhavilath, Zakira Farheen, Sandhya Kumari V, "Content Based Image Retrieval Using Independent Component Analysis", IJCSNS International Journal ofComputer Science and Network Security, VOL.8 No.4, April 2000.

[18] H.B. Kekre, V.A. Bharadi, S.D. Thepade, B.K. Mishra, S.E. Ghosalkar, S.M. Sawant, "Content Based Image Retreival Using Fusion of Gabor Magnitude and Modified Block Truncation Coding," icetet, pp.140-145, 2010 3rd International Conference on Emerging Trends in Engineering and Technology, 2010.

[19] Flickner M. et al, "Query by image and video content: the QBIC system", IEEE Computer 1995, Volume 28, Number 9, pp 23-32, 1995.

[20] Sinora Banker Ghosalkar, Vinayak A.Bharadi, Sanjay Sharma, Asif Ansari, "Feature Extraction using Overlap Blocks for Content based Image Retreival" International Journal of Computer Applications (0975-8887), Volume 28-No.7, August 2011.

[21] Markos Zampoglou, Theophilos Papadimitriou, IEEE Member, and Konstantinos I. Diamantaras, IEEE Member, "Support Vector Machines Content-Based Video Retrieval basedsolely on Motion Information", IEEE, ISSN : 1551-2541, Print ISBN: 978-1-4244-1566-3, 2007.