USER'S VISUAL INTENTION DRIVEN RETRIEVAL MODEL FOR SEMANTIC VIDEO SEARCH AND RETRIEVAL

***Lochandaka Ranathunga*[1]*, Nor Aniza Abdullah*[2] *and Roziati Zainuddin*[3]**
[1]*Faculty of Information Technology, University of Moratuwa, Katubedda, Moratuwa, Sri Lanka*
[2, 3]*Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia*
[1]*lochandaka@uom.lk,* [2]*noraniza@um.edu.my,* [3]*roziati@um.edu.my*

ABSTRACT

There are various models of retrieval mechanisms for effective video retrieval, however, 'semantic gap' still persists. This study aims to reduce the gap by providing a graphical user interface that allows user to express the visuals in their mind into a query using semantic association of spatial and temporal visual concepts. To accomplish this, we have developed a video content depiction schema that serves as the foundation to our semantic search mechanisms, namely a generic visual concept search, Spatial Concept Association Based Search (SCABS) and Spatial Concept Transition Based Search (SCTBS). In our approach, a clear understanding between the user and the system has been established through the depiction schema and the visual retrieval mechanisms in contrast to common tag based retrieval systems. This retrieval model employs visual experiences and context knowledge of the users to generate search results. Our experiment result shows that our approach produces positive result against other approaches.

Keywords: *CDPC, Visual concepts depiction, Semantic gap, Semantic video retrieval, Semantic video search*

1.0 INTRODUCTION

The availability of various video capturing devices contributes to the exponential growth of video repositories each day. However, the expansion of such repositories put forth a problem of retrieving meaningful video files. In retrieving video files, current internet-based video search engines such as Google, Yahoo and YouTube extensively use metadata or tag to describe video content in the Web [2,3]. However, the process of metadata tagging is often done manually thus consumes tremendous amount of time and effort at high cost. In addition to that, inconsistency and sparse nature of human generated tags makes the differences in the semantic video content to the generated tags unavoidable. This has led to the issues of dissimilarity in semantic interaction of multimedia retrieval systems when compared to human level of visual understanding [1, 2, 7].

Recent efforts in video content retrieval have put forward many techniques to bridge this semantic gap. Among which, many researchers aimed to combine tags with content-based techniques in order to improve data navigation and search [5]. However, most content-based retrieval systems are constrained to syntactic information for classifying and indexing multimedia data. Therefore it is importance to fill the gap between the classified syntactic features of multimedia data and the semantic level of human's perception [4, 7, 19-21]. To address this issue, researchers like [1] has utilized the motion behaviours of videos for querying, and [18] has provided a communication interface for users to express their query expectations through motion sketches and 'query by example'.

This study offers an alternative mechanism to the existing querying interface for video data. The mechanism is derived from the spatial and temporal concept associations that permit the incorporation or user's visual querying intention into the querying process. This novel visual content retrieval mechanism implements our visual description schema that maintains the semantic relationships among possible visual concepts and their instances.

The rest of the article organised as follows: Section 2 highlights the review of related literature and underlined studies. Section 3 introduces syntactic visual feature and classifier used in this study with reasons. Section 4 presents a novel video visual depiction schema and algorithms for video description and search. Section 5 proposes the novel semantic video visual search mechanisms with demonstrations. Section 6 describes the experimental setup and section 7 presents the results with comparative discussions. Finally, section 8 concludes findings of this study.

2.0 SEMANTIC VIDEO SEARCH AND RETRIEVAL

Semantic video visual search gains prominent importance due to the needs of World Wide Web consumers and ever increasing video applications. The popular image search system QBIC uses dominancy of syntactic features such as colour, texture and location to retrieve images from database [13]. The VIRAGE system is widely used for automated video searching, creation of metadata and categorisation, which includes key frames, face recognition, speaker information and a full transcript of the audio stream which is used in cross language platform with mainly sonic features [14]. Blinkx search which is built on theories of content based video retrieval and Artificial Intelligence basically uses video indexing on metadata [3, 15]. The core technologies used by Blinkx are speech recognition, subtitle identification, and annotation. But still these techniques have deficiencies in video content analysis or visual feature extraction where visual semantics cannot be fully catered using the above techniques. Popular YouTube search technology on the Web is totally depended on low level metadata. Some querying systems work with user inputs with expectation that the user has some understanding of the expected visuals, for example in VisualSEEK and Massachusetts Institute of Technology (M.I.T) Chabot [16]. Those applications use users' prior knowledge about visuals.

The goal of semantic video search is to deliver the information queried by a user rather than having a user to sort through a list of loosely related keyword results. Therefore, it is important for semantic video search to improve search accuracy by bridging the searcher's aim and the contextual meaning of query terms to obtain relevant results. However, the 'semantic gap' is the forefront barrier in realisation of semantic video visual search on the web. It is worthwhile to put on an research effort in the web's user interfaces to identify challenges and opportunities for interactions that the semantic web would afford [17]. The challenge that remains is how to afford these kinds of explorations through web-based users' interaction tools to support user to query and represent information in order to build new knowledge over wild data on demand, rather than over well tamed, well accurate, integrated heterogeneous datasets? In line with minimizing of above challenge, approaches such as 'query by example' has already been experimented, where the system is fed with a number of example images and it should indicate the closest match. Various features of the chosen image are evaluated and matched against the images in the database [18]. Here, features that have been commonly used are colour, shape, texture and spatial distribution. Video summarisation such as 'Video College' attempted to synthesise most important frames of a video by extracting salient visuals through regions of interest (ROI) approach and representing a summary through a static image canvas [6]. Similarly, some research projects have used concept on query by icon to feed clearer requirements for the system to obtain images [18]. These approaches use pre-processing of syntactical information given in user interface such as icons with locations, objects and colours to generate query results from databases.

3.0 VISUAL FEATURE EXTRACTION AND CLASSIFICATION

Syntactic visual feature extraction is a major phase in automatic visual concept depiction. These visual features are derived from complex chromatic data. The quality of the extracted feature would determine the success of the visual concepts classification. The ultimate video visual depiction accuracy depends on both the extracted features and classifier. High level semantic features are derived after visual concept classification where these features can be used to search for semantic information. These search mechanisms could be more intuitive when it is integrated with additional domain knowledge and perceptual interactivity [8].

Recent projects on visual content analysis reveal the problem of dealing with high level resolution and colour depth which usually lead to extraction of high dimensional syntactic visual features that are computationally expensive with poor efficiency. High dimensional feature space requires additional effort in training as it requires large amount of training samples [10-12]. Even though, the precision of visual data mining can be improved by combining different descriptors, such combination can produce high dimensionality in the resulting feature space which drastically reduces efficiency of storage and retrieval. The complexity of extraction can also increase as the number of attributes in the algorithms involved is high in both training and matching (classification) processes.

To overcome the issues of high dimensional and computationally expensive colour features, this study used a compact form of visual feature called Compacted Dither Pattern Code (CDPC) in visual feature extraction, and

Bhattacharyya classification for its visual concept depiction [9]. CDPC promotes low complexity of feature space representation and low depth of the training space. The compactness of the visual feature space can further simplify the classification process thus improves the effectiveness and efficiency of generating video visual description for latter use in semantic search and retrieval [9]. Generally, this CDPC-based approach processes video visual regions and extracts micro-level compact patterns to describe video frames. In this mechanism of video visual content analysis, probability distribution of CDPCs is used to identify and classify visual concepts.

4.0 VIDEO SCENE CONCEPT DETECTION AND DEPICTION

The main functions of the scene concept detection and depiction in this system are to classify the extracted syntactic visual features into known visual concepts in a trained knowledge base, and describe the video shots. CDPC syntactic visual feature is employed in visual feature extraction, and Bhattacharyya classification is used in the visual concept classification [9]. In feature extraction and concept depiction, the video frames are divided into 16 rectangular sections as shown in Fig. 1.

Fig. 1: Sectional view of a frame

The visual concepts and special concepts (instances) detected in each section of the video frame are then sent for description. After obtaining the section based frame description, it is compared with previously depicted section based shot description for differences. If a minimum of four sectional concepts differ in comparison, then the available frame description is depicted as a new shot according to the XML depiction schema structure. The elements of the organisation structure of the Depiction Schema used in the Video Scene / Shot Descriptions is shown in Fig. 2. This schema was used to create XML descriptions for video files. The description files in XML format are named according to the video file names. Each video file content description has two sections for low level descriptions and high level concept description.

High level description is organised according to shot sequence identified by the system and one frame of each shot is described according to the sectional concepts and special concepts. Therefore, it able to maintain semantic relationship among visual concepts and their instances in a way that maps to human semantic visual understanding. In addition, each sectional concept is accompanied with its Bhattacharyya coefficient which has been used for ranking purpose in retrieval.

The description method used in this program module detects the changes that take place in sectional descriptions when shots are derived. In other words, this description mechanism uses the changes in the meaning of video frames for shot separation. This way of shot segmentation is a direction towards semantic shot segmentation rather than conventional low level colour feature based shot segmentations. This shot segmentation uses the changes that take place in sectional concepts and also its special instances. Most existing shot segmentation methods depend on statistical distributions of chromatic data where visual concepts are not considered.

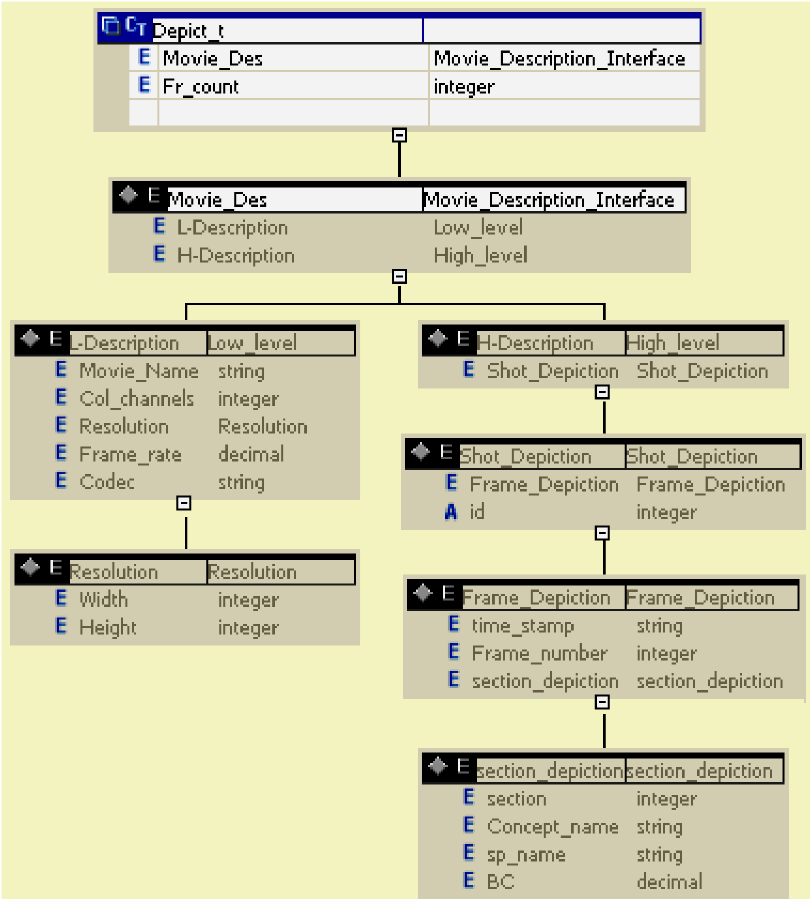


Fig. 2: Element Structure of the XML depiction schema used in video scene / shot descriptions

The main algorithmic steps of the visual concept detection and depiction are listed below.

Step 1: If it is not the end at the video file, then obtain video filename, frame number, section number and extract the probability distribution of CDPC in the specified section. Otherwise go to ***Step 11***

Step 2: If the pointer is not at the end of the given visual concept list, and then obtain a trained visual concept knowledge data file name from the concept list. Otherwise go to ***Step 5***

Step 3: Obtain each CDPC probability distribution presented under trained visual concepts' knowledge data file and perform the classification according to the Bhattacharya classification. Among the successful Bhattacharyya coefficients returned by each classification task keep the name of the visual concept and the special instance name respect to the highest Bhattacharyya coefficient.

Step 4: If classification is finished for the current trained visual concepts' knowledge data represented, then go to ***Step 2***. Otherwise go to ***Step 3***

Step 5: Update the current video frames' visual concepts list with visual concepts and their special concepts for the current frame according to the section numbers given

Step 6: If section number is less than 16, then go to ***Step 1***. Otherwise go to ***Step 7***

Step 7: Compare the list of current frames' visual concepts with the list of previous shots' visual concepts according to their respective sections (Comparison includes special instances)

Step 8: If comparison results in ***Step 7*** represents ¼ or higher dissimilarity, then go to ***Step 9***. Otherwise clear the list of current frames' visual concepts and go to ***Step 1***

Step 9: If description XML data file already exist, then update the high level description section by appending the current frame description with a new shot description according to the depiction schema structure in Fig. 2. Otherwise create a new description XML file (according to video file name) according to the depiction schema structure in Fig. 2 and add the high level description section with the current frame description as a new shot description.

Step 10: Copy the list of current frames' visual concepts to the list of previous shots' visual concepts list and clear the list of current frames' visual concepts and go to ***Step 1***

Step 11: Save and close the current description XML file. Clear the list of previous shots' visual concepts

The above steps are called by the system when describing video visuals. The video content descriptions produced by the scene concepts detection and depiction are used in semantic video visual search and retrieval system. Therefore the retrieval mechanism also uses the same XML schema to access information about the descriptions. According to the XML depiction schema used in visual description, there are different possible ways of retrieving visual information of video visual contents.

5.0 VIDEO VISUAL SEARCH AND RETRIEVAL DRIVEN BY USER'S INTENTION

The proposed retrieval mechanism provides interfaces and methods to apply the video visual depiction schema, and generate semantic search results. To emphasise the usage and the effectiveness of the schema, we have implemented three novel semantic search mechanisms, namely a generic visual concept search (one concept with its special instance), Spatial Concept Association Based Search (SCABS) and Spatial Concept Transition Based Search (SCTBS). These search mechanisms provide are built on the association and organisation of spatial and temporal video visual search and humans' visual query intention in deriving significant query results. The novel interfaces help to reduce the requirement to hardcode visual ontology for visual concept analysis and ranking. The inclusion of visual relationships for each visual concept in a program is a difficult task. Therefore, the usage of users' visual cognition in video content retrieval saves computational cost and time. Also, it provides an opportunity to present users' expectations more explicitly by reducing the gap between users' level of expectations and the system's level of understanding of the requirements involved in search mechanism.

The proposed visual concept search uses an interface to obtain textual visual concept to search within video shots. In the generic visual concept search, as in Fig. 3, a user can perform queries with or without any special instance name of a visual concept.

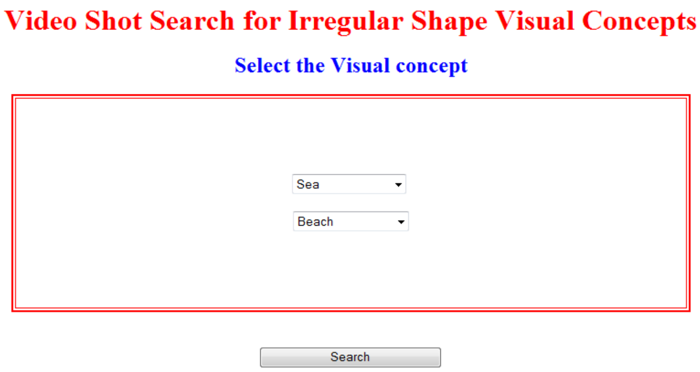


Fig. 3: One Visual Concept Search Interface

One visual concept search activates a search for the given concept within the XML descriptions. In the search process, each video file description is accessed by a search agent and every shot description is searched without any sectional constraints. A special XSL style sheet is used to format the search results according to their spatial concept relation based ranks, number of sectional hit counts and Bhattacharyya coefficient. This type of search method only provides information on the existence or absence of a given concept or special instance inside a video visual content. As the depiction schema is able to keep semantic relationships of visual concepts, this search mechanism can also extends the ability of semantic video visual search to include semantic visual concept relationships.

5.1. Spatial Concept Association Based Search (SCABS)

Spatial Concept Association Based Search (SCABS) mechanism provides a convenient interface to express a user's intention (image in mind) for retrieving relevant videos. This kind of search method expresses the query and the retrieval results according to the user's specification. Here, context oriented knowledge of human visual perception is fed into the query to improve the performance of retrieval results. This inclusion of human logic of visual understanding according to each of the visual concept and its context into query is an adaptive solution to get rid of hard coded human cognition based visual description. In addition to that, hard coding of such visual concept association for each visual concept can increase the computational cost. Most importantly, the inclusion of spatial relationships and the expected associations of visual concepts by the user's intention can provide the necessary logic to obtain semantic search.

Fig. 4 shows the interface of SCABS. This interface gives an impression of a frame arrangement and provides an opportunity to select the required visual concept association according to the four spatial regions compared to previous generic visual concept search. Normally video visuals consist of multiple concepts within the same shot. The concept association adds the semantic to the overall visual. Therefore, the representation of user's image in a semantic manner improves the search towards semantic way of retrieval.

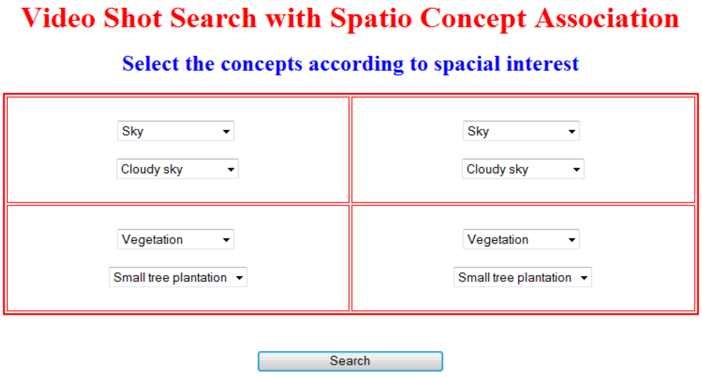


Fig. 4: Interface of SCABS

Fig. 5: An example result set with a view of retrieved records for the query in Fig. 4.

The example in Fig. 4 requests for a cloudy sky in the upper part of the frame and a small tree plantation in the lower part of the frame. This expression of user's intention is common for human as vegetation is normally visualised below the sky. Therefore, these types of logics also make it easy to filter the queries with additional accuracy. Results generated for the above request is shown in Fig. 5 with a view of two retrieved shots. The SCABS specialises video shot search with more specific and targeted queries.

5.2. Spatial Concepts Transition Based Search (SCTBS)

Spatial Concepts Transition Based Search (SCTBS) is a special type of video content retrieval introduced in this study. This search mechanism grasps semantic and episodic (change from one scene to another) expectations of a user in generating retrieval results. The search interface comes with a representation of two conceptual frames where each is divided into two regions. Each region represents a group of sections according to the frame section specification shown in Fig. 1. The experiences of scenery changes (semantic and episodic memory together) can be effectively represented through this nature of interfaces. Fig. 6 shows the interface that is used in the SCTBS. The search is performed in adjoin shots for specific visual concepts association.

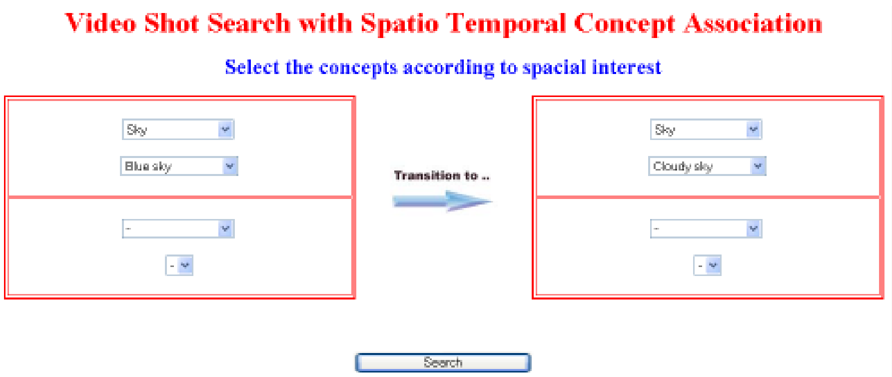


Fig. 6: Search interface for SCTBS

Fig. 7: Two screen shots extracted from the result generated for query displayed in Fig. 6

According to the representation given in the interface, the specified frame sections in video shots are examined by the search agent in adjoin shots in the video descriptions. The search result produces video shots that match the specified transition. User’s video visual experience is logically used to match the expectation and minimise filter errors in producing results. This mechanism reduces the communication gap between the system and the user through a more semantic understanding between the parties. An example of search result produced for the query displayed in Fig. 6 is shown in Fig. 7 with two extracted screen shots of retrieved result. Operational steps of the search mechanism based on Spatial Concepts Transition program (only the search agent) are given below.

Step 1: Obtain the visual concepts organisation of intended first shot and second shot according to their regions (ex: frame 1 and frame 2 shown in Fig. 6)

Step 2: If pointer is not at the end of video description file list, then access next video description xml file according to video description file list in the Video Scene / Shot Descriptions. Otherwise view the result set xml file according to given XSL style sheet (Sort according to Rank, Average number of sections matched, Bhattacharyya Coefficient)

Step 3: If at least two shots are not remaining to match, then close the current video description xml file and go to ***Step 2.*** Otherwise access the frame information of current shot

Step 4: If the visual concept of the top region of frame 1 is non empty, then compare the visual concept given (with or without special instance according to availability) in the top region of frame 1 with sectional descriptions given in number 1-8 in the frame of current shot and increase region value by 1. Increase the similar count by 1 and update the highest Bhattacharyya coefficient for each matched section

Step 5: If the visual concept of bottom region of frame 1 is not empty, then compare the visual concept given (with or without special instance according to availability) in bottom region of frame 1 with sectional descriptions given in number 9-16 in the frame of current shot and increment region value by 1. Increment the similar count by 1 and update the highest Bhattacharyya coefficient for each matched section

Step 6: If the region value is not zero, then divide the similar count by region value, set region value and similar counts to 0. Otherwise show error message and terminate (Concept missing in the first shot)

Step 7: If result produced by ***Step 6*** is larger than or equal to 2, then go to ***Step 8.*** Otherwise increase current shot by 1, clear highest Bhattacharyya coefficient and go to ***Step 3.***

Step 8: Access the frame information in the shot specified by the current shot + 1.

Step 9: If the visual concept of the top region of frame 2 is not empty, then compare the visual concept given (with or without special instance according to availability) in the top region of frame 2 with sectional descriptions given in number 1-8 in the frame of current shot +1 and increase region value by 1. Increase the similar count by 1 and update the highest Bhattacharyya coefficient (2) for each matched section

Step 10: If the visual concept of bottom region of frame 2 is non empty, then compare the visual concept given (with or without special instance according to availability) in bottom region of frame 2 with sectional descriptions given in number 9-16 in the frame of current shot +1 and increment region value by 1. Increase the similar count by 1 and update the highest Bhattacharyya coefficient (2) for each matched section

Step 11: If region value is not zero, then divide the similar count by region value, set region value and similar counts to 0. Otherwise show error message and terminate (Concept missing in the second shot)

Step 12: If result produced by ***Step 11*** is larger than or equal to 2, then go to ***Step 13.*** Otherwise increase current shot by 1, clear highest Bhattacharyya coefficient, Bhattacharyya coefficient (2) and go to ***Step 3.***

Step 13: Add results produced by ***Step 7*** and ***Step 12*** then divide it by 2, and divide the sum of Bhattacharyya coefficient and Bhattacharyya coefficient (2) by 2

Step 14: Categorise the resultant shot in to a rank according to the result produced by ***Step 13*** (value) as follows,
value > 5 – Rank 1
5 > value > 4 – Rank 2
4 > value > 3 – Rank 3
Otherwise Rank 4

Step 15: Obtain video file name, frame numbers of current shot and current shot + 2, average Bhattacharyya Coefficient produced in ***Step 13*** and rank

Step 16: Append the result set with information obtained in ***Step 15*** according to given XML Schema. Clear video file name, frame numbers of current shot and current shot + 2, average Bhattacharyya Coefficient and rank. Go to ***Step 3***.

The algorithmic information given in the above steps can be altered with a variety of options and different search strategies. Fig. 8 shows another example video visual transition query and Fig. 9 illustrates the extracted adjacent shots according to the query.

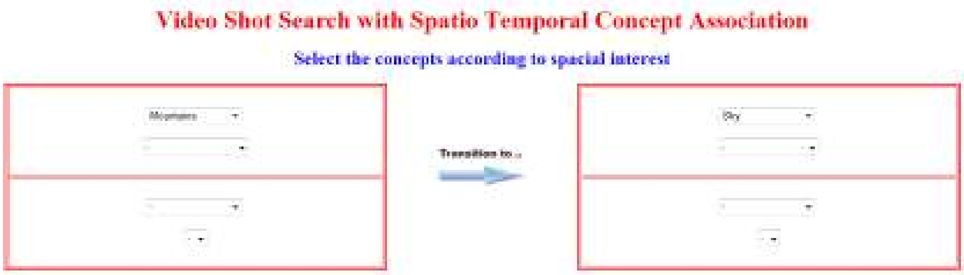


Fig. 8: SCTBS query for visual concept of Mountain to Sky on top part of video frame

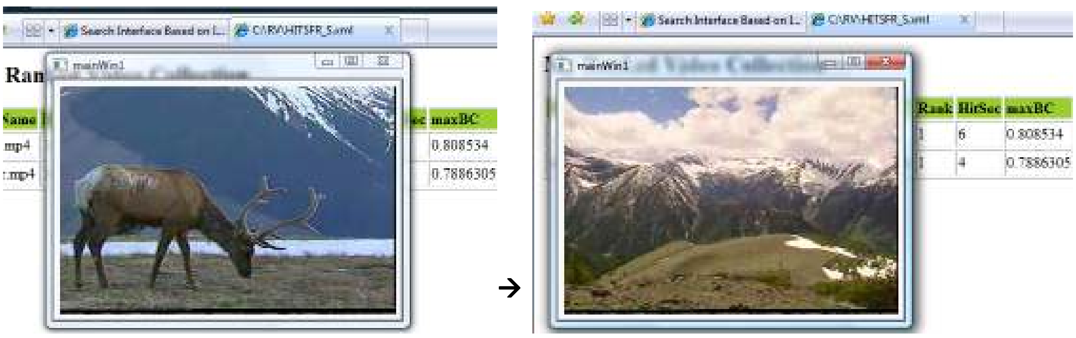


Fig. 9: Two frames captured from retrieved shot transition resulted by the query in Fig. 8

6.0 EXPERIMENTAL SETUP

6.1. Automated Concept Depiction

The experiment of concept depiction was done using TRECVID 2005 keyframe test dataset. The CDPC based system was trained by using selected samples from diverse sources, such as the open video project, adobe premier sample video clips, and our own video clips. In this experiment, the CDPC syntactic visual feature with Bhattacharyya classifier was tested for ten irregular shapes based visual concepts (with their special instances). 30 training samples were used for each of the visual concept separately to detect and depict video visuals automatically. This sample population has significantly low training space compared to most state-of-art research studies. 2000 shots were tested for the ten visual concepts. Depictions were generated for each video shot and repository was created. The generated video depictions (according to the schema in the Fig. 2) were kept for the video search in retrieving videos from the repository. Based on the generated depictions, the following two experiments under section 6.2 and 6.3 were conducted.

6.2. Generic Search Experiment

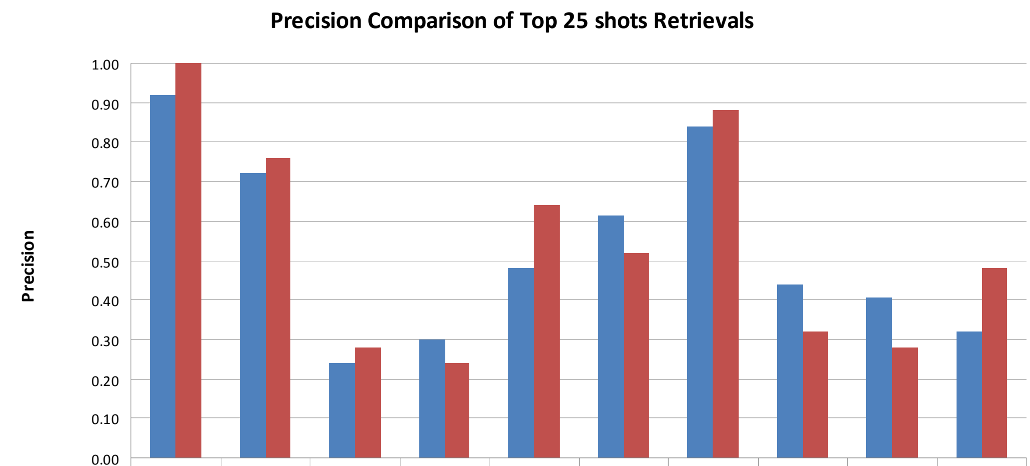
Subsequently, the one concept search mechanism was experimented with generated visual descriptions with the help of the depiction schema. The search was conducted by considering whole repository with 2000 shot descriptions. In search query, only the main concept was searched without any special instances. The generic search interface can be found in the Fig. 3. All ten visual concepts were tested by considering the accuracy (Precision) of top 25 and 50 retrievals in the search result set of each visual concept. The result generated from this experiment was used for the comparison. This result is indicated by the General Rank category in Fig. 10 and Fig. 11 respectively.

6.3. SCABS Experiment

The SCABS mechanism was experimented with generated visual descriptions according to the depiction schema. When using the SCABS mechanism, the most relevant spatial area for each visual concept was employed according to generic visual understanding. In this approach, most probable visual area related to the visual concept was employed in each concept search. For an example, when searching for the visual concept of sky, the upper two search options were selected as sky concept (refer the Fig. 4.). Further, the special concepts and other non-related spatial areas for the subjected concept were not selected (kept blank). The retrieval accuracies (precision) of top 25 and 50 retrievals were calculated for each category of the visual concepts retrieval.

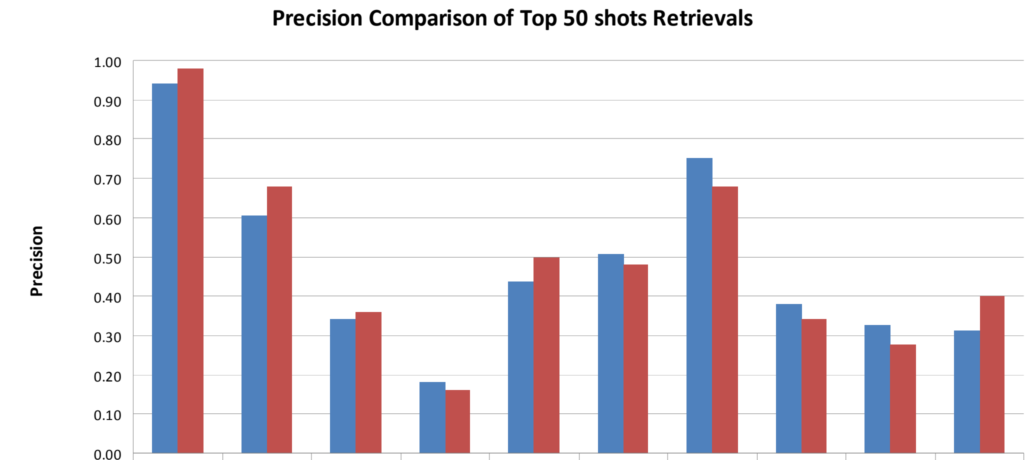
7.0 RESULTS AND DISCUSSION

When exploring Spatial Concept Association Based Search (SCABS) results, the General Rank category in Fig. 10 and Fig. 11 represents the retrieval result set taken from the result set of current approaches (considering the first 25 and 50 result set respectively). SCABS in Fig. 10 and Fig. 11 represent the semantic search mechanism based on Spatial Concept Association (SCA) driven results. When searching for the visual concepts using SCA based semantic search interface, the search queries were formed by assigning the visual concepts into its most natural spatial areas to obtain result sets. Further, the special instances and non-related spatial areas were excluded even though the option is available. This decision was taken to obtain a fare benchmark.



	Sky	Sea	Mountains	Roads	Desert	Buildings	Vegetation	Water Resource	Fire	Snow
General Rank-25	0.9200	0.7200	0.2400	0.3000	0.4800	0.6143	0.8400	0.4400	0.4052	0.3200
SCABS-25	1.0000	0.7600	0.2800	0.2400	0.6400	0.5200	0.8800	0.3200	0.2800	0.4800

Fig. 10: Precision comparison of first 25 retrievals using general raked method and SCABS mechanism



	Sky	Sea	Mountains	Roads	Desert	Buildings	Vegetation	Water Resource	Fire	Snow
General Rank-50	0.9400	0.6042	0.3400	0.1800	0.4375	0.5081	0.7500	0.3800	0.3275	0.3111
SCABS-50	0.9800	0.6800	0.3600	0.1600	0.5000	0.4800	0.6800	0.3400	0.2759	0.4000

Fig. 11: Precision comparison of first 50 retrievals using general raked method and SCABS mechanism

When using SCA based search mechanism, an improved precision is shown for the concepts of Sky, Sea, Mountains, Desert, Vegetation and Snow as in Fig. 10. In Fig. 11 with the first 50 retrievals, improved results are shown for the concepts of Sky, Sea, Mountains, Desert and Snow. As most top results contain higher impact, this nature of user's visual intention driven search mechanisms is helpful to filter video shots with false classifications.

The mean values of the averaged precisions of the two methods presented in Fig. 10 and Fig. 11 prove the improvement gained by SCABS mechanism, where SCABS reports 0.54, 0.4856 and previous ranking mechanism (General Rank) reports 0.528, 0.4778 in these two figures respectively.

SCTBS interface further provides search mechanisms for combination of visual concepts within space and time. The description schema designed in this study provides opportunity to maintain semantic relationships of concepts. The special events of general concepts are described under each visual concept. Therefore, retrieval of the required special events is possible. In some popular studies related to LSCOM and VIREO-374, these events are considered as separate concepts without keeping their semantic relationships [22]. In SCA, the description schema and search interface provides a more semantic way for the visual concept combination retrieval by keeping the semantic relationships of visual concepts. It has been seen that, when combining spatially related concept together in the search interface, the retrieval accuracy could be further improved.

In temporal segmentation, by considering concepts composition and association including special events, it is possible to segment video shots according to the changes that occur in video visuals. This nature of temporal segmentation provides division of shots according to the accompanied meaning of visuals. However, the accuracy of shot segmentation depends on the accuracy of visual identification and therefore it may not be more accurate as statistical techniques. But this nature of shot segmentation is a good starting point for a semantic shot segmentation of videos.

8.0 CONCLUSION

The CDPC combined with Bhattacharyya classifier has produced depiction result set with adequate precision in irregular shapes based visual concepts detection and depiction while maintaining low computational cost. The depiction schema used in this study has extended the description ability while maintaining the semantic relationships of visual concepts where most well known studies have not maintained as such. This schema has opened up novel ways of video visual search and retrieval mechanisms that enabled the bridging of the 'semantic gap'. The SCABS comes with user-friendly interface where users can express the visuals in their mind. Further this mechanism able to absorb the relationship between visual concepts and their spatial dominancy. The user's visual intention driven query helps to further filter the most relevant visual concepts without additional cost. This has proven through the average accuracy improvement of visual retrieval shown by the SCABS mechanism compared to generic visual ranking which are 2.28% and 1.62% respectively for the first 25 and 50 retrieved shots. When comparing this mechanism with hard coded visual relevance based filtering, this mechanism shows computational advantages especially when programming user's visual experience is a hard problem. The SCTBS is another dimension of video visual search where users can perform their queries to retrieve episodic expectations. The semantic way of shot segmentation empowered by the depiction schema has drawn a path for the mechanisms of retrieving semantic video visuals.

This study has contributed to drive towards semantic video visual search and retrieval by minimising semantic gap in two main processes. First is by improving the semantic depiction phase by contributing to the efficiency and effectiveness of automatic video visual description by keeping conceptual semantic relationships. The second is by introducing novel semantic search mechanism, which is accompanied with user's visual intention to improve the semantic search and retrieval for video search engines. We believe that these contributions would make path towards the realization of an effective and efficient semantic video visual search and retrieval in mitigating semantic gap.

REFERENCES

[1] A. Mittal, "An Overview of Multimedia Content-Based Retrieval Strategies," *Informatica,* vol. 30, pp. 347–356, 2006.

[2] C. G. M. Snoek and A. W. M. Smeulders, "Visual-Concept Search Solved?," *IEEE Computer,* vol. 43, pp. 76-78, 2010.

[3] Blinkx, "Blinkx SEO (Search Engine Optimisation) white paper," Blinkx 2007.

[4] F. Bailan, C. Juan, B. Xiuguo, B. Lei, Z. Yongdong, L. Shouxun, and Y. Xiaochun, "Graph-based multi-space semantic correlation propagation for video retrieval," *Vis. Comput.,* vol. 27, pp. 21-34, 2010.

[5] K. Bischoff, C. S. Firan, W. Nejdl, and R. Paiu, "Bridging the gap between tagging and querying vocabularies: Analyses and applications for enhancing multimedia IR," *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 8, pp. 97-109, 2010.

[6] Mei, Tao, Yang, Bo, Q. Shi, Hua, and G. Xian-Sheng, "Video collage: presenting a video sequence using a single image," *Vis. Comput.,* vol. 25, pp. 39-51, 2009.

[7] M. S. Lew, N. Sebe, C. D. Jebara, and R. Jain, "Content-based Multimedia Information Retrieval: State of the Art and Challenges," *ACM Transactions on Multimedia Computing, Communications and Applications,* vol. 2, pp. 1-19, 2006.

[8] W. Zheng, J. Li, Z. Si, F. Lin, and B. Zhang, "Using High-Level Semantic Features in Video Retrieval " *Image and Video Retrieval - 5th International Conference, CIVR 2006,* vol. 4071/2006, pp. 370-379, June 30, 2006 2006.

[9] L. Ranathunga, R. Zainuddin, and N. A. Abdullah, "Performance evaluation of the combination of Compacted Dither Pattern Codes with Bhattacharyya classifier in video visual concept depiction," *Multimedia Tools and Applications,* vol. 54 pp. 263-289, 2011.

[10] E. F. Badjio and F. Poulet, "Dimension reduction for visual data mining," in *In International symposium on applied stochastic models and data analysis (ASMDA-2005)*, Brest, France, 2005, pp. 266-275.

[11] Y. Gao and J. Fan, "Semantic image classification with hierarchical feature subset selection," in *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval* Hilton, Singapore: ACM, 2005.

[12] E. Mrowka, A. Dorado, W. Pedrycz, and E. Izquierdo, "Dimensionality Reduction for Content-Based Image Classification," in *Proceedings of the Information Visualisation, Eighth International Conference*: IEEE Computer Society, 2004.

[13] IBM-Corporation, "QBIC(TM) -- IBM's Query By Image Content ". vol. 2008: IBM Corporation, 2003.

[14] A. Virage, "Technology." vol. 2010: Autonomy Virage, 2009.

[15] Blinkx, "Blinkx Technology White Paper," Blinkx 2008.

[16] J. R. Smith and S.-F. Chang, "VisualSEEk: a fully automated content-based image query system," in *Proceedings of the fourth ACM international conference on Multimedia* Boston, Massachusetts, United States: ACM, 1996.

[17] M. Schraefel and L. Rutledge, "User interaction in semantic web research," *Web Semantics: Science, Services and Agents on the World Wide Web,* vol. 8, pp. 375-376.

[18] M. S. Lew, "Next-Generation Web Searches for Visual Content," *Computer,* vol. 33, pp. 46-53, 2000.

[19] R. Datta, J. Li, and J. Z. Wang, "Content-based image retrieval: approaches and trends of the new age," in *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval* Hilton, Singapore: ACM, 2005.

[20] N. Sebe, "Multimedia Information Retrieval: Promises and Challenges," in *5th ACM SIGMM International Workshop on Multimedia Information Retrieval*, New York, 2003.

[21] Z. Xiong, X. S. Zhou, Q. Tian, Y. Rui, and T. S. Huangm, "Semantic retrieval of video - review of research on video retrieval in meetings, movies and broadcast news, and sports," *Signal Processing Magazine, IEEE,* vol. 23, pp. 18-27, 2006.

[22] S.-F. Chang, D. Ellis, W. Jiang, K. Lee, A. Yanagawa, A. C. Loui, and J. Luo, "Large-scale multimodal semantic concept detection for consumer video," in *Proceedings of the international workshop on Workshop on multimedia information retrieval* Augsburg, Bavaria, Germany: ACM, 2007, pp. 255-264.