Ontology-based Automatic Video Annotation Technique in Smart TV Environment

Jin-Woo Jeong, Hyun-Ki Hong, and Dong-Ho Lee, *Member*, IEEE

Abstract — *Due to the recent explosive growth of smart TV and its applications, the need for multimedia contents analysis techniques for smart TV environment has dramatically increased. In particular, automatic video annotation and summarization technique is considered as one of the most important technique for smart TV because an enormous number of social videos are now becoming available as well as traditional digital videos such as TV programs and VODs.*

In this paper, we present an automatic video annotation technique which employs the ontologies to facilitate the video retrieval and sharing process in smart TV environment. In our work, ontologies are exploited to hierarchically describe the contents of video and facilitate the sharing of video contents among heterogeneous users or devices (e.g., smart phone, PDA, and so on). Through various experimental results, we show the effectiveness of the proposed approach[1].

Index Terms — IPTV, Smart TV, Video annotation, Ontology.

I. INTRODUCTION

With the great success of the smart phone in recent years, a new paradigm "*smart"* has been emerging in digital TV market. A new kind of TV called Smart TV was released in 2010 by leading smart phone companies.

Smart TV is generally defined as a medium that provides TV broadcasting, Internet, application, and intelligent services via the mounting of a CPU and operating platform on the set-top box or display [1]. In fact, a TV platform with a CPU and Internet service called IPTV was already developed in early 2000's with the vendors attempting to provide TV programs and VOD contents to TV through Internet protocol suite. However, despite the significant researches to facilitate the access, sharing, and retrieval of TV contents for IPTV environment in past years [2]-[6], they could not successfully draw customers' attention due to the limitations of IPTV environment such as very limited range of available contents, the lack of various free Apps, and the inconvenient interface for accessing media contents.

As commonly known, smart TV consists of TV, Internet access, and PC. The main difference between smart TV and IPTV is that smart TV has its own operating systems and software platforms, so that it can provide not only TV broadcasting and VODs but also various useful application services. The users are now able to freely access a number of contents through a TV platform. These features extend the reach of the smart TV service and allow the active participation of users. Furthermore, the large screen of smart TV makes it more attractive for users to enjoy various multimedia contents.

By combining the characteristics from the smart phone, smart TV is expected to become a catalyst to lead innovation of TV, broadcasting, advertising, and even home life. Smart TV as a digital hub in home network will provide users with various intelligent media services like *N-screen* by interacting with multiple home appliances such as smart phone and tablet PC. In order to realize a smart TV centric environment and facilitate the access, search, and sharing of multimedia contents in such an environment, the techniques for analysis of multimedia contents considering smart TV environment and its applications should be studied.

In this paper, we mainly focus on the video analysis technique in smart TV environment. In particular, an automatic video annotation technique for smart TV based on ontology infra-structure is presented. Various ontologies are exploited to build a common interface among both smart devices and users. We also propose efficient video concept detection approaches using semantic inference rules and SVM classifier for efficient video search, sharing, and browsing.

The rest of this paper is organized as follows. In Section II, we briefly review the characteristics of smart TV and its applications. In Section III, the overall architecture of the proposed work is described. Section IV presents the detailed description of the exploited ontologies and Section V describes the proposed video annotation approaches. Section VI shows the experimental results and analysis. Finally, we conclude our work in Section VII.

II. FEATURES OF SMART TV AND ITS APPLICATIONS

According to the survey of Korean domestic market for smart TV [1], a number of experts expect that the accumulated diffusion rate of smart TV in Korea will be 8.9 million households, which amounts to 52.6% of all households in Korea by 2022. The reason for this positive prospect comes from the following attractive features of smart TV.

- Open contents

  In traditional digital TV environment, contents providers are limited to only broadcasting companies and the kinds of available contents are also very limited. In contrast, thanks to Apps (widely consumed in the smart phone market), anyone can produce and distribute TV contents and unlimited types of TV contents are available

[1] All the authors are with the department of computer science and engineering, Hanyang University, Korea (e-mail: selphyr@hanyang.ac.kr, route@hanyang.ac.kr, dhlee72@hanyang.ac.kr).

Contributed Paper
Manuscript received 09/15/11
Current version published 12/27/11
Electronic version published 12/27/11.

in smart TV environment. This is the most remarkable characteristic of smart TV, which can create a service with newer concepts and diverse revenue sources.

- Entertainment & Communication

  Compared to previous digital TV environment, users of smart TV can more likely to be active participants due to the powerful Apps and the web accessibility of smart TV. A number of users already experienced the advantages of making use of various entertainments and communication features from smart phones. With more powerful environment and equipment of smart TV, these features will still be attractive to users.

- *N-screen* service

  In fact, *N-screen* service is not newly introduced technique in smart TV environment. Even though IPTV tried to support *N-screen* service between a mobile phone, table PC, and TV, the performance was not satisfactory due to the limitations such as the lack of well-structured interfaces among the devices, inconvenient connection, and a slow network bandwidth. One of the most important features of smart TV is that they have their own operating platforms. *N-screen* service will be more efficiently supported with the help of the common operating platform of smart TV and well-organized Wi-Fi infra-structure.

- Smart multi-tasking

  The efficiency of *N-screen* service is maximized when the devices can perfectly support a multi-tasking functionality. A traditional digital TV provides a simple multi-tasking service such as displaying information about another channel in a small window while watching, listing preferable channels at the moment of watching. However, multi-tasking on one screen can disturb customer's behaviors. If users can watch a TV program on TV screen, search the preferable channel list on his smart phone, and visit a shopping mall dealing in something appeared on the TV program via his tablet PC at the same time, it would be more attractive scenario. Fortunately, smart TV and current smart devices share a common operating platform and are expected to realize a smart multi-tasking service also known as one-source multi-use.

- Smart advertisement

  In IPTV or digital TV environment, an opportunity for the advertisement was very limited. The product vendors could only use regular advertisements or particular home shopping channels. On the other hand, the vendors will be able to promote their products by using Apps or embedded information in TV programs in smart TV environment.

- Smart home server

  In [1], most researchers and experts expect that smart TV will take the center position in smart life innovation. They also anticipate that smart TV as a digital hub will be able to provide a 3D, augmented reality, personalized broadcasting, u-Health, smart media, smart home, and smart work service as well as general TV programs.

In addition to the above features of smart TV, a large screen and large storage capability of smart TV is one of the most attractive characteristics in the perspective of video displaying, browsing, and retrieval. In contrast to smart phone environment with a small screen and small storage medium, smart TV is more suitable for displaying high quality video contents. Also, archiving and sharing of the large scale video contents is more feasible on smart TV.

Figure 1 depicts the overall architecture of the possible smart TV centric environment. The users will be able to access the unlimited types of multimedia contents from standard broadcasting, social network, and Apps through the smart devices connected with smart TV. With an easy access to Internet and the powerful Apps, smart TV will be the most attractive *lean-forward* media ever. However, without a well-structured interface for describing and searching multimedia contents, the users may feel again confused from the application overload as they felt from the information overload within early Internet. The more important problem is that, although we can make use of a well-structured interface for describing contents, the users will not tend to annotate multimedia contents by themselves through their limited interface (i.e., remote controller). Also, for efficient sharing of the video contents among smart devices and users, information of the video contents should be easily accessible via a well-organized interface.

From this reason, automatic annotation and summarization technique for multimedia contents is expected to become a core module for video sharing and search in smart TV environment. However, traditional video annotation approaches are heavily dependent on computer vision techniques and statistics-based algorithms. These techniques require high computational cost in terms of computing power and memory capacity. This means that traditional video annotation methods cannot be feasible on current smart TV because the computing power of smart TV is still similar to that of smart phone. Therefore, a new efficient video annotation technique with less computing overhead is required for smart TV and its applications.

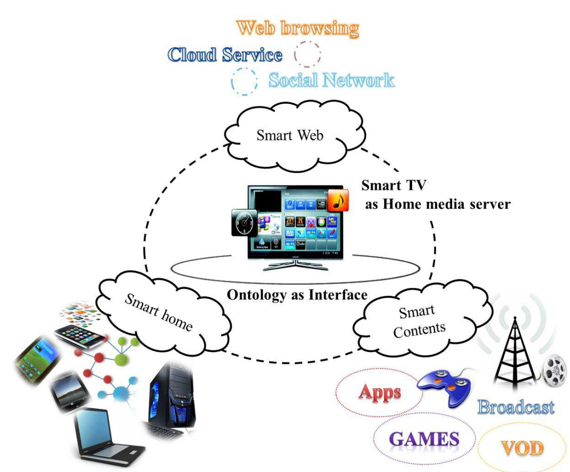

Fig. 1. Architecture of the smart TV centric environment

III. Overview of the Proposed Approach

We present a novel framework for representing multimedia contents through a well-structured ontology and automatically annotating video contents through the semantic inference rules and machine learning techniques. Figure 2 shows a system organization of the proposed approach. The proposed framework consists of the knowledge representation module including various ontologies and the automatic video annotation module. Suppose that smart TV can search video contents through the web with regard to user's preference or queries. The retrieved contents from various resources may then be stored into the storage of smart TV or streamed to other smart home appliances. In this stage, video contents are analyzed and the important semantic concepts are automatically extracted by the automatic video annotation module. In our annotation module, visual features of the video contents are extracted by MPEG-7 visual descriptors [7]. The extracted visual features are then mapped to the semi concept values for learning the semantic inference rules and support vector machine to detect the high-level concepts of the video contents. After analyzing the video contents, meta-data for video contents and various semantic concepts are represented in a form of ontology instances and stored into the ontologies in the knowledge representation module.

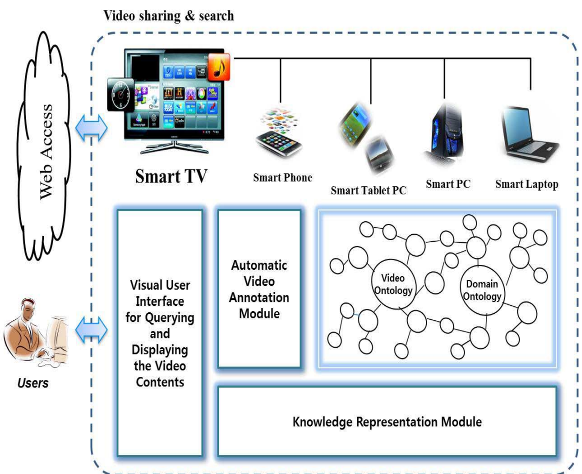

Fig. 2. System organization of the proposed approach

IV. Semantic Web Technologies for Video Annotation

A. Ontologies for Video Annotation

An ontology is commonly defined as an explicit specification of a conceptualization [8]. That is, ontology provides a description of the concepts and their relationships existing in a certain domain. For the aspect of the video retrieval, the use of ontologies can make it possible to efficiently and effectively retrieve the desired contents. In our work, the two kinds of ontologies are exploited.

- Ontology for content description: this ontology provides a formal and sharable scheme for storing information of the video contents. For this, we propose *VideoAnnotation* ontology which can hierarchically describe video contents.
- Ontologies for domain knowledge acquisition: this ontology provides a knowledge-base of a specific domain for annotating video contents. For sharable video annotation, LSCOM ontology [9] and object ontology are exploited.

B. VideoAnnotation Ontology

The video contents can be generally represented by using a hierarchy consisting of the five layers (video, scenes, groups, shots, frames) from top to bottom in increasing granularity for content expression [10]. Consequently, for efficient video annotation, a knowledge management scheme comprehensive for the structure of video contents is required.

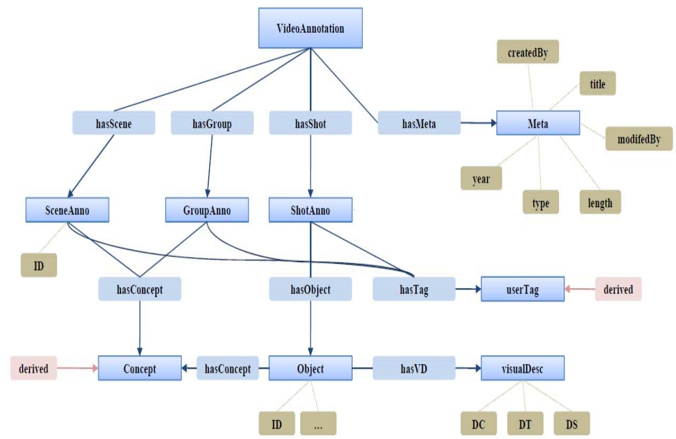

Fig. 3. A partial structure of *VideoAnnotation* ontology

The partial structure of *VideoAnnotation* ontology is depicted in Fig. 3, where four important classes, *ShotAnno*, *GroupAnno*, *SceneAnno*, and *Meta* are adopted with each description defined as follows:

- *ShotAnno* class: it describes objects that exist in the representative frame (key-frame) of each video shot. Visual information of an object is also described in visualDesc class and the high-level concept of an object is described in Concept class. Each class is concerned with ShotAnno class by hasVD relationship and hasConcept relationship, respectively.
- *GroupAnno* class: it describes information about video groups. A video group consists of several video shots. In this class, information about dominant concepts in the group and group meta-data are described.
- *SceneAnno* class: it describes the concepts of each video scene. A video scene represents more sophisticated and semantic information of the video content. For example, in case of a tiger documentary, we can use high-level concepts such as "animal-tracking", "interview with tamer" for the scene annotation. These keywords are derived from LSCOM ontology which provides standard index terms to annotate the video contents.
- *Meta* class: it presents the meta data of a video sequence such as a title, a type of the video contents.

In annotation process, information with different granularity is automatically extracted from the video contents by using the proposed high-level concept extraction methods or initially embedded video information provided by contents vendors. And then, the semantic information of the video contents is stored as instances of each class in *VideoAnnotation* ontology.

C. Domain Ontology

As mentioned in Section IV-A, domain ontology is required for providing a sharable and reusable vocabulary when annotating the video contents of a specific domain. In our work, we exploit the two types of domain ontologies: 1) LSCOM ontology, and 2) object ontology.

The LSCOM (Large Scale Concept Ontology for Multimedia) ontology provides a set of standardized concepts for video annotation. In their full ontology, approximately 3,000 high-level concepts for video annotation are provided. However, we exploited a light-weight version of the LSCOM ontology which provides approximately 400 high-level concepts. In our work, the concepts in this ontology are used for both a group-level and scene-level annotation.

The object ontology is a knowledge-base for a specific domain. This type of ontology can be made manually or derived from an existing ontology such as WordNet [11]. The concepts in this ontology are used for a shot-level annotation. In our work, the result of the shot-level concept extraction is mapped to the corresponding concept of WordNet.

V. High-level Concept Extraction

For the efficient high-level concept extraction of the video shots, we propose the two methods where the semantic inference rules and machine learning technique are combined.

Figure 4 depicts the overall procedure for the extraction of the high-level concepts from the video shot. Given a video shot and its salient object, low-level visual features of the object are automatically extracted by MPEG-7 visual descriptors. And then, these visual features are mapped to their corresponding semi-concept values. Semi-concept value is a simple keyword assigned according to the quantity of each visual feature. As shown in Fig. 4, the low-level color feature is mapped to its corresponding semi-concept value like “Red-Orange”. Afterwards, semantic inference rules and support vector machine (SVM) are learned based on a set of semi-concept values. The final classification is performed through the semantic inference rules and SVM classifiers.

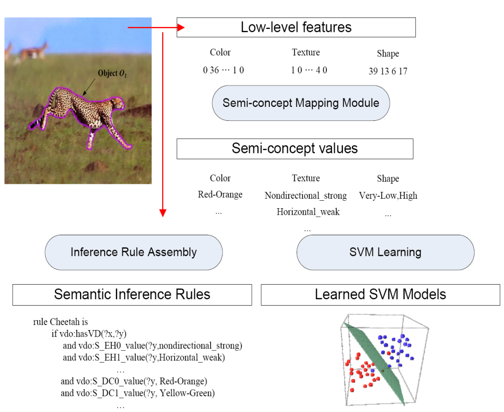

Fig. 4. Overall procedure for high-level concept extraction

A. Key Frame Detection and Object Segmentation

The one video shot consists of multiple video frames. To annotate all the frames in each shot is not meaningful, so that the key frame of each video shot is determined and the concept extraction procedure is performed only on the key frame.

The detection of a video shot and its corresponding key frame are performed based on the visual similarity between adjacent video frames. We adopted color structure descriptor (CSD) [13] for representing color feature of each video frame. The CSD represents an image by both the color distribution of images and the local spatial structure of the color using HMMD color space. To calculate the visual similarity between video frames, we use Euclidean distance measure. The shot boundary is detected by the following simple procedure:

1. For each video frame v_i (i = 0 ... *N,* where *N* is a total length of a video sequence), calculate a visual similarity between v_x and its subsequent frame v_y.
2. Determine frame v_y as a shot boundary if the similarity between frame v_x and v_y exceeds a threshold.
3. Repeat 1 and 2 until all frames are examined.

For simplicity, we choose the first frame of each shot as a key frame. After selecting all key frames from the video sequence, the region of interest is segmented from the key-frames using the segmentation algorithm proposed in [12].

B. Semi-concept Mapping

In this stage, the low-level visual features of an object are extracted by MPEG-7 visual descriptors and then mapped to their corresponding semi-concept values. For the color feature, we exploit MPEG-7 color structure descriptor (CSD). The detailed description of the semi-concept mapping for the color feature is as follows:

Algorithm 1 (Color semi-concept mapping)

Input : a region of interest for an object
Output : a set of semi-concept values for color feature

1. CSD feature of an object is extracted and quantized to a 128-bin CSD histogram
2. The proportion of each semi-concept color is calculated
3. Select *k* most frequently appeared semi-concept colors and return as an output

A set of semi-concept values for color feature C_i (i = 0… 10) is defined as {*Red-Orange*, *Yellow-YellowGreen,* …, *Gray*, *White*}. The proportion of each semi-concept C_i is calculated by summation of the values for each bin h_i in the quantized CSD histogram. The detailed equation is as follows:

$$C_{j\in\{0,\ldots,7\}} = \sum_{i=0}^{2} S_{ij},\ C_8 = S_{28} + S_{29},\ C_9 = S_{30} + S_{31} + S_{32} + S_{33},$$

$$C_{10} = S_{34} + S_{35},\ \text{where}\ S_{00} = (h_0 + \ldots + h_3),\ S_{01} = (h_4 + \ldots + h_7),$$

$$S_{35} = h_{124} + h_{125} + h_{126} + h_{127}$$

The texture feature of an object is extracted by MPEG-7 Edge Histogram Descriptor (EHD) which represents the distribution of 5 types of edges [14]. The edges in the image are categorized into 5 types: vertical, horizontal, 45-degree diagonal, 135-degree diagonal and non-directional edges. The result of the texture extraction by EHD is an 80-bin feature vector and each bin has its own semantics in terms of location and edge type. The semi-concept mapping algorithm for the texture feature is described in *Algorithm* 2.

Algorithm 2 (Texture semi-concept mapping)

Input : a region of interest for an object
Output : a set of semi-concept values for texture feature

1. EHD feature of an object is extracted and quantized to an 80-bin EHD histogram.

2. The proportion of each semi-concept texture E_j (j=0..4) is calculated by $E_{j\in\{0,\dots,4\}} = \sum_{i=0}^{15} h_{ij}$, where

$$semi-concept\ texture\ E_j = \begin{Bmatrix} Vertical & j=0 \\ Horizontal & j=1 \\ 45-degree\,diagonal & j=2 \\ 135-degree\,diagonal & j=3 \\ Nondirectional & j=4 \end{Bmatrix}$$

3. Select k most frequently appeared semi-concept textures and return as an output**.**

In order to describe a shape feature of a region in the image, MPEG-7 provides region-based shape descriptor and contour shape descriptor (contour-SD). The region-based shape descriptor basically represents a pixel distribution within a region. This descriptor is able to describe the complex objects consisting of multiple disconnected regions such as company logo and trademark. On the other hand, contour-SD efficiently describes the objects with a single contour. In particular, animal objects such as tiger, horse are much better captured by the contour-SD. Therefore, we make use of the contour-SD to describe the shape information of salient objects in an image. The mapping procedure of the elements in contour-SD, which represent the eccentricity and circularity of the shape of an object, to the semi-concept value is briefly depicted in Fig 5.

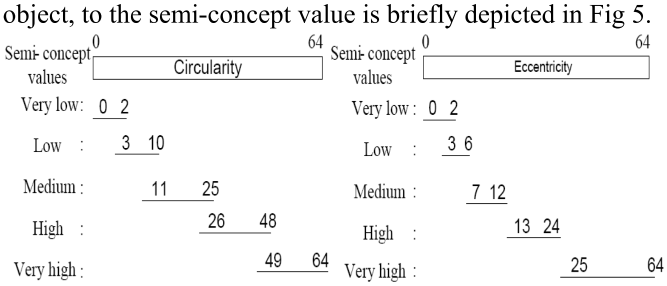

Fig. 5. Mapping contour-SD to semi-concept values

C. Semantic Inference Rules for High-level Concept Extraction

The aim of applying the inference rules is to derive new knowledge from the existing knowledge in a specific domain. Currently, various rule engines for OWL reasoning have been proposed [15, 16]. In our work, we use *Bossam* rule engine [17] which provides a rule language called *Buchingae*.

The key idea for the detection of the semantic inference rules for high-level concept extraction in a video shot is based on the observation that the visual features of the same objects are very similar to each other. Based on this observation, the most common semi-concept values for an object are defined as the terms of the shot-level semantic inference rules.

In order to intuitively show the procedure for extracting high-level concept from a video shot, we take the object in Fig. 4 as an example. The inference procedure is performed as follows:

1. The low-level visual features of an object in the video shot are extracted by MPEG-7 visual descriptors and then mapped to the corresponding semi-concept values. Assume that the extracted semi-concept values are:
   Color = {"Red_Orange", "Yellow_Green", "Black"}
   Texture = {"Non-directional", "Horizontal"}
2. Apply rule Cheetah to determine the high-level concept of the object in the video shot.

Suppose that rule Cheetah is found by the observation of the training data as described in Fig 6. In this case, the object in Fig. 4 satisfies rule *Cheetah* because it has non-directional, horizontal for texture feature and red-orange, yellow-green for color feature. Therefore, the concept of the object is determined as a cheetah.

```
prefix rdf = http://www.w3.org/1999/02/22-rdf-syntax-ns#;
prefix owl = http://www.w3.org/2002/07/owl#;
prefix xsd = http://www.w3.org/2001/XMLSchema#;
prefix rdfs = http://www.w3.org/2000/01/rdf-schema#;
prefix vdo = http://example.org/videoAnnotation#,
prefix obj = http://example.org/Object#
rulebase Cheetah
{
      rule Cheetah is
        if vdo:hasVD(?x,?y) and
           vdo:S_EH0_value(?y, "Nondirectional") and
           vdo:S_EH1_value(?y, "Horizontal") and
                                     ...
           vdo:S_DC0_value(?y, "Red-Orange") and
           vdo:S_DC1_value(?y, "YellowGreen") and
                                     ...
           then obj:Cheetah(?x);
}
```

Fig. 6. An example of semantic inference rule

D. SVM Learning for High-level Concept Extraction

Based on the semi-concept values, the training procedure for an object SVM classifier is performed. The feature vector of the training set for a particular concept C_i is represented as $T_i = \{L_i, DC_{0-3}, DT, DS\}$, where L_i is a label, $DC_{0\text{-}3}$ are four color semi-concepts, DT is a texture semi-concept, and DS is a shape semi-concept for C_i, respectively. In order to take into account the limited computing power of smart TV, we exploit a feature vector consisting of a set of semi-concept values rather than high-dimensional low-level visual features.

Figure 7 shows an example of the annotation results performed by the proposed approaches. Each video shot is represented by a concept (instance) of WordNet and LSCOM ontology. Finally, the annotation results and meta information of the video contents are stored in *VideoAnnotation* ontology.

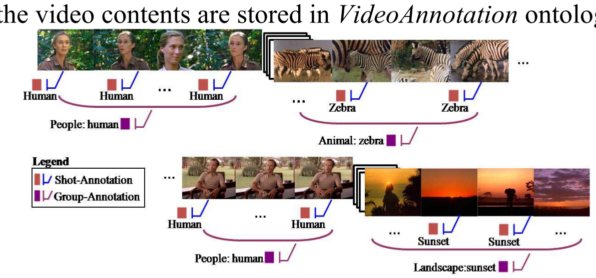

Fig. 7. An example of video annotation

VI. EXPERIMENTS AND ANALYSIS

In this section, we show the effectiveness of the proposed approach and analyze the experiment results. In this work, we focus on the accuracy of annotation for the video shots because the result of the further processes is mainly dependent on the accuracy of the shot-level annotation.

We performed the three kinds of evaluations for the proposed approaches to investigate 1) the effectiveness of the semantic inference rule-based approach, 2) the effectiveness of the semi-concept-based SVM classifiers, and 3) the relative effectiveness of the both proposed approaches. For the performance evaluation, we measured precision and recall.

We have implemented the proposed system by using Java 1.6 (ImageJ library and MPEG-7 XM) on a barebone system equipped with Atom 1.8GHz processor, 2 GB main memory, and 250 GB secondary storage. Also, all experiments were performed on the same system. The evaluation is performed on the parts of *National Geographic* videos containing various concepts such as animals, landscape scene.

A. Effectiveness of Semantic Rule-based Approach

In order to investigate the effectiveness of the semantic inference rule based concept detection approach (SemRule), Bayes-classifier (Bayes) is selected as a baseline method. The evaluation set consists of various video shots containing a horse, tiger, wolf, dolphins, snow, forest, sky, and so on.

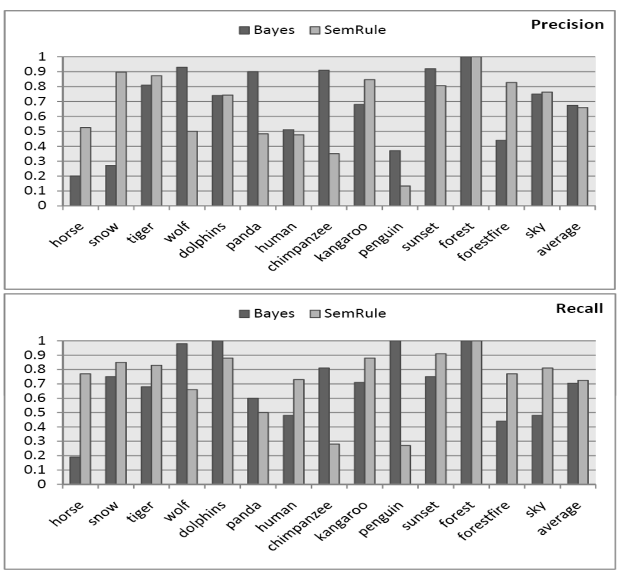

Fig. 8. Performance evaluation between *Bayes* classifier and SemRule

Figure 8 shows the experimental results between the proposed approach and Bayes-classifier. Compared to Bayes classifier, the semantic inference rule based approach does not outperform, but still shows a reasonable performance with respect to precision. Specifically, the proposed approach shows a high precision for the concepts whose visual features are distinct. On the other hand, with respect to recall, the proposed approach shows better performance compared to Bayes-classifier. It is worth noting that the proposed semantic rule inference based approach shows a competitive performance without high computational cost. This point must be considered especially important in smart TV environments because the computing power of CPU used in smart TV is still similar to that of smart phone.

B. Effectiveness of Semi-concept based SVM Classifier

The proposed semi-concept based SVM classifier is evaluated with one of the lazy unsupervised classification algorithms, *k-nn* classifier. For this experiment, some additional complex objects such as a boat, castle, and sculpture are included in the data set. Figure 9 depicts the experimental results in terms of precision and recall.

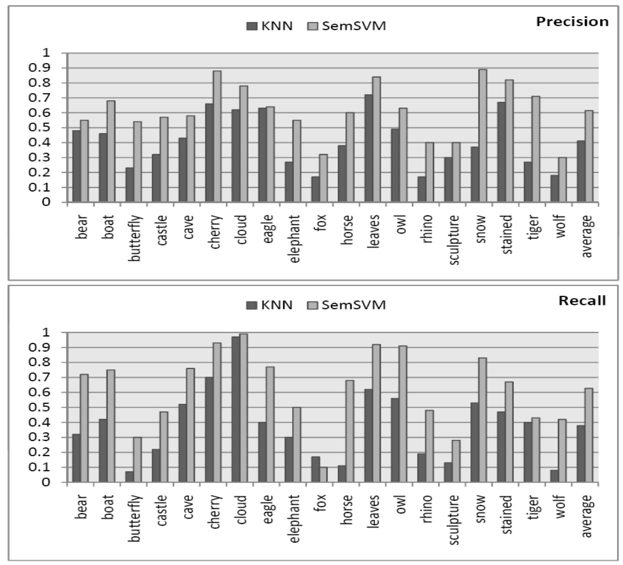

Fig. 9. Performance evaluation between *k-nn* classifier and SemSVM

As shown in Fig. 9, semi-concept based SVM classifier outperforms *k-nn* classifier for the all cases. Although the feature space of the training set is drastically reduced because of the semi-concept values, SVM classifier shows a reasonable performance for even complicated objects such as a castle, cave, and boat. If we exploit high-dimensional low-level visual features for training SVM classifiers, we may get higher precision and recall. However, the size of the learned model and computational cost will increase exponentially. This can also cause a critical problem in smart TV environments.

C. Relative Effectiveness of the Proposed Approaches

Finally, we evaluated the performance of the semantic rule based concept detection approach and the semi-concept based SVM classifier. The two approaches are learned based on the semi-concept values from the data sets used in our first experiment.

Figure 10 shows the experimental results. As expected, the semi-concept based SVM classifier outperforms the rule-based approach. The precision and recall of the SVM classifier reached almost 80%. However, this result does not indicate that the semi-concept based SVM classifier is always the best solution for smart TV. As discussed before, although the SVM learner shows better performance compared to other approaches, it requires high computational cost (even though we reduce the feature space by means of the semi-concept values) and a large memory space. On the other hand, the semantic inference rule based approach shows relatively lower but a reasonable performance without high computational cost that can be caused by statistics-based algorithms. Therefore, there is a trade-off between the proposed approaches in smart TV environment. If one can use a high-end smart TV equipped with a high performance CPU and a large-sized memory, the SVM-based approach is a better choice. Otherwise, semantic inference rule based approach is a more attractive solution.

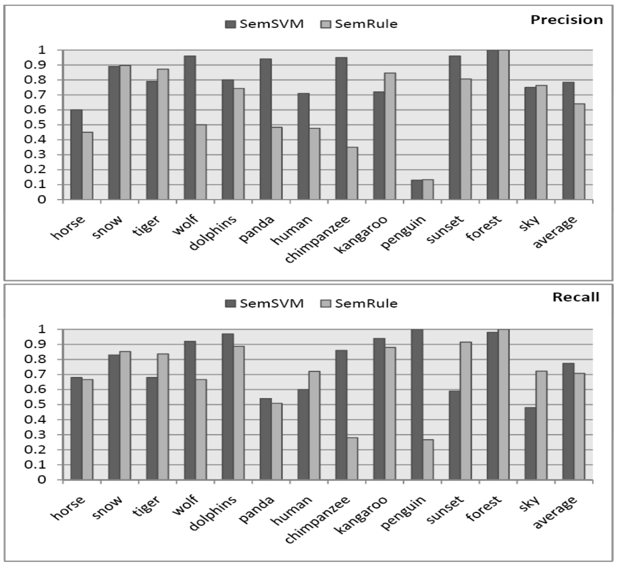

Fig. 10. Performance evaluation between the proposed methods

VII. Conclusions

We have presented an automatic video annotation technique which employs the ontologies to facilitate the video retrieval and sharing process in smart TV environment. Also, we proposed the two kinds of high-level concept extraction methods: 1) semantic inference rule based approach, and 2) semi-concept based SVM classifier. From the experimental results, we show that the proposed semi-concept based approaches achieve a competitive performance compared to conventional classification algorithms.

Also, the proposed approach has the following advantages. First, our system is fully automatic. Second, the proposed methods can be easily applied to other domains without a serious modification. Third, our approach provides a reasonable performance in terms of video annotation with a low complexity. Fourth, our approaches provide high inter-operability by adopting MPEG-7 descriptors and LSCOM standard ontology. Consequently, the proposed method can be successfully applied to smart TV environment due to the aforementioned features.

References

[1] MoonKoo Kim, JongHyun Park, “Demand forecasting and strategies for the successfully development of the smart TV in Korea*” Proceeding of* International Conference on Advanced Communication Technology, pp.1475-1478, 2011

[2] Choonsung Shin, Woontack Woo, “Socially aware tv program recommender for multiple viewers”, *IEEE Trans. Consumer Electron.,* vol. 55, no. 2, pp. 927-932, May. 2009

[3] Matsubara, F. M. , Kawamori, M. ,”Lightweight interactive multimedia environment for TV”, *IEEE Trans. Consumer Electron.,* vol.57, no. 1, pp.283-287, Feb. 2011

[4] Meng-Huang Lee, “The design of a heuristic algorithm for IPTV web page navigation by using remote controller”, *IEEE Trans. Consumer Electron.,* vol. 56, no. 3, pp.1775-1781, Aug. 2010

[5] Meng-Huang Lee, “Apply relay recording and video segment annotatino for IPTV network personal video recorder”, *IEEE Trans. Consumer Electron.,* vol.56, no. 4, pp.2364-2372, Nov. 2010

[6] Yolanda Blanco-Fernandez, Jose J. Pazos-Arias, Alberto Gil-Solla, Manuel Ramos-Cabrer, Martin Lopez-Nores, “Providing entertainment by content-basd filtering and semantic reasoning in intelligent recommender systems”, *IEEE Trans. Consumer Electron*., vol.54, no. 2, pp.574-735, May. 2008

[7] ISO/IEC 15938-5 FDIS Information Technology: MPEG-7 Multimedia Content Description Interface - Part 5: Multimedia Description Schemes. 2001

[8] Thomas R. Gruber. “A translation approach to portable ontology specifications”, Knowledge Acquisition, vol. 5, no. 2, pp.199-220, 1993

[9] L. Kennedy and A. Hauptmann "LSCOM lexicon definitions and annotations version 1.0", DTO Challenge Workshop on Large Scale Concept Ontology for Multimedia. ADVENT Technical Report #217-2006-3, Columbia University, March 2006

[10] Xingquan Zhu, Jianping Fan, Xiangyang Xue, Lide Wu and Ahmed K. Elmagarmid,“Semi-automatic video content annotation”, Proceeding of Third IEEE Pacific Rim Conference on Multimedia, pp. 37-52, 2002

[11] George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.

[12] M. Jacob, Blu and M. Unser, “Efficient energies and algorithms for parametric snakes”, IEEE Transactions on Image Processing; vol. 13, no. 9, pp.1231-1244, 2004

[13] Messing, D. S. ; van Beek, P.; Errico, J. H., “The MPEG-7 colour structure descriptor: image description using colour and local spatial information”, Proceedings of International Conference on Image Processing, pp.670-673, October 2001

[14] Chee Sun Won, Dong Kwon Park, and Soo-Jun Park., “Efficient use of Mpeg-7 edge histogram descriptor”, ETRI Journal 2002; vol. 24, no. , pp.23-30

[15] Harold Boley, Said Tabet, and Gerd Wagner, “Design Rationale of RuleML: A Markup Language for Semantic Web Rules”, Proceeding of International Semantic Web Working Symposium, pp.381-401, 2001

[16] Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean, “SWRL: A semantic web rule language combining owl and ruleMl”, W3C Member submission, May, 2004

[17] Minsu Jang and Joo-Chan Sohn, Bossam: An Extended Rule Engine for OWL Inferencing, Proceedings of RuleML (Lecture Notes in Computer Science, vol.3323), pp.128-138, 2004

Biographies

Jin-Woo Jeong received his BS and MS degree in computer science and engineering from Hanyang University, South Korea, in 2006 and 2008, respectively. He is currently enrolled for PhD degree in Hanyang University and His current interests include contents-based image retrieval and social image search.

Hyun-Ki Hong received his BS degree and MS degree from Hanyang University, South Korea, in 2009 and 2011, respectively. He is currently enrolled for PhD degree in knowledge and data engineering lab of Hanyang University. His current interests include database, multimedia information retrieval, and so on.

Dong-Ho Lee received his BS degree in computer engineering from Hongik University, and his MS and PhD degrees in computer engineering from Seoul National University, South Korea, in 1995, 1997 and 2001, respectively. From 2001 until 2004, he worked in software center, SAMSUNG Electronics Co. Ltd., where he was involved in several research projects for next-generation digital appliances. He is currently an associate professor in the department of computer science and engineering at Hanyang University, South Korea. His research interests include system software for flash memory, embedded DBMS, and multimedia systems.