A collaborative algorithm for semantic video annotation using a consensus-based social network analysis

Trong Hai Duong a, Ngoc Thanh Nguyen b,∗, Hai Bang Truong c, Van Huan Nguyen d

aInternational University, Vietnam National University-HCM, Viet Nam
bInstitute of Informatics, Wroclaw University of Technology, Poland
cUniversity of Information Technology, Ho Chi Minh City, Viet Nam
dNguyen Tat Thanh University, Viet Nam

A R T I C L E  I N F O

Article history:
Available online 7 August 2014

Keywords:
Video annotation
Semantic web
Social network
Consensus

A B S T R A C T

Social TV represents a new form of shopping that enables consumers to view, select and buy products. This highlights the need for a collaborative video annotation technique. This paper proposes a collaborative algorithm for semantic video annotation using consensus-based social network analysis (SNA). The collaborative video annotation process is organized based on social networks. Here, the media content is shared with friends of friends who collaboratively annotate it. This study used an ontology-based approach to semantically describe the media content and allow sharing between users. A consensus-based method was used to reconcile conflicts between participants’ annotations. The experimental results indicated that the use of SNA-based collaboration criteria to evaluate the collaborative process enhances the completeness and consistency of collaborative annotation. The more the collaboration criteria are satisfied by the collaborative group, the faster the group will reach a consensus. In addition, the consensus-based method is an effective approach for resolving conflicts on collaborative annotation.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Social TV represents a new form of shopping that enables consumers to view, select and purchase products. For this, the sellers annotate videos and associate items with information from e-commerce systems in a semantic manner. On the other hand, manufacturers annotate the way items are produced. In addition, buyers provide their opinions on the items received. Therefore, there is a need for a collaborative video annotation method. Collaborative video annotation is a process, in which a group of participants contribute annotations to specific videos of interest. Annotation can be improved by collaboration between participants because they can benefit from each others’ skill and knowledge. According to Duong and Jo (2010) and Duong, Nguyen, and Jo (2010), any effective collaboration should be satisfied criteria including inclusive, egalitarian, interactive, coordinated, and trustworthy. Here, the term inclusive means that there should be a sufficient number of participants. The term egalitarian indicates that these participants should be provided with as many opportunities as possible for their collaboration. Interactive suggests that there should be an easy way to establish contact with all collaborators within a collaborative group. The term coordinated means that any annotation information shared within the group should be easy to access and the term trustworthy indicates that the participants should be allowed to edit annotations based their behavior. Therefore, the workflow of a collaborative video annotation process is complex. According to Palau, Montaner, and Lopez (2004), SNA can highlight many relationships based on the aforementioned collaboration criteria and entail a range of centrality measures, such as density, degree centrality, closeness centrality, and between centrality. If the collaborative relationships in video annotation systems can be represented by means of social network, these centrality measures can be used to organize the collaborative annotation process and evaluate the collaboration criteria. For example, the closeness centrality can be used to identify those annotators occupying more advantageous positions for obtaining annotation support. The density is useful for evaluating the conditions for accessing and sharing annotation information between collaborators in a coordinated manner. In addition, consensus-based methods seem to be an effective approach for resolving conflicts in collaborative annotation.

This paper proposes a collaborative algorithm for semantic video annotation using a consensus-based SNA. The ontologies for semantically describing media content and facilitating sharing

∗ Corresponding author.
E-mail addresses: haiduongtrong@gmail.com (T.H. Duong), ngoc-thanh.nguyen@pwr.edu.pl (N.T. Nguyen), bangtbh@uit.edu.vn (H.B. Truong), congthuan_bkstudent@yahoo.com (V.H. Nguyen).

http://dx.doi.org/10.1016/j.eswa.2014.07.046
0957-4174/© 2014 Elsevier Ltd. All rights reserved.
between heterogeneous users or devices (e.g., smart-phone and PDAs) were considered. This study employed two ontologies. One ontology describes visual features of media objects, such as their color, texture, shape, motion and position (e.g., MPEG-7), and the other provides a knowledge base for a specific domain (domain ontology) for annotating the video content. The LSCOM ontology reported by Naphade et al. (2006), which is used in broadcast news video annotation, was considered for the domain ontology. Each media object in a specific video can be considered as an instance of a domain concept belonging to the LSCOM ontology. Here, the annotator chooses a relevant concept to describe a media object of interest and the collaborative video annotation process is organized based on a social network. The media content is shared within a large network of friends of friends who annotate the content based on the SNA. Annotators collaborate only with reliable users. Reliability is expressed through trust scores and other collaboration criteria. Once the annotator has a set of reliable users, he or she can benefit from their knowledge or skill for annotation purposes. When the annotator is unsure of the annotation information, he or she asks reliable users for their knowledge and opinions, and uses their trust scores to determine if the annotated object should be of interest. This study suggests that reliable users provide pertinent opinions while evaluating the entire collaborative process using collaboration criteria but that they can provide insufficient or conflicting knowledge. A consensus choice was used to resolve conflicts over annotation information between participants.

The remainder of this paper is organized as follows. Section 2 provides a literature review. Section 3 proposes a collaborative semantic annotation methodology using a consensus-based SNA. Section 4 presents the experimental results, and Section 5 reports the conclusions.

2. Related work

With the recent explosive growth of smart or social TV services, there have been considerable increases in social videos as well as traditional digital videos, such as TV programs and VOD offerings. On the other hand, it is difficult to find interesting and relevant media content and encourage interactions between users and Smart TV platforms due to the lack of semantic media content. This highlights the urgent need for semantic techniques to analyze the multimedia content, particularly those for a semantic video annotation.

2.1. Multimedia ontology

Ontologies have the following attributes: provide machine-processable information sources; enable users to organize information according to taxonomical concepts based on their own attributes; allow for a semantic description of the media content; and facilitate the sharing of media content. Each media object can be considered as an instance of a domain concept belonging to multimedia ontology, which offers advanced search functions that can include searches for similar or related information integrated from a range of social media sites (Hakeem, Lee, Javed, & Haering, 2009; Sebastine, Thuraisingham, & Prabhakaran, 2009; Wei, Zhao, Zhu, & Liu, 2010). Several multimedia ontologies have been proposed. Among them, the LSCOM ontology was proposed by Naphade et al. (2006) includes a set of 1,000 concepts built collaboratively by IBM, Carnegie Mellon University, and Columbia University in conjunction with CyC Corporation and various research communities. The aim is to create a framework for ongoing research on the semantic analysis of multimedia content. The Video Event Representation Language (VERL) of Francois, Nevatia, Hobbs, Bolles, and Smith (2005) is an event ontology used to describe the events and objects in videos. Here, video events and objects were annotated using the Video Event Markup Language (VEML). Annotations draw on the VEML for organizing content. Unambiguous data sharing between users enables marking up data streams using the VERL and the VEML. In addition, annotation data is accessible to automatic machine manipulations for indexing or inferring. The MPEG-7 multimedia content description standard already provides tools for representing media content fragments. The MPEG-7 Visual Part supports the objects color (e.g., dominant colors and color layouts), texture, shape (e.g., region/contour-based), and motion (local or global) descriptors (Sikora, 2011). The ultimate goal and objective of the MPEG-7 Visual Standard is to provide standardized descriptions of streamed or stored images or header bits that assist users or applications to identify, categorize or filter images or videos. Similarly, the MPEG-7 Multimedia Description Scheme (MDS) supports spatial (directional or topological) and temporal multimedia segment relationships as well as the hierarchical structures for the decomposition of multimedia segments.

2.2. Automatic semantic video annotation

In general, automatic semantic video annotation also referred to as video concept detection (Jeong, Hong, & Lee, 2011; Naphade & Smith, 2004; Park, Lee, Moon, Park, & Lee, 2007), video semantic analysis (Snoek, Worring, & Smeulders, 2005), and high-level feature extraction (Kraaij & Over, 2005; Wu & Li, 2011) can be performed using machine learning methods. This paper presents the representative learning-based video annotation methods as follows. According to Park et al. (2007) and Jeong et al. (2011), an automatic semantic video annotation system employs ontologies and semantic inference rules to facilitate video retrieval and consists of four functional modules: shot-level annotation, group-level annotation, scene-level annotation and video-level summarization. The visual features of objects in a video shot are first extracted using MPEG-7 visual descriptors, which are then mapped to the corresponding semi-concept values that are used as the basic units for inferring high-level concepts. The high-level concepts of objects are extracted automatically by applying shot-level inference rules to ontologies, and semantically-similar groups are parsed into one semantic scene based on the frequency and similarity of high-level concepts from each group. Finally, a video-level summary is generated by integrating and analyzing the high-level concepts of groups and scenes.

2.3. Collaborative semantic video annotation

Collaborative semantic video annotation methods focus on a web-based annotation interface. The participants collaboratively create and share video annotations through social networks. Several approaches can be used to achieve this. The IBM Efficient Video Annotation (EVA) system is a server-based tool for the semantic annotation of large video and image collections, and allows user access through web browsers Vollmer, Smith, and Natsev, 2005. The users can be assigned to specific tasks such that each annotator can have a personalized view of the collection. Tasks can be shared by a variety of users. The novel features of the system include its ability to collect aggregate-level user data during the annotation process and support inter-annotator analyses.

However, this system was designed with a few simplifying assumptions to promote consistency, simplicity, and speed of annotation:

- All annotations use an upper ontology and no free text is allowed;
– All annotations are for static visual concepts only;
– This system exclusively allows the users’ contributions to the entire shot without object-identifiable;
– This study has not yet considered conflicts among participants’ annotation.

A video scene annotation method based on a social media system provides collaborative tagging for video annotation with two elements, an annotation interface based on the tag-cloud and a function for sharing tags with other users (Yamamoto, Masuda, Ohira, & Nagao, 2008a, 2008b). The user comments associated with a video are first collected from existing video sharing services, and a tag cloud is then generated from these comments. The tag cloud is displayed on the video window of the web browser. When the users click on a tag in the cloud while watching the video, the tag becomes associated with the video time point. The users can share information on the tags that are already clicked. Here, this method provides a wider coverage of annotations than other methods. Moreover, the users are motivated to add tags using tag-sharing and tag-cloud methods. Owing to this the development of high-quality and advanced video applications can be facilitated.

Another work (Juan et al., 2014) which addressed an approach to assessing semantic annotation based on formal concept analysis (FCA). Annotators use a predefined ontology of domain experts. The annotated collections are automatically analyzed using FCA to be a lattice-based graphical representation which allows domain experts can assess how the proposed ontology is being used by annotators. Based on the usage of the ontology, the domain experts can refine the ontology and give a guidance to annotators for annotation.

However, the aforementioned methods have not yet explored the following questions:

– How can annotation information be shared effectively between the users of a social network, and how can the users adequately understand the shared annotation in order to contribute to the collaboration?
– Who should share the annotating object to afford as many opportunities as possible for the collaboration and to benefit as much as possible from the user’s social skills and collaboration?
– Conflicts among the annotated versions of the object are unavoidable. How well can the conflict-annotated versions of the object being contributed by participants be integrated in order to obtain the best version of the annotations for the object?

How can annotation information be shared effectively between the users of a social network, and how can the users adequately understand the shared annotation in order to contribute to the collaboration? Ontologies have been developed to provide machine-processable semantic information resources that can be communicated between different agents (software and humans). The idea of ontologies is to allow users to organize information according to the concept taxonomies, with their own attributes, in order to describe the relationships between the concepts. In this way, a multimedia ontology is exploited to semantically describe the media content and facilitate the sharing of media content among heterogeneous users. Each media object can be considered to be an instance of a domain concept belonging to the multimedia ontology. It also offers advanced search functionality that includes searches for similar or related information integrated from different social media websites.

Who should share the annotating object to afford as many opportunities as possible for the collaboration and to benefit as much as possible from the user’s social skills and collaboration? To the best of our knowledge, there is no work exploring the aforementioned problem. Here, the collaborative video annotation process is organized via social networking. Beyond profiles, friends, comments, and private messaging, social networking sites vary greatly in their features and user base. Some have photo-sharing or video-sharing capabilities; others have built-in blogging and instant messaging technology. There are mobile specific SNSs (e.g., Dodgeball), but some web-based SNSs also support limited mobile interactions (e.g., Facebook, MySpace, and Cyworld). For these social networking advantages, the semantic media content can be shared with a large network of friends of friends who are annotating on it based on their available social profiles. Through this, a participant does not have enough knowledge/experience to annotate an interesting object in a specific video, he/she will share the object to his/her friends who are also interested in it. Thus, there may be many users participating in the collaborative annotation workflow via social networking. However, the participant does not simply share the object of interest to every friend in his/her network. The shared friends need to satisfy the collaboration criteria. Egalitarianism, one of the most important criteria, means that participants need to be afforded as many opportunities for collaboration as possible. To select an egalitarian candidate, we assume that each user in the social network has a profile that describes his/her interests, knowledge, and experience. Therefore, the egalitarian degree of each user is determined by the similarity between his/her profile with the annotating object profile. Another important criterion is Trust. How can someone trust a social user’s contribution? That is why we do not allow a participant to share the object annotation to everyone in their network. In addition, the degree of trust is determined by the relationship between the annotating participant and his/her friends. If they are more closed, they can better satisfy the Interactive criterion.

How well can the conflict-annotated versions of the object being contributed to by participants be integrated in order to obtain the best version of the annotations for the object? There may be many people annotating the same object in a specific video. This means there may be many annotated versions/profiles of the same video being contributed by participants. Even though these participants are trusted via social networking analysis, the conflicts among the annotated versions are unavoidable. Fortunately, one well-known conflict situation is called the conflict profile, and a consensus method is an effective approach that can be used to solve this type of conflict. In a conflict profile, there are different sets of knowledge that explain the same goal or elements in the real world. The consensus aims to determine a reconciled version of knowledge which best represents the given versions. Therefore, the consensus choice can be applied to determine the best annotation of the object from the conflicting participants’ points of view. In this work we extend the concept presented in the conference paper (Duong, Tran, Jung, & Nguyen, 2014). We present in details the consensus-based model for collaborative video annotation.

3. Collaborative semantic annotation using a consensus-based SNA

3.1. A collaborative framework for semantic video annotation

There are two well-known methods for collaborative work: the Delphi method worked out by Pill (1971) and Nominal Group Technique (NGT) issued by Gallagher, Hares, Spencer, Bradshaw, and Webb (1993). The former is a consensus-based collaborative method and is more popular than the latter. The Delphi method is used for distributed discussion requiring non-complex communication between experts including face-to-face meetings. In this paper, this method was used to implement a collaborative framework for semantic video annotation, which is divided into the following phases:
• Phase 1-Preparatory: This phase provides the collaboration criteria for guiding and evaluating the collaborative process. In this phase, the media objects are introduced to the participants, and a domain ontology is used to describe the media content semantically. Each media object in a specific video can be considered as an instance of a domain concept belonging to the ontology. The annotator chooses a relevant concept to describe a media object of interest. For example, an annotator may choose the concept Car to describe a Mercedes automobile.

• Phase 2- Contribution: Processes occur in rounds, allowing individuals to contribute or change their opinions on or understanding of the current annotated version for objects. The annotators collaborate only with reliable users, and the reliability is expressed through trust scores, where each user labels its relationships. Once the annotator has a set of reliable social users, he/she can benefit from their knowledge or skill for annotation purposes. When the annotator is unsure of the annotation information, the annotator asks his/her reliable users for their knowledge and opinions.

• Phase 3-Controlled Feedback: If the collaboration criteria are not satisfied, a combination of prior versions is used as a new version for the object, and the system shows this version of the group’s contribution to each group participant. Phase 2 is repeated until the collaboration criteria are satisfied, and the consensus choice is then applied to make the final version of the object annotation.

Using the ontologies for collaborative annotation, annotation information can be shared between collaborative participants in an effective manner, and the participants have an adequate understanding of the shared annotation. Section 3.2 presents the ontology along with its role, and Section 3.3 shows an SNA-based collaboration and various types of collaborative groups. Section 3.4 presents the corresponding algorithms for the consensus choice by a range of collaborative groups. Section 3.5 proposes a novel collaborative algorithm using consensus-based SNA.

3.2. A semantic model of collaborative processes based on social networks

Collaborators belonging to the same network cooperate with one another through the Internet. In the sharing mechanism for facilitating mutual understanding among social users, there is a need for the structured representation of shared knowledge to allow efficient collaboration between users (Dengler, Lamparter, Hefke, & Abecker, 2009). A media object within a certain context can be modeled in different ways to allow for different interpretations. To address the problem of ambiguity, it is important to specify the shared media object through conceptualization, an abstract or a simplified view of the world to represent for some purpose (Gruber, 1995). To build this conceptualization, ontologies were used as logic vehicles. Ontologies allow for a mutual understanding within some domain of interest, which can then used as a unifying framework, and be communicated between users within a social network. In this way, all collaborative partners can communicate in the same semantic manner, making it easier for them to understand the content of shared media objects as well as one another. Using semantics in conjunction with ontologies, the collaborative activities can be specified in a structured manner and arranged in a meaningful way. Advanced search functions, including the searches for similar or related information integrated from various social media sites, become possible with all content produced through collaborative processes and a correctly classified and structured social network.

A real world \((A, V)\) was assumed, where \(A\) is a finite set of attributes and \(V\) is the domain of \(A\). Here, \(V\) can be expressed as a set of attribute values, and \(V = \bigcup_{a \in A} V_a\), where \(V_a\) is the domain of attribute \(a\). In this paper, the following assumptions were considered based on Duong, Nguyen, and Jo (2009), Duong, Jo, Jung, and Nguyen (2009):

Definition 1 (Ontology). An ontology is a triplet that can be expressed as

\[ O = (C, \bigcup \mathcal{R}) \]  

where,

- \(C\): a set of concepts (classes).
- \(\mathcal{R}\): a set of binary relationships between concepts from \(C\).
- \(\bigcup\mathcal{R}\): the taxonomic structure of concepts from \(C\), where \(\sum\) is a collection of subsumption \(\subseteq\), equivalence \(\equiv\), and disjointness \(\sqcup\) relationships between two concepts from \(C\).

Definition 2 (Concept). The concept \(c\) of an \((A, V)\)-based ontology is defined as a pair that can be expressed as

\[ c = (A^c, V^c) \]  

where \(A^c \subseteq A\) is a set of attributes describing the concept and \(V^c \subseteq V\) is the attributes’ domain: \(V^c = \bigcup_{a \in A^c} V_a\).

The pair \((A^c, V^c)\) is called the possible world or the structure of concept \(c\). Note that within an ontology, there might be two or more concepts with the same structure.

Definition 3 (Instance). An instance of concept \(c\) is described by the attributes from set \(A^c\) with the values from set \(V^c\). Therefore, an instance of concept \(c\) can be defined as a pair:

\[ \text{instance} = (i_d, v) \]  

where \(i_d\) is a unique identifier of the instance in world \((A, V)\) and \(v\) is the value of the instance, which is a tuple of type \(A^c\) and can be presented as the following function:

\[ v: A^c \rightarrow V^c \]  

such that \(v(a) \in V^c\) for \(a \in A^c\). All instances of the same concept in an ontology are different from one another.

By \(\text{Ins}(O, c)\) a set of instances belonging to concept \(c\) in ontology \(O\) can be denoted. A set of ontological instances can be expressed as

\[ I = \bigcup_{c \in C} \text{Ins}(O, c) \]  

Two types of ontologies were used to annotate videos. One describes the visual features of media objects, such as their color, texture, shape, motion, and position (i.e. MPEG-7), and the other provides a knowledge base for a specific domain for annotating the video content (i.e. the LSCOM ontology (Naphade et al., 2006)).

3.3. SNA-based collaboration

3.3.1. SNA

A social network is a social structure comprised of individuals (or organizations) called “nodes”, which are connected by one or more specific types of interdependent relationships, such as friendships, kinships, common interests, financial exchanges, dislikes, sexual relationships, beliefs, knowledge and prestige (Barnes, 1972). The social network is defined in the following graph:

Definition 4 (Social Networking). A social network is a directed loop graph with a pair that can be expressed as

\[ G = (U, R) \]
where

- \( U \) is a set of nodes representing social users.
- \( R \) is an incidence matrix of \( G, R(G) \), and is an \( n \)-by-\( m \) matrix in which \( m \) is the number of edges (repeated relationships) in \( G \); the entry \( m_{ij} \) is the weight of a relationship represented by the arc if \( u_i \) is the starting point of \( e_j \); the entry \( m_{ij} \) is \(-1\) if \( u_i \) is the second point of \( e_j \) and \( 0 \) otherwise.
- If the vertex \( u \) is the starting point of the edge \( e \), \( u \) and \( e \) are incident values.
- The degree of the vertex \( u \), \( d(u) \), is the number of incident value for edges.

The SNA views social connectedness and distance in terms of mathematical network theory such as size, density, and centrality (Scott, 1991; Wasserman & Galaskiewicz, 1994), and SNA measures can be used to analyze the collaboration between users (Palau et al., 2004). Size is indexed by counting the number of nodes and it is often an important factor when calculating other measures. This measure provides a basic understanding of the network status. In a network of size \( n \), the number of possible directed ties is \((n \times (n - 1))\). The density is the proportion of all possible ties that are actually present. In a social network based on collaboration, it is restrictive to establish collaborative relationships between participants in a low density network. On the other hand, in a high density network, the participants can form relationships easily. The network density facilitates information sharing between participants in a collaborative group. Centrality measures, including degree centrality, closeness centrality, and betweenness centrality, describe how close a node is to the “center” of the network based on dispersion. A centralized network has many of its ties connected around one or several nodes (star network), whereas a decentralized network shows a hierarchical structure composed of a set of stars connected in the form of a larger star with additional links forming a loop.

### 3.3.2. Collaboration within a social network

**Definition 5 (Collaborative Group).** A collaborative group can be expressed as subgraph \( G_c = (U_c, R_c, C) \). Here, each user, \( u_i \in U_c \subseteq U, i = 1 \ldots n \) is represented by a node in \( G_c \), and each user has a directed relationship with the users represented by \( R_c \subseteq R \). A user’s collaborative membership in a group can be expressed by the n-by-m matrix \( C(u,g) \). Here, \( M(u,g) \) indicates that user, \( u \in U_c \), is a member of a centralized group, \( g \). Each group includes a centroid member \( u_1 \), and his or her directed relationships are identified by his or her matrix of neighbors. The neighbor matrix of vertex \( u_i \), \( L(u_i) \), is limited by the left column \( \sum_{w_i \in U_c - u_i} d(u) + 1 \) and the right column \( \sum_{w_i \in U_c - u_i} d(u) + d(u_i) \), where \( u_i \) is a vertex in row \( i \) of the matrix \( R(G) \). There is a core group \( g_1 \), whose centroid is the first annotator sharing the annotated object. Other groups have a corresponding group level determined by the depth of their centroid nodes from the core-centroid node.

Fig. 1 shows a collaborative group; there are four centralized groups. According to the SNA, the centrality includes centralization, decentralization and distribution. Centralization refers to networks with a low level of connectedness and a high degree of individual/group control over the others. Decentralization refers to networks with a high level of connectedness between members and a low degree of individual/group control. The distribution refers to networks with no individual/group control. Fig. 2 shows the various types of collaborative groups.

**Definition 6 (Collaborative Annotation).** Collaborative annotation is the best representative annotation from the range of annotated versions of an object, contributed to by a collaborative group. Here, the term “best” suggests that the representation should be the minimized sum of the distances between the consensus of the groups belonging to a collaborative group.

Therefore, two methods were considered for collaborative annotation based on centrality:

1. If a collaborative group reflects a centralized network (see Fig. 2(a)), it considers each member in the group at the same level and normally includes a centroid member serving as a moderator forming a group of experts who participate in the process to annotate a media object in a video. Therefore, the version of the collaborative annotation should be consensus based on the individuals’ versions and satisfy as many consensus criteria as possible (Nguyen, 2008).

2. If a collaborative group is a decentralized network (see Fig. 2(c)), it calculates a version of the consensus annotation for each centralized sub-group belonging to the group, and assigns it to the corresponding centroid. The consensus version of the core group is the collaborative annotation version. Each centralized sub-group participates in the annotation process. Here, the contributed network (see Fig. 2(b)) can be considered decentralized (see Fig. 1).

### 3.3.3. SNA-based collaboration criteria

This subsection presents a set of criteria for the entire process and provides an objective evaluation of collaborative video annotation. These criteria are based on the SNA measures, i.e., inclusion, egalitarian, coordinated, interactive, and trust.

#### 3.3.3.1. Inclusion

Consensus is difficult to reach if only a few individuals participate in a collaborative process, and their opinions are different or even contradictory. In this regard, as many contributors/participants as possible should be involved in the collaboration process to address this problem. The number of participants should be large enough to reach a consensus, and the size of a collaborative group is the number of annotators present in a sub-network. This is useful not only for identifying the inclusive criterion, but also for calculating the other criteria.

#### 3.3.3.2. Egalitarianism

The participants need to be provided with as many opportunities as possible for their collaboration. For example, the more competent an expert is in his or her field, the more likely he or she is to be offered opportunities for collaboration. Therefore, any collaboration should consider this criterion when assigning tasks to the participants according to their area of expertise.

#### 3.3.3.3. Coordination

This criterion measures how easy it is to access and share annotation information in a coordinated manner. Fortunately, the closeness centrality measures the number of connections that an individual should attempt to contact in the network. The closer a participant is to the other participants, the faster he or she will make contact with their collaborative partners. Those participants with a high level of closeness centrality have more opportunities to coordinate with others in a collaborative group. In this manner, they can enhance the knowledge and skills of their collaborative partners to obtain support for accomplishing certain tasks, solving problems and generating ideas.

#### 3.3.3.4. Interactivity

Interactivity refers to the capacity to establish contact within a collaboration group. Any interaction between group members can involve a high or low level of collaboration.

---

If the density of a network with a collaborative group is low, the analyzed network is considered to be restrictive when the participants establish relationships with other participants. On the other hand, the participants can better establish new collaborative relationships if the density is high. In addition, the closeness centrality facilitates the measurements of the interactivity criterion. The closer a participant is to the other participants, the faster he or she will interact with their collaborative partners.

### 3.3.3.5. Trust
Any successful collaborative network operates based on trust. Any user with writing privileges can edit anything he or she contributes to. On the other hand, users should have more refined control. The degree centrality is an effective measure for evaluating the trust criterion. According to SNA theory, if a participant has many ties, i.e., high in-degree centrality, he or she enjoys a high level of prestige among the other participants because many other participants refer directly to him or her. In this manner, the participant can influence the other participants through his or her opinions. On the other hand, a participant with high out-degree centrality often trusts a large number of other participants and has more opportunities to ask for opinions and advice. In this regard, those participants with high out-degree centrality may have advantages because they have more options to satisfy their needs, are less dependent on the other participants, or may have access to more resources.

### 3.4. Consensus choice for conflict resolution
The goal of consensus-based decision making is to find common ideas and explore these issues until everyone’s viewpoint is recognized and understood by the group. Discussions that lead to a consensus aim for mutual understanding by addressing all the
participant’s concerns. A consensus does not require unanimity, but the participants must make a commitment to engage in honest cooperation for the final decision. One well-known conflict is the conflict profile, and a consensus-based method is an effective approach for solving this conflict. In the conflict profile, there is a set of various versions of knowledge that explain the same goal or elements in the real world. Therefore, a consensus aims at determining a reconciled version of knowledge that best represents the given version. The scheme of using a consensus-based method was proposed previously (Kemeny, 1959; Nguyen, 2007; Nguyen, 2008; Nguyen, 2009).

Denote the consensus chosen by these functions as an O1-consensus and O2-consensus. The first function provides the closest representation of the given version, whereas the second gives a consensus representing a good compromise (Nguyen, 2007). The conflict regarding the video annotation can be formulated as follows:

**Definition 7.** Let O be an \((A, V)\)-based ontology. Let concept \((c, A^c, V^c)\) belong to ontology, O, and assume that the same media object, \(i\), be annotated by two users, \(u_1\) and \(u_2\), based on c in O with values of \(v_1\) and \(v_2\), respectively. That is, \((c, v_1) \in \text{Annotation}(u_1, c)\) and \((c, v_2) \in \text{Annotation}(u_2, c)\). A conflict occurs if \(v_1 \neq v_2\).

Consensus-based methods are useful for solving conflicts in the aforementioned annotation. Previous studies proposed various criteria, data structures and algorithms for conflict resolution using consensus (Nguyen, 2008). For this type of conflict, the consensus problem can be defined by the following:

Given a set of values, \(X = \{v_1, \ldots, v_n\}\), where \(v_i\), a tuple of the same type, \(A^c\), is an annotated version contributed to participant, \(i\), for the same media object using concept c, that is,

\[
v_i : A^c \rightarrow V^c
\]

for \(i = 1, \ldots, n; A^c \subseteq A\) and \(V = \bigcup_{a \in A^c} V_a\), we must find the tuple, \(v_i\), of type \(A^c\), such that one or more selected postulates for the consensus are satisfied (Nguyen, 2008).

One popular postulate requires minimization of the following sum:

\[
\sum_{i=1}^{n} \delta_a(v_i, t_i) = \min_{\forall t' \in D} \sum_{i=1}^{n} \delta_a(v'_i, t_i)
\]

where \(T(A^c)\) is a set of all tuples of type \(A^c\).

This postulate is a standard condition for a consensus choice and is useful for reconciling annotation conflicts. This criterion is popular for reaching a consensus, and its justification is based on the requirement that the annotation of a media object should best represent the given opinions of the participants on the object and there should be little difference between the consensus opinion and these opinions.

An integration that satisfies all postulates simultaneously can be derived using the distance functions of the type, \(\rho\), or the proportional distance functions of type, \(\delta\), reported by Nguyen (2008) for a given profile. In general, all postulates except for P2 should be satisfied. Therefore, it is important to determine an integration satisfying P1 and P6.

Here, a consensus in a social network with various levels of participants was considered. The following three corresponding algorithms can be derived according to the type of collaborative group.

### 3.4.1. Centralization algorithm

The centroid participant serves as a moderator and forms a group of experts who participate in the annotation process for a media object in a video. The participants are considered to be at the same level. Here, it is assumed that reliable users offer pertinent opinions and knowledge, but they may also provide insufficient knowledge or conflicting opinions. Therefore, an object can be associated with a range of conflicting versions of the annotation. The collaborative annotated version should be a consensus-based version from the participants’ conflicting versions and satisfy as many of the consensus criteria reported by Nguyen (2008) as possible. This algorithm is based on determining the sub-profiles for the attributes and then for each sub-profile to determine its integration.

**Algorithm 1. A centralization algorithm for reconciling annotation conflicts**

input : A set of values \(X = \{v_1, \ldots, v_n\}\) where \(v_i\) is a tuple of type \(A^c\) and distance functions \(\delta_a\) for attributes \(a \in A^c\).

output: A tuple \(v^*\) of type \(A^c\) that best represents tuples from \(X\).

1. \(A^c = \bigcup A^c_i\);
2. **foreach** each \(a \in A^c\) **do**
   3. Determine a set with repetitions \(X_v = \{v_1 : v_i \in X\} \quad i = 1, 2, \ldots, n\);
3. **end**
4. **foreach** each \(a \in A^c\) **do**
   5. Using distance function \(\delta_a\) determine a value \(v^*_a\) such that \(\sum_{i=1}^{n} \delta_a(v, v_i) = \min_{v^*_a} \sum_{i=1}^{n} \delta_a(v^*_a, v_i)\);
7. **end**
8. Create a tuple \(v^*\) consisting of values \(v^*_a\) for all \(a \in A^c\);
9. Return(\(v^*\)).

### 3.4.2. Decentralization algorithm

A decentralized group is normally formed for collaboration within a social network. If an annotator is unsure about the annotation information on a given object, he or she asks reliable friends for their knowledge and opinions. This collaborative group can be decomposed into centralized sub-groups with each centralized sub-group having a centroid member who plays the role as a moderator. This algorithm calculates the consensus annotated version for each centralized sub-group belonging to the collaborative group and assigns it to the corresponding centroid participant. The consensus version of the core group is the collaborative annotated version.

**Algorithm 2. A decentralization algorithm for reconciling annotation conflicts**

input : A collaborative group \(G = (U, R)\) and the core-centroid participant \(u^* \in U\).

output: A consensus version of the participant \(u^*\).

1. \(S_{GCD} = \{u_i : d(u_i) > Centrality degree\}\) is a set of centroid participants;
2. Create a set of centralized subgroups \(S_g = \{g(u_i), i = 1..|S_{GCD}|\}\);
3. **foreach** \(u_i \in S_{GCD}\) **do**
   4. Let \(L(u_i)\) be a matrix of neighbors of the vertex \(u_i\) or be \((R(G)\) limited by the left column \(\sum_{j \in L(u_i)} d(u_j) + 1\) and the right column \(\sum_{j \in L(u_i)} d(u_j) + d(u_i)\), where \(u_i\) is a vertex in row \(i\) of the matrix \(R(G)\).
5. **end**
6. Sort \(S_g\) in descending order for level of group \(g \in S_g\);
7. **foreach** \(g \in S_g\) **do**
   8. **foreach** \(u_j \in L(u_i)\) where \(u_j\) is an annotated version contributed to by \(u_j \in L(u_i)\);
9. Annotation of centroid participant \(u_i = \text{CentralizationAlgorithm}(X)\);
10. **end**
11. **end**
12. Return(the core-centroid \(u^*\)’s annotation);

### 3.5. Collaborative annotation algorithm using consensus-based SNA

This section presents a simple collaborative annotation algorithm using a consensus-based SNA. An annotator shares a media...
object for annotation and then joins a collaborative group to annotate the object (see lines 4–8 in Algorithm 3). If the annotator is unsure about the information on the object, he or she can ask reliable users for their knowledge and opinions (lines 9–14). Here, trust is measured using the degree centrality, as discussed earlier. If the collaboration criteria are satisfied, then all participants’ annotations are used to make a consensus version of the object annotation (line 22). Otherwise, the annotation information is shared to relevant users who are determined to be trustworthy based on the SNA (lines 15–21). Here, line 23 calculates the quality of the consensus version. If this quality is sufficient for reaching a consensus, the consensus version is considered to be the final annotated version, and the collaborative annotation process is complete (line 24). Otherwise, the consensus version is sent to all participants belonging to the collaborative group, and the contribution round (phase 2–contribution) is repeated (lines 25–27).

**Algorithm 3. A collaborative annotation algorithm**

```plaintext
input : G(U, R) as a social network and a media object for annotation using concept c, and collaboration criteria
output: τ as the annotation of a given object and a tuple of type \( A^c \)
1. Denote \( G(U', R') \subset G(U, R) \) as a current collaborative network;
2. Denote \( X = \{v_i : v_i \text{ an annotation value, is provided by } u_i \in U'\}; \)
3. Label*;
4. Denote \( u_i \) as a current participant who is shared a media object;
5. if \( u_i \in U' \) is knowledgeable about the shared object then
   6. \( u_i \) annotates the object with the value \( v_i \);
   7. \( X = \cup\{v_i\}; \)
else
   8. \( F^c = \{u_1 : R(u_1, u_1) = 1\} \) is a set of friends of participant \( u_1; \)
   9. \( U^c = \{u_1 : u_1 \in F^c \& \text{CentralityDeg}(u_1) \geq \text{Threshold}_c\}; \)
   10. \( u_1 \) shares the object with reliable friends \( U^c \) through the social network;
   11. \( G(U', R') = G(U, R') \cup G(U', R') \);
end
if \( \text{Size}(G) < \text{Threshold}_S \text{ or } \text{Density}(G) < \text{Threshold}_D \) then
   13. UnShared := \{u_i : u_i \in U' \text{ is a participant not yet sharing the object with friends}\};
else
   15. Share the object with \( U' \) and \( u_i \in \text{UnShared}; \) Goto Label*;
end
\( \text{else} \) Share the object with \( U'[u_i \in U' \cup U'; \) Goto Label*;
\( \text{end} \)
\( \text{else} \) Share the object with \( V\); Goto Label*;
\( \text{end} \)
decentralizationAlgorithm\( G(U', R') \);
\( d^c(\tau, X) = 1 - \frac{d(\tau, X)}{\text{card}(X)} \);
21. if \( d^c(\tau, X) > \text{Threshold} \ast \& d^c(\tau, X) \) is unchanged then Return(\( \tau \));
\( \text{else} \) Share \( \tau \) to \( U' \); Goto Label*;
\( \text{end} \)
```

### 4. Experiments

#### 4.1. Data analysis

The TRECVID\(^1\) data set reported by Kraaij and Over (2005), which consists of broadcast news video sources, was used to confirm the performance of the proposed approach. The domain ontology used here is the LSCOM (Naphade et al., 2006). Along with a taxonomy of 1000 concepts, the LSCOM produces a set of use cases and queries as well as a large annotated broadcast news video data set. The aim is to construct a framework for achieving mutual understanding for multimedia content. Each media object in a specific video can be considered as an instance of a domain concept belonging to the LSCOM ontology. The LSCOM was too large for the experiment. Therefore, an ontology based the LSCOM was created using the 374 concepts reported by Yanagawa, Chang, Kennedy, and Hsu (2007) and Naphade et al. (2006). For those concepts with no attributes, 50 attributes were supplemented for the ontology.

In this experiment, the following two aspects of the proposed approach were evaluated: the effects of SNA measures on the collaboration process and an evaluation of the consensus choice for solving conflicts in collaborative annotation. For this, two collaborative groups, A (see Fig. 3) and B (see Fig. 4), were considered for video annotation. Each group included 30 members. The difference is that group A has 43 relationships between students, whereas group B has only 30. All members were asked to annotate 100 shots (50 shots each for group A and B) for the 374 concepts. The students are allowed to use any agent, their knowledge, or their own predictions to provide annotation information. In both networks, the initial sharer is Thanh Hien. Each individual can share annotation information only with those with whom he or she has relationships.

#### 4.2. Evaluation method

To evaluate the effect of the SNA on the collaborative quality, the consensus quality in groups A and B was compared. According to Nguyen (2008), the consensus quality can be expressed as the following theorem:

**Theorem 1.** Let \( X \in \prod(U), \ C \in \text{Con}(U), \) and \( x \in C(X) \). By the quality of consensus \( x \) in profile \( X \) we have the following:

\[
d(x, X) = 1 - \frac{d(x, X)}{\text{card}(X)}
\]

where \( U \) is a universe and \( \text{Con}(U) \) is the set of all consensus functions for conflict profiles from \( \prod(U) \).

Here, it is postulated that the collaborative quality will increase with increasing consensus quality. This was used to analyze how the collaborative process proceeds under the proposed methodology.

In addition, to evaluate the accuracy of the collaborative annotation, the collaborative annotation by the collaborative group (a consensus based on the members’ opinions and understanding) was compared with the proper annotation (expertly verified annotation). The feature vector of collaborative annotation provided by group \( i \) for a specific media object based on concept \( c \) can be expressed as follows:

\[
\vec{F}_i = (id, \psi)
\]

where \( id \) is a unique identifier of an instance in world \( (A, V) \) and \( \psi \) is the value of the instance, which is a tuple of type \( A^c \). In addition, \( F_e \) denotes the proper annotation verified by experts.

The accuracy of the collaborative annotation provided by group \( i \) is defined as follows:

\[
\text{accuracy}(u, p) = \frac{\text{similarity}(\vec{F}_i, \vec{F}_e)}{\sqrt{\sum(F_i)^2 + \sqrt{\sum(F_e)^2}}}
\]

#### 4.3. SNA-based evaluation of collaboration

To evaluate the effects of the SNA on the collaborative quality, the improvement in consensus quality during the contribution round was considered (phase 2). The consensus quality for each contribution round was calculated (see line 23 in Algorithm 3) and the threshold for this quality was set to 0.8. Fig. 5 shows the mean consensus quality for each round. These results show general

---

\(^1\) <http://trecvid.nist.gov/>.
improvement in consensus quality as the rounds progress because the consensus version is shared by the participants, who take advantage of their group and only supplement insufficient information or modify incorrect information. On the other hand, for individual shots, the consensus quality is sometimes worse than that in the previous round. As shown in Fig. 5, group A reaches consensus faster and shows better consensus quality than group B. UCINet reported by Borgatti, Everett, and Freeman (2002), which is a software package designed to represent and analyze social networks, was used to explain this.

The size and density of the two collaborative groups were first analyzed. The size of both groups was 30, and the largest number of ties in a fully saturated network of size 30 was 870. Network B consisted of 30 real ties, whereas network A consisted of 43. Therefore, the density of network A exceeds that of network B, which suggests that the former creates new collaborative relationships.
faster than the latter. Therefore, group B has an insufficient number of participants in the first round to reach consensus.

The closeness centrality was considered to evaluate the effects of the SNA on collaboration. As shown in Figs. 6 and 7, the members of network A are closer to one another than those of network B. Therefore, the former develops relationships and interacts with one another faster. As a result, group A reaches a consensus faster than group B from rounds 2–8.

In terms of the degree centrality (see Fig. 7), Thanh Hien showed the highest degree centrality in groups A and B, suggesting that others trust her and she enjoys greater prestige. In addition, she has more power because she can influence others through her opinions. In other words, she can communicate annotated objects to many others and many trust her, which provides her with more opportunities to ask for opinions and advice. On the other hand, Thanh Hien shows a higher degree centrality in group A than in group B, i.e. collaborative evolution is faster in group A than in group B from the first round.

Fig. 6 shows that actors with a high closeness centrality are more likely to obtain support for accomplishing tasks, solving problems and generating ideas. Here, a highly connected network is ideal.

4.4. Consensus choice for collaborative annotation

According to the experimental results, many participants annotate the same object in a specific video. Therefore, many annotated versions/profiles are associated with the same object. Although these participants are based on the SNA, conflicts over the annotated versions are unavoidable. This type of conflict situation is referred to as a conflict profile, and the consensus-based method reported by Nguyen (2008) is an effective approach for solving this conflict. This method aims to determine a reconciled version of knowledge that best represents a given version. Therefore, a consensus choice was applied to determine the version that best reflects/represents the annotation for an object by considering those participants with conflicting points of view. For an evaluation of the consensus choice for a collaborative annotation, the accuracy and quality of the consensus in the first 12 contribution rounds in groups A and B was compared. According to the results, the accuracy of the annotation improves with increasing consensus quality. In addition, the consensus accuracy depends on the SNA. Here, the time the network takes to achieve consensus accuracy decreases with increasing density or closeness centrality. (See Fig. 8)

4.5. Discussion

The aforementioned automatic systems for semantic video annotation depend strongly on the quality and availability of a large collection of training videos (Jeong et al., 2011; Naphade & Smith, 2004; Park et al., 2007). The annotation of large collections, however, is a time-consuming and error-prone task. Therefore, it should be performed manually. In addition, the literature indicates that there is no automatic object description annotation. Moreover, the existing collaborative approaches do not mention the collaboration criteria for evaluating the collaboration process and do not consider the question of who should share the annotating object to maximize the opportunities for collaboration and the benefits from others’ skills and knowledge. Conflicts between annotated versions of an object are unavoidable. Therefore, the question is how well can a participant integrate conflicting annotated versions of an object with the contributions by other participants to construct the best-annotated version of the object? Accordingly, a collaborative video annotation approach is proposed using consensus-based SNA to benefit from others’ skills and knowledge (Volkmer et al., 2005; Yamamoto, Masuda, Ohira, & Nagao, 2008b).

Overall, who should share an annotated object to maximize the opportunities for collaboration and benefit from the users social skills and knowledge? To the authors’ knowledge, no study has addressed this issue. In this regard, the collaborative video annotation process was organized based on social networks. Beyond profiles, friends, comments and private messages, the social networking sites (SNSs) can vary widely in terms of their features and user base. Some have photo- or video-sharing capabilities, and others feature built-in blogging and instant-messaging technologies. Examples include mobile-specific SNSs (e.g. Dodgeball). Some SNSs also support limited mobile interactions (e.g. Facebook, MySpace, and Cyworld). For these social network advantages, semantic media content can be shared through a large network of friends of friends who annotate it based on their available social profiles. In this way, a participant has insufficient knowledge or experience to annotate an interesting object in a specific video, he or she can share the object with friends who are also interested in it. Therefore, many users can participate in the collaborative annotation workflow through social networks. On the other hand, the participant does not simply share the object of interest with every friend in his or her network. Here, any shared friend needs to satisfy the collaboration criteria. Trust, which is one of the most important criteria, means that participants need to be afforded as many opportunities for collaboration as possible. Accordingly,
how can a user trust another user's contribution? This is why a participant is not allowed to share his or her object annotation with everyone in his or her network. The level of trust is determined by the relationship between the annotating participant and his or her friends. If they are closer, they can better satisfy the interactivity criterion.

One question remains: how well can conflict-annotated versions of an object being contributed to by participants be integrated to obtain the best annotated version? Many users can annotate the same object in a specific video, which means that there may be many annotated versions/profiles of the same video being contributed to by the participants. Even when these participants are trustworthy based on the SNA, conflicts between annotated versions are unavoidable. Here the consensus-based method is useful for resolving the conflict profile, which is a well-known conflict situation. In a conflict profile, there are

![Comparison of closeness centrality of groups A and B.](image1)

![Comparison of degree centrality of groups A and B.](image2)

![Consensus accuracy vs. quality in first 12 contribution rounds.](image3)

different sets of knowledge that explain the same goals or elements in the real world. Here the consensus-based method aims to determine a reconciled version of knowledge that best represents a given version. According to the experimental results, the consensus choice can be applied effectively to determine the best annotation of an object based on the conflicting participants’ knowledge.

The strengths and weaknesses of the proposed research method are clearly addressed as follows:

**Strengths:**

- The proposed method is able to indicate the collaborative participants who should share the annotated object to his/her friends instead of blindly or exhaustively share among participants in the network to maximize the opportunities for collaboration and benefit from the users social skills and knowledge. Social network analysis is used as criteria for participant collection.
- The proposed method resolves conflict-annotated versions of an object being contributed by participants. Here the consensus-based method is useful for resolving the conflict, which is a well-known conflict profile. In a conflict profile, there are different sets of knowledge that explain the same goals or elements in the real world. Here the consensus-based method aims to determine a reconciled version of knowledge that best represents conflict-annotated versions. According to the experimental results, the consensus choice can be applied effectively to determine the best annotation of an object based on the conflicting participants’ knowledge.

**Weaknesses:**

- In the proposed method, the collaboration is ended when consensus is reached. The method has not yet considered the quality of the collaboration for each contribution round. Based on the quality of collaboration for each contribution round, we can get to know when the collaboration should be finished. This is our future work.

5. Conclusion

An increasing number of social videos are now available as well as traditional digital videos such as TV programs and video on demand (VOD). However, it is difficult to find relevant content due to a lack of semantic content. Therefore, there is a great need for multimedia content analysis techniques. In this work, we proposed a novel method to embed media content with available information from the Internet in a semantic manner. A multimedia ontology is exploited to semantically describe the media content. Each media object can be considered as an instance of a domain concept belonging to the multimedia ontology. This novel idea is to provide semantic for media content that connects pieces of data available in Internet. It also offers advanced semantic search functionality and facilitates the sharing of media content among heterogeneous agents. Within the scope of the research topic, we intend to develop effective methods and algorithms for collaborative ontology-based video annotation using consensus-based social network.

In this work, the collaborative video annotation process was organized based on social networks. Here, the media content of a video is shared through a large network of friends who annotate it based on the SNA, and the consensus choice is used to address conflicts over a collaborative annotation. According to the experimental results, a collaborative group in a network can form new mutual relationships faster with increasing network density. In addition, a collaborator with a large number of connections can be an important (power) user, who can facilitate the collaborative annotation process by maximizing the benefits from the other users’ skills and knowledge. This is because many other users trust this user. Therefore, he or she can have a considerable influence on other users through his or her opinions and knowledge. In addition, closeness centrality plays an important role in the evolution of the collaborative process. Users with high closeness centrality are more likely to obtain support for accomplishing tasks, resolving problems and generating ideas. In addition, ontologies should be exploited to semantically describe the media content to facilitate its sharing by heterogeneous users in a social network. This can be considered mutual understanding, and can guide their opinions on shared objects. According to the experimental results, conflicts between collaborative annotators are unavoidable. This situation is referred to as a conflict profile, and the consensus-based method provides an effective approach to resolving this conflict. In this paper, SNA-based collaboration criteria were used to evaluate the collaborative process. Overall, the more collaboration criteria a collaborative group can satisfy, the faster the group can reach a consensus.

The novel of the proposed collaborative algorithm for semantic video annotation using consensus-based SNA can be mainly stated as follows:

- The previous related works have not yet explored the criteria/methods to select the collaborative participants who can influence other participants to maximize the opportunities for collaboration and benefit from the users social skills and knowledge. The proposed algorithm used social network analysis to form collaborative criteria for participant collection who share the annotated object to his/her friends instead of blindly or exhaustively share among participants in the network. This novel idea leads to fast satisfy the consensus and to enhance the quality of collaboration.
- The conflict problem on collaborative annotation has not yet considered before. The previous works on collaborative annotation usually allow a current participant refines the pre-annotated version contributed by other participants. Here the consensus-based method is applied to determine a reconciled version of knowledge that best represents conflict-annotated versions. According to the experimental results, the consensus choice can be applied effectively to determine the best annotation of an object based on the conflicting participants’ knowledge.

In future work, we will consider the computational quality of the collaboration for each contribution round. Based on the quality of collaboration for each contribution round, we can get to know when the collaboration should be finished. The semantic video annotation can be also applied for developing intelligent systems such as an intelligent e-commerce system representing new forms of shopping that enable consumers to view, select, and buy products from Smart/Internet TV. To do so sellers should annotate videos and associate items with information from e-commerce systems. On the other hand, manufacturers annotate the way items are produced. In addition, buyers provide their opinions on received items. The way for image representation proposed in Chohra, Kanaoui, Amarger, and Madani (2014) can be used for this process.

Acknowledgments

This research is partially funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 102.01-2013.12 and Vietnam National University Ho Chi Minh City (VNU-HCM) under Grant No. C2014-26-05.